

Парадигма развития науки

Методологическое обеспечение

А. Е. Кононюк

**ОБОБЩЕННАЯ ТЕОРИЯ
МОДЕЛИРОВАНИЯ**

Книга 1

Начала

Часть 3

**Киев
Освіта України
2012**

А. Е. Кононюк Обобщенная теория моделирования

УДК 51 (075.8)

ББК В161.я7

К 213

Рецензент: *Н.К.Печурин* - д-р техн. наук, проф. (Национальный авиационный университет).

Кононюк А. Е.

К65 Обобщенная теория моделирования. Начала. К.1.Ч.3

К.4:"Освіта України", 2012. - 568 с.

ISBN 978-966-7599-50-8

Настоящая работа является систематическим изложением обобщенной теории моделирования. Основное внимание уделяется идейным основам теории моделирования, их сравнительному анализу и примерам использования. Рассмотрен значительный круг задач моделирования — от общих задач моделирования до частных задач моделирования, а именно: моделирование объектов по выполняемым функциям, по составу, по структуре, по форме, по организации, по управлению. Обсуждается методика постановки и решения проблем моделирования. Рассматриваются средства математического описания объектов и процессов моделирования. Описываются системы автоматизированного моделирования.

Работа предназначена для магистров, аспирантов, докторантов, инженеров, экономистов, статистиков, вычислителей и всех тех, кто сталкивается с задачами моделирования, прежде всего, математического.

ББК В161.я7

ISBN 978-966-7599-50-8

©А.Е. Кононюк, 2012

Оглавление

1. Введение в статистическое моделирование.....	7
Введение.....	7
1.1. Метод Монте-Карло.....	23
1.2. Планирование натуральных испытаний опытных образцов сложной системы.....	23
1.2.1. Задачи, решаемые при подготовке и проведении натуральных испытаний.....	23
1.2.2. Постановка задачи об оптимальном плане проведения натуральных испытаний с учетом экономических факторов.....	25
1.2.3. Алгоритм вычисления $\alpha(s_i)_2$ методом Монте — Карло и алгоритм решения задачи оптимального распределения средств.....	27
1.2.4. Методические указания по подготовке исходных данных и решению задачи оптимального распределения средств.....	32
1.3. Разработка математических моделей для опытно-теоретической оценки показателей сложных систем.....	35
1.3.1. Общие принципы построения моделей.....	35
1.3.2. Характеристика методов математического моделирования.....	40
1.3.3. Ограничения на сложность математических моделей.....	43
1.3.4. Распределение требований к точности моделирования подсистем.....	44
1.3.5. Разработка программ моделей.....	48
1.4. Калибровка математических моделей по результатам натуральных испытаний.....	52
1.4.1. Общие положения о калибровке моделей сложных систем.....	52
1.4.2. Критерии и условия оптимальности.....	58
1.4.3. Методы решения уравнений оптимальности.....	61
1.4.4. Особенности и вычислительная реализуемость методов.....	64
1.4.5. Приближенная методика калибровки моделей сложных систем.....	65
1.4.6. Оценка статистической совместимости результатов моделирования с результатами натуральных испытаний.....	68
1.5. Оценка точности результатов статистического моделирования.....	75
1.5.1. Классификация ошибок.....	75
1.5.2. Ошибки дискретной реализации моделирующих алгоритмов.....	77
1.5.3. Оценка случайных ошибок, обусловленных конечным числом реализаций на модели.....	83
1.5.4. Особенности и последовательность задач, решаемых при	

оценке влияния ограниченного объема реальной информации на точность результатов моделирования.....	89
1.5.5. Корректировка результатов статистического моделирования.....	93
1.5.6. Оценка методических ошибок моделирования в пределах всей области рассматриваемого факторного пространства.....	102
1.6. Методы обработки и анализа результатов испытаний сложных систем.....	105
1.6.1. Требования к методам обработки результатов измерений.....	105
1.6.2. Первичная обработка результатов измерений.....	107
1.6.3. Оценка дисперсий и математических ожиданий.....	113
1.6.4. Анализ качества измерений от различных источников.....	122
1.6.5. Оценка корреляционных моментов и коэффициентов регрессии.....	151
1.6.6. Байесова процедура оценки.....	155
1.6.7. Принцип максимального правдоподобия.....	162
1.6.8. Алгоритмы метода наименьших квадратов.....	166
1.7. Методы статистической обработки информации.....	173
1.7.1. Общая характеристика и формализация задач обработки информации.....	173
1.7.2. Показатели эффективности алгоритмов обработки сигналов.....	179
1.7.3. Классификация методов синтеза алгоритмов статистической обработки сигналов.....	187
1.7.4. Специальные методы устойчивой обработки сигналов.....	193
1.7.5. Особенности реализации устойчивых алгоритмов обработки сигналов.....	195
1.8. Устойчивые алгоритмы измерения (оценки) параметров сигналов на фоне помех.....	197
1.8.1. Основные понятия и определения.....	197
1.8.2. Параметрические алгоритмы оценки.....	201
1.8.3. Алгоритмы, использующие М-оценки Хьюбера.....	211
1.8.4. Алгоритмы, использующие порядковые статистики.....	220
1.8.5. Алгоритмы, основанные на ранговых статистиках.....	224
1.8.6. Оценивание с расслоением выборки.....	227
1.8.7. Алгоритмы устойчивого оценивания корреляционных функций.....	234
1.8.8. Сравнение эффективности алгоритмов оценивания.....	236
1.9. Устойчивые алгоритмы обнаружения сигналов.....	239
1.9.1. Задачи и методы их решения.....	239
1.9.2. Классификация и методы синтеза параметрических (адаптивных) алгоритмов.....	245
1.9.3. Параметрические алгоритмы, использующие L-оценки.....	258

1.9.4. Ранговые непараметрические алгоритмы	262
1.9.5. Квантильные непараметрические алгоритмы	270
1.9.6. Алгоритмы обнаружения, использующие расслоение выборки.....	277
1.9.7. Комбинированные адаптивно-непараметрические алгоритмы в сложных помеховых ситуациях.....	287
1.9.8. Сравнение эффективности алгоритмов обнаружения.....	292
1.10. Устойчивые алгоритмы фильтрации сигналов.....	295
1.10.1. Устойчивые адаптивные алгоритмы фильтрации квазидетерминированных сигналов.....	296
1.10.2. Устойчивые алгоритмы рекурсивной фильтрации сигналов..	299
1.11. Моделирование алгоритмов обработки сигналов радиоло- кационных информационно-измерительных системах (ИИС).....	312
1.11.1. Постановка задачи и критерии эффективности ИИС.....	312
1.11.2. Алгоритмы моделирования сигнально-помеховой ситуации	314
1.11.3. Имитационная модель ИИС АС УВД и ее применение.....	318
2. Введение в имитационное моделирование.....	329
2.1. Введение в имитацию и моделирование.....	329
2.1.1. Общие замечания.....	329
2.1.2. Модели и их построение.....	331
2.1.3. Примеры имитационных моделей.....	335
2.1.4. Процесс имитационного моделирования.....	341
2.2. Элементы теории вероятностей и математической статистики..	350
2.2.1. Введение.....	350
2.2.2. Эксперимент, пространство выборки и результат.....	351
2.2.3. Вероятность.....	352
2.2.4. Случайные величины и вероятностные распределения.....	353
2.2.5. Математическое ожидание и моменты.....	357
2.2.6. Функции случайных величин.....	359
2.2.7. Генераторы.....	360
2.2.8. Закон больших чисел и центральная предельная теорема.....	361
2.2.9. Распределения.....	362
2.2.10. Генерация псевдослучайных чисел.....	372
2.2.11. Тесты на случайность.....	376
2.2.12. Сбор и анализ данных.....	377
2.2.13. Статистический вывод.....	383
2.2.14. Проверка гипотез.....	385
2.2.15. Статистические проблемы имитационного моделирования..	388
2.3. Методологические подходы в имитационном моделировании..	391
2.3.1. Введение.....	391
2.3.2. Системы и модели.....	392
2.3.3. Дискретное имитационное моделирование.....	396

2.3.4. Непрерывное имитационное моделирование модели.....	402
2.3.5. Комбинированные дискретно-непрерывные модели.....	404
2.3.6. СЛАМ: объединенная схема моделирования.....	405
2.4. Формализация описаний процесса функционирования сложной технической системы.....	407
2.4.1. О понятии имитационного моделирования.....	407
2.4.2. Алгоритмическая модель процесса.....	409
2.4.3. Формы описания процессов.....	415
2.5. Программные имитационные модели.....	419
2.5.1. Анализ одновременных событий.....	419
2.5.2. Управляющий алгоритм.....	425
2.5.3. Практическая реализация моделирующих алгоритмов.....	429
2.6. Статистические процедуры в имитационном моделировании	431
2.6.1. Генерирование случайных чисел.....	431
2.6.2. Оценка точности результатов.....	436
2.6.3. Оценки типовых характеристик.....	438
2.6.4. Начальный интервал моделирования.....	440
2.7. Система имитационного моделирования СТАМ-КЛАСС.....	442
2.7.1. Принципы построения.....	442
2.7.2. Описание элементов языка.....	447
2.7.3. Вычислительные и моделирующие операторы.....	452
2.7.4. Пример программы имитационной модели.....	457
2.8. Система имитационного моделирования GPSS.....	462
2.8.1. Принципы построения имитационных программ.....	462
2.8.2. Синтаксис элементов языка.....	466
2.8.3. Типовые конструкции программ.....	491
2.9. Статистические аспекты имитационного моделирования.....	506
2.9.1. Математическая статистика и разработка имитационных моделей.....	506
2.9.2. Дисперсия выборочного среднего $\text{VAR}[\bar{X}_t]$ и ее значение.....	510
2.9.3. Процедура оценки $\text{VAR}[\bar{X}_t]$	515
2.9.4. Методы понижения дисперсии.....	526
2.9.5. Стратегии запуска.....	534
2.9.6. Правила останова.....	538
2.9.7. Планирование эксперимента.....	541
2.10. Применения имитационного моделирования.....	547
2.10.1. Области применения.....	547
2.10.2. Анализ потока запросов в филиале фирмы по страхованию жизни и имущества.....	548
Литература.....	562

1. Введение в статистическое моделирование

Введение

Статистическое моделирование - численный метод решения математических задач, при котором искомые величины представляют вероятностными характеристиками какого-либо случайного явления, это явление моделируется, после чего нужные характеристики приближённо определяют путём статистической обработки «наблюдений» модели. Например, требуется рассчитать потоки тепла в нагреваемой тонкой металлической пластине, на краях которой поддерживается нулевая температура. Распределение тепла описывается тем же уравнением, что и расплывание пятна краски в слое жидкости. Поэтому моделируют плоские частицы «краски» по пластине, следя за их положениями в моменты $k_t, k = 0, 1, 2, \dots$. Приближённо принимают, что за малый интервал t частица перемещается на шаг h равновероятно во всех направлениях. Каждый раз направление выбирается случайным образом, независимо от всего предыдущего. Соотношение между t и h определяется коэффициентом теплопроводности. Движение начинается в источнике тепла и кончается при первом достижении края (наблюдается налипание «краски» на край). Поток Q (C) тепла через участок C границы измеряется количеством налипшей краски. При общем количестве N частиц согласно *закону больших чисел* такая оценка даёт случайную

относительную ошибку порядка $\frac{1}{\sqrt{N}}$ (и систематическую ошибку порядка h из-за дискретности выбранной модели).

Искомую величину представляют *математическим ожиданием* числовой функции f от случайного исхода ω явления:

$E_f(\omega) = \int f(\omega) dP$, т. е. интегралом по вероятностной мере P . На оценку $E_f(\omega) \approx [f(\omega_1) + \dots + f(\omega_N)] / N$, где $\omega_1, \dots, \omega_N$ - смоделированные исходы, можно смотреть как на квадратурную формулу для указанного интеграла со случайными узлами ω_k и случайной погрешностью R_N обычно принимают $|R_N| \leq 2\sqrt{Df} / \sqrt{N}$,

считая большую погрешность пренебрежимо маловероятной; дисперсия Df может быть оценена в ходе наблюдений.

В разобранный выше примере $f(\omega) = 1$, когда траектория кончается на C ; иначе $f(\omega) = 0$. Дисперсия

$$Df = [1 - \rho(C)]\rho(C) \leq \frac{1}{4}$$

Интеграл берётся по пространству ломаных со звеньями постоянной длины; он может быть выражен через кратные интегралы.

Проведение каждого «эксперимента» распадается на две части: «розыгрыш» случайного исхода ω и последующее вычисление функции $f(\omega)$. Когда пространство всех исходов и вероятностная мера P слишком сложны, розыгрыш проводится последовательно в несколько этапов. Случайный выбор на каждом этапе проводится с помощью случайных чисел, например генерируемых каким-либо физическим датчиком; употребительна также их арифметическая имитация - псевдослучайные числа. Аналогичные процедуры случайного выбора используются в математической статистике и теории игр.

Статистическое моделирование широко применяется для решения на ЭВМ интегральных уравнений, например при исследовании *больших систем*. Они удобны своей универсальностью, как правило, не требуют большого объёма памяти. Недостаток - большие случайные погрешности, слишком медленно убывающие при увеличении числа экспериментов. Поэтому разработаны приёмы преобразования моделей, позволяющие понижать разброс наблюдаемых величин и объём модельного эксперимента.

1.1. Метод Монте-Карло

Как мы уже отмечали, статистическое моделирование — базовый метод моделирования, заключающийся в том, что модель испытывается множеством случайных сигналов с заданной плотностью вероятности. Целью является статистическое определение выходных результатов. В основе статистического моделирования лежит *метод*

Монте-Карло. Напомним, что имитацию используют тогда, когда другие методы применить невозможно.

Рассмотрим метод Монте-Карло на примере вычисления интеграла, значение которого аналитическим способом найти не удастся.

Задача 1. Найти значение интеграла:

$$y = \int_{x_1}^{x_2} f(x) dx$$

На рис. 1 представлен график функции $f(x)$. Вычислить значение интеграла этой функции — значит, найти площадь под этим графиком.

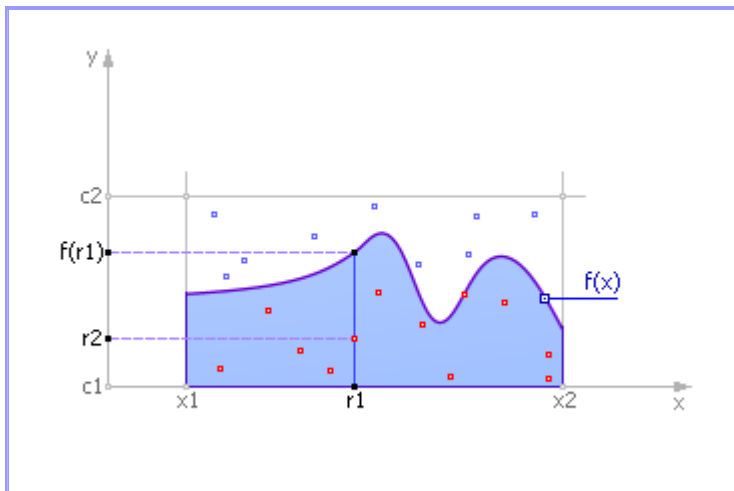


Рис. 1. Определение значения интеграла методом Монте-Карло

Ограничиваем кривую сверху, справа и слева. Случайным образом распределяем точки в прямоугольнике поиска. Обозначим через N_1 количество точек, принятых для испытаний (то есть попавших в прямоугольник, эти точки изображены на рис. 1 красным цветом), и через N_2 — количество точек под кривой, то есть попавших в закрашенную площадь под функцией (эти точки изображены на

рис. 1 красным цветом). Тогда естественно предположить, что количество точек, попавших под кривую по отношению к общему числу точек пропорционально площади под кривой (величине интеграла) по отношению к площади испытываемого прямоугольника. Математически это можно выразить так:

$$\frac{N_2}{N_1} = \frac{y}{(x_2 - x_1)(c_2 - c_1)}$$

Рассуждения эти, конечно, статистические и тем более верны, чем большее число испытываемых точек мы возьмем. Фрагмент алгоритма метода Монте-Карло выглядит так, как показано на рис. 2.

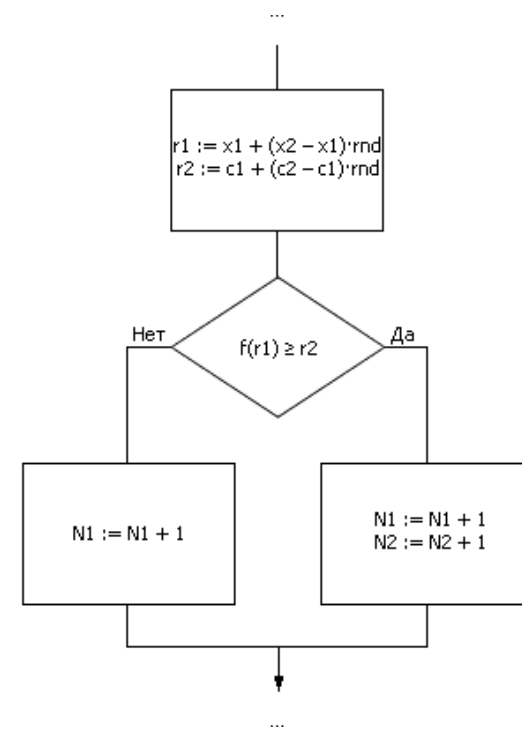


Рис. 2. Фрагмент алгоритма реализации метода Монте-Карло

Значения r_1 и r_2 на рис. 2 являются равномерно распределенными случайными числами из интервалов $(x_1; x_2)$ и $(c_1; c_2)$ соответственно. Метод Монте-Карло чрезвычайно эффективен, прост, но необходим «хороший» генератор случайных чисел. Вторая проблема применения метода заключается в определении объема выборки, то есть количества точек, необходимых для обеспечения решения с заданной точностью. Эксперименты показывают: чтобы увеличить точность в 10 раз, объем выборки нужно увеличить в 100 раз; то есть точность примерно пропорциональна корню квадратному из объема выборки:

$$\text{точность} \approx \sqrt{\text{объем выборки}}$$

Схема использования метода Монте-Карло при исследовании систем со случайными параметрами

Построив модель системы со случайными параметрами, на ее вход подают входные сигналы от генератора случайных чисел (ГСЧ), как показано на рис. 3. ГСЧ устроен так, что он выдает *равномерно распределенные* случайные числа r_{pp} из интервала $[0; 1]$. Так как одни события могут быть более вероятными, другие — менее вероятными, то равномерно распределенные случайные числа от генератора подают на преобразователь закона случайных чисел (ПЗСЧ), который преобразует их в *заданный* пользователем закон распределения вероятности, например, в нормальный или экспоненциальный закон. Эти преобразованные случайные числа x подают на вход модели. Модель обрабатывает входной сигнал x по некоторому закону $y = \varphi(x)$ и получает выходной сигнал y , который также является случайным.

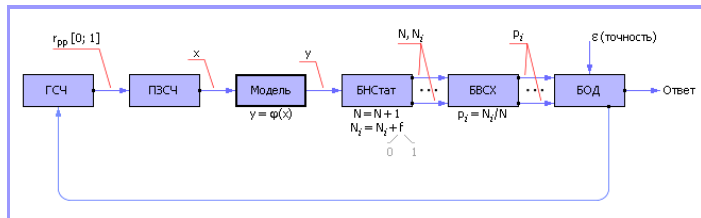


Рис. 3. Общая схема метода статистического моделирования

В блоке накопления статистики (БНССтат) установлены фильтры и счетчики. Фильтр (некоторое логическое условие) определяет по

значению y , реализовалось ли в конкретном опыте некоторое событие (выполнилось условие, $f = 1$) или нет (условие не выполнилось, $f = 0$). Если событие реализовалось, то счетчик события увеличивается на единицу. Если событие не реализовалось, то значение счетчика не меняется. Если требуется следить за несколькими разными типами событий, то для статистического моделирования понадобится несколько фильтров и счетчиков N_i . Всегда ведется счетчик количества экспериментов — N .

Далее отношение N_i к N , рассчитываемое в блоке вычисления статистических характеристик (БВСХ) по методу Монте-Карло, дает оценку вероятности p_i появления события i , то есть указывает на частоту его выпадения в серии из N опытов. Это позволяет сделать выводы о статистических свойствах моделируемого объекта.

Например, событие А совершилось в результате проведенных 200 экспериментов 50 раз. Это означает, согласно методу Монте-Карло, что вероятность совершения события равна: $p_A = 50/200 = 0.25$. Вероятность того, что событие не совершится, равна, соответственно, $1 - 0.25 = 0.75$.

Обратите внимание: когда говорят о вероятности, полученной экспериментально, то ее называют частотью; слово вероятность употребляют, когда хотят подчеркнуть, что речь идет о теоретическом понятии.

При большом количестве опытов N частота появления события, полученная экспериментальным путем, стремится к значению теоретической вероятности появления события.

В блоке оценки достоверности (БОД) анализируют степень достоверности статистических экспериментальных данных, снятых с модели (принимая во внимание точность результата ε , заданную пользователем) и определяют необходимое для этого количество статистических испытаний. Если колебания значений частоты появления событий относительно теоретической вероятности меньше заданной точности, то экспериментальную частоту принимают в качестве ответа, иначе генерацию случайных входных воздействий продолжают, и процесс моделирования повторяется. При малом числе испытаний результат может оказаться недостоверным. Но чем более

испытаний, тем точнее ответ, согласно центральной предельной теореме.

Заметим, что оценивание ведут по худшей из частот. Это обеспечивает достоверный результат сразу по всем снимаемым характеристикам модели.

Пример 1. Решим простую задачу. Какова вероятность выпадения монеты орлом вверх при падении ее с высоты случайным образом?

Начнем подбрасывать монетку и фиксировать результаты каждого броска (см. табл. 1).

Таблица 1.
Результаты испытаний бросания монеты

Количество опытов N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Значение счетчика выпадения орла N_o	0	0	1	1	2	3	4
Значение счетчика выпадения решки N_p	1	2	2	3	3	3	3
Частость выпадения орла $P_o = N_o/N$	0	0	0.33	0.25	0.4	0.5	0.57
Частость выпадения решки $P_p = N_p/N$	1	1	0.66	0.75	0.6	0.5	0.43

Будем подсчитывать частость выпадения орла как отношение количества случаев выпадения орла к общему числу наблюдений. Посмотрите в табл. 1. Случаи для $N=1$, $N=2$, $N=3$ — сначала значения частости нельзя назвать достоверными. Попробуем построить график зависимости P_o от N — и посмотрим, как меняется частость выпадения орла в зависимости от количества проведенных опытов.

Разумеется, при различных экспериментах будут получаться разные таблицы и, следовательно, разные графики. На рис. 4 показан один из вариантов.

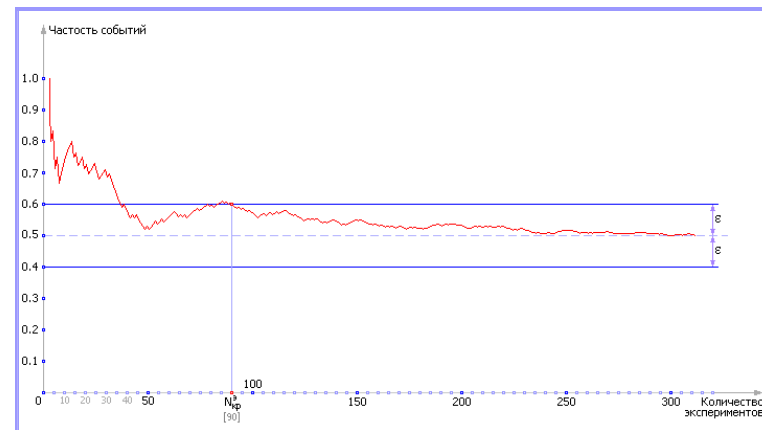


Рис. 4. Экспериментальная зависимость частоты появления случайного события от количества наблюдений и ее стремление к теоретической вероятности

Сделаем некоторые выводы.

1. Видно, что при малых значениях N , например, $N=1$, $N=2$, $N=3$ ответу вообще доверять нельзя. Например, $P_o = 0$ при $N=1$, то есть вероятность выпадения орла при одном броске равна нулю! Хотя всем хорошо известно, что это не так. То есть пока мы получили очень грубый ответ. Однако, посмотрите на график: в процессе **накопления** информации ответ медленно, но верно приближается к правильному (он выделен пунктирной линией). К счастью, в данном конкретном случае правильный ответ нам известен: в идеале, вероятность выпадения орла равна 0.5 (в других, более сложных задачах, ответ нам, конечно, будет неизвестен). Допустим, что ответ нам надо знать с точностью $\epsilon = 0.1$. Проведем две параллельные линии, отстоящие от правильного ответа 0.5 на расстояние 0.1 (см. рис. 4). Ширина образовавшегося коридора будет равна 0.2. Как только кривая $P_o(N)$ войдет в этот коридор так, что уже никогда его не покинет, можно остановиться и посмотреть, для какого

значения N это произошло. Это и есть **экспериментально вычисленное критическое значение** необходимого количества опытов $N_{кр}^3$ для определения ответа с точностью $\varepsilon = 0.1$; ε -окрестность в наших рассуждениях играет роль своеобразной трубки точности. Заметьте, что ответы $P_o(91)$, $P_o(92)$ и так далее уже не меняют сильно своих значений (см. рис. 4); по крайней мере, у них не изменяется первая цифра после запятой, которой мы обязаны доверять по условиям задачи.

- Причиной такого поведения кривой является действие **центральной предельной теоремы**. Пока здесь мы сформулируем ее в самом простом варианте «Сумма случайных величин есть величина неслучайная». Мы использовали среднюю величину P_o , которая несет в себе информацию о сумме опытов, и поэтому постепенно эта величина становится все более достоверной.
- Если проделать еще раз этот опыт сначала, то, конечно, его результатом будет другой вид случайной кривой. И ответ будет другим, хотя примерно таким же. Проведем целую серию таких экспериментов (см. рис. 5). Такая серия называется ансамблем реализаций. Какому же ответу в итоге следует верить? Ведь они, хоть и являются близкими, все же разнятся. На практике поступают по-разному. Первый вариант — вычислить среднее значение ответов за несколько реализаций (см. табл. 2).

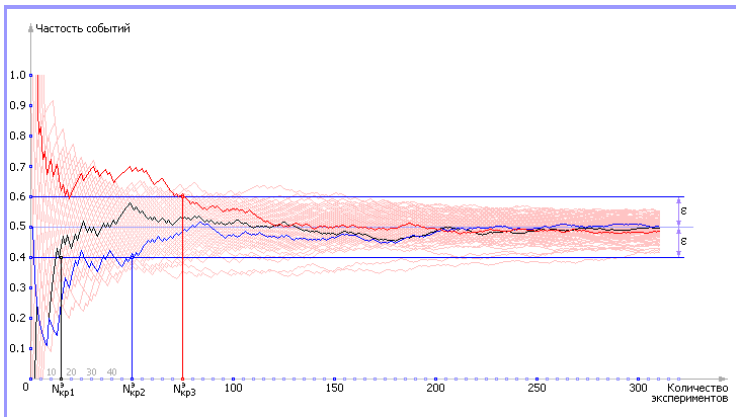


Рис. 5. Экспериментально снятый ансамбль случайных зависимостей частоты появления случайного события от количества наблюдений

Мы поставили несколько экспериментов и определяли каждый раз, сколько необходимо было сделать опытов, то есть $N_{кр}^3$. Было проделано 10 экспериментов, результаты которых были сведены в табл. 2. По результатам 10-ти экспериментов было вычислено среднее значение $N_{кр}^3$.

Таблица 2.

Экспериментальные данные необходимого количества бросков монеты для достижения точности $\varepsilon = 0.1$ при вычислении вероятности выпадения орла

Опыт	$N_{кр}^3$
1	288
2	95
3	50
4	29
5	113
6	210
7	30
8	42
9	39
10	48
Среднее $N_{кр}^3$	94

Таким образом, проведя 10 реализаций разной длины, мы определили, что достаточно в **среднем** было сделать 1 реализацию длиной в 94 броска монеты.

Еще один важный факт. Внимательно рассмотрите график на рис. 5. На нем нарисовано 100 реализаций — 100 красных линий. Отметьте на нем абсциссу $N = 94$ вертикальной чертой. Есть какой-то процент красных линий, которые не успели пересечь ε -окрестность, то есть $(P^{экссп} - \varepsilon \leq P^{теор} \leq P^{экссп} + \varepsilon)$, и войти в коридор точности до момента $N = 94$. Обратите внимание, таких линий 5. Это значит, что 95 из 100, то есть 95%, линий достоверно вошли в обозначенный интервал.

Таким образом, проведя 100 реализаций, мы добились примерно 95%-ного доверия к полученной экспериментально величине вероятности выпадения орла, определив ее с точностью 0.1. Для сравнения полученного результата вычислим теоретическое значение $N_{кр}^T$ теоретически. Однако для этого придется ввести понятие доверительной вероятности Q_F , которая показывает, насколько мы готовы верить ответу. Например, при $Q_F = 0.95$ мы готовы верить ответу в 95% случаев из 100. Формула теоретического расчета числа экспериментов имеет вид: $N_{кр}^T = k(Q_F) \cdot p \cdot (1-p)/\varepsilon^2$, где $k(Q_F)$ — коэффициент Лапласа, p — вероятность выпадения орла, ε — точность (доверительный интервал). В табл. 3 показаны значения теоретической величины количества необходимых опытов при разных Q_F (для точности $\varepsilon = 0.1$ и вероятности $p = 0.5$).

Таблица 3.

Теоретический расчет необходимого количества бросков монеты для достижения точности $\varepsilon = 0.1$ при вычислении вероятности выпадения орла

Доверительная вероятность Q_F	Коэффициент Лапласа $k(Q_F)$	Требуемое число опытов $N_{кр}^T = k(Q_F) \cdot p \cdot (1-p)/\varepsilon^2$
0.90	2.72	68
0.95	3.84	96
0.99	6.66	167

Как видите, полученная нами оценка длины реализации, равная 94 опытам очень близка к теоретической, равной 96. Некоторое несовпадение объясняется тем, что, видимо, 10 реализаций недостаточно для точного вычисления $N_{кр}^T$. Если вы решите, что вам нужен результат, которому следует доверять больше, то измените значение доверительной вероятности. Например, теория говорит нам, что если опытов будет 167, то всего 1-2 линии из ансамбля не войдут в предложенную трубку точности. Но имейте в виду, количество экспериментов с ростом точности и достоверности растет очень быстро. Второй вариант, используемый на практике — провести **одну** реализацию и **увеличить полученное для нее $N_{кр}^T$ в 2 раза**. Это считают хорошей гарантией точности ответа (см. рис. 6).

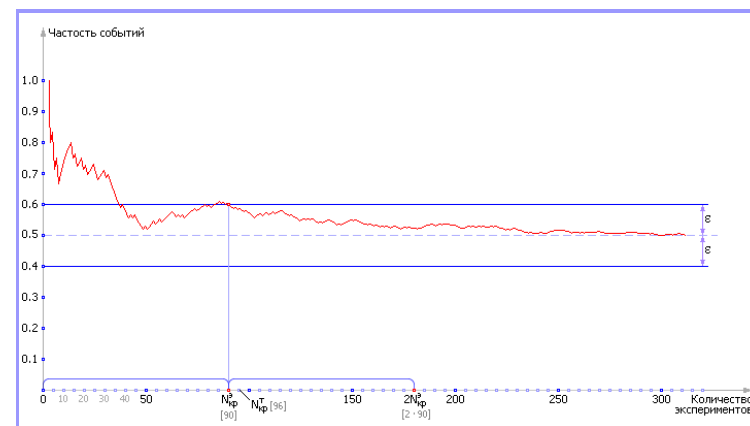


Рис. 6. Иллюстрация экспериментального определения $N_{кр}^T$ по правилу «умножь на два»

Если присмотреться к ансамблю случайных реализаций, то можно обнаружить, что сходимость частоты к значению теоретической вероятности происходит по кривой, соответствующей обратной квадратичной зависимости от числа экспериментов (см. рис. 7).

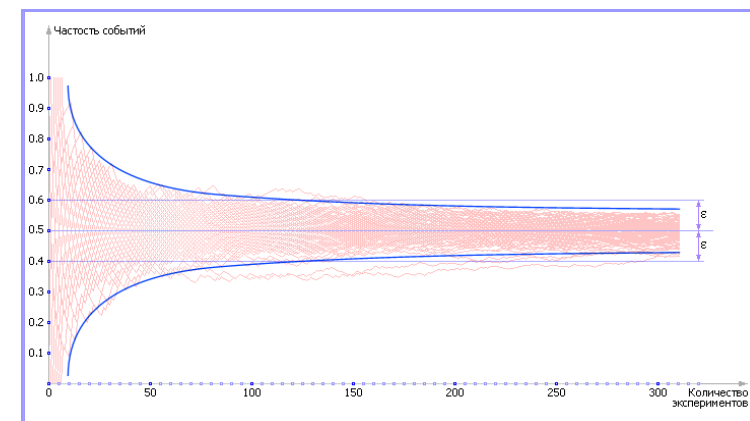


Рис. 7. Иллюстрация скорости схождения экспериментально получаемой частоты к теоретической вероятности

Это действительно так получается и теоретически. Если изменять задаваемую точность ε и исследовать количество экспериментов, требуемых для обеспечения каждой из них, то получится табл. 4.

Таблица 4.

Теоретическая зависимость количества экспериментов, необходимых для обеспечения заданной точности при $Q_F = 0.95$

Точность ε	Критическое количество экспериментов $N_{кр}^T$	число
0.1	96	
0.01	9600	
0.001	960000	

Построим по табл. 4 график зависимости $N_{кр}^T(\varepsilon)$ (см. рис. 8).

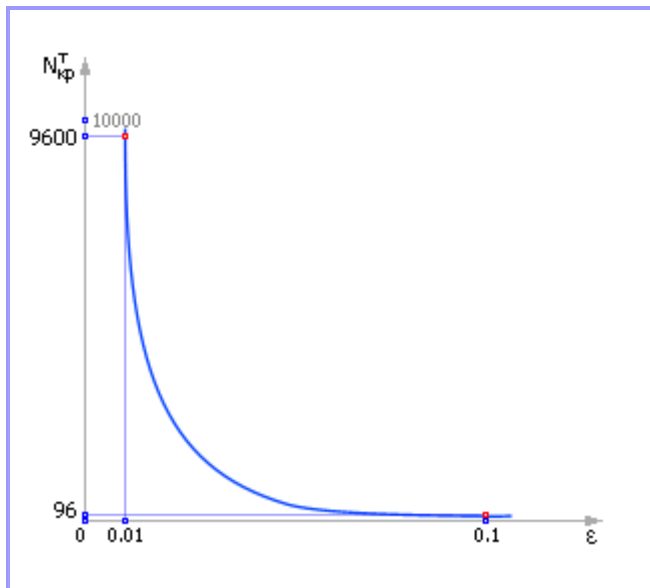


Рис. 8. Зависимость числа экспериментов, требуемых для достижения заданной точности ε при фиксированном $Q_F = 0.95$

Итак, рассмотренные графики подтверждают приведенную выше оценку:

$$\text{точность} \cong \sqrt{\text{объем выборки}}$$

Заметим, что оценок точности может быть несколько.

Пример 2. Нахождение площади фигуры методом Монте-Карло. Определите методом Монте-Карло площадь пятиугольника с координатами углов $(0, 0)$, $(0, 10)$, $(5, 20)$, $(10, 10)$, $(7, 0)$.

Нарисуем в двумерных координатах заданный пятиугольник, вписав его в прямоугольник, чья площадь, как нетрудно догадаться, составляет $(10 - 0) \cdot (20 - 0) = 200$ (см. рис. 9).

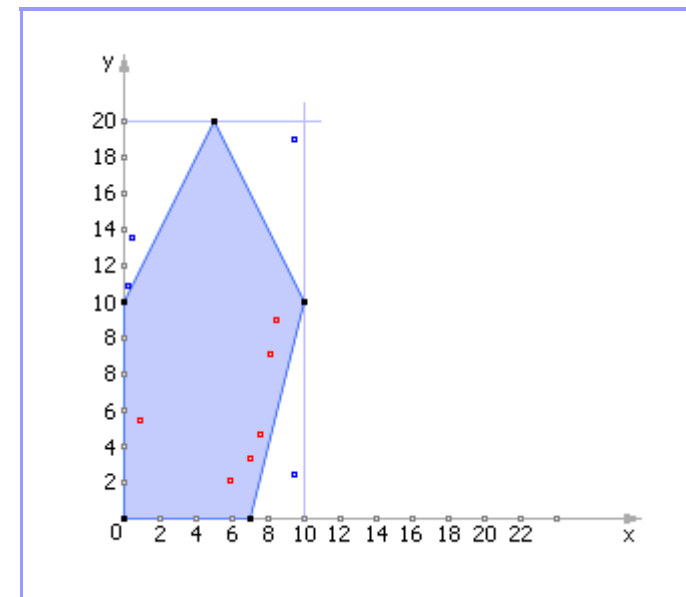


Рис. 9. Иллюстрация к решению задачи о площади фигуры методом Монте-Карло

Используем таблицу случайных чисел для генерации пар чисел R, G , равномерно распределенных в интервале от 0 до 1. Число R будет

имитировать координату X ($0 \leq X \leq 10$), следовательно, $X = 10 \cdot R$. Число G будет имитировать координату Y ($0 \leq Y \leq 20$), следовательно, $Y = 20 \cdot G$. Сгенерируем по 10 чисел R и G и отобразим 10 точек $(X; Y)$ на рис. 21.9 и в табл. 5.

Таблица 5.
Решение задачи методом Монте-Карло

Номер точки	R	G	X	Y	Точка (X; Y) попала в прямоугольник ?	Точка (X; Y) попала в пятиугольник ?
1	0.8109	0.3557	8.109	7.114	Да	Да
2	0.0333	0.5370	0.333	10.740	Да	Нет
3	0.1958	0.2748	1.958	5.496	Да	Да
4	0.6982	0.1652	6.982	3.304	Да	Да
5	0.9499	0.1090	9.499	2.180	Да	Нет
6	0.7644	0.2194	7.644	4.388	Да	Да
7	0.8395	0.4510	8.395	9.020	Да	Да
8	0.0415	0.6855	0.415	13.710	Да	Нет
9	0.5997	0.1140	5.997	2.280	Да	Да
10	0.9595	0.9595	9.595	19.190	Да	Нет
Всего:					10	6

Статистическая гипотеза заключается в том, что количество точек, попавших в контур фигуры, пропорционально площади фигуры:

$6:10 = S:200$. То есть, по формуле метода Монте-Карло, получаем, что площадь S пятиугольника равна: $200 \cdot 6/10 = 120$.

Проследим, как менялась величина S от опыта к опыту (см. табл. 6).

Оценка точности ответа

Таблица 6.

Количество испытаний N	Оценка вероятности попадания случайной точки в испытываемую область	Оценка площади S методом Монте-Карло
1	1/1 = 1.00	200
2	1/2 = 0.50	100
3	2/3 = 0.67	133
4	3/4 = 0.75	150
5	3/5 = 0.60	120
6	4/6 = 0.67	133
7	5/7 = 0.71	143
8	5/8 = 0.63	125
9	6/9 = 0.67	133
10	6/10 = 0.60	120

Поскольку в ответе все еще меняется значение второго разряда, то возможная неточность составляет пока больше 10%. Точность расчета может быть увеличена с ростом числа испытаний (см. рис. 10).

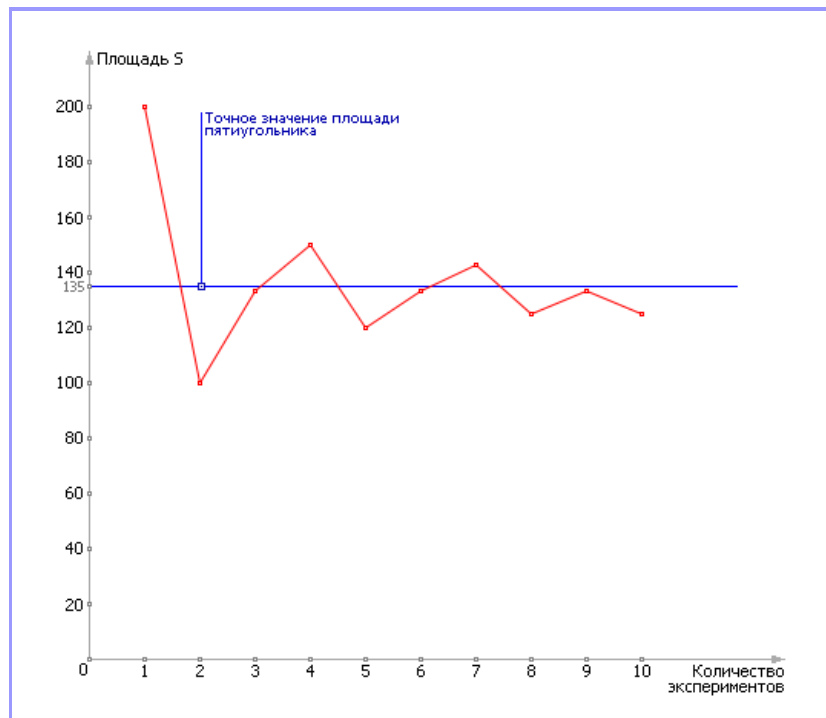


Рис. 10. Иллюстрация процесса сходимости определяемого экспериментально ответа к теоретическому результату

1.2. Планирование натуральных испытаний опытных образцов сложной системы

1.2.1. Задачи, решаемые при подготовке и проведении натуральных испытаний

Опытный образец сложной системы по своему составу обычно создается в сокращенном варианте и предназначен для проверки функционирования отдельных типовых устройств узлов и деталей системы, выполненных в соответствии с технической документацией, и оценки принятых новых конструктивных и технических решений. Натурные эксперименты на таком опытном образце используют в

целях отработки математических моделей, предназначенных для всесторонней оценки испытываемой сложной системы. Такой подход характерен именно для сложных систем и имеет существенно важное значение, так как во многом предопределяет выбор условий испытаний и планирование количества натуральных экспериментов.

При натуральных испытаниях опытных образцов сложной системы:

- 1) проверяют функционирование элементов, средств и устройств опытного образца;
- 2) проверяют принципы, заложенные в основу принятых технических решений, и правильность их конструктивного воплощения;
- 3) обрабатывают (калибруют) математические модели по результатам натуральных испытаний;
- 4) получают характеристики испытываемых средств, представляющих дополнение к составу исходных данных, которые необходимы для оценки эффективности.

Решение этих задач сводится к организации и планированию натуральных испытаний на опытном образце и представляет достаточно серьезную проблему.

Успех эксперимента во многом зависит от правильного учета множества различных факторов, например, порядка проведения работ для отработки функционирования; квалификации специалистов; эксплуатационных характеристик создаваемой аппаратуры и т. д.

Поскольку для этапа отработки на опытном образце характерна, как правило, недостаточно высокая эксплуатационная надежность аппаратуры, то большая часть времени, отведенного на организацию взаимодействия отдельных средств и устройств, тратится на устранение неисправностей. Чем сложнее система, тем большее время требуется на стыковку отдельных узлов и устройств, а значит, тем больше будет вероятность возникновения неисправностей, на устранение которых требуется дополнительное время. Это в свою очередь приводит к длительной работе аппаратуры, что опять же способствует увеличению числа появляющихся неисправностей, а следовательно, приводит к новым задержкам и увеличению времени, отводимого на испытания. Поэтому очень важно правильно определить состав средств, испытываемых на том или ином этапе, а также последовательность их наращивания. Последовательное наращивание функционально законченных устройств предопределяет количество возможных промежуточных этапов, которые могут носить характер комплексных испытаний, проводимых с целью последовательной отработки взаимодействия между элементами системы, начиная с

совместной работы нескольких функционально-законченных устройств и кончая всей системой.

При разработке планов проведения натуральных экспериментов необходимо учитывать существующие ограничения на затраты материальных средств. Поэтому планирование и выбор необходимого числа натуральных экспериментов с учетом достижения требуемой достоверности при определенных ограничениях на материальные затраты представляют собой одну из самых важных задач.

1.2.2. Постановка задачи об оптимальном плане проведения натуральных испытаний с учетом экономических факторов

Обычно задачу планирования числа экспериментов по этапам испытаний стремятся решить, руководствуясь интуитивными соображениями, на основе опыта с некоторым учетом выделенных материальных фондов. Естественно, что такие планы не могут гарантировать достижение требуемой достоверности искомых оценок (в лучшем случае можно лишь ожидать более или менее благоприятного исхода), так как они не учитывают вероятности успешной реализации эксперимента, а это означает, что планируемое число экспериментов может оказаться недостаточным.

Правильное решение задачи планирования натуральных экспериментов должно обеспечить оптимальное распределение выделенных средств по этапам испытаний с учетом вероятности успешной реализации экспериментов и дать ответ на вопрос — достаточно ли выделено материальных средств для достижения требуемой достоверности искомых оценок.

Минимальное число экспериментов для обеспечения заданной достоверности однозначно определяет требуемый минимум материальных затрат. Можно сформулировать и обратную задачу обеспечения максимальной достоверности при условии реализации заранее заданной суммы ассигнований.

Положим, что испытания некоторого опытного образца сложной системы проводятся в n этапов. На каждом i -м этапе ($i = \overline{1, n}$) необходимо получить оценку векторного параметра

$$\Theta_i = (\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(l)}, \dots, \theta_i^{(r_i)}),$$

где $l = \overline{1, r_i}$. При этом задается точность статистической оценки $D(x_i^*)$ в виде семейства областей r_i -мерного пространства D :

$$D(x_i^*) = \{x_i : (x_i - x_i^*) \in D_i\},$$

где x_i^* — оценка параметра θ_i , D_i — доверительная область.

При рассмотрении задачи ограничимся случаем, когда отклонения искомых оценок не превышают наперед заданных предельных значений:

$$D(x_i^*) = \{x_i^{(l)} : |x_i^{(l)} - x_i^{*(l)}| \leq a_i^{(l)} \quad (1 \leq l \leq r_i)\}.$$

Пусть для получения искомых оценок на каждом этапе планируют s_i независимых экспериментов, в результате проведения которых должны быть получены s_i независимых наблюдений:

$$x_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)}, \dots, x_{ij}^{(r_i)}), \quad j = \overline{1, s_i}.$$

Каждому наблюдению x_{ij} (при $j = 1, 2, \dots, s_i$) можно приписать соответствующий случайный параметр $\omega_{i1}, \omega_{i2}, \dots, \omega_{i, s_i}$, который характеризует влияние различных случайных факторов на результат наблюдения. Область возможных значений этого параметра ω_{ij} очевидно, будет конечной: $0, 1, 2, \dots, t$. Например, можно принять, что $\omega_{ij} = 0$, если эксперимент в силу влияния различных случайных факторов оказался неудачным, и $\omega_{ij} = 1$, — если эксперимент удален.

Положим, что вероятность $P(\omega_{ij})$ для всех $j = 1, 2, \dots, s_i$ задана и все этапы испытаний и соответствующие значения ω_{ij} независимы. Если в основу дальнейших рассмотрений положить указанную модель, то сформулированная ранее задача может быть изложена следующим образом.

Требуется определить число экспериментов на каждом этапе, т. е. значения s_1, s_2, \dots, s_n таким образом, чтобы обеспечить выполнение условия

$$\sum_{i=1}^n c_i s_i \leq C \quad (1)$$

и максимизировать вероятность

$$\prod_{i=1}^n P\{(\Theta_i^{(l)} - x_i^{*(l)}) \in D_i, (i = \overline{1, n})\} = \max, \quad (2)$$

где c_i — стоимость одного эксперимента на i -м этапе; C — выделенные ассигнования на проведение испытаний опытного образца (за вычетом постоянных расходов, не зависящих от s_1, s_2, \dots, s_n);

$$x_i^{*(l)} = \sum_{j=1}^{s_i} \frac{x_{ij}^{(l)}}{[\sigma_i^{(l)}(\omega_{ij})]^2} / \sum_{j=1}^{s_i} \frac{1}{[\sigma_i^{(l)}(\omega_{ij})]^2} -$$

выборочная средняя оценка.

Найдем для данного эксперимента вероятность совместного удовлетворения всех искомых оценок $x_i^{*(l)}$, при $l = \overline{1, r_i}$ условиям требуемой достоверности и выразим ее через функцию Лапласа:

$$P \{ |\Theta_i^{(l)} - x_i^{*(l)}| \leq a_i^{(l)}; l = \overline{1, r_i} \} = \prod_{l=1}^{r_i} \Phi \left(z_i^{(l)} = a_i^{(l)} \sqrt{\rho_i^{(l)}} \right), \quad (3)$$

где

$$\Phi(z_i^{(l)}) = \sqrt{\frac{2}{\pi}} \int_0^{z_i^{(l)}} t^{-z^{l/2}} dz; \quad \rho_i^{(l)} = \frac{1}{D(x_i^{*(l)})} = \sum_{j=1}^{s_i} \frac{1}{[\sigma_i^{(l)}(\omega_{ij})]^2}.$$

Рассматривая результаты наблюдений для i -го этапа с учетом случайности параметров $\rho_i^{(l)}$ можно написать

$$P \{ (\Theta_i - x_i^*) \in D_i \} = E \left\{ \prod_{l=1}^{r_i} \Phi \left(a_i^{(l)} \sqrt{\rho_i^{(l)}} \right) \right\} = \alpha(s_i), \quad (4)$$

где $\alpha(s_i)$ — вероятность того, что при заданном числе испытаний s_i все оценки параметров находятся в требуемых пределах относительно их истинных значений.

Так как алгоритм нахождения $\alpha(s_i)$ определен, то задача выбора s_1, s_2, \dots, s_n из условий максимизации (2) с учетом ограничений (1) может быть решена с помощью метода динамического программирования.

1.2.3. Алгоритм вычисления $\alpha(s_i)$ методом Монте — Карло и алгоритм решения задачи оптимального распределения средств

По определению $\alpha(s_i)$ — это вероятность того, что при заданном числе испытаний s_i все оценки параметров, определяемых на i -м этапе, находятся в требуемых пределах относительно истинных значений.

Для расчета значений $\alpha(s_i)$ методом Монте — Карло можно предложить следующий алгоритм.

1. С помощью датчика случайных чисел, равномерно распределенных в интервале $[0, 1]$, получается случайное число, которое сравнивается с заданной вероятностью P_i успешного исхода экспе-

римента. При $\zeta \leq P_i$ принимаем $\omega_{ij} = 1$ и в дальнейших вычислениях участвует $\sigma_i^{(l)}(1) = \sigma_{ip}^{(l)}$. При $\zeta_j > P_i$ реализуется $(\omega_{ij} = 0)$ и тогда $\sigma_i^{(l)}(0) = \infty$.

2. Вычисляется

$$\rho_{ij}^{(l)} = \rho_{ij-1}^{(l)} + 1/[\sigma_i^{(l)}(\omega_{ij})]^2,$$

где i и j — соответственно номера этапа испытаний и эксперимента.

Смысл этой формулы сводится к следующему: j -й эксперимент либо уточняет информацию об измеряемом параметре, полученную по результатам $(j-1)$ -го эксперимента $(\omega_{ij} = 1)$, либо не уточняет ее $(\omega_{ij} = 0)$. В соответствии с этим к величине ρ_{ij-1} либо добавляется

второе слагаемое, либо нет.

3. Находятся значения

$$z_{ij}^{(l)} = a_i^{(l)} \sqrt{\rho_{ij}^{(l)}},$$

$$\Phi(z_{ij}^{(l)}) = \frac{1}{\sqrt{2\pi}} \int_{-z_{ij}^{(l)}}^{z_{ij}^{(l)}} e^{-z^{l/2}} dz, \quad l = \overline{1, r_i},$$

где $a_i^{(l)}$ — заданные величины.

По окончании вычислений $\Phi(z_{ij}^{(l)})$ для всех $l = \overline{1, r_i}$ подсчитывается величина

$$y_{ij} = \prod_{l=1}^{r_i} \Phi \left(a_i^{(l)} \sqrt{\rho_{ij}^{(l)}} \right).$$

4. Рассчитывается выборка

$$y_{i1}; y_{i2}; \dots; y_{is_i}.$$

5. $\alpha(s_i)$ получится в виде среднего арифметического:

$$\alpha(s_i) = \frac{y_{i1} + y_{i2} + \dots + y_{is_i}}{s_i}.$$

Блок-схема программы для вычисления $\alpha(s_i)$ приведена на рис. 1.

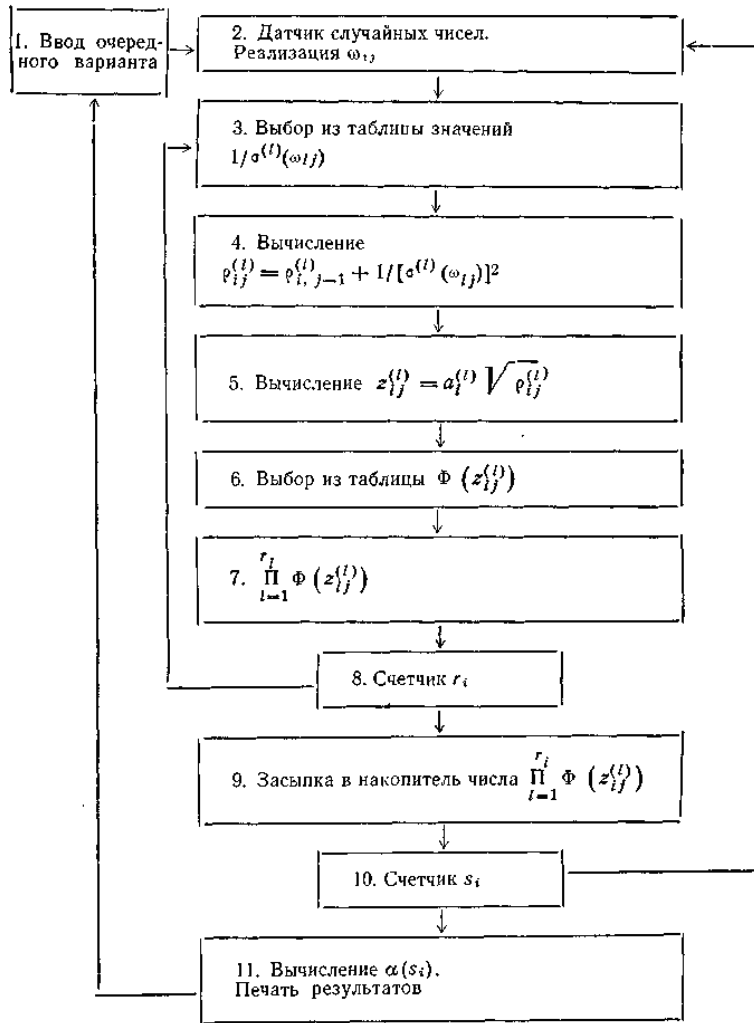


Рис. 1

Рассмотрим алгоритм решения задачи оптимального распределения средств. Эта задача заключается в том, чтобы найти

$$\max_{s_1, \dots, s_n} \prod_{i=1}^n \alpha(s_i)$$

с учетом ограничений

$$\sum_{i=1}^n s_i c_i \leq x = C,$$

где n — число этапов испытаний.

Вместо решения одной задачи с данным количеством ресурсов $x = C$ и фиксированным числом этапов рассмотрим целое семейство таких задач, в которых x может принимать любые положительные значения, а n и s_i — любые целые значения. Пусть

$$g = \prod_{i=1}^n \alpha_i(s_i).$$

Построим таблицу значений

$$g_k^*(x) = \max_{s_k} [\alpha_k(s_k) g_{k-1}^*(x - c_k s_k)], \quad (1)$$

где x принимает значения от 0 до заданной величины C с шагом h , и таблицу значений $s_k(x)$, соответствующих $g_k^*(x)$.

Вычисление можно провести по следующей схеме. Весь ресурс x назначается сначала на один этап и решается следующая задача: найти $\max g_1(x)$ при условии $s_1 c_1 \leq x$ (s_1 может принимать значения от 0 до $\min\{x/c_1, S_1\}$), где S_1 — допустимое количество экспериментов с учетом ограничений.

Составляем табл. 1, где найденный максимум $g_1^*(x)$ записываем в один столбец, а в следующем столбце записываем значение переменной, которое обеспечивает этот максимум.

Таблица 1

x	$g_1^*(x)$	$s_1(x)$	$g_2^*(x)$	$s_2(x)$
0	$g_1^*(0)$	$s_1(0)$	$g_2^*(0)$	$s_2(0)$
h	$g_1^*(h)$	$s_1(h)$	$g_2^*(h)$	$s_2(h)$
$2h$	$g_1^*(2h)$	$s_1(2h)$	$g_2^*(2h)$	$s_2(2h)$
...
kh	$g_1^*(kh)$	$s_1(kh)$	$g_2^*(kh)$	$s_2(kh)$
...
C	$g_1^*(C)$	$s_1(C)$	$g_2^*(C)$	$s_2(C)$

Данную процедуру следует повторить для всех значений x от $x=0$ до $x = C$ с шагом h . Далее число этапов увеличивается на единицу и распределение имеющегося ресурса уже проводится на два этапа. Расчетная формула принимает вид:

$$g_2^*(x) = \max_{s_2} [\alpha_2(s_2) g_1^*(x - c_2 s_2)], \quad (2)$$

где $s_2 = 0 \dots S_2$; $c_2 s_2 \leq x$.

Значения $g_1^*(x - c_2 s_2)$ выбираются из первого столбца таблицы (для нетабличных значений можно произвести линейную интерполяцию). В результате решения (2) заполняются еще два столбца.

Затем вновь увеличиваем число этапов на единицу, заполняем следующие два столбца таблицы и т. д. (до конца, по числу этапов). Расчет ведется по рекуррентному соотношению (3.3.1).

В целях экономии объема памяти ЭВМ и упрощения алгоритма расчета вновь полученные результаты записывают на носитель и печатают для визуального контроля, т. е. всей таблицы целиком в оперативной памяти не будет, так как из нее на каждом следующем этапе вычислений нужен только столбец значений $g_{k-1}^*(x)$.

После заполнения всей таблицы задача оптимального выбора количества испытаний на каждом этапе при заданных значениях C и n осуществляется в порядке, обратном построению таблицы.

Для получения большей точности результатов можно уменьшать шаг таблицы. Это, естественно, приведет к перераспределению памяти ЭВМ и увеличению времени счета.

1.2.4. Методические указания по подготовке исходных данных и решению задачи оптимального распределения средств

Предположим, что необходимо найти оптимальное распределение средств по трем этапам испытаний, осуществляемых последовательно. На каждом этапе испытаний по результатам соответствующих измерений определяют заново или же уточняют некоторые параметры.

Пусть на первом этапе предполагается определить параметры $\Theta_1^{(l)}$ (при $l = \overline{1, r_1}$), на втором этапе — Θ_2 (при $l = \overline{1, r_2}$), на третьем этапе — $\Theta_3^{(l)}$ (при $l = \overline{1, r_3}$). Значение каждого параметра нужно оценить с некоторой наперед заданной (при

точностью $\alpha_i^{(l)} = \sigma_i^{(l)} \text{доп}$ (при $l=1,2,3$).

Назначая допустимые точности, естественно исходить из заданных требований по техническому заданию, согласованному между заказчиком и исполнителем. Однако на практике в технических заданиях зачастую оговаривают лишь максимальные отклонения от номинальных значений. В подобных случаях можно принять, что заданные максимальные отклонения соответствуют утроенной величине средней квадратической ошибки.

Значения вероятностей, соответствующие успешной реализации проводимых экспериментов, и реально достижимые точности определения параметров на различных этапах получают в результате теоретических исследований или по результатам ранее проведенных испытаний аналогичных систем.

С учетом высказанных рекомендаций рассмотрим конкретный пример подготовки исходных данных и проведения расчета.

Будем считать, что на этапе I оценке подлежат шесть параметров, на этапе II — три, на этапе III — один. Пусть на основании результатов теоретических исследований и априорных сведений:

1) для каждого параметра выбрано возможное соотношение между $\sigma_{ip}^{(l)}$ и $\sigma_i^{(l)} \text{доп}$; (табл. 1).

Таблица 1

Параметры	Этап I	Этап II	Этап III
	$\sigma_p / \sigma_{\text{доп}}$	$\sigma_p / \sigma_{\text{доп}}$	$\sigma_p / \sigma_{\text{доп}}$
Θ_1^1	1,3		
Θ_1^2	1,6		
Θ_1^3	1,3		
Θ_1^4	2,0		
Θ_1^5	1,4		
Θ_1^6	1,4	$\left\{ \begin{array}{l} 1,1; 1,2 \\ 1,5; 2,0; 5,0 \end{array} \right.$	
Θ_2^7			
Θ_2^8			1,2
Θ_2^9		1,3	
Θ_3^{10}			1,6

2) в связи с недостаточной статистикой для параметра $\Theta_2^{(7)}$ назначено пять вариантов соотношений $\sigma_{ip}^{(l)} / \sigma_{\text{доп}}^{(l)}$;

3) соотношение стоимостей по этапам выбрано в следующем виде:

$$C_I = C_{III} = 1,5C_{II};$$

4) вероятности успешной реализации экспериментов приняты равными 0,6; 0,7 и 0,8 соответственно на этапах I, II и III;

5) накопление информации на различных этапах испытаний для выбранных параметров отсутствует.

Необходимо найти оптимальное распределение по трем этапам 13 комплектов испытываемого средства и 24 комплектов на этапах II и III, но для двух различных точек факторного пространства.

Результаты расчета сведем в табл. 2 и 3.

Таблица 2

Вариант	Значение $\sigma_{р/с доп}$ для θ_2^7	Количество экспериментов			Достоверность оценки		
		этап I	этап II	этап III	этап I	этап II	этап III
1	1,1	7	4	2	0,32	0,56	0,55
2	1,2	7	4	2	0,32	0,56	0,55
3	1,5	7	4	2	0,32	0,48	0,55
4	2,0	7	4	2	0,32	0,40	0,55
5	5,0	5	6	2	0,20	0,25	0,55

Как следует из табл. 2, распределение комплектов будет оптимальным, если на первый этап назначить семь экспериментов, на второй — четыре и на третий — два. Вследствие явно малого количества выделенных комплектов испытываемого средства достоверность оценок, которые могут быть получены по результатам экспериментов, крайне недостаточна и колеблется в пределах 0,3-0,6. Влияние неточности исходных данных на распределение количества комплектов по этапам, судя по результатам расчета, для первых четырех вариантов заметно не проявилось (можно лишь отметить некоторое снижение достоверности оценки искомых параметров). Даже при значительной, умышленно введенной, грубой ошибке в пятом варианте распределение экспериментов по этапам изменилось незначительно. Эти результаты дают основание для заключения о практической приемлемости полученных рекомендаций, несмотря на некоторые неточности, допущенные из-за отсутствия необходимых материалов при выборе исходных данных.

Из табл. 3 следует, что распределение экспериментов для первых четырех вариантов остается неизменным: для первой проверяемой точки на этапах II и III соответственно по восемь и пять комплектов и для второй точки — соответственно по семь и четыре комплекта.

Таблица 3

Вариант	Значение $\sigma_{р/с доп}$ для θ_2^7	Количество экспериментов				Достоверность оценки			
		1-я точка		2-я точка		1-я точка		2-я точка	
		этап II	этап III	этап II	этап III	этап II	этап III	этап II	этап III
1	1,1	8	5	7	4	0,84	0,78	0,79	0,73
2	1,2	8	5	7	4	0,83	0,78	0,78	0,73
3	1,5	8	5	7	4	0,77	0,78	0,71	0,73
4	2,0	8	5	7	4	0,69	0,78	0,62	0,73
5	5,0	9	4	8	3	0,64	0,73	0,61	0,66

Достоверность оценки параметров при этом может быть обеспечена 0,6-0,8.

Результаты расчета при решении обратной задачи — задачи о необходимом количестве экспериментов для достижения заданной достоверности, приведены в табл. 4 (при тех же исходных данных для этапа II).

Таблица 4

Варианты	Необходимое количество экспериментов	
	$\alpha(s_i) = 0,95$	$\alpha(s_i) = 0,90$
1	13	10
2	13	10
3	15	12
4	21	17
5	132	98

Если параметры могут уточняться на последующих этапах, то достоверность оценок будет несколько выше по сравнению с той достоверностью, что получена в расчетах.

1.3. Разработка математических моделей для опытно-теоретической оценки показателей сложных систем

1.3.1. Общие принципы построения моделей

Разработка математических моделей сложных систем связана с проведением очень трудоемких и разнообразных исследований. Обычно эти исследования начинают на этапе проектирования и заканчивают либо в процессе эксплуатации, либо после экспериментальных работ, организованных на средствах системы. На этапе проектирования в первую очередь решают вопросы выбора методов и способов реализации математических моделей при помощи вычислительных машин. При решении этих вопросов определяющим фактором является ожидаемая сложность модели. В большинстве случаев этот фактор приводит к тому, что на этапе теоретических исследований приходится рассматривать целый комплекс вопросов, связанных с поисками наиболее точных и в то же время достаточно простых в программном исполнении форм математического описания процессов в исследуемых системах управления. На практике при изучении указанных вопросов обычно предполагается, что процессы в элементах исследуемой системы могут быть описаны с помощью решений дифференциальных, разностных или функциональных уравнений.

Среди названных способов наиболее **общей формой описания процесса функционирования элементов сложных систем является непосредственное представление выходных характеристик с помощью функционалов, определенных на некотором известном множестве входных функций.**

При использовании такого явного описания зависимости между входом $x(t)$ и выходом $z(t)$ характеристика широкого класса нелинейных элементов может быть выполнена с помощью **функциональных рядов Вольтерра**:

$$z(t) = \sum_{j=0}^{\infty} \int \dots \int K_j(\tau_1, \tau_2, \dots, \tau_j) x(t-\tau_1), \dots, \times \\ \times x(t-\tau_j) d\tau_1, \dots, d\tau_j. \quad (1)$$

В области определения каждого ядра $K_j(\tau_1, \dots, \tau_j)$ его ортогональное разложение по собственным функциям $\Psi_{m_j}(\tau_j)$ обычно записывают в виде:

$$K_j(\tau_1, \dots, \tau_j) = \lim_{\substack{M_1 \rightarrow \infty \\ \dots \\ M_j \rightarrow \infty}} \sum_{m_1=0}^{M_1} \dots \sum_{m_j=0}^{M_j} c_{m_1 \dots m_j} \Psi_{m_1}(\tau_1) \dots \Psi_{m_j}(\tau_j). \quad (2)$$

Однако при достаточно больших значениях j и M_j реализация ядер Вольтерра высокого порядка очень трудоемка. Поэтому при построении моделей элементов с использованием подобных функциональных рядов нужно: 1) знать перспективы определения ядер по результатам натуральных экспериментов; 2) найти такую конечную систему ортонормированных функций для каждого ядра $K_j(\tau_1, \dots, \tau_j)$, чтобы достичь требуемой точности аппроксимации каждого оператора преобразования и в то же время стремиться к возможной простоте описаний, ибо от этого в значительной степени зависит время последующего моделирования.

Выбор типа разложения ядер $K_j(\tau_1, \dots, \tau_j)$ заключается в том, что наилучшая аппроксимация состоит из первых упорядоченных собственных значений и им соответствующих собственных функций. Однако каких-либо общих рекомендаций по методам решения всей задачи в целом пока еще не имеется, за исключением того, что для гауссовых сигналов задачу определения ядер второго порядка можно свести к нахождению собственных функций некоторого интегрального уравнения. В общем случае вопросы выбора конечной системы собственных функций ядер $K_j(\tau_1, \dots, \tau_j)$, а также принципы конструирования ядер высших порядков для гауссовых и негауссовых распределений из-за трудностей решения возникающих сложных функциональных уравнений пока еще до конца не ясны.

Многие реальные элементы удобнее, а иногда просто и необходимо описывать в дискретной области с помощью дискретных функциональных рядов Вольтерра. Для таких систем ядра K_1, K_2, \dots, K_{N_0} обычно аппроксимируют конечной суммой:

$$K_j(m_1, m_2, \dots, m_j) = \sum_{i=1}^N c_{ji} \varphi_i[m_1, \dots, m_j]. \quad (3)$$

При использовании такого описания наиболее трудоемкой и наименее изученной операцией является задача определения для каждого элемента совокупности линейно независимых функций φ_i , где $i=1, 2, \dots, N$ (первый член разложения ряда Вольтерра позволяет описывать процессы, происходящие в линейных элементах).

Часто для описания процессов используют системы дифференциальных или разностных уравнений:

$$dz/dt = f(z(t), x(t)), t \in [0, T], \quad (4)$$

$$z[n] = g(z[n-1], x[n-1]), n = 0, 1, \dots \quad (5)$$

Для определения параметров, входящих в уравнения (1)-(5), в процессе разработки системы проводят экспериментальные исследования на элементах, а если удастся, то и на средствах всей системы в условиях ее нормального функционирования. Исследования отличаются друг от друга охватом реальных средств системы и задачами, которые могут быть решены при их проведении, но несмотря на это, в результате получают информацию, в определенной степени характеризующую свойства всей сложной реальной системы. Чтобы объединить полученную информацию и тем самым подготовить условия для определения оцениваемых показателей с максимальной точностью, на практике стремятся работать сложную систему на такую совокупность подсистем, которая бы наилучшим образом отображала работу и функциональное взаимодействие всех ее элементов, участвующих при постановке того или иного вида физического эксперимента. Структурное объединение математических описаний этих подсистем с теми подсистемами, которые по каким-либо причинам не исследовались в ходе физических экспериментов, но определяют процессы принятия решений в ходе выполнения системой своего целевого назначения, и составляет моделирующий алгоритм системы.

Например, если каждый элемент системы, включая и модель взаимодействия подсистем, описывается с помощью линейных разностных или дифференциальных уравнений, то при составлении модели всей системы применим **метод типовых звеньев**. Суть этого метода состоит в том, что, используя описание каждого элемента, на основании формальных правил, которые соответствуют некоторым типовым соединениям (параллельное, последовательное и т. д.), **определяют на основании законов операционного исчисления передаточную функцию всей замкнутой системы**.

Обычно такой метод построения моделей сложных систем используют тогда, когда удается в достаточно малой области, чаще всего около установившегося режима работы системы, применить методы линеаризации и описать каждый элемент системы линейными разностными или дифференциальными уравнениями. Однако такой способ создания динамических систем в некоторой степени условен, хотя и очень широко применяется при изучении процессов в сложных системах.

Более точным является метод описания, основанный на непосредственном использовании тех нелинейных дифференциальных или разностных уравнений, которые на

основании теоретических исследований являются наиболее адекватным описанием свойств каждого реального элемента системы. При таком описании систем не возникает ошибок из-за метода линеаризации, а если и усложняется программная реализуемость моделей, то она окупается уверенностью в том, что из-за принятого описания не возникает ошибок в определении выходных характеристик всей сложной системы. Но рассмотренные способы описания реальных элементов, а тем более способы образования моделей сложных систем, не охватывают многих практически важных случаев. Данный вывод является следствием того, что обычно сложная система состоит из очень большого количества разнотипных элементов и подсистем, которые в процессе функционирования выполняют различные функции. По этим причинам при разработке моделирующего алгоритма сложной системы приходится пользоваться более сложными математическими конструкциями: для описания процессов функционирования и взаимодействия подсистем, наряду с функциональными операторами, использовать логические операторы; для имитации случайных процессов и последовательностей разрабатывать различного рода датчики случайных чисел; при анализе многоканальных систем привлекать схемы и методы описания, которые относятся к теории массового обслуживания и т. д.

Несмотря на такое значительное многообразие способов описания реальных процессов, практически все реальные системы могут быть описаны математическими схемами, предложенными Н. П. Бусленко и образующими класс агрегатированных систем (подсистемами подобных систем являются агрегаты). Процессы преобразования входной информации в агрегатированных системах осуществляются с учетом текущего состояния каждого агрегата. В агрегатах формирование выходных сигналов происходит в соответствии с некоторым заданным алгоритмом, который учитывает не только вероятностную природу функционирования элементов агрегата, но и реально существующие обратные связи. **Частными случаями агрегатированных систем являются системы: динамические, массового обслуживания, кусочно-линейные**, введенные в практику И. Н. Коваленко и т. д.

В общем случае при использовании того или иного способа описания реальных подсистем моделирующий алгоритм может быть записан с помощью операторных уравнений вида

$$z_i(t) = H_i \{t, t_0, z_i(t_0), (t, x_{L_i})_{t_0}^t\}, \quad (6)$$

где z_i — текущее состояние i -й подсистемы в момент t ; $z_i(t_0)$ — начальное состояние i -й подсистемы в момент начала ее функцио-

рования t_0 ; $\mathbf{x}_{L_i}(t)$ — вектор-функция, определяющая входной процесс i -й подсистемы; $(t, \mathbf{x}_{L_i})_{t_0}^t$ — входное сообщение для i -й подсистемы (входное сообщение определяется совокупностью упорядоченных пар (t, \mathbf{x}_{L_i}) для всех $t \in T_i$, где T_i — множество моментов времени, в которых рассматривается функционирование i -й подсистемы). В каждой j -й реализации на модели i -й подсистемы вектор-функцию $\mathbf{x}_{L_i}(t)$ выбирают из некоторого известного множества функций $\mathbf{L}_i(t)$, $t \in T_i$.

Для различных подсистем функциональные зависимости (6) будут получаться, естественно, отличными друг от друга. Совокупность всех функций $\mathbf{x}_{L_i}(t)$ в пространстве их определения $\mathbf{L}_1 \times \mathbf{L}_2 \times \dots \times \mathbf{L}_n = \mathbf{L}$ можно рассматривать как множество входных воздействий для модели сложной системы. При таком подходе модель как математический эквивалент реальной системы по некоторому показателю качества ее работоспособности $\mathbf{R}(t|\mathbf{c}) =$

$= E\{\mathbf{R}(t, \mathbf{c})$ может быть охарактеризована при фиксированном входном сообщении $(t, \mathbf{x}_{L_i})_{t_0}^t$ выражением

$$\mathbf{R}(t, \mathbf{c}) = \mathbf{H}^* \{t, t_0, \mathbf{z}(t_0), (t, \mathbf{x}_{L_i})_{t_0}^t\}; \quad (7)$$

$$\mathbf{x}_{L_i}(t) = \{\mathbf{x}_{L_1}(t), \dots, \mathbf{x}_{L_n}(t)\}. \quad (8)$$

Процессы смены состояний в такой системе описываются соотношениями

$$\mathbf{z}(t, \mathbf{c}) = \mathbf{H} \{t, t_0, \mathbf{z}(t_0), (t, \mathbf{x}_{L_i})_{t_0}^t\}; \quad (9)$$

$$\mathbf{z}(t, \mathbf{c}) = \{\mathbf{z}_1(t), \dots, \mathbf{z}_n(t)\}, \mathbf{z}(t, \mathbf{c}) \in Z,$$

где \mathbf{H} — оператор функционирования сложной системы, определяющий алгоритм взаимодействия ее подсистем.

Разработка алгоритма математической модели системы на этих принципах позволяет создать программу на ЭВМ, состоящую из субблоков, которые можно при необходимости заменить или скорректировать по результатам физических экспериментов другими более точными аналогами. При этом модель взаимодействия подсистем, которая обычно гораздо сложнее моделей элементов, остается без изменений, если в системе не нарушено функциональное взаимодействие ее реальных элементов.

1.3.2. Характеристика методов математического моделирования

Опытнo-теоретический метод оценки показателей эффективности сложных систем основан на использовании результатов моделирования и потому предусматривает создание математических моделей, которые должны с достаточной точностью описывать изучаемые процессы и явления. Для сложных систем разрабатываемые модели обычно получаются чрезвычайно громоздкими. В значительной степени сложность создаваемых моделей зависит и определяется математическим обеспечением моделей, разрабатываемым для обработки результатов моделирования. Основу алгоритмов математического обеспечения моделей составляют расчетные формулы методов, используемых при постановке экспериментов на модели и обработке всех полученных результатов. Поэтому на этапе создания моделей сложных систем стремятся выбрать метод математического моделирования, удовлетворяющий следующим требованиям:

1. Расчет оценок выходных показателей должен осуществляться с использованием достаточно простых алгоритмов обработки;
2. Определение необходимого объема моделирования из условий достижения заданной точности оценок выходных показателей должно происходить на основании простых и в то же время достаточно точных соотношений;
3. Методика организации экспериментов на модели должна быть по возможности простой и реализуемой на средствах используемой вычислительной техники.

Наиболее полно всем указанным требованиям удовлетворяет метод моделирования, основанный на методе статистических испытаний. Этот метод обладает высокой помехозащищенностью к случайным ошибкам, возможным при проведении отдельных опытов.

Расчетные формулы метода статистических испытаний базируются на основном законе теории вероятностей — законе «больших чисел». Практическое использование этого закона гарантирует при увеличении числа статистических испытаний на модели получение все более и более точных оценок (алгоритмы обработки обеспечивают при $N \rightarrow \infty$ сходимость по вероятности, т. е.

$$\mathbf{R} = \lim_{N \rightarrow \infty} \mathbf{R}^*).$$

Стремление достичь максимальной информации в результатах статистического моделирования обычно приводит к очень простой

схеме организации экспериментов на модели: для удовлетворения этому требованию эксперименты на модели осуществляют так, чтобы они были независимы по ансамблю реализаций. С практической точки зрения реализация метода статистических испытаний при оценке характеристик сложных систем связана с разработкой и включением в математическое обеспечение создаваемых моделей датчиков случайных чисел с такими законами распределения, имитация которых не предусмотрена в математическом обеспечении используемых ЭВМ. При разработке датчиков анализируют довольно широкий круг вопросов, среди которых основными являются:

1. Оценка качества генерирования случайных величин и случайных последовательностей (проверка законов распределения, определение циклов периодичности и случайности генерируемых чисел и т. д.);
2. Анализ возможности повышения быстродействия используемых датчиков.

Изучение последнего вопроса особенно важно при статистическом моделировании сложных систем.

К недостаткам метода статистических испытаний следует отнести тот факт, что этот метод для точностей, предъявляемых при практических оценках, приводит к довольно большому объему моделирования. Этот недостаток является следствием того, что метод статистических испытаний не полностью использует все априорные сведения о динамике функционирования системы и не учитывает особенностей законов распределения параметров системы. Более полно указанные свойства учитывают интерполяционный метод и метод Б. Г. Доступова, которые наиболее широко применяют при анализе точностных нелинейных автоматических систем сравнительно невысокой сложности. При определенных условиях, в частности, когда рассматривают задачи оценки первого и второго моментов выходных характеристик, а функциональная зависимость выходного показателя выражается квадратичной формой, удается достичь такого положения, при котором необходимое число интегрирований исходной системы уравнений линейно зависит от числа параметров, определяющих закон распределения входных случайных параметров и случайных начальных условий. При анализе более сложных систем, когда модель системы описывается сложными математическими конструкциями, преимущества рассматриваемых методов по сравнению с методом статистических испытаний с точки зрения выигрыша в объеме моделирования резко уменьшаются.

Например, для метода Б. Г. Доступова, если не применять никаких специальных мер, порядок роста числа реализаций, которые

необходимы для оценки математического ожидания выходного показателя, нетрудно установить из соотношения $N = C^q_{r+q}$, где q — число членов, учитываемых в разложении выходного показателя R в ряд Маклорена по величинам c_i ($i = 1, 2, \dots, r$).

При специальном задании решения вспомогательной системы уравнений зависимость N от r можно несколько ослабить и при $q = b$ получить $N = 2r^2 + 1$.

В этом случае квадратичная зависимость N от r приводит к тому, что при оценке характеристик сложных систем, для которых обычно $r > 100$, необходимое число реализаций уже получается сравнимым с тем, которое нужно при использовании метода статистических испытаний. По этой причине и в силу того, что метод статистических испытаний более прост в программном исполнении, область эффективного применения метода Б. Г. Доступова, так же как и интерполяционного, получается несколько уже, чем метод статистических испытаний, в том отношении, что они дают выигрыш в объеме моделирования на уровне подсистем, а порой и чаще всего при рассмотрении элементов реальных сложных систем.

Внимание широкого круга исследователей обращено на разработку аналитических методов анализа сложных систем

Интерес к подобным вопросам в значительной степени обусловлен тем, что аналитические методы позволяют абстрагироваться от некоторых частных и за счет этого с более общих позиций охарактеризовать изучаемые процессы и явления

Успех практического применения аналитических методов во многом зависит от обоснованности допущений, принимаемых при описании реальных систем. В этом отношении к интересным и практически важным результатам следует отнести результаты формализации процесса функционирования сложных систем с использованием кусочно-линейных систем, введенных и теоретически обоснованных И. Н. Коваленко.

Выделение такого подкласса агрегатированных систем целесообразно, так как для анализа кусочно-линейных систем можно применить методы аналитического исследования и с помощью их изучить общие закономерности в сложных системах, которые с достаточной для практики точностью могут быть описаны подобными математическими конструкциями.

Если считать, что процессы изменения переменных осуществляются в соответствии с некоторыми дифференциальными или другого рода функциональными уравнениями, то логическая структура кусочно-линейных систем дает возможность описывать более широкий класс

реальных систем. Однако в последнем случае возможности применения аналитических методов исследования резко сокращаются.

1.3.3. Ограничения на сложность математических моделей

При анализе сложных систем возможность применения математических моделей в значительной степени зависит от сложности их программной реализации и времени моделирования, необходимого для расчета искомых характеристик. Если оценку показателей качества реальной системы осуществляют по результатам статистического моделирования, то достоверность принимаемых статистических выводов определяется точностью имитации процессов в реальной системе, временем проигрыша одной случайной ситуации и тем количеством реализаций, которые нужно провести на модели. Если точность рассчитываемых оценок задана, а время моделирования ограничено рядом технических условий или соображений, разработка алгоритмов моделей, сравнительно просто реализуемых на средствах используемой вычислительной техники, приобретает важное практическое значение. Однако стремление к простоте математических описаний находится в известном противоречии с точностью имитации исследуемых процессов. Поэтому при разработке допустимых вариантов структурного описания каждого оператора модели системы нужно учитывать:

- 1) требования к точности оценок характеристик качества работоспособности или эффективности системы;
- 2) возможности практической реализации моделей на используемых ЭВМ;
- 3) ограничения на интервал времени, необходимый для получения оценок.

Указанные требования и ограничения определяют некоторую совокупность условий, которые необходимо реализовать при выборе наилучшего варианта построения математической модели системы. Обычно эти ограничения, по своему физическому смыслу характеризующие пределы изменения параметров системы, относят к ограничениям второго рода и записывают для каждого структурного описания модели системы в виде системы неравенств:

$$g_{2v}(\mathbf{c}) < 0, \quad v = 1, 2, \dots, N_2, \quad (1)$$

где g_{2v} — некоторые функции вектора параметров модели \mathbf{c} .

Кроме того, при разработке моделей сложных систем учитывают ограничения первого рода, которые выражают в виде некоторой системы равенств относительно известных функций:

$$g_{1v}(\mathbf{c}) = 0, \quad v = 1, 2, \dots, N_1. \quad (2)$$

К ограничениям первого рода относят уравнения, описывающие процессы в реальной системе, а также некоторые другие условия, которые могут быть выражены с помощью подобных соотношений.

Часто, чтобы учесть ограниченный объем информации, получаемой при проведении физических экспериментов, в системы уравнений (1), (2) вводят равенства и неравенства математических ожиданий от соответствующих функций:

$$\left. \begin{aligned} g_{1v}(\mathbf{c}) &= E_{\mathbf{y}} \{h_{1v}(\mathbf{y}, \mathbf{c})\}, \quad v = 1, 2, \dots; \\ g_{2v}(\mathbf{c}) &= E_{\mathbf{y}} \{h_{2v}(\mathbf{y}, \mathbf{c})\}, \quad v = 1, 2, \dots \end{aligned} \right\} \quad (3)$$

где \mathbf{y} — вектор случайных последовательностей или процессов, полученный при проведении физических экспериментов.

Однако при недостаточной априорной информации записать в явной форме все ограничения не удастся (такой случай характерен для сложных систем). В связи с этим многие ограничения удастся сформулировать только в виде некоторых рекомендаций, в форме словесных формулировок, а иногда и в виде общих пожеланий относительно допустимой сложности структурного описания разрабатываемой модели. Такая неопределенность значительно усложняет процедуру выбора наилучшего варианта построения математической модели.

1.3.4. Распределение требований к точности моделирования подсистем

При разработке моделей важно организовать работу так, чтобы программирование моделирующих алгоритмов подсистем велось параллельно и была уверенность в том, что точность описания процессов в подсистемах обеспечивает требуемую точность расчета выходных показателей эффективности всей сложной системы.

Чтобы удовлетворить этим требованиям на практике рассматривают целый комплекс задач, связанных с определением допустимых ошибок в имитации процессов в каждой подсистеме. Причем на начальном этапе в условиях неполной информации при постановке этих задач обычно используют очень упрощенное описание для всей сложной системы, но такое, чтобы оно достаточно полно отражало вероятностную природу функционирования реальной системы. Для

этих условий, если не вводить новых обозначений для упрощенного оператора системы и предположить, что ошибки моделирования можно выразить через суммарные ошибки задания вектора параметров \mathbf{c} , то на основании (7) п.1.3.1 для каждого фиксированного входного сообщения $\{t, \mathbf{x}_L\}_{t_0}^t$ нетрудно получить уравнение, связывающее вариации параметров модели $\delta \mathbf{c}_k$ ($k=1, 2, \dots, r$) с отклонениями критерия качества от некоторого номинального значения:

$$\delta R(t|\mathbf{c}) = E \left\{ \frac{\partial \mathbf{H}^* T}{\partial \mathbf{c}} \delta \mathbf{c} \right\} = E \left\{ \sum_{k=1}^r \frac{\partial \mathbf{H}}{\partial c_k} \delta c_k \right\}. \quad (1)$$

Если компоненты вектора состояния $\mathbf{z}(t)$ в моменты времени t_i ($i=1, 2, \dots, s$) терпят разрывы, что возможно при рассматриваемом описании систем (выход из строя элементов системы, разрывы непрерывности в моменты включения и выключения системы, в некоторых случаях процессы преобразования $\mathbf{z}(t)$ в $\mathbf{z}(t+\Delta t)$ для произвольных t и т. д.), то полная вариация $\delta R(t|\mathbf{c})$, когда оператор

\mathbf{H}^* аддитивен и однороден по t ($\mathbf{H}^*(t_s, t_0) = \sum_{i=1}^s \mathbf{H}^*(t_i, t_{i-1})$,

$t_i \in [t_s, t_0]$ при $i=1, 2, \dots, s$ — свойство дистрибутивности оператора \mathbf{H}^*) может быть записана так:

$$\begin{aligned} \delta R(t_s|\mathbf{c}) = E \left\{ \sum_{i=1}^s \left(\frac{\partial \mathbf{H}^* \{t_i, t_{i-1}, \mathbf{z}(t_{i-1}), (t, \mathbf{x}_L)\}_{t_{i-1}}^{t_i}}{\partial \mathbf{x}} \frac{\partial x_{i-1}}{\partial \mathbf{c}} + \right. \right. \\ \left. \left. + \frac{\partial \mathbf{H}^*}{\partial \mathbf{z}(t_{i-1})} \cdot \frac{\partial z(t_{i-1})}{\partial \mathbf{c}} + [\mathbf{H}^* \{t_i^-, t_{i-1}, \mathbf{z}(t_i^-), (t, \mathbf{x}_L)\}_{t_{i-1}}^{t_i^-}] - \right. \right. \\ \left. \left. - \mathbf{H}^* \{t_i^+, t_{i-1}, \mathbf{z}(t_i^+), (t, \mathbf{x}_L)\}_{t_{i-1}}^{t_i^+} \} \frac{dt_i}{d\mathbf{c}} \right\} \delta \mathbf{c}, \end{aligned} \quad (2)$$

$(t_i^- = t_i - 0, t_i^+ = t_i + 0).$

При записи (2) предполагалось, что текущие моменты процесса функционирования системы $t = t_i$ (при $i=1, 2, \dots, s$) и вектор ее состояния $\mathbf{z}(t)$ являются функциями от вектора параметров модели \mathbf{c} , а момент включения t_0 системы не зависит от \mathbf{c} . Вероятностная природа процесса функционирования рассматриваемых систем приводит к тому, что производную $\partial \mathbf{z}/\partial \mathbf{c}$ необходимо понимать как производную случайной вектор-функции $\mathbf{z}(\mathbf{c}, t)$ по вектору параметров \mathbf{c} . В дальнейшем будем определять указанную производную как производную Гаю (слабая производная):

$$\frac{d\mathbf{z}(t, \mathbf{c})}{d\mathbf{c}} = \lim_{\mu \rightarrow 0} \frac{\mathbf{z}(t, \mathbf{c} + \mu \Delta \mathbf{c}) - \mathbf{z}(t, \mathbf{c})}{\mu}, \quad (3)$$

где μ — некоторое вещественное число.

Для определения производной $dt_i/d\mathbf{c}$, входящей в (2), необходимо знать уравнение поверхности

$$F(\mathbf{z}(t), t, \mathbf{c}) = 0, \quad (4)$$

на которой в моменты времени t_i ($i=1, 2, \dots, s$) вектор состояния системы $\mathbf{z}(t)$ изменяется скачком. Используя правила дифференцирования неявной функции, нетрудно получить:

$$\frac{\partial t}{\partial c_k} = - \frac{(\partial F / \partial \mathbf{z})(\partial \mathbf{z} / \partial c_k) + \partial F / \partial c_k}{(\partial F / \partial \mathbf{z})(\partial \mathbf{z} / \partial t) + \partial F / \partial t} \Big|_{t=t_i^-}. \quad (5)$$

При практических расчетах необходимо помнить, что для большинства реальных элементов эти уравнения определяют процессы скачкообразного изменения переменных $F(t), \mathbf{z}(t)$ в моменты времени $t_i^- = t_i - 0$. Производные, входящие в (5), как и производная $\partial \mathbf{z} / \partial \mathbf{c}$ — слабые производные.

Так как ошибки моделирования являются в общем случае случайными величинами, то при разработке моделирующего алгоритма системы с учетом принятых методов его дискретной реализации целесообразно потребовать, чтобы вероятность его последующей доработки после проведения натуральных испытаний не превышала некоторой заданной величины P_Δ , т. е.

$$P\{\delta R(\mathbf{c}, t | (t, \mathbf{x}_L)\}_{t_0}^t \geq \Delta\} \leq P_\Delta, \quad (6)$$

где Δ — величина, характеризующая требуемую точность расчета показателя эффективности системы $R(t|\mathbf{c})$.

При разработке моделей реальных систем величину Δ задают на основании опыта, с учетом целевого назначения системы и тех требований, которые предъявляют технические задания на точность оцениваемых показателей качества. Но так как уравнение связи между $\delta R(t|\mathbf{c})$ и $\delta \mathbf{c}$ одно, то решений уравнения о распределении требований к точности моделирования подсистем может быть получено бесчисленное множество. Поэтому при разработке моделей сложных систем привлекают экспертов и на основании их оценок строят совокупность весовых коэффициентов $\beta_1, \beta_2, \dots, \beta_l$, определяющих алгоритм распределения Δ по каждой подсистеме:

$$\Delta_i = \beta_i \Delta, \quad i=1, 2, \dots, l. \quad (7)$$

Кроме рассмотренного выше способа, можно предложить и другие способы, среди которых нужно выделить один простой и в то же время достаточно наглядный способ деления Δ на части Δ_i . По этому способу весовые коэффициенты β_i рассчитывают как суммы относительных ошибок оценки параметров c_{ih} каждой i -й подсистемы:

$$\beta_i = \sum_{k=1}^{r_i} \left| \frac{\delta c_{ik}}{c_{ik}} \right|, \quad \sum_{i=1}^l r_i = r, \quad i=1, 2, \dots, l \quad (8)$$

Когда значения β_i найдены, расчет Δ , нетрудно осуществить по (7).

Необходимость постановки и решения подобных задач определяется тем, что при известных требованиях значительно упрощается выбор метода моделирования и способов дискретной реализации операторов, описывающих процессы в реальных элементах системы. Если модель разрабатывают для оценки векторного показателя $\mathbf{R}(i|\mathbf{c})$, то задача распределения требований к точности моделирования подсистем получается более сложной, чем в рассмотренном выше примере.

1.3.5. Разработка программ моделей

Для сложных систем программная реализация моделирующих алгоритмов обычно начинается на этапе проектирования, когда степень разработанности моделирующего алгоритма является недостаточной, чтобы полностью исключить вероятность его последующей доработки. В связи с этим на этапе анализа возможных способов построения блок-схемы программы необходимо предусматривать требования, связанные с оперативным вводом изменений в созданную программу. Для этой цели программа модели системы разбивается на ряд логически законченных модулей. На практике широко распространенным способом разбиения программы является трехуровневое разбиение на: укрупненные модули (**макромодули**), соответствующие принятому разбиению системы на подсистемы; **модули**, соответствующие функционально обособленным элементам каждой подсистемы; простые модули (**микромодули**), определяющие некоторые логически замкнутые операции, которые выполняют в соответствии с моделирующим алгоритмом для рассматриваемого элемента системы. Такое разбиение позволяет сравнительно просто реализовывать процессы доработки программ, если в этом есть необходимость после проведения испытаний на реальной системе.

Кроме указанного способа разбиения программ моделей сложных систем, существуют и другие, которые при определенных условиях могут стать доминирующими. Так, на начальной стадии критериями оценки эффективности создаваемых программ являются результаты, получаемые при решении задач:

1) определения времени окончания разработки программы математических моделей;

- 2) оценки экономических затрат, необходимых для разработки программ;
- 3) оценки необходимого количества исполнителей для разработки программ заданной сложности.

Сложность решения этих задач определяется недостаточной априорной информацией как о завершенности моделирующего алгоритма, так и о необходимых количественных оценках временных и экономических затрат, затрачиваемых программистом на кодировку и отладку программы различного размера. Кроме того, немаловажное значение имеет если не полное отсутствие методик, то во всяком случае очень приближенный способ оценки сложности разрабатываемой программы.

Предложенная шкала логической сложности программы (1 — простые программы, 2 — умеренные, 3 — трудные, 4 — сложные, 5 — очень сложные), несмотря на кажущуюся простоту и обоснованность используемых терминов, является все же субъективной, так как оставляет значительный произвол в оценке сложности конкретной программы. Но тем не менее эта шкала вносит определенный порядок в процесс оценки сложности программы и потому может быть использована при сравнительных оценках сложности двух или нескольких программ.

По своему содержанию первые две задачи являются задачами анализа, а третья задача относится к задачам синтеза.

Для постановки и решения первых двух задач необходимо иметь следующие исходные данные: N — размер программы (число команд); Q — сложность программы; S — стоимость программы; a — число команд, разрабатываемых одним программистом высшей квалификации в единицу времени; ΔS_i — затраты на программиста i -й квалификации в единицу времени; ΔS_M — стоимость эксплуатации машины в единицу времени.

Предлагается шкала квалификации программистов в зависимости от времени его работы:

Таблица 1

Занимаемая должность	Срок работы	Коэффициент производительности P_i
Ученик	0 ÷ 0,5 лет	0,30
Младший программист	0,5 ÷ 1,5 лет	0,65
Программист	1,5 ÷ 3 года	0,90
Старший программист	3 года и выше	1,00

Если все исходные данные известны, то можно привести расчетные формулы, которые будут отражать временные и экономические затраты на кодировку программы программистом i -й квалификации:

$$t_{i \text{ код}} = NQ_i(nP_i); \quad (1)$$

$$S_{i \text{ код}} = t_i \Delta S_i. \quad (2)$$

Чтобы определить суммарные временные и экономические затраты на разработку программы, нужно знать этапы ее создания и долю каждого этапа в процессе разработки программы. Для грубых оценок времени создания программы можно рекомендовать следующие результаты:

Изучение моделирующего алгоритма.....	10%
Составление блок-схемы программы.....	25%
Кодировка программы.....	25%
Проверка программы на рабочем месте.....	5%
Подготовка исходных данных для контрольных расчетов.....	10%
Отладка программы с выходом на машину.....	20%
Оформление программы.....	5%

Используя эти результаты, нетрудно для программиста i -й квалификации записать общее время и экономические затраты, которые ему необходимы для разработки программы:

$$t_i = (4 \div 6) t_{i \text{ код}}; \quad (3)$$

$$S_i = \Delta S_i t_i + \frac{1}{(4 \div 6)} t_i \Delta S_m. \quad (4)$$

Когда программа каждой подсистемы разрабатывается одним программистом, общее время и стоимость программной реализации моделирующего алгоритма сложной модели выражаются соотношениями:

$$T = \max_i t_i; \quad (5)$$

$$S = \sum_{i=1}^l S_i, \quad (6)$$

где l — число программистов, привлекаемых к разработке программы.

Если для разработки программ моделей подсистем привлекают несколько программистов, то соотношения (3)-(6) необходимо видоизменить, что нетрудно сделать в каждом конкретном случае.

При практических расчетах, соответствующих начальному этапу создания программы, можно пользоваться следующими ориентировочными данными: $n = 30 \div 40$, $\Delta S_m = 60$, $P_i = 0,9$.

Использование этих данных не противоречит результатам, приведенным в ряде работ и, как показывает практика, довольно точно отражает существо рассматриваемых задач, когда программист имеет сравнительно высокую квалификацию и разбирается в физическом смысле процессов, происходящих в данной подсистеме.

На рис. 1 приведена зависимость времени разработки T (в человеко-днях) от числа команд программы N (тысяч команд).

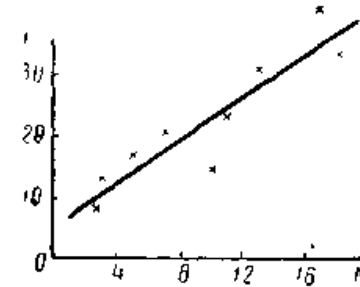


Рис. 1.

Более сложная задача возникает тогда, когда необходимо обосновать штатный состав программистов из условий минимизации стоимости разработки S программы при заданном директивном сроке ее создания. В общем случае решение этой задачи может быть найдено с помощью методов целочисленного программирования. При практических расчетах очень часто, исходя из каких-либо других соображений, удается сузить область возможных решений в том смысле, что выбор наилучшего варианта состава программистов нужно осуществить, зная, что число вариантов очень ограничено. В этих случаях проще всего для всех анализируемых вариантов рассчитать по формулам (3)-(6) значения T_j , S_j , где j - номер варианта, сравнить полученные значения и рекомендовать тот вариант, который удовлетворяет условиям рассматриваемой задачи. Подобные расчеты являются приближенными, а порой и очень грубыми оценками существа рассматриваемых задач, это можно подчеркнуть хотя бы тем, что при их проведении не учитывают такой важный фактор, как существенное повышение квалификации программиста за время создания программы модели сложной системы.

На рис. 2 дан график роста производительности начинающего программиста, где n — число команд, разрабатываемых

программистом за один день; k — число программ, разработанных программистом после обучения.

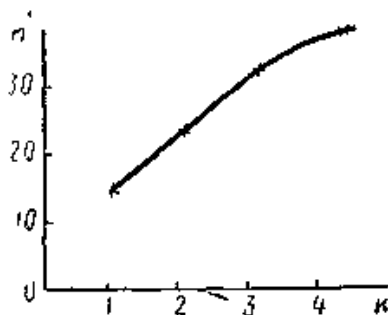


Рис. 2

Несмотря на грубость подобных расчетов, нужно отметить их пользу хотя бы потому, что они позволяют руководителю охарактеризовать объем работ и возможные сроки создания программы, а на этапе разработки программы в случае невыполнения каких-либо намеченных сроков принять обоснованные меры по ликвидации различного рода отставаний от разработанного плана графика работ.

1.4. Калибровка математических моделей по результатам натурных испытаний

1.4.1. Общие положения о калибровке моделей сложных систем

При разработке моделей сложных систем большое значение имеют результаты теоретических исследований и априорные сведения о характере и свойствах изучаемых процессов. Эффективность использования этих результатов базируется на том, что при создании реальных систем применяют технические решения и реальные элементы, ранее освоенные при проведении каких-либо экспериментальных работ. Однако при проектировании вновь создаваемых систем часто используют новые элементы и более совершенные, но не проверенные на практике, методы организации взаимодействия элементов реальной системы. Поэтому при анализе подобных систем

всегда есть некоторые сомнения в правильности функционирования исследуемой системы в рассматриваемых условиях. По этой причине на средствах системы и на всей системе до сдачи ее в эксплуатацию проводят эксперименты, основная цель которых — получить заключения, что сложность структурного описания модели и найденные законы распределения ее параметров позволяют с достаточной точностью описывать исследуемые процессы управления.

В общем случае задачу построения моделирующего алгоритма можно свести к задаче идентификации некоторого неизвестного оператора взаимосвязи двух случайных процессов в постановке, которую сделал А. Н. Колмогоров. Но в такой постановке задача получается достаточно общей и не в полной мере учитывает априорные сведения о динамике работы исследуемой системы и возможности по организации и проведению измерений промежуточных параметров и процессов. Если же учесть эти сведения, то можно отказаться от предположения, что исследуемый объект — «черный ящик», и считать его «серым» в том смысле, что **математические конструкции, описывающие процессы взаимодействия элементов в системе, известны с точностью до некоторой группы параметров.** Для сложных систем эти параметры чаще всего удается идентифицировать в каждой единичной реализации по результатам непосредственных или косвенных измерений, а законы их распределения по ансамблю реализаций построить по результатам серии натурных испытаний. Чтобы получить необходимый статистический материал, в процессе проведения экспериментов осуществляют телеметрические, радиотехнические и оптические измерения измерительными устройствами, которые в большинстве случаев фиксируют скалярные переменные в заданные моменты времени с ошибками (статистические свойства ошибок считаются известными).

Наиболее типичные условия, встречающиеся на практике, — это условия, когда процесс измерения в каждом единичном эксперименте можно описать уравнениями вида:

$$x_u[n] = G(x[n]); \quad (1)$$

$$y[n] = x_u[n] + \xi[n], \quad (2)$$

где G — известный, в общем случае, нелинейный оператор преобразования измерительного устройства; $x[n]$, $x_u[n]$ — соответственно сигналы на входе и выходе измерительного устройства; $y[n]$ — измеренные значения; $\xi[n]$ — помеха с нулевым средним значением и конечным вторым моментом.

Для обработки измеренных значений $y[n]$ с целью определения некоторых параметров их распределения на практике разрабатывают частные методики обработки, при этом основными задачами, подлежащими решению, являются:

- 1) определение интервалов стационарности наблюдаемых процессов;
- 2) выбор критериев оптимальности, на основании которых хотят реализовать оценку неизвестных параметров;
- 3) оценка возможностей получения и эффективного использования априорной информации о характеристиках анализируемых процессов;
- 4) разработка конкретных алгоритмов обработки и оценки сложности их программной реализации на ЭВМ.

Успешное решение последней задачи особенно важно при обработке результатов испытаний сложных систем, так как при проведении натуральных испытаний необходимо перерабатывать большое количество информации, что возможно только с помощью ЭВМ. В связи с этим при постановке экспериментов и выборе способов регистрации промежуточных данных нужно учитывать и анализировать условия программной реализуемости алгоритмов обработки и изучать возможности их введения в математическое обеспечение используемых ЭВМ.

Для удовлетворения указанных требований на практике стремятся создать методики обработки результатов наблюдений, которые по возможности охватывали бы достаточно широкий круг задач, связанных с обработкой наиболее типичных процессов управления.

По этой причине часто разрабатываемые методики предполагают, что наблюдаемый сигнал может быть зааппроксимирован следующим рядом:

$$x[n] = \sum_{v=1}^s c_v \varphi_v[n] = \mathbf{c}^T \boldsymbol{\varphi}[n], \quad (3)$$

а задача идентификации заключается в нахождении параметров c_v ($v=1, 2, \dots, s$), удовлетворяющих уравнению оптимальности

$$\rho(\mathbf{c}, \vec{y}_n | \boldsymbol{\varphi}[n]) = \min_{\mathbf{c}} \quad (4)$$

$$\vec{y}_n = (y[1], y[2], \dots, y[n]).$$

Если априорных сведений нет, то выражения для уравнений оптимизации получают на основании метода максимального правдоподобия или метода наименьших квадратов. Эти случаи характерны при решении задач идентификации параметров в системах с неполной информацией о динамике их функционирования.

При обработке результатов испытаний сложных систем наиболее часто в качестве системы функций $\varphi_v(n)$ ($v=1, 2, \dots, s$) используют разложения в ряды по ортогональным функциям (ортогональные полиномы Чебышева, разложения по тригонометрическим функциям и т. д.), что объясняется рядом свойств, полезных при обработке сигналов с плохо изученными свойствами. Так, практическое применение метода наименьших квадратов в сочетании с разложениями по ортогональным функциям позволяет сравнительно просто организовать дополнительные вычисления, если реализованная точность аппроксимации ниже требуемой. Это свойство объясняется тем, что при переходе к более точным аппроксимациям заново рассчитывают лишь коэффициенты, стоящие перед вновь введенными членами разложения.

В универсальных ЭВМ в математическое обеспечение машины обычно уже включены программы обработки на основании метода наименьших квадратов. Поэтому когда методики обработки предполагают анализ результатов наблюдений с помощью указанных методов, необходимость их включения в математическое обеспечение создаваемых моделей отпадает.

Часто при анализе сложных систем обработка результатов наблюдений на основании метода наименьших квадратов оказывается недостаточной, поскольку она не в полной мере учитывает все имеющиеся априорные сведения о законах распределения оцениваемых параметров в математических конструкциях, описывающих исследуемые процессы. В этих случаях разрабатывают дополнительные методики обработки и включают их в математическое обеспечение создаваемых моделей. Обычно при разработке этих методик используют байесову процедуру переоценки априорных распределений в апостериорные.

Общим свойством оценок, получаемых с помощью частных методик обработки, является их неоптимальность в том смысле, что при их нахождении использовалась не вся доступная информация и что они, будучи введенными в модель, не гарантируют отсутствия смещения в результатах статистического моделирования. Поэтому при рассмотрении вновь создаваемых систем такой способ определения параметров модели не дает полной уверенности в том, что математическая модель с достаточной точностью позволяет рассчитывать искомые оценки показателей эффективности при всех условиях работы реальной системы. Сказанное является следствием того, что при обработке результатов непосредственных измерений не анализируются процессы преобразования входных воздействий в выходные. Чтобы определить достаточность структурного описания

модели, необходимо получить количественные характеристики о взаимной зависимости входных и выходных процессов, получаемых на модели и зафиксированных при проведении натурных испытаний.

По своему физическому смыслу эта задача, как уже отмечалось выше, сводится к задаче, связанной с разработкой конкретных методик проверки адекватности моделей реальным объектам. При анализе реальных систем такие задачи исследования формулируются по отношению к следующей схеме обработки и использования результатов натурных испытаний (рис. 1).

Для схемы сравнения операторов модели и реальной системы, показанной на этом рисунке, анализу подлежат процессы $z_M(t)$, $z_p(t)$ (при практических оценках характеристики их сечений в некоторые определенные моменты времени) и задача сравнения, следовательно, состоит в разработке решающего правила $\bar{\Gamma}\{u_{mn} | \bar{Z}_M^{(m)}, \bar{Z}_P^{(n)}\}$. Это правило на основании выборок $\bar{Z}_M^{(m)} = Z_{M1}, Z_{M2}, \dots, Z_{Mm}$; $\bar{Z}_P^{(n)} = Z_{P1}, Z_{P2}, \dots, Z_{Pn}$ должно с учетом требований к точности искомых оценок показателей эффективности определять условия, при которых модель можно считать адекватной исследуемой системе.

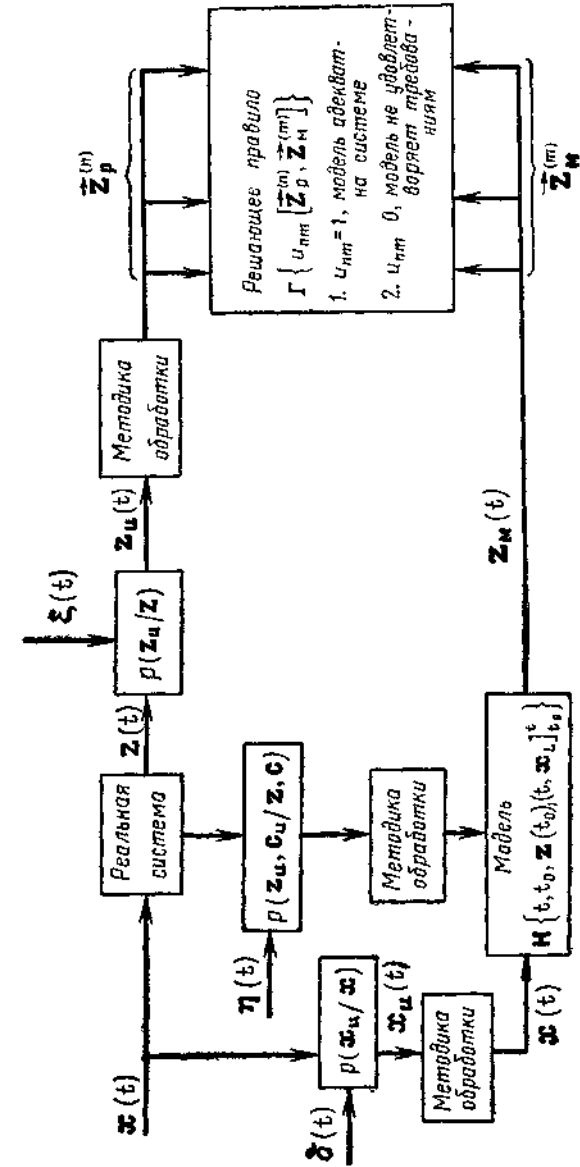


Рис. 1

Если принять, что расстояние между выборками $\vec{Z}_m^{(m)}(t)$ и $\vec{Z}_p^{(n)}(t)$ и обозначить его через $F\{\vec{Z}_m^{(m)}(t), \vec{Z}_p^{(n)}(t)\}$, то математическую формулировку рассматриваемой задачи можно дать с использованием методов теории статистических решений и синтезируемый алгоритм записать так:

1. Если $F\{\vec{Z}_m^{(m)}(t), \vec{Z}_p^{(n)}(t)\} \leq F(\Delta)$, где $F(\Delta)$ — величина, характеризующая допустимое расстояние между выборками $\vec{Z}_m^{(m)}(t)$ и $\vec{Z}_p^{(n)}(t)$, а Δ — заданная точность расчета на модели оценок выходного показателя системы, то $u_{mn}(t)=1$ и, следовательно, модель адекватна реальной системе;
2. Если справедливы соотношения

$$F\{\vec{Z}_m^{(m)}(t), \vec{Z}_p^{(n)}(t)\} > F(\Delta); u_{mn}(t)=0,$$

то модель необходимо либо доработать, либо продолжить эксперименты на системе и модели из-за недостаточности статистики.

Если проанализировать комплекс возникающих задач, то среди них можно выделить такие, без решения которых невозможно создать алгоритм принятия решений, определенный указанными выше соотношениями. К таким задачам следует отнести:

- 1) разработку методов организации и проведение измерений в процессе экспериментальных работ на системе;
- 2) разработку методов обработки наблюдаемых процессов;
- 3) выбор методов решения возникающих уравнений оптимальности;
- 4) выбор объема и условий проведения экспериментов, на основании которых можно с требуемой достоверностью получить интересующие нас статистические выводы.

Последовательность решения указанных задач для конкретных систем может быть различной. Однако только их комплексное рассмотрение может дать ответ о целесообразности последующего использования математической модели для оценки характеристик исследуемых сложных систем. По результатам подобных исследований, называемых на практике **режимом калибровки моделей**, разрабатывают паспорт, в котором указывается:

- 1) назначение и общие принципы использования математической модели;
- 2) точность и достоверность результатов моделирования по каждой оцениваемой характеристике;
- 3) объем натуральных экспериментов и критерии, на основании которых установлена адекватность модели реальной системе;

- 4) рекомендации по математическому обеспечению, необходимому для обработки результатов моделирования и натуральных испытаний.

1.4.2. Критерии и условия оптимальности

При построении математических моделей выбором критерия оптимальности стремятся добиться такой эквивалентности модели реальной системе, чтобы достичь некоторых вполне определенных свойств оценок, характеризующих показатели качества исследуемых систем. Наиболее общие требования к получаемым оценкам состоят в том, чтобы они обладали достаточной информативностью и позволяли с максимально возможной степенью точности установить выполнение системой требований, предъявленных при ее разработке.

В самом общем случае критерий оптимальности определяет степень близости оценки выбранного показателя эффективности R^* к его истинному значению R . Обычно качество принятой оценки характеризуют с помощью некоторого заданного функционала $\rho(c^*, \vec{y})$. В рассматриваемом случае оценку R^* получают в результате статического моделирования на математической модели, т. е. справедливо формальное равенство $R^* = R(t|c^*)$. Поэтому для каждого варианта построения модели функционал $\rho(c^*, \vec{y})$ характеризует величину, зависящую от оценок вектора параметров модели c^* . В свою очередь оценки вектора параметров модели функционально или статистически связаны с величинами или процессами $\vec{y} = (y_1, y_2, \dots, y_n)$, которые наблюдались при проведении физических экспериментов на элементах системы. Очевидно, что оценка вектор-параметров модели c^* является случайной величиной. Поэтому целесообразно определить правило получения оценок $c^*(\vec{y})$ из условий минимизации функционала

$$\rho\{c^*(\vec{y}), \vec{y}\} = \min_{c^*(\vec{y})} \quad (1)$$

Если учесть необходимость перебора всех допустимых структурных описаний, то задачу построения математической модели можно свести к нахождению эффективных методов решения следующих уравнений:

$$\rho\{c_v^*(\vec{y}), \vec{y}\} = \min_{c_v^*, v} \quad (2)$$

где \forall — множество всех допустимых структурных описаний модели.

При решении уравнений (1) и (2) обычно считают известными:

- 1) ограничения на сложность математической модели и на время, отведенное на весь процесс моделирования;
- 2) объем физических экспериментов, проведенных с целью нахождения оценок вектора параметров модели и ее структуры (известны выборки величин или процессов $\vec{y} = (y_1, y_2, \dots, y_n)$);
- 3) области изменения и размерность вектора параметров модели для каждого ее структурного описания;
- 4) функциональные зависимости оцениваемого показателя $R(t|\mathbf{c})$ с выходом математической модели \mathbf{c}_v в каждой единичной реализации;
- 5) функционал, характеризующий тип ошибок, выбранных для оценки точности определения истинного значения показателя $R(t|\mathbf{c})$.

Необходимые условия оптимальности параметров математической модели для каждого фиксированного ее структурного описания с учетом ограничений первого и второго рода можно записать на основании теоремы Куна-Таккера:

$$\left. \begin{aligned} \text{grad } \rho \{ \mathbf{c}^*(\vec{y}), \vec{y} \} + \lambda \mathbf{G} \{ \mathbf{c}^*(\vec{y}) \} &= 0; \\ \mathbf{g}(\mathbf{c}^*) + \delta &= 0; \lambda \geq 0; \delta \geq 0; \lambda^T \delta = 0; \\ \mathbf{G}(\mathbf{c}^*) &= \left\| \frac{\partial g_\nu(\mathbf{c}^*)}{\partial c_{\mu}^*} \right\|, \nu = 1, 2, \dots, N; \\ &\mu = 1, 2, \dots, N_1, \end{aligned} \right\} \quad (3)$$

где $\mathbf{G}(\mathbf{c}^*)$ — матрица размерности $N \times N_1$; $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{N_1})$, $\delta = (\delta_1, \delta_2, \dots, \delta_{N_1})$.

Для выпуклых функционалов $\rho \{ \mathbf{c}^*(\vec{y}), \vec{y} \}$ и функций $g_\nu(\mathbf{c}^*)$, $\nu = 1, 2, \dots, N_2$ условия (3) являются в то же время и достаточными условиями.

Чтобы определить наилучший вариант построения математической модели, нужно найти значения $\rho \{ \mathbf{c}^*_v(\vec{y}), \vec{y} \}$ для всех допустимых ее структурных описаний, сравнить их и в качестве оптимального структурного описания модели выбрать то, для которого значение $\rho \{ \mathbf{c}^*_v(\vec{y}), \vec{y} \}$ наименьшее.

Анализ математических аспектов рассматриваемых вопросов показывает, что наиболее трудоемкими при вычислении являются

задачи:

- 1) определения наиболее удобных форм математического описания операторов $\mathbf{H}_i \{ t, t_0, \mathbf{z}_i(t_0), (t, \mathbf{x}_{L_i}]_{t_0}^t \}$ с учетом условий их последующей реализации в модели системы;
- 2) нахождения зависимостей $g_{1\nu}, g_{2\nu}$ ($\nu = 1, 2, \dots$) для каждого допустимого структурного описания модели;
- 3) разработки эффективных методов поиска глобального экстремума функционала (3).

Трудности решения указанных задач определяются:

- 1) сложностью математического описания разрабатываемых моделей;
- 2) многообразием видов и способов проведения измерений при физическом экспериментировании;
- 3) частичным или полным отсутствием априорной информации о структуре и законах распределения параметров некоторых элементов системы;
- 4) значительной размерностью вектора параметров модели.

1.4.3. Методы решения уравнений оптимальности

Математические уравнения (3), определяющие условия оптимальности параметров разрабатываемой модели при фиксированном точно известном ее структурном описании, в большинстве практических случаев не могут быть записаны в явной форме, как некоторые функционалы, зависящие от вектора \mathbf{c} . Это определяется как сложностью исследуемых систем, так и трудностями, связанными с нахождением точных аналитических зависимостей, определяющих ограничения, которые необходимо учитывать при разработке моделей. В таких условиях обычно трудно на основании априорных сведений оценить дифференцируемость анализируемого функционала. В связи с этим при анализе конкретных задач приходится применять методы оптимизации, не использующие подобных свойств исследуемых функционалов.

При рассмотрении сложных систем возникает дополнительное требование, обычно предъявляемое к используемым методам оптимизации. Это требование можно сформулировать в виде условия, которое требует разработки алгоритмов, максимально использующих результаты моделирования и приводящих к различного рода итерационным схемам.

В значительной степени всем предъявляемым требованиям удовлетворяют алгоритмы, основанные на вероятностных итеративных методах оптимизации. По своему содержанию эти методы относятся к

градиентным методам с той лишь разницей, что градиент исследуемого функционала $\rho\{\mathbf{c}^*(\vec{y}), \vec{y}\}$ оценивают на основании результатов статистического моделирования по $\mathbf{R}(\mathbf{c}_1, t), \dots, \mathbf{R}(\mathbf{c}_p, t)$, полученных для некоторого заранее выбранного набора векторов $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p$. Способ выбора векторов \mathbf{c}_i ($i=1, 2, \dots, p$) определяется алгоритмом, на основании которого хотят получить оценку градиента функционала в точке $\mathbf{c} = \mathbf{c}[n-1]$, соответствующей $n-1$ -шагу итераций. В общем случае алгоритм нахождения оптимального значения вектора параметров \mathbf{c}^* можно записать в виде следующего поискового алгоритма оптимизации:

$$\left. \begin{aligned} \mathbf{c}_n &= \mathbf{c}_{n-1} - \gamma_n [\nabla \rho(\mathbf{c}_{n-1}) + \mathbf{G}(\mathbf{c}_{n-1}) \lambda]; \\ \lambda_n &= \max \{0, \lambda_{n-1} + \gamma_n \mathbf{g}(\mathbf{c}_{n-1})\}; \lambda_0 \geq \lambda \end{aligned} \right\} \quad (1)$$

где Δ_p — оценка градиента функционала $\rho\{\mathbf{c}(\vec{y}), \vec{y}\}$ на n -шаге итераций; γ_n — матрица, характеризующая алгоритм уточнения $n-1$ -приближения; $\mathbf{V}_1 n$ — матрица, определяющая влияние ограничений второго рода на способ формирования n -приближения для множителей Лагранжа $\lambda[n]$.

При реализации алгоритмов типа (1) надо оценивать:

- 1) условия и скорость сходимости итерационного процесса поиска;
- 2) методику выбора первоначального приближения;
- 3) единственность и оптимальность получаемых решений;
- 4) способы формирования оценок градиента функционала;
- 5) алгоритмы расчета элементов матриц $\mathbf{V}[\cdot], \mathbf{V}_1[\cdot]$ для каждого n такта итераций.

Если некоторые общие свойства анализируемых уравнений оптимальности известны, то указанные выше вопросы могут быть решены на основании результатов теоретических исследований.

Если априорных сведений нет, то для решения уравнений (3) нужно применять метод «проб и ошибок», что позволяет с помощью методов регрессионного анализа установить многие важные закономерности и с использованием их довести исследование уравнений оптимальности до практически приемлемых решений. Иногда процесс предварительного анализа уравнений удастся алгоритмизировать и объединить с процессом поиска экстремума исследуемого функционала. Формально такое объединение можно выполнить ведением зависимости элементов матрицы γ_n от последовательности значений векторов $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p$.

При изучении вероятностных методов поиска понятие сходимости обычно включает в себя и отражает процесс последовательного вероятностного уточнения оценок вектора \mathbf{c}^* . Для некоторых частных случаев условия сходимости подобных алгоритмов приведены в литературе.

К примеру, если градиент

$$\nabla_{\mathbf{c}} \rho\{\mathbf{c}(\vec{y}), \vec{y}\}$$

оценивать приближенным способом

$$\left. \frac{\partial \rho\{\mathbf{c}, \vec{y}\}}{\partial e_i} \right|_{\mathbf{c}} \approx \frac{\rho\{\mathbf{c} + a\mathbf{e}_i\} - \rho\{\mathbf{c} - a\mathbf{e}_i\}}{2a}, \quad (2)$$

где \mathbf{e}_i — базисные векторы

$$\mathbf{e}_1 = (1, 0, 0, \dots); \mathbf{e}_2 = (0, 1, 0, 0, \dots); \mathbf{e}_r = (0, 0, \dots, 1);$$

a — скаляр, то можно получить условия сходимости, которые будут гарантировать близость найденных оценок \mathbf{c}_n к оптимальному значению \mathbf{c}^* в среднеквадратическом смысле.

Для всякого итерационного процесса необходимым элементом является операция проверки, заключающаяся в том, что на каждом шаге итераций найденное приближение анализируется на предмет его оптимальности. Обычно эта проверка связана с разработкой некоторых правил остановки процесса поиска. Причем, исходными данными для работы подобных алгоритмов служат условия оптимальности, найденное значение вектор-параметров модели \mathbf{c}_n и требования, при котором точность полученных оценок достаточна для их практического использования.

Для рассматриваемого класса математических моделей, когда оценку показателя эффективности $\mathbf{R}(\mathbf{c}^*, t)$ рассчитывают по результатам статистического моделирования, алгоритм остановки процесса поиска можно построить с использованием методов статистической проверки гипотез. Например, если под оптимальными оценками вектора \mathbf{c}^* понимать значения, при которых результаты моделирования и натуральных испытаний получаются однородными в смысле тождественности функций их распределения, то правило остановки при заданных уровнях значимости ошибок первого и второго рода может быть сформулировано на основании известных критериев проверки гипотез (критерий Смирнова, Уилкоксона и т. д.).

Разработка подобных алгоритмов особенно важна при анализе результатов испытаний и моделирования сложных систем. Это объясняется тем, что указанным способом можно оценить возможность практического использования в математической модели оценок

вектора $\hat{\mathbf{c}}$, рассчитанных на основании методик обработки, разработанных для каждого конкретного элемента системы.

1.4.4. Особенности и вычислительная реализуемость методов

Как уже отмечалось, математические модели сложных систем строят на принципах функционального объединения подсистем, отображающих и характеризующих некоторые вполне определенные процессы принятия решений в ходе выполнения реальной системой своего целевого назначения. При создании подобных моделей общее число допустимых вариантов ее построения зависит от количества разрабатываемых структурных описаний m для каждой подсистемы и, естественно, от их общего числа l . Если при решении уравнений (3) рассматривать все варианты со строгим учетом всех ограничений (1), (2) п. 1.3.1, то число сравниваемых структурных описаний модели системы можно охарактеризовать значением m^l , например, при $l \geq 10$, $m = 3$ получаем, что для определения наилучшего варианта описания модели системы нужно проанализировать более 50000 возможных решений. Однако из этого вовсе не следует, что нужно разрабатывать такое же количество разных моделей системы. При блочном построении программ многие варианты структурного описания модели системы можно получить без радикального изменения модели взаимодействия подсистем.

Такое положение открывает определенные перспективы по анализу возможно большего числа возможных вариантов построения модели системы, но не снимает вопросов, связанных с необходимостью проведения моделирования с целью получения значений $\rho\{\mathbf{c}^*(\vec{\mathbf{y}}), \vec{\mathbf{y}}\}$ для всех рассматриваемых ее структурных описаний

Более серьезные затруднения возникают из-за высокой размерности вектора параметров модели системы. Размерность вектора параметров моделей сложных систем характеризуется несколькими сотнями, а порой и несколькими тысячами параметров. Найти минимум (1) относительно всех параметров модели, даже для одного точно известного структурного описания, практически невозможно.

Если в математической модели системы использовать оценки параметров \hat{c}_i ($i = 1, 2, \dots, r$), полученные на основании методов, развитых только для каждого отдельного средства, то оценки выходного показателя эффективности всей системы будут получаться неоптимальными и при малом объеме физических экспериментов могут

привести к неправильным выводам относительно выполнения системой заданных требований. Несмотря на это возникает вопрос: а нельзя ли все же провести статистическое моделирование при найденных указанным выше образом оценках вектора параметров модели, а потом скорректировать оценки показателей эффективности системы так, чтобы они обладали всеми свойствами, которые предопределены выбранным критерием оптимальности?

Возможности практической реализации такого подхода определяются сложностью выбранного критерия оптимальности и объемом моделирования, которое нужно провести для нахождения зависимости показателя эффективности $\mathbf{R}(t|\mathbf{c})$ как функции от вектора параметров \mathbf{c} . Например, если выбором критерия оптимальности стремятся достичь определения точечных несмещенных оценок показателя \mathbf{R}^* , то можно скорректировать результаты моделирования, полученные на математической модели при введении в нее вектора параметров $\hat{\mathbf{c}}$, рассчитанного на основании методов, развитых для каждого отдельного средства системы. В этом случае исследованию подлежат уравнения типа

$$\nabla_{\mathbf{R}} = \mathbf{E} \{ \mathbf{R}(t|\hat{\mathbf{c}}) \} - \mathbf{R}(t|\mathbf{c}). \quad (1)$$

В (1) операция математического ожидания осуществляется по распределению $p(\hat{\mathbf{c}}|\mathbf{c})$, определяющему множество всех возможных оценок $\hat{\mathbf{c}}$, которые могут быть получены в ходе физических экспериментов на элементах системы того же объема. Из-за высокой размерности вектора \mathbf{c} поиск точных решений уравнений (1) связан так же, как и при решении уравнений (3) с довольно значительными трудностями вычислительного характера. Поэтому при разработке моделей конкретных систем задачу параметрической обработки моделей по результатам физических экспериментов, связанную с корректировкой результатов моделирования, рассматривают только относительно некоторых наиболее существенных факторов:

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k), \quad k \ll r.$$

Это приводит к необходимости изучения вопросов, связанных с разработкой эффективных методов определения наиболее существенных параметров $\lambda_1, \lambda_2, \dots, \lambda_k$ и установления зависимости оцениваемых показателей $\mathbf{R}(t|\mathbf{c})$ как функций от элементов вектора $\boldsymbol{\lambda}$.

Кроме того, из-за невозможности нахождения экстремума уравнения (1) по всем элементам вектора \mathbf{c} требуется иной подход к обработке и использованию результатов, полученных при экспериментальных работах на средствах всей системы. А именно, для определения степени доверия к результатам моделирования проводят

дополнительные исследования по оценке статистической совместимости моделируемых и реальных процессов к одной и той же генеральной совокупности.

В совокупности реализация указанных выше процедур позволяет развить приближенную методику калибровки моделей сложных систем.

1.4.5. Приближенная методика калибровки моделей сложных систем

Трудности, возникающие при поиске точных решений функциональных уравнений оптимальности, в большинстве случаев приводят к необходимости разработки приближенных методов калибровки математических моделей по результатам натуральных экспериментов.

При анализе конкретных систем могут встретиться различные случаи, но рассмотрим лишь схему проведения натуральных испытаний, изображенную на рис. 1. При этом будем считать, что:

1. Оценка параметров модели c_i ($i=1, 2, \dots, r$) осуществляется по результатам непосредственных или косвенных измерений;
2. Результаты, характеризующие степень выполнения реально!¹¹ системой своего целевого назначения (выборка значений $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n$), можно наблюдать, но из-за сложности вычислений их нельзя использовать при оценке параметров модели;
3. Операторы преобразования \mathbf{G}_i ($i=1, 2, \dots, n$) и законы распределения единичных ошибок для всех используемых измерительных устройств точно известны;

4. Априорная информация о законах распределения существует только для некоторой группы параметров модели (обозначим эту границу параметров через \mathbf{c}_δ , а остальные параметры через \mathbf{c}_m);

5. Моделирование случайных значений параметров модели \mathbf{c}_δ по ансамблю реализаций осуществляется на основании апостериорных законов $p(\mathbf{c}_\delta | \bar{\mathbf{y}})$, рассчитанных по результатам реальных испытаний.

Поскольку подобные условия наиболее характерны при рассмотрении сложных систем, то методика калибровки моделей должна учитывать указанные особенности и приводить к четко определенным и реализуемым операциям, позволяющим установить степень соответствия результатов моделируемых и реальных процессов.

На практике математическая модель считается адекватной (в смысле возможностей практического использования результатов моделирования) реальной системе в том случае, когда доказано, что выборки моделируемых и реальных значений показателей эф-

фективности системы имеют одинаковые законы распределения или принадлежат к одной и той же генеральной совокупности.

Для достижения требуемой достоверности принимаемых статистических выводов должны быть определенного объема выборки результатов натуральных испытаний и моделирования. Если критерий проверки рассматриваемых статистических гипотез выбран, то объемы выборок могут быть получены на основании теоретических расчетов.

Охарактеризуем методику, наиболее полно отражающую специфику калибровки моделей сложных систем. Для рассматриваемых условий методика калибровки моделей должна включать в себя следующие операции.

Обработка результатов натуральных испытаний по единичным реализациям и по ансамблю реализаций на основании частных методик с целью получения исходных данных для моделирования (оценка параметров модели по результатам проведенных измерений). Оценку параметров, для которых имеются априорные законы распределения, осуществляют на основании байесового подхода (определяются апостериорные законы распределения названных параметров), а при отсутствии априорной информации обработку результатов наблюдений производят с использованием методов максимального правдоподобия или наименьших квадратов. В последнем случае получают совокупность оценок, характеризующих законы распределения параметров модели по ансамблю обрабатываемых реализаций.

Моделирование при исходных данных, полученных на основании частных методик обработки. Основная цель такого моделирования — получить выборку значений выходного показателя $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$ необходимого объема (объем моделирования обычно выбирают на основании результатов теоретических расчетов, но он может быть изменен в ходе проведения калибровки моделей).

Корректировка результатов статистического моделирования из условий достижения их несмещенности. Необходимость этой операции объясняется хотя бы тем, что введение в модель оценок параметров $\hat{\mathbf{c}}_m$, полученных на основании метода максимального правдоподобия, приводит к тому, что оценки выходных показателей

$\mathbf{R}(t | \hat{\mathbf{c}}_m)$ оказываются смещенными, даже если исходные параметры были несмещенными. Этот результат является следствием того, что модель осуществляет нелинейное преобразование над входными воздействиями. Оценки выходного показателя $\mathbf{R}(t | \hat{\mathbf{c}}_m)$ максимально правдоподобны, но вследствие указанной выше причины не удовлетворяют условию несмещенности оценок:

$$E \{R(t|\hat{c}_m)\} \neq R(t|c_m),$$

где $R(t|c_m)$ — истинное значение выходного показателя в момент окончания процесса функционирования t системы (выходной показатель от реализации к реализации изменяется как случайная величина с определенным законом распределения).

Если бы для всех параметров модели удалось построить апостериорные распределения, то принципиально при бесконечном числе реализаций с помощью модели можно было бы найти апостериорное распределение выходного показателя $p(R|y, t)$ и уже на основании его рассчитать оценки, которые характеризуют качество функционирования или эффективность исследуемой системы. Но это означает, что найденные таким способом оценки выходных показателей будут байесовскими и, следовательно, смещенными. Поэтому перед тем, как оценивать статистическую совместимость результатов моделирования с результатами натуральных испытаний, нужно исключить смещение в результатах моделирования, что является вполне обоснованной теоретической операцией, которая практически необходима, так как позволяет повышать точность получаемых результатов.

Проверка статистической совместимости результатов моделирования с результатами натуральных испытаний. При проверке гипотез считаются заданными уровни значимости и тип используемого критерия. При выборе критерия обычно стремятся достичь такого положения, чтобы он для рассматриваемых условий позволял реализовать максимум вероятности правильного опознавания гипотезы в том случае, когда она верна. К сожалению, этот общий принцип выбора, связанный с использованием, так называемых, наиболее мощных критериев, довольно редко удается воплотить при рассмотрении реальных задач.

Если разрабатываемая модель предназначена для оценки одного показателя, то при проверке рассматриваемых гипотез обычно используют для произвольных распределений критерий Смирнова и критерий Уилкоксона, а для нормальных распределений — критерий Стьюдента (более полная характеристика методов проверки гипотез дана ниже).

Чтобы полнее использовать результаты физических экспериментов, оценку статистической совместимости осуществляют на основании результатов, которые невозможно использовать при оценке параметров модели в рамках частных методик обработки. При анализе сложных систем решение подобных задач обычно осуществляют с

использованием результатов, характеризующих состояние реальной системы в момент окончания процесса ее функционирования. При необходимости могут привлекаться результаты, определяющие текущее состояние системы в некоторые характерные моменты времени (смена режимов работы системы, моменты выхода из строя подсистем и т. д.).

Принятие решения о пригодности модели к практическому использованию. Такое решение обычно принимают по результатам, получаемым при проверке статистических гипотез о тождественности распределения моделируемых и реальных выборок. В случае, когда проверка гипотез дает положительный результат, модель считается в принятом смысле адекватной реальной системе и на ней можно проводить моделирование с целью определения показателей, характеризующих эффективность исследуемой системы. Если же результат проверки отрицателен, то имеют место два случая.

Первый случай характерен тем, что рассчитанный уровень значимости превышает заданный уровень незначительно и исследователи убеждены в достаточной точности модели и считают полученный результат случайным, либо из-за недостаточности информации, полученной в ходе натуральных испытаний, либо из-за небольшого объема, реализованного при моделировании. В этом случае продолжают экспериментирование на системе и модели.

Второй случай соответствует тому, что моделируемые выборки значений выходного показателя и его значения, полученные в результате натуральных экспериментов, существенно отличаются друг от друга. Это говорит о том, что при разработке моделирующего алгоритма системы допущены серьезные просчеты, которые не выявлены в процессе предварительного анализа точности каждой подсистемы.

Для вновь создаваемых систем такой случай возможен и потому, что процессы в некоторых подсистемах не могут быть описаны с необходимой точностью на основании результатов теоретических исследований. При этом разработанная модель нуждается в значительной переработке.

1.4.6. Оценка статистической совместимости результатов моделирования с результатами натуральных испытаний

Практическая необходимость проверки статистической совместимости результатов статистического моделирования с результатами натуральных испытаний обусловлена тремя причинами.

1. Для сложных систем практически невозможно при определении параметров моделей подсистем воспользоваться результатами измерений, полученными при испытаниях всей системы и характеризующими степень выполнения системой своих целевых функций и каждом единичном испытании;
2. Из-за сложности возникающих функциональных уравнений трудно реализовать обработку результатов испытаний на основании критериев, удовлетворяющих требованиям, предъявляемым к точности расчета выходных показателей всей сложной системы;
3. Корректировка результатов статистического моделирования, которая обычно осуществляется только по отношению к значимым параметрам модели, несмотря на то, что позволяет повысить точность получаемых оценок, не гарантирует полного отсутствия ошибок в выходных результатах.

Вследствие названных причин в результатах моделирования присутствуют методические и случайные ошибки, влияющие на точность определения выходных характеристик всей сложной системы и достоверность тех статистических выводов, которые принимаются на основании этих оценок.

Для повышения достоверности статистических выводов на практике проводят исследования, основная цель которых состоит в доказательстве факта принадлежности моделируемых и реальных выборок к одной и той же генеральной совокупности. Причем, чтобы использовать наиболее полно результаты натуральных испытаний, указанные исследования проводят на основании материалов, по тем или иным причинам не привлекаемых при оценке параметров создаваемой модели (например, измерения значений выходных характеристик при испытаниях всей системы). В зависимости от точности априорных сведений, результатов обработки натуральных испытаний и применяемых статистических критериев проверки гипотез объем и глубина проводимых исследований в каждом конкретном случае будут различными.

Примерный комплекс вопросов, возникающих при анализе результатов испытаний и моделирования сложных систем, состоит в решении следующих задач:

- 1) обнаружения и отбраковки аномальных натуральных испытаний;
- 2) отождествления типа распределения случайных величин по результатам наблюдаемых выборок;
- 3) оценки значимости величины смещения в результатах моделирования;
- 4) проверки статистической совместимости результатов моделирования и натуральных испытаний на предмет тождественности функций их распределения.

Методы решения указанных задач и примеры, иллюстрирующие возможности их решения для моделей, предназначенных для оценки одного показателя эффективности, приведены разделе 1.6.

Более сложные задачи возникают, когда математические модели разрабатывают с таким расчетом, чтобы по результатам моделирования можно было произвести расчет некоторой совокупности характеристик реальной системы. При этом оценку качества моделируемых процессов дают на основании нескольких показателей, имеющих некоторое совместное распределение. Если законы распределения оцениваемых показателей подчиняются многомерному нормальному распределению, то достаточно широкий круг задач, связанных с проверкой статистической совместимости результатов моделирования с результатами натуральных испытаний, может быть доведен до конечных строго обоснованных цифр и результатов. К таким задачам следует отнести следующие задачи.

Задача 1. На математической модели в процессе моделирования получена выборка p -мерных векторов ограниченного объема. Известно, что выборка $R_{m1}, R_{m2}, \dots, R_{mm} = \vec{R}_m^{(m)}$ принадлежит генеральной нормально распределенной совокупности

$$N\{\mathbf{m}_{R_m}, \mathbf{K}_{R_m}\},$$

где \mathbf{m}_{R_m} — вектор математического ожидания; \mathbf{K}_{R_m} ковариационная матрица. Результаты натуральных испытаний представлены выборкой нормально распределенных p -мерных векторов $(R_{p1}, R_{p2}, \dots, R_{pn}) = \vec{R}_p^{(n)}$ с неизвестными вектором математического ожидания \mathbf{m}_{R_p} и ковариационной матрицей \mathbf{K}_{R_p} . Априорные сведения позволяют считать, что справедливо соотношение

$$\mathbf{K}_{R_m} = \mathbf{K}_{R_p}^*$$

На основании выборок $\mathbf{R}_{M_1}, \mathbf{R}_{M_2}, \dots, \mathbf{R}_{M_m}; \mathbf{R}_{P_1}, \mathbf{R}_{P_2}, \dots, \mathbf{R}_{P_n}$ требуется проверить гипотезу об отсутствии смещения в результатах моделирования, в качестве которых выступают оценки вектора математического ожидания \mathbf{R}_M^* , полученные при моделировании

Для проверки указанной гипотезы запишем ее символически в виде $H: \mathbf{m}_{R_M} = \mathbf{m}_{R_P}$, рассчитаем значение T^2 -статистики:

$$\left. \begin{aligned} T^2 &= \frac{mn}{m+n} (\mathbf{R}_M^* - \mathbf{R}_P^*)^T \mathbf{S}^{-1} (\mathbf{R}_M^* - \mathbf{R}_P^*); \\ \mathbf{S} &= \frac{1}{m+n-2} \left\{ \sum_{i=1}^m (\mathbf{R}_{M_i} - \mathbf{R}_M^*) (\mathbf{R}_{M_i} - \mathbf{R}_M^*)^T + \right. \\ &\quad \left. + \sum_{i=1}^n (\mathbf{R}_{P_i} - \mathbf{R}_P^*) (\mathbf{R}_{P_i} - \mathbf{R}_P^*)^T \right\}, \end{aligned} \right\} (1)$$

где

$$\mathbf{R}_P^* = \frac{1}{n} \sum_{i=1}^n \mathbf{R}_{P_i} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n r_{P_i}^{(1)} \\ \dots \\ \frac{1}{n} \sum_{i=1}^n r_{P_i}^{(p)} \end{pmatrix} = \begin{pmatrix} r_P^{(1)*} \\ \dots \\ r_P^{(p)*} \end{pmatrix};$$

$$\mathbf{R}_M^* = \frac{1}{m} \sum_{i=1}^m \mathbf{R}_{M_i} = \begin{pmatrix} \frac{1}{m} \sum_{i=1}^m r_{M_i}^{(1)} \\ \dots \\ \frac{1}{m} \sum_{i=1}^m r_{M_i}^{(p)} \end{pmatrix} = \begin{pmatrix} r_M^{(1)*} \\ \dots \\ r_M^{(p)*} \end{pmatrix};$$

$$\mathbf{R}_{P_i} = (r_{P_i}^{(1)}, r_{P_i}^{(2)}, \dots, r_{P_i}^{(p)}), \quad i=1, 2, \dots, n;$$

$$\mathbf{R}_{M_i} = (r_{M_i}^{(1)}, r_{M_i}^{(2)}, \dots, r_{M_i}^{(p)}), \quad i=1, 2, \dots, m.$$

Для всякого фиксирования уровня значимости α гипотезу $H: \mathbf{m}_{R_P} = \mathbf{m}_{R_M}$ следует считать справедливой, если выполнено соотношение:

$$T^2 \leq \frac{(m+n-2)p}{m+n-p-1} F_{p, m+n-p-1}(\alpha), \quad (2)$$

де $F(\alpha)$ — значение функции F — распределения с p и $m+n-p-1$ степенями свободы для заданного уровня значимости α .

Проверка условия (2) достаточно проста и для нее не требуется большого объема вычислений.

Задача 2. Даны две выборки p -мерных векторов $\vec{\mathbf{R}}_M^{(m)}, \vec{\mathbf{R}}_P^{(n)}$, принадлежащих нормально распределенным совокупностям $N\{\mathbf{m}_{R_P}, \mathbf{K}_{R_P}\}, N\{\mathbf{m}_{R_M}, \mathbf{K}_{R_M}\}$.

Требуется установить эквивалентность указанных совокупностей.

Эта задача для рассматриваемых распределений состоит в проверке гипотезы о равенстве одновременно средних значений и ковариаций, т. е. в проверке гипотезы $H: \mathbf{m}_{R_P} = \mathbf{m}_{R_M}; \mathbf{K}_{R_P} = \mathbf{K}_{R_M}$.

Для проверки гипотезы H рекомендуется использовать распределение случайной величины

$$V = V_1 V_2, \quad (3)$$

$$\text{где } V_1 = \frac{|\mathbf{A}|^{(1/2)(n-1)} |\mathbf{B}|^{(1/2)(m-1)}}{|\mathbf{A}|^{(1/2)(n+m-2)}}; \quad V_2 = (|\mathbf{A}| / |\mathbf{B}|)^{(1/2)(n+m-2)}.$$

В свою очередь

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2,$$

где

$$\mathbf{A}_1 = \sum_{i=1}^n (\mathbf{R}_{P_i} - \mathbf{R}_P^*) (\mathbf{R}_{P_i} - \mathbf{R}_P^*)^T; \quad \mathbf{A}_2 = \sum_{i=1}^m (\mathbf{R}_{M_i} - \mathbf{R}_M^*) (\mathbf{R}_{M_i} - \mathbf{R}_M^*)^T;$$

$$\mathbf{B} = \mathbf{A} + \frac{mn}{m+n} (\mathbf{R}_M^* - \mathbf{R}_P^*) (\mathbf{R}_M^* - \mathbf{R}_P^*)^T.$$

Статистика V_1 используется для проверки гипотезы

$$H_1: \mathbf{K}_{R_P} = \mathbf{K}_{R_M},$$

а статистика V_2 — для проверки гипотезы

$$H_2: \mathbf{m}_{R_P} = \mathbf{m}_{R_M}$$

при $\mathbf{K}_{R_P} = \mathbf{K}_{R_M}$. Кроме того, показано, что статистики V_1 и V_2 независимы в том случае, когда гипотезы H_1 и H_2 верны. Независимость статистик V_1 и V_2 означает, что проверку гипотезы H можно осуществлять в два этапа: сначала проверить гипотезу H_1 с использованием V_1 , а затем гипотезу H_2 с помощью V_2 .

Для произвольных p в аналитической форме найти выражения для законов распределения статистик V_1 и V_2 невозможно. Поэтому на практике при проверке гипотез H_1, H_2, H для вычисления плотностей

распределения статистик V_1, V_2, V применяют асимптотические разложения.

Если ввести переменные

$$\left. \begin{aligned} W_1 &= \frac{(n+m-2)^{(1/2)p(n+m-2)}}{(n-1)^{(1/2)p(n-1)}(m-1)^{(1/2)p(m-1)}} V_1; \\ W_2 &= (n+m-2)^{(1/2)p(n+m-2)} (n-1)^{-(1/2)p(n-1)} \times \\ &\quad \times (m-1)^{-(1/2)p(m-1)} V, \end{aligned} \right\} \quad (4)$$

то можно доказать справедливость следующих соотношений:

а) при проверке гипотезы H_1

$$P\{-\rho \ln W_1 \leq \chi^2(\alpha)\} = P\{\chi_f^2 \leq \chi^2(\alpha)\} + \omega_2 [P\{\chi_{f+4}^2 \leq \chi^2(\alpha)\} - P\{\chi_f^2 \leq \chi^2(\alpha)\}] + O(n^{-3}), \quad (5)$$

где $\rho = 1 - \left[\frac{1}{m-1} + \frac{1}{n-1} - \frac{1}{m+n-2} \right] \frac{2p+1}{6}$;

$$f = \frac{1}{2}(p+1)p; \quad \omega_2 = \frac{p(p+1)\left\{(p-1)(p+2)\left[\frac{1}{(n-1)^2} + \frac{1}{(m-1)^2} - \frac{1}{(n+m-2)^2}\right] - 6(1-\rho^2)\right\}}{48\rho^2};$$

б) при проверке гипотезы H

$$P\{-2\rho \ln W \leq \chi^2(\alpha)\} = P\{\chi_f^2 \leq \chi^2(\alpha)\} + \omega_2 [P\{\chi_{f+4}^2 \leq \chi^2(\alpha)\} - P\{\chi_f^2 \leq \chi^2(\alpha)\}] + O(n^{-3}), \quad (6)$$

где

$$\rho = 1 - \left[\frac{1}{m-1} + \frac{1}{n-1} - \frac{1}{m+n-2} \right] \frac{2p+1}{6} - \frac{p}{(n+m-2)(p+3)}; \\ f = \frac{1}{2}p(p+3); \quad \omega_2 = \frac{p}{288\rho^2} \left\{ 6 \left[\frac{1}{(n-1)^2} + \frac{1}{(m-1)^2} - \frac{1}{(n+m-2)^2} \right] \times \right. \\ \times (p+1)(p+2)(p-1) - \left. \left(\frac{1}{n-1} + \frac{1}{m-1} - \frac{1}{m+n-2} \right)^2 \times \right. \\ \times \frac{(2p^2+3p-1)^2}{p+3} - 12 \left(\frac{1}{n-1} + \frac{1}{m-1} - \frac{1}{m+n-2} \right) \frac{(2p^2+3p-1)p}{(p+3)(n+m-2)} - \\ \left. - 36 \frac{p^2}{(n+m-2)^2(p+3)} + \frac{24(p^2-1)}{n+m-2} \right\}.$$

Если ω_2 мало по сравнению с $P\{\chi_f^2 \leq \chi^2(\alpha)\}$, то проверка рассматриваемых гипотез достаточно проста и основывается на том,

что величины $-2\rho \ln W$ и $-\rho \ln W_1$ имеют χ^2 — распределение с числом степеней свободы, соответственно равным

$$f = \frac{1}{2}p(p+3); \quad f = \frac{1}{2}p(p+1). \quad (7)$$

Задача 3. Если на математической модели удастся получить достаточно большой объем статистического материала, то возникают задачи проверки статистических гипотез вида:

$$H: \mathbf{m}_{R_p} = \mathbf{m}_{R_M}^0; \quad \mathbf{K}_{R_p} = \mathbf{K}_{R_M}^0. \quad (8)$$

При рассмотрении подобных задач считают точно известными $\mathbf{m}_{R_M}^0$ и $\mathbf{K}_{R_M}^0$, а оценки истинных значений \mathbf{m}_{R_p} и \mathbf{K}_{R_p} рассчитывают на основании наблюдаемой выборки p -мерных векторов

$$\mathbf{R}_{p_1}, \mathbf{R}_{p_2}, \dots, \mathbf{R}_{p_n} = \vec{\mathbf{R}}_p^{(n)}.$$

Для решения поставленной задачи найдем значение

$$\lambda = (e|n)^{pn/2} |\mathbf{BK}_{R_M}^{0-1}|^{n/2} \times \exp\left\{-\frac{1}{2}\left[sp(\mathbf{BK}_{R_M}^{0-1}) + n(\mathbf{R}_p^* - \mathbf{m}_{R_M}^0)^T \mathbf{K}_{R_M}^{0-1}(\mathbf{R}_p^* - \mathbf{m}_{R_M}^0)\right]\right\}, \quad (9)$$

где $\mathbf{B} = \sum_{i=1}^n (\mathbf{R}_{p_i} - \mathbf{R}_p^*)(\mathbf{R}_{p_i} - \mathbf{R}_p^*)^T$; $\mathbf{R}_p^* = \frac{1}{n} \sum_{i=1}^n \mathbf{R}_{p_i}$;

$sp(\cdot)$ — след матрицы.

Предположив, что нулевая гипотеза H справедлива, проверяем неравенство

$$-2 \ln \lambda \leq \chi^2(\alpha), \quad (10)$$

где $\chi^2(\alpha)$ — определяют из условий, что асимптотическое распределение величины $-2 \ln \lambda$ стремится к χ^2 -распределению с $[p+1]p/2 + p$ степенями свободы.

Если неравенство (10) выполнено, то проверяемая гипотеза справедлива.

Рассмотренные выше задачи довольно часто ставятся при анализе результатов моделирования сложных систем. Успешное их решение в значительной степени определяется принятым предположением о нормальности распределения анализируемых выборок. Если бы законы распределения указанных выборок не подчинялись многомерному гауссову распределению, то степень адекватности модели реальной системе в смысле точности имитации некоторых процессов может

быть установлена только с помощью приближенных методов проверки статистических гипотез.

К этим методам относится широко используемый на практике прием, когда проверка выдвигаемых гипотез относительно некоторой статистической эквивалентности двух p -мерных выборок сводится к последовательной проверке тех же гипотез, но уже для каждой фазовой компоненты векторов $\mathbf{R}_m, \mathbf{R}_p$ в отдельности. Причем, исходная многомерная гипотеза считается справедливой, если все результаты проверки одномерных гипотез положительны.

Описанный выше способ проверки многомерных гипотез теоретически необоснован и, вероятно, можно придумать примеры, когда он будет просто недопустимым. Однако практика показывает, что такой способ проверки позволяет экспериментатору эффективнее использовать все результаты натурных испытаний и получить более качественные выводы о статистической эквивалентности результатов моделирования и натурных испытаний.

1.5. Оценка точности результатов статистического моделирования

1.5.1. Классификация ошибок

Практическое использование математических моделей для расчета показателей эффективности сложных систем неразрывно связано с оценкой точности результатов статистического моделирования. Анализ этих вопросов необходим, так как статистические выводы, принимаемые на основании получаемых оценок, должны быть достоверными с некоторой наперед заданной вероятностью. Изучение этих вопросов состоит в определении полной группы ошибок, которые возникают при оценке показателей эффективности реальных систем с помощью математических моделей. Если классификацию осуществляют с точки зрения причин, приводящих к появлению ошибок моделирования, то деление ошибок на классы можно выполнить следующим образом:

1. Ошибки моделирования, возникающие из-за несоответствия операторов модели реальной системе;
2. Ошибки расчета оцениваемых показателей из-за неточностей дискретной реализации моделирующего алгоритма на средствах используемой цифровой вычислительной техники;

3. Ошибки моделирования, возникающие из-за нелинейных эффектов преобразования моделями неточных входных данных;
4. Случайные ошибки моделирования, обусловленные ограниченностью статистики, которую получают при проведении статистических испытаний на модели;
5. Ошибки моделирования, являющиеся результатом ограниченного числа испытаний, проведенных на реальной системе.

Ранее отмечалось, что суммарная ошибка моделирования состоит из регулярной и случайной составляющих. Причем каждая (оставляющая зависит как от начального состояния системы $\mathbf{z}(t_0)$, так и от входного сообщения $(t, \mathbf{x}_L]_{t_0}^t$ реализуемого в данном эксперименте. В связи с этим суммарную ошибку моделирования $\delta\mathbf{R}(\mathbf{c}|t)$ необходимо оценивать в условиях, в которых выполняется неравенство

$$\inf_{\mathbf{z}(t_0), \mathbf{x}(t)} P \{ \delta\mathbf{R}(\mathbf{c}|t, \mathbf{z}(t_0), (t, \mathbf{x}_L]_{t_0}^t) \leq \Delta \} > 1 - P_\Delta, \quad (1)$$

где Δ — область предельно допустимых ошибок расчета показателя эффективности системы; P_Δ — вероятность, определяющая степень доверия к результатам моделирования на созданной модели.

Другими словами, ошибки моделирования должны оцениваться для входных сообщений $(t, \mathbf{x}_L]_{t_0}^t$ и начальных состояний системы $\mathbf{z}(t_0)$, на которых достигается точная нижняя грань значений вероятности

$$P \{ \delta\mathbf{R}(\mathbf{c}|t, \mathbf{z}(t_0), (t, \mathbf{x}_L]_{t_0}^t) \leq \Delta \}.$$

Если удастся представить каждую компоненту входного сигнала $x(t)$ в виде некоторой комбинации линейно-независимых ортонормированных функций, то уравнение (6.1.1) можно преобразовать к виду

$$\inf_{\mathbf{z}(t_0), \mathbf{a}} P \{ \delta\mathbf{R}(\mathbf{c}|t, \mathbf{z}(t_0), (t, \mathbf{x}_L]_{t_0}^t) \leq \Delta \} > 1 - P_\Delta, \quad \mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n), \quad (2)$$

где $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in})$ — вектор параметров, характеризующий разложения $\mathbf{x}_i(t) = \mathbf{a}_i^T \Phi(t); \Phi(t) = (\Phi_1(t), \Phi_2(t), \dots, \Phi_n(t))$.

Переход от (1) к (2) целесообразен, так как позволяет применять более простые методы анализа уравнений чувствительности, связанные с определением условий, в которых должны оцениваться ошибки моделирования.

Для нахождения решения $z(t)$ на интервале $[0, t]$ с помощью ЭВМ необходимо уравнение (5) записать в виде разностного уравнения:

$$\mathbf{H}_h z^{(h)} = x^{(h)}, \quad (6)$$

где \mathbf{H}_h — разностный оператор; $z^{(h)}$ — решение разностного уравнения; $x^{(h)}$ — правая часть разностного уравнения.

Практически операцию преобразования (5) в (6) можно осуществить различными способами, которые будут отличаться друг от друга как точностью представления непрерывного дифференциального оператора, так и временем, необходимым для нахождения решения на используемой ЭВМ. В общем случае точность и время счета являются противоречивыми требованиями. Чтобы между этими требованиями найти практически приемлемый компромисс, нужно знать, каким образом оценивать сходимость решений разностных уравнений к решениям исходных дифференциальных уравнений.

Для выяснения этого вопроса нужно установить, что необходимо понимать под разностью между решением дифференциального уравнения $z(t)$, определенным во всех точках отрезка времени $[0, t]$, и решением разностного уравнения $z^{(h)}$ полученным только в конечном числе точек разностной сетки. Если каждой функции $z(t)$ на отрезке времени $[0, t]$ поставить каким-либо способом в соответствие сеточную функцию $[z(t)]_h$, то разность между решениями $z(t)$ и $z^{(h)}(t)$ можно оценивать в точках сетки по разности $[z(t)]_h - z^{(h)}(t)$. На основании этого под близостью сеточной функции $z^{(h)}(t)$ и функции $z(t)$ следует понимать малость величины нормы:

$$\|z^{(h)}(t) - [z(t)]_h\|_{Z_h}, \quad (7)$$

где Z_h — пространство определения сеточных функций $z^{(h)}(t)$.

Если полагать, что аналогичные нормы установлены для оценки точности аппроксимации начальных данных и правых частей, то чтобы исследовать сходимость, необходимо установить порядок аппроксимации и оценить устойчивость решений для выбранной разностной схемы. В ряде работ показано, что порядки аппроксимации дифференциального оператора \mathbf{H} , граничных условий и правой части уравнения совпадают с порядком точности разностной схемы, если решение разностного уравнения удовлетворяет условию

$$\|z^{(h)}\|_{Z_h} \leq N_1 \|x^{(h)}\|_{X_h} + N_2 \|f^{(h)}\|_{F_h}, \quad (8)$$

где X_h, F_h — соответственно функциональные пространства определения сеточных функций $x^{(h)}$ и $f^{(h)}$; N_1, N_2 — постоянные, не зависящие от шага интегрирования h .

Если условие (8) выполнено, то используемая разностная схема устойчива.

Пример. Рассмотрим вопросы определения порядка точности аппроксимации обыкновенного дифференциального оператора второго порядка с помощью разностных операторов вида:

$$H_h^{(1)} z = T^2 \frac{z(t+h) - 2z(t) + z(t-h)}{h^2} + 2\xi T \frac{z(t+h) - z(t-h)}{2h} + z(t); \quad (9)$$

$$H_h^{(2)} z = T^2 \frac{z(t) - 2z(t-h) + z(t-2h)}{h^2} + 2\xi T \frac{z(t) - z(t-h)}{h} + z(t). \quad (10)$$

Вопросы аппроксимации решений дифференциального уравнения $z(t)$ решениями разностного уравнения $z^{(h)}(t)$ будем рассматривать по норме пространства Z_h , которая определяет условия достижения равномерной близости функций $z(t)$ и $z^{(h)}(t)$ во всех точках сетки.

Для проведения последующих оценок найдем:

$$\begin{aligned} z(t+h) &\approx z(t) + hz^I(t) + \frac{h^2}{2} z^{II}(t) + \frac{h^3}{6} z^{III}(t); \\ z(t-h) &\approx z(t) - hz^I(t) + \frac{h^2}{2} z^{II}(t) - \frac{h^3}{6} z^{III}(t); \\ z(t-2h) &\approx z(t) - 2hz^I(t) + 2h^2 z^{II}(t) - \frac{4h^3}{3} z^{III}(t); \end{aligned} \quad (11)$$

$$\begin{aligned} z(t+h) &\approx z(t) + hz^I(t) + \frac{h^2}{2} z^{II}(t) + \frac{h^3}{6} z^{III}(t) + \frac{h^4}{24} z^{IV}(t); \\ z(t-h) &\approx z(t) - hz^I(t) + \frac{h^2}{2} z^{II}(t) - \frac{h^3}{6} z^{III}(t) + \frac{h^4}{24} z^{IV}(t); \\ z(t-2h) &\approx z(t) - 2hz^I(t) + 2h^2 z^{II}(t) - \frac{4h^3}{3} z^{III}(t) + \frac{h}{24} z^{IV}(t). \end{aligned}$$

На основании этих формул при условии, что

$$|z^{III}(t)| + |z^{IV}(t)| < a, \quad t \in [0, t] \quad (12)$$

можно получить для разностного оператора следующую оценку:

$$|H_h^{(1)} z| \leq \max_t \left| T^2 \frac{h^2 A}{12} + 2\xi T \frac{h^2 A}{6} \right| = \mu_1 h^2, \quad (13)$$

где

$$\mu_1 = \max_t \left| \frac{T^2 A}{12} + \frac{\xi T A}{3} \right|;$$

A — некоторая константа.

Отсюда следует, что

$$|H_h^{(1)} z - H z| < \mu_1 h^2. \quad (14)$$

На основании соответствия, установленного между функциями $z(t)$ и $z^{(h)}(t)$ и неравенства (14), получаем:

$$\|H_h^{(1)} [z]_h - [H z]_h\| < \mu_1 h^2. \quad (15)$$

Следовательно, разностный оператор I аппроксимирует дифференциальный оператор H на функции $z(t)$ со вторым порядком аппроксимации.

Проведя аналогичные выкладки для разностного оператора II , нетрудно получить

$$\|H_h^{(2)} [z]_h - [H z]_h\| < \mu_2 h, \quad (16)$$

т. е. разностный оператор II аппроксимирует дифференциальный оператор на функции $z(t)$ с первым порядком аппроксимации.

Итак, разностный оператор I с точки зрения порядка аппроксимации на функции $z(t)$ имеет существенное преимущество перед разностным оператором II .

Для завершения исследований по определению порядка точности рассматриваемых разностных схем оценим устойчивость решений по отношению к значениям коэффициентов, входящих в уравнения (9), (10). Для простоты рассмотрим случай, когда $x(t)=0$ для всех t . Полагая $t = nh$, общее решение разностных уравнений (9) и (10) можно записать в виде:

$$a z_{n-1} + b z_n + c z_{n+1} = 0; \quad (17)$$

$$z_n = \gamma_1 (\sqrt{a/c})^n \cos n\varphi + \gamma_2 (\sqrt{a/c})^n \sin n\varphi;$$

$$\gamma_1 = z_0;$$

$$\gamma_2 = [z_{-1} - z_0 (\sqrt{a/c})^{-1} \cos \varphi] / [- (\sqrt{a/c})^{-1} \sin \varphi];$$

$$\cos \varphi = b/2 \sqrt{ac}.$$

Для начальных условий $z_0=1, z_{-1}=1$ параметры γ_1, γ_2 рассчитываем по формулам:

схема 1:

$$\gamma_1 = 1;$$

$$\gamma_2 = (2\xi T - h) / \sqrt{4T^2 - 2\xi^2 T^2 - h^2},$$

схема 2:

$$\gamma_1 = 1;$$

$$\gamma_2 = (\xi h^2) / \sqrt{4\xi T + h(1 - \xi^2)}.$$

Поскольку интерес представляют действительные решения уравнений (17), то из (15) получаем нижнюю грань шага вычислений по схеме 1:

$$h^2 < 4T^2 (1 - \xi^2). \quad (18)$$

Из (18) видно, что для схемы 2 действительные решения существуют при любых h , так как всегда

$$1 - \xi^2 > 0. \quad (19)$$

Для схемы 1 ввиду того, что $T^2 > 0$, можно получить более простое соотношение для нижней грани шага вычислений:

$$T > \xi h. \quad (20)$$

Если условия (18), (19) выполнены, то обе разностные схемы устойчивы, а следовательно, из (15) и (16), следует, что при $x(t)=0, t \in [0, t_1]$ порядок точности k схемы 1 равен 2, а схемы 2 равен 1.

Рассмотренный пример говорит о том, что для сложных дифференциальных уравнений оценить таким образом сходимость и порядок точности получаемых решений очень трудно. Сказанное является следствием того, что для изучения подобных вопросов нужно знать общее решение рассматриваемых разностных уравнений. Для разностных уравнений высокого порядка задача нахождения общих решений довольно сложна и в ряде случаев не разрешима аналитическими методами.

По этой причине при анализе ошибок дискретной реализации обыкновенных дифференциальных уравнений прибегают к приближенным методам. Наиболее широкое применение на практике находит метод, основанный на принципе Рунге. Чтобы реализовать этот метод, необходимо:

1. Знать порядок точности разностной схемы;
2. Располагать значениями решений разностного уравнения для двух шагов h и $2h$, т. е. необходимо получить $z^{(h)}(t)$ и $z^{(2h)}(t)$.

При этих исходных данных приближенная оценка погрешности может быть рассчитана по формуле:

$$\|z^{(h)}(t) - [z(t)]_h\| = \frac{|z^{(2h)}(t) - z^{(h)}(t)|}{2^{k-1}}. \quad (21)$$

Практическая реализация описанной выше процедуры достаточно проста и может быть полезной как на этапе разработки моделирующего алгоритма, так и на этапе окончательной оценки точности спроектированной модели. В сказанном нетрудно убедиться,

если принять во внимание, что соотношение (21) устанавливает взаимосвязь между основными показателями, определяющими точность дискретной реализации элементов, описываемых дифференциальными операторами.

Ошибки дискретного представления чисел в ЭВМ. Природа возникновения этой группы ошибок обусловлена конечной точностью представления чисел в ЭВМ и естественными операциями округления промежуточных результатов. Характер проявления этих ошибок в значительной степени зависит от принятого способа программной реализации моделирующего алгоритма. Для одного и того же моделирующего алгоритма, но программно реализованного различными способами, влияние таких ошибок на выходной результат будет различным. Поэтому оценку влияния этой группы ошибок на результаты моделирования нужно проводить для каждого случая отдельно с учетом того способа реализации алгоритмов, который принят при разработке программ моделей.

Некоторые методы оценки ошибок округления подробно изложены в ряде работ. Суть этих методов составляют алгоритмы расчета, основанные на различных предположениях о вероятностном характере проявления указанных ошибок.

Часто при анализе сложных систем влиянием ошибок округления из-за высокой точности представления чисел в ЭВМ пренебрегают. Нужно сказать, что априори поступать таким образом нельзя, так как при определенных условиях эти ошибки могут существенно исказить выходные результаты.

Итак, при дискретной реализации моделирующих алгоритмов на используемых ЭВМ возникают ошибки, которые влияют на результаты моделирования, внося в них как случайные, так и методические ошибки. Для сложных систем охарактеризовать влияние этих ошибок на точность получаемых оценок очень трудно. Поэтому при калибровке модели нужно проводить исследования с целью определения возможностей компенсации методических ошибок, а также оценки влияния случайных ошибок на точность результатов моделирования.

1.5.3. Оценки случайных ошибок, обусловленных конечным числом реализаций на модели

Метод статистических испытаний предусматривает многократный проигрыш различных ситуаций. Поскольку на модели получить выборку неограниченного объема нельзя, то в результатах модели-

рования будут присутствовать случайные ошибки, обусловленные конечным числом реализаций. Характеристики распределения этих ошибок определяют выбранным методом планирования статистических испытаний и принятым способом обработки получаемых результатов.

На практике для получения возможно простых алгоритмов обработки стремятся к тому, чтобы наблюдаемая выборка результатов моделирования была линейно связана с оцениваемыми показателями. При выполнении этого условия, а также при соблюдении некоторых правил построения моделей, иногда удается сравнительно просто установить свойства получаемых оценок, а порой до моделирования гарантировать их оптимальность в некотором классе оценок.

Так, если разрабатываемая модель предназначена для оценки математического ожидания показателя $R(t|\mathbf{c})$, то среди всех линейных несмещенных оценок среднее арифметическое

$$R_N^*(t|\mathbf{c}) = \frac{1}{N} \sum_{i=1}^N R_i(t, \mathbf{c}), \quad (1)$$

где R_i — значение выходного показателя, реализовавшееся в i -м эксперименте на модели; N — количество реализаций, будет эффективной оценкой истинного значения в том смысле, что оценка R_N^* имеет наименьшую дисперсию.

Обработка результатов моделирования с использованием (1) включает в себя тот практически важный случай, когда необходимо вычислить вероятность выполнения некоторого логического условия. При этом расчет оценок искомых вероятностей осуществляют по формуле:

$$P\{R(\mathbf{c}, t) \in \Omega_R\} \approx m/N, \quad (2)$$

где m — число опытов, при которых значения $R(\mathbf{c}, t)$ принадлежали некоторой заданной области Ω_R .

Если по результатам моделирования необходимо определить некоторые моменты для законов распределения векторного показателя эффективности, то искомые оценки могут быть найдены по следующим формулам:

$$E\{r_i^p\} \approx \frac{1}{N} \sum_{k=1}^N (r_{ik}^p); \quad (3)$$

$$E \{r_{i_1}^{p_1}(t_1), r_{i_2}^{p_2}(t_2), \dots, r_{i_n}^{p_n}(t_n)\} \approx \frac{1}{N} \sum_{k=1}^N [r_{i_1,k}(t_1)]^{p_1} [r_{i_2,k}(t_2)]^{p_2} \dots [r_{i_n,k}(t_n)]^{p_n}, \quad (4)$$

где p_i — целые положительные числа; $r_{i_k}(t)$ — значения фазовых компонент вектора $\mathbf{R}(\mathbf{c}, t)$, полученные и зафиксированные в различных реализациях на модели.

Выражения (1)-(4) можно сравнительно просто преобразовать к рекуррентным соотношениям, что важно при обработке результатов статистического моделирования с использованием быстродействующих машин.

Обработка результатов статистического моделирования на основании алгоритмов (1)-(4) наиболее часто встречается при практических исследованиях. Поэтому очень важно знать зависимость точности получаемых оценок от объема статистического моделирования.

Принимая во внимание то, что алгоритмы (1)-(4) реализуют одну и ту же операцию усреднения, проиллюстрируем методику определения указанных зависимостей на примере расчета фазовых компонент $E\{r_i\}$ ($i=1, 2, \dots, n$) вектора $\mathbf{R}(t|\mathbf{c})$.

Грубую оценку погрешности расчета r_i^* для независимых экспериментов на модели нетрудно получить с помощью неравенства Чебышева

$$P \{ |r_i - E\{r_i\}| \geq \varepsilon \} \leq \frac{E \{ |r_i - E\{r_i\}|^2 \}}{\varepsilon^2} = \frac{D\{r_i\}}{\varepsilon^2}, \quad (5)$$

где $\varepsilon > 0$ — некоторое заданное сколь угодно малое число, получим, что

$$D\{r_i^*\} = D\{r_i\}/N \quad (6)$$

и как следствие этого

$$P \left\{ \frac{|r_i - E\{r_i\}|}{3\sqrt{D\{r_i\}}} \geq \varepsilon \right\} \leq \frac{1}{9\varepsilon^2 N}. \quad (7)$$

Можно найти более точную зависимость, если при выводе использовать центральную предельную теорему Ляпунова, которая для оценок, рассчитываемых по формулам (1)-(3), гарантирует асимптотическую нормальность их распределения. Прделав все необходимые преобразования, можно получить

$$P \left\{ \frac{|r_i - E\{r_i\}|}{3\sqrt{D\{r_i\}}} \leq \varepsilon \right\} \approx \Phi(x = 3\varepsilon\sqrt{N/2}), \quad (8)$$

где $\Phi(x)$ — функция Лапласа:

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (9)$$

Оценки, найденные по формуле (8), оказываются достаточно точными не только при больших N , но и в том случае, когда по тем или иным причинам на модели удастся получить выборку результатов моделирования сравнительно малого объема. При $x=2$ взаимосвязь между числом реализаций на модели и достигаемой точностью результатов моделирования выражается широко используемым на практике соотношением

$$N \approx 1/\varepsilon^2, \quad (10)$$

где ε — заданная точность моделирования.

Характерной особенностью полученных выше соотношений является то, что порядок убывания погрешности равен $N^{-1/2}$ и не зависит от размерности n пространства Ω_R .

Если исследование анализируемой системы возможно аналитическими методами или с помощью упрощенных моделей, то точность оценок r_i^* ($i=1, 2, \dots, n$) может быть существенно повышена при том же объеме статистического моделирования N . Для реализации этого положения можно воспользоваться алгоритмами, которые по терминологии, приведенной в ряде работ, позволяют находить сверхэффективные оценки.

В рассматриваемом случае, когда на модели рассчитывают оценки математического ожидания составляющих показателя $\mathbf{R}(t|\mathbf{c})$, алгоритм обработки можно записать следующим образом:

$$\tilde{r}_i^* = \frac{1}{\eta_i N} \sum_{j=1}^N (r_{ij} - \tilde{r}_i) + \tilde{r}_i, \quad (11)$$

где \tilde{r}_i — оценка, найденная аналитическим методом или на упрощенной модели; $\eta_i > 1$ — коэффициент, определяющий область, в которой можно реализовать процедуру построения сверхэффективных оценок.

Если закон распределения r_{ij} (при $j=1, 2, \dots, N$) гауссов с параметрами $N\{r_i, D_{i_i}\}$, то выигрыш в точности оценок r_i^{**} можно получить тогда, когда выполняется неравенство

$$\left[\frac{(r_i - \tilde{r}_i)(1 - \eta_i)}{\eta_i} \right]^2 + \frac{1}{\eta_i^2} \frac{D_{r_i}}{N} \leq \frac{D_{r_i}}{N}. \quad (12)$$

Это неравенство имеет простой физический смысл, если принять во внимание, что в его левой части записано выражение для дисперсии суммарной ошибки оценки \tilde{r}_i^{**} , рассчитываемой по (11), а в правой части — дисперсия оценки как среднего арифметического, найденного с помощью (1). При фиксированных η_i , D_{r_i} , N неравенство (12) позволяет определить требования, которые нужно предъявлять к точности аналитических расчетов, чтобы результирующая точность оценки $r_i^*(t|\mathbf{c})$ была бы большей, чем оценок, рассчитываемых на основании алгоритма (1).

Практическое использование соотношения (11) в сочетании с (1) дает возможность записать общий алгоритм совместной обработки результатов аналитических расчетов и статистического моделирования:

$$\tilde{r}_i^{**} = \begin{cases} r_i^* = \frac{1}{N} \sum_{j=1}^N r_{ij} & \text{при } \left| \frac{r_i^* - \tilde{r}_i}{\sigma_{r_i}} \right| \geq \frac{\eta_i + 1}{N(\eta_i - 1)}; \\ \tilde{r}_i^{**} & \text{при } \left| \frac{r_i^* - \tilde{r}_i}{\sigma_{r_i}} \right| < \frac{\eta_i + 1}{N(\eta_i - 1)}. \end{cases} \quad (13)$$

Для этого алгоритма характерно то, что при $r_i = \tilde{r}_i$ асимптотическая дисперсия оценки \tilde{r}_i обратно пропорциональна величине $\eta_i^2 N$ (справедливо равенство $D_{r_i^*} = D_{r_i} / (\eta_i^2 N)$), а при $r_i \neq \tilde{r}_i$ асимптотическая дисперсия равна $D_{r_i^*}$.

Когда известна корреляционная взаимосвязь результатов аналитических расчетов с результатами статистического моделирования, комбинированные оценки могут быть найдены с использованием известных алгоритмов.

Кроме этого способа уменьшения дисперсии $D_{r_i^*}$ сравнительно просто можно реализовать процедуру, соответствующую методу существенной выборки. Физический смысл этого метода заключается в том, что исходный интеграл, определяющий математическое ожидание оцениваемого показателя по распределению вероятностей $\mu(d\mathbf{z})$

$$E\{\mathbf{R}(t, \mathbf{c})\} = \mathbf{R}(t|\mathbf{c}) = \int \mathbf{H}^*[\mathbf{z}(t)] \mu(d\mathbf{z}), \quad (14)$$

преобразуется к виду:

$$\mathbf{R}(t|\mathbf{c}) = \int \mathbf{H}^*[\mathbf{z}(t)] \frac{d\mu}{d\lambda} \lambda(d\mathbf{z}), \quad (15)$$

где $\lambda(d\mathbf{z})$ — специально выбранное распределение вероятностей случайных векторов \mathbf{z} в момент времени t .

Эффективность применения этого метода гарантируется в том случае, если справедливо неравенство

$$D_{r_{i\lambda}} > D_{r_{i\mu}};$$

$$D_{r_{i\mu}} = \frac{1}{N} \left\{ \int \mathbf{H}^*[\mathbf{z}(t)] \mu(d\mathbf{z}) \right\}_i;$$

$$D_{r_{i\lambda}} = \frac{1}{N} \left\{ \int \left[\mathbf{H}^*[\mathbf{z}(t)] \frac{d\mu}{d\lambda}(\mathbf{z}) - E\{\mathbf{R}(\mathbf{c}, t)\} \right]^2 \lambda(d\mathbf{z}) \right\}_i.$$

Переход от распределения вероятностей $\mu(d\mathbf{z})$ к $\lambda(d\mathbf{z})$ широко используется в случае, когда моделирование $\lambda(d\mathbf{z})$ проще, чем моделирование $\mu(d\mathbf{z})$.

Уменьшить дисперсию оценок $D_{r_i^*}$ можно и тогда, когда исходная область интегрирования Ω соответствующим образом разбивается на ряд непересекающихся подмножеств Ω_j ($j = 1, 2, \dots, m$)

$$\mathbf{R}(t|\mathbf{c}) = \int_{\Omega} \mathbf{H}^*[\mathbf{z}(t)] \mu(d\mathbf{z}) = \sum_{j=1}^m \int_{\Omega_j} \mathbf{H}^*[\mathbf{z}(t)] \mu(d\mathbf{z}). \quad (16)$$

Этот способ вычисления интегралов носит название метода расщепленной выборки.

При $m=2$, $n_1 = p_1 N$, $n_2 = p_2 N$ (p_1 и p_2 — вероятности принадлежности случайных векторов \mathbf{z} областям Ω_1 и Ω_2 соответственно) не трудно показать, что для функций $\mathbf{H}^*[\mathbf{z}(t)]$, принадлежащих гильбертову пространству \mathbf{L}_2 , метод расщепленной выборки дает результаты более точные, чем прямой метод статистических испытаний.

Рассмотренные методы уменьшения погрешностей в результатах статистического моделирования приводят к сравнительно простым алгоритмам обработки и планирования испытаний на создаваемых математических моделях. Поэтому данные алгоритмы наиболее часто включают в моделирующий алгоритм сложной системы.

Общим свойством методов уменьшения погрешностей является более тонкий учет характера изменения подинтегральной функции и вида распределения, по которому осуществляют интегрирование.

1.5.4. Особенности и последовательность задач, решаемых при оценке влияния ограниченного объема реальной информации на точность результатов моделирования

Планирование экспериментальных работ на средствах и элементах системы осуществляется исходя из условий достижения конечной цели всего испытательного процесса при наименьших или заданных экономических затратах. В процессе планирования, а также после завершения испытаний большое значение имеют вопросы по определению влияния ограниченного объема реальной информации, получаемой при физических экспериментах, на точность расчета характеристик эффективности или показателей качества работоспособности всей сложной системы. Сложность решения указанных вопросов имеет место из-за значительной разновидности экспериментальных работ, чрезвычайной громоздкости расчетных формул и других факторов, обусловленных значительной разнотипностью измерительных устройств, которые привлекают для регистрации характеристик исследуемых процессов управления.

Для широко распространенного способа задания требований на систему, когда по результатам моделирования нужно осуществить точечную оценку одного показателя эффективности, алгоритм корректировки можно получить из следующего уравнения:

$$E\{R(t|c^*)\} - R(t|c) = \Delta_{R^*}(c, y, t), \quad (1)$$

а точность найденных оценок охарактеризовать величиной дисперсии

$$D\{R(t|c^*)\} = D_{R^*}(c, y, t), \quad (2)$$

где $R(t|c)$ — значение показателя эффективности системы, соответствующее истинному значению вектора c ; y — результаты экспериментальных работ, организованных на элементах системы.

В более общем случае оценка влияния ограниченного объема реальной информации на точность расчета величины $R(t|c)$ при заданной структуре модели состоит в выводе функциональных зависимостей:

$$E\{W[R(t|c), R(t|c^*)]\} = W^*(c, y, t), \quad (3)$$

где W — функция, характеризующая тип ошибок, которые выбраны для измерения точности оценок $R(t|c^*)$.

Анализ зависимостей (1)-(3) показывает, что для определения величины смещения Δ_{R^*} и дисперсии необходимо D_{R^*} знать

истинные значения вектора параметров c . Но по результатам физических экспериментов найти истинный вектор параметров c практически невозможно. Поэтому при анализе сложных систем оценка влияния ограниченного объема реальной информации на точность получаемых оценок состоит в решении ряда задач, связанных с определением $\Delta_{R^*}(c, y, t)$, $D_{R^*}(c, y, t)$ для некоторой ожидаемой области Ω_c возможных значений истинного вектор-параметров c .

Когда зависимости $\Delta_{R^*}(c, y, t)$, $D_{R^*}(c, y, t)$ как функции вектор-параметров c определены, то анализ характера их изменения в исследуемой области Ω_c позволяет охарактеризовать ожидаемые ошибки определения этих зависимостей, если за истинное значение вектор-параметров c принять его оценку c^* , полученную по результатам физических экспериментов.

При анализе зависимостей (1)-(3) могут встретиться три случая:

1. Оценка параметров модели осуществлена на основании байесового подхода; моделирование значений вектора параметров c осуществляется на основании апостериорного закона распределения

$$p(c|y, t);$$

2. Априорных распределений $p(c)$ нет и обработка результатов натурных испытаний произведена на основании частных методик обработки с использованием метода максимального правдоподобия;

3. Часть параметров модели рассчитывают на основании найденных апостериорных распределений, а другие параметры определяют из условий максимизации функции правдоподобия.

Третий случай наиболее часто встречается при анализе реальных систем и получается как комбинация двух первых случаев.

Условия оценки параметров модели, соответствующие первому случаю, позволяют записать алгоритмы корректировки и оценки точности результатов статистического моделирования с использованием известных апостериорных распределений.

Во втором случае из-за отсутствия априорных сведений найти точное распределение $p(c|y, t)$ практически невозможно. Поэтому приходится прибегать к приближенным методам оценивания, либо использовать фидуциальный подход, который позволяет записать $p(c|y, t)$ в следующем виде:

$$p^*(c|y, t) = [p(y|c, t)] / [p(y|t)], \quad (4)$$

что соответствует заданию $p(c)$ на основании постулата Байеса, т. е. $p(c) = 1$ при $c \in \Omega_c$, где Ω_c — область определения вектора параметров c .

Если распределения $p^*(\mathbf{c}|\mathbf{y}, t)$, $p(\mathbf{c}|\mathbf{y}, t)$ определены, то последовательность задач, решаемых при оценке влияния ограниченного объема реальной информации на точность результатов моделирования, состоит в реализации операций:

1. Разработки модели системы и проведения на ней статистического моделирования для некоторого набора векторов $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p$

Число рассчитываемых значений $R^*(t|\mathbf{c}_i)$, $i=1, 2, \dots, p$ и выбор значений векторов $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p$ определяются методом стохастической аппроксимации многомерных поверхностей. Для сложных систем задачи исследования уравнений типа (1)-(3) из-за вычислительных трудностей могут быть решены только относительно наиболее существенных параметров системы;

2. Определения коэффициентов a_1, a_2, \dots, a_s для уравнения регрессии:

$$R^*(t|\mathbf{c}) \approx G(\mathbf{c}, a_1, a_2, \dots, a_s) = G(\mathbf{c}, \mathbf{a}).$$

Широко распространенной формой записи $G(\mathbf{c}, \mathbf{a})$ являются функциональные разложения по некоторой совокупности функций $\varphi_i(\mathbf{c})$ ($i=1, 2, \dots, s$) с независимыми переменными в виде фазовых компонент вектора параметров:

$$\mathbf{c} = (c_1, c_2, \dots, c_p);$$

3. Нахождения законов распределения $p(\mathbf{c}|\mathbf{y}, t)$, $p^*(\mathbf{c}|\mathbf{y}, t)$, как функций от реализованного объема экспериментальных работ.

4. Определения значений $\Delta_{R^*}(\mathbf{c}, \mathbf{y}, t)$, $D_{R^*}(\mathbf{c}, \mathbf{y}, t)$ для некоторого набора значений истинного вектора параметров \mathbf{c}_i ($i=1, 2, \dots, k$).

Число рассчитываемых значений Δ_{R^*} , D_{R^*} должно определяться в каждом конкретном случае, так как оно зависит от степени гладкости функций $\Delta_{R^*}(\mathbf{c}, \mathbf{y}, t)$, $D_{R^*}(\mathbf{c}, \mathbf{y}, t)$ по параметру \mathbf{c} .

Итак, полученные результаты позволяют охарактеризовать изучаемые явления только при условии, что форма и все характеристики закона распределения $p(\mathbf{c}, \mathbf{y}, t)$ точно известны, хотя и влияние некоторых из них, в частности значений вектора параметров \mathbf{c} , удастся проанализировать только в некотором диапазоне его изменения. Однако можно предположить, что если зависимости $\Delta_{R^*}(\mathbf{c}, \mathbf{y}, t)$, $D_{R^*}(\mathbf{c}, \mathbf{y}, t)$ оказываются достаточно гладкими функциями от \mathbf{c} , то решение задач чувствительности в рассмотренной выше постановке позволит получить практически приемлемые решения.

Пример. Пусть математическая модель описывается выражениями:

$$R(t|\mathbf{c}) = \int_{-\infty}^{+\infty} R(x) p(x|\mathbf{c}) dx,$$

$$p(x|\mathbf{c}) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{(x-m_x)^2}{2\sigma_x^2}\right\};$$

$$R(x) = R_0 + \sum_{i=1}^s \frac{d^i R}{dx^{(i)}} x^{(i)};$$

$$\mathbf{c} = m_x \text{ или } \sigma_x^2,$$

а уравнение регрессии с необходимой точностью аппроксимируется рядом:

$$R(t|\mathbf{c}) = G(\mathbf{a}, \mathbf{c}) = a_0 + a_1 c + a_2 c^2.$$

Как величина смещения точность оценки $R(t|\mathbf{c}^*)$ зависят от объема выборки x_1, x_2, \dots, x_n , если оценки m_x^* , σ_x^{*2} рассчитывают по результатам физических экспериментов:

$$m_x^* = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma_x^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x^*)^2.$$

Если априори точно известно, что $m_x=0$, то при $s=2$ нетрудно убедиться в том, что $a_2=0,5 (d^2 R/dx^2)=0$ и как результат этого

$$\Delta_{R^*}(n) = E\{R(t|\mathbf{c}^*)\} - R(t|\mathbf{c}) = 0.$$

Для тех же условий при $s=4$ величину смещения оценки $R(t|\mathbf{c}^*)$ как функцию от объема реальной информации можно охарактеризовать зависимостью

$$\Delta_{R^*}(n) = \frac{3}{4!} \frac{d^4 R}{dx^{(4)}} D_{\sigma_x^{*2}};$$

$$D_{\sigma_x^{*2}} \approx \frac{2}{n-1} \sigma_x^{*2}, \text{ так как } E\{R(t|e^*)\} =$$

$$= R_0 + \frac{1}{2} \frac{d^2 R}{dx^{(2)}} \sigma_x^2 + \frac{3}{4!} \frac{d^4 R}{dx^{(4)}} [D_{\sigma_x^{*2}} + \sigma_x^4];$$

$$R(t|e) = R_0 + \frac{1}{2} \frac{d^2 R}{dx^{(2)}} \sigma_x^2 + \frac{3}{4!} \frac{d^4 R}{dx^{(4)}} \sigma_x^4.$$

Если точно известна величина дисперсии σ_x^2 , а оценка математического ожидания m_x^* рассчитывается по результатам измерений, то при $s = 2$

$$\Delta_{R^*}(n) = E\{R(t|m_x^*)\} - R(t|m_x) = \frac{1}{2} \frac{d^2 R}{dx^{(2)}} D_{m_x^*};$$

$$E\{R(t|m_x^*)\} = R_0 + \frac{1}{2} \frac{d^2 R}{dx^{(2)}} \sigma_x^2 + \frac{dR}{dx} m_x + \frac{1}{2} \frac{d^2 R}{dx^{(2)}} [D_{m_x^*} + m_x^2];$$

$$R(t|m_x) = R_0 + \frac{1}{2} \frac{d^2 R}{dx^{(2)}} \sigma_x^2 + \frac{dR}{dx} m_x + \frac{1}{2} \frac{d^2 R}{dx^{(2)}} m_x^2;$$

$$D_{m_x^*} = \sigma_x^2/n.$$

Рассмотренные задачи просты, но результаты их решения могут оказаться полезными при инженерных расчетах.

В общем случае для многомерных распределений ожидаемую величину смещения Δ_{R^*} и величину дисперсии D_{R^*} нужно рассчитывать для каждого фиксированного объема реальной информации по формулам (1) и (2).

1.5.5. Корректировка результатов статистического моделирования

Корректировку результатов статистического моделирования сложных систем в большинстве практических случаев проводят с использованием алгоритмов, которые основаны на квадратичной трансимации оцениваемого показателя эффективности:

$$R(t|\mathbf{c}) = R(t|\mathbf{c}^*) + \left. \frac{dR}{d\mathbf{c}} \right|_{\mathbf{c}^*} \cdot (\mathbf{c} - \mathbf{c}^*) + \left. \frac{1}{2} \frac{d^2 R}{d\mathbf{c}^{(2)}} \right|_{\mathbf{c}^*} (\mathbf{c} - \mathbf{c}^*)^2, \quad (1)$$

где

$$\frac{dR}{d\mathbf{c}} = E \left\{ \sum_{i=1}^s \left[\frac{\partial \mathbf{H}^*}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{c}} + \frac{\partial \mathbf{H}^*}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial \mathbf{c}} + \frac{\partial \mathbf{H}^*}{\partial \mathbf{e}} \right]_{\mathbf{r}_{i-1}} + \right.$$

$$\left. + [\mathbf{H}^*(t_i^-) - \mathbf{H}^*(t_i^*)] \frac{dt_i}{d\mathbf{c}} \right\};$$

$$\frac{d^2 R}{d\mathbf{c}^{(2)}} = E \left\{ \sum_{i=1}^s \left[\frac{\partial^2 \mathbf{H}^*}{\partial \mathbf{x}^{(2)}} \left(\frac{\partial \mathbf{x}}{\partial \mathbf{c}} \right)^2 + \frac{\partial \mathbf{H}^*}{\partial \mathbf{x}} \cdot \frac{\partial^2 \mathbf{x}}{\partial \mathbf{c}^{(2)}} + \frac{\partial^2 \mathbf{H}^*}{\partial \mathbf{z}^{(2)}} \left(\frac{\partial \mathbf{z}}{\partial \mathbf{c}} \right)^2 + \right. \right.$$

$$\left. + \frac{\partial \mathbf{H}^*}{\partial \mathbf{z}} \cdot \frac{\partial^2 \mathbf{z}}{\partial \mathbf{c}^{(2)}} + \frac{\partial^2 \mathbf{H}^*}{\partial \mathbf{c}^{(2)}} \right]_{t_{i-1}} + \left[\frac{\partial \mathbf{H}^*}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{c}} + \frac{\partial \mathbf{H}^*}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial \mathbf{c}} + \right.$$

$$\left. + \frac{\partial \mathbf{H}^*}{\partial \mathbf{c}} \right]_{t_i^-} - \left[\frac{\partial \mathbf{H}^*}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{c}} + \frac{\partial \mathbf{H}^*}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial \mathbf{c}} + \frac{\partial \mathbf{H}^*}{\partial \mathbf{c}} \right]_{t_i^+} \frac{dt_i}{d\mathbf{c}} +$$

$$\left. + [\mathbf{H}^*(t_i^-) - \mathbf{H}^*(t_i^+)] \frac{\partial^2 t_i}{\partial \mathbf{c}^{(2)}} \right\};$$

$$\frac{dt}{dc_k} = - \left. \begin{array}{l} \frac{\partial F}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial c_k} + \frac{\partial F}{\partial c_k} \\ \frac{\partial F}{\partial \mathbf{z}} \cdot \frac{\partial \mathbf{z}}{\partial t} + \frac{\partial F}{\partial t} \end{array} \right|_{t=t_i}; \quad k=1, 2, \dots, r,$$

где $F(\mathbf{z}(t), t, \mathbf{c})$ —уравнение поверхности, на которой в моменты времени $t_i (i=1, 2, \dots, s)$ вектор \mathbf{z} изменяется скачком.

Характерная особенность практически используемых алгоритмов корректировки заключается в том, что коэффициенты разложения обычно оцениваются не в точке, соответствующей истинному вектору параметров \mathbf{c} , а в его оценке, полученной при обработке результатов натурных испытаний. Если такое положение допустимо, то процедура корректировки результатов статистического моделирования состоит в исключении из них смещения, равного

$$\Delta_{R^*}(\mathbf{y}, t, \mathbf{c}^*) = \frac{1}{2} \int \left. \frac{d^2 R}{d\mathbf{c}^{(2)}} \right|_{\mathbf{c}^*} (\mathbf{c} - \mathbf{c}^*)^2 p(\mathbf{c}|\mathbf{y}, t) d\mathbf{c}. \quad (2)$$

Вычислительная реализуемость описанной выше процедуры определяется временем, которое необходимо затратить на моделирование с целью расчета производных, входящих в уравнение (1) На практике для определения указанных производных часто используют методы регрессионного анализа. С этой целью начальный вектор состояния системы включают в вектор параметров модели, а входной сигнал $x(t)$ представляют в виде функционального ряда

$$x(t) = \sum_{i=1}^{r_1} c_i \varphi_i(t), \quad (3)$$

где c_i — коэффициенты разложения, $\varphi_1(t), \dots, \varphi_{r_1}(t)$ — некоторая совокупность координатных функций.

Если коэффициенты разложения (3) также отнести к параметрам модели, то зависимость оцениваемого показателя эффективности можно записать в виде некоторой функции:

$$R(t|\mathbf{c}) \approx G(c_1, c_2, \dots, c_r; a_1, a_2, \dots, a_s) \quad (4)$$

от независимых переменных c_1, c_2, \dots, c_r и коэффициентов a_1, a_2, \dots, a_s , которые называют параметрами искомого уравнения регрессии.

Для определения оценок параметров a_1, a_2, \dots, a_s обычно проводят моделирование и на основании полученных результатов рассчитывают значения $R(t|c_j)$ для такого набора векторов c_j ($j=1, 2, \dots, p$), который необходим для применения выбранного метода аппроксимации.

При организации и проведении экспериментов на модели нужно помнить, что методы регрессионного анализа позволяют достичь некоторых вполне определенных свойств оценок параметров a ($i=1, 2, \dots, s$) при выполнении следующих предпосылок:

- а) значения $R(t|c_j)$ (при $j=1, 2, \dots, p$), полученные в результате статистического моделирования, должны представлять собой независимые, нормально распределенные случайные величины;
- б) дисперсии $D_{R_j^*}$ (при $j=1, 2, \dots, p$), которые являются следствием ограниченного числа экспериментов на модели, должны быть равны друг другу или представлять собой известную функцию от вектора c ;
- в) значения независимых переменных c_1, c_2, \dots, c_r должны задаваться абсолютно точно или с очень малой ошибкой по сравнению с ошибкой в расчете ($R_j^* = R(t|c_j^*)$).

При решении практических задач наиболее часто не выполняется первое требование. Однако и в этом случае, даже если нормальности распределения оценок R_j^* ($j=1, 2, \dots, p$) не удается достичь ни при каком преобразовании случайных величин, методы регрессионного анализа все же можно применять для расчета параметров искомого аппроксимирующей поверхности. Но в этом случае оценки коэффициентов уравнения регрессии будут неоптимальными в том смысле, что не всю информацию извлекают из результатов статистического моделирования и свойства получаемых коэффициентов уравнения регрессии, особенно при малых выборках на модели, охарактеризовать очень трудно.

При нормальном же распределении величин R_j^* (при $j=1, 2, \dots, p$) оценки уравнения регрессии a_i ($i=1, 2, \dots, s$) получают наиболее эффективными среди всех возможных линейных оценок.

В общем случае задача определения оценок параметров a_i ($i=1, 2, \dots, s$) на основании метода наименьших квадратов может быть сведена к поиску корней следующей системы уравнений:

$$\sum_{j=1}^p [R_j^* - G(c_j, \mathbf{a})] \frac{\partial G(c_j, \mathbf{a})}{\partial a_i} = 0, \quad i=1, 2, \dots, s. \quad (5)$$

Если искомая поверхность аппроксимируется рядом

$$G(\mathbf{c}, \mathbf{a}) = \sum_{i=1}^s a_i f_i(\mathbf{c}), \quad (6)$$

где $f_i(\mathbf{c})$ — известные функции, то система уравнений становится линейной относительно неизвестных коэффициентов a_1, a_2, \dots, a_s :

$$\sum_{i=1}^s a_i \sum_{j=1}^p f_i(c_j) f_k(c_j) = \sum_{j=1}^p R_j^* f_k(c_j), \quad k=1, 2, \dots, s \quad (7)$$

и ее можно представить матричным соотношением вида

$$\mathbf{B}\mathbf{a} = \mathbf{B}^T \mathbf{y}, \quad (8)$$

где \mathbf{B} — матрица размерности $s \times p$; \mathbf{y} — p -мерный вектор наблюдений $\mathbf{y} = (y_1, y_2, \dots, y_p)$.

Элементы матрицы \mathbf{B} и вектора \mathbf{y} нетрудно получить из соотношений (7).

Характерная особенность системы уравнений (7) следующая: если матрица $\|\mathbf{B}^T \mathbf{B}\|^{-1}$ (\mathbf{B}^T — транспонированная матрица) является невырожденной, то выбранная система независимых переменных c_1, c_2, \dots, c_r линейно независима и решение системы (7) будет единственным.

При этом диагональные элементы матрицы $\sigma_{y^2} = \|\mathbf{B}^T \mathbf{B}\|^{-1}$, определяют дисперсии ошибок расчета коэффициентов регрессии, а остальные элементы характеризуют ковариации соответствующих им коэффициентов регрессии.

При ортогональном планировании экспериментов на модели корреляционная матрица $\|\mathbf{B}^T \mathbf{B}\|^{-1}$ получается диагональной. Поэтому коэффициенты уравнения регрессии можно определять независимо друг от друга по простым расчетным формулам.

Практическая реализуемость методов регрессионного анализа зависит от суммарного времени, необходимого для определения числа

наиболее существенных факторов и расчета требуемого количества значений R_j^* ($j = 1, 2, \dots, p$).

Применительно к задачам исследования сложных систем наиболее типичным случаем является полиномиальная аппроксимация истинной поверхности с помощью уравнения регрессии, которое включает в себя линейные и квадратичные члены разложения, также учитывает все смешанные произведения факторов.

Для такого описания уравнений регрессии общее время, необходимое для расчета ряда значений R_j^* ($j = 1, 2, \dots, p$) при центральном композиционном планировании экспериментов, определяют в формуле

$$t = \Delta t N (2^r + r + 1),$$

где Δt — время проигрыша на модели одной случайной реализации.

Анализ приведенного соотношения показывает, что время моделирования в значительной степени зависит от выбранного числа наиболее существенных факторов. Поэтому выбор и обоснование числа значимых параметров модели c_i ($i = 1, 2, \dots, r$) должен производиться очень тщательно с привлечением всех доступных средств анализа.

Широко распространенный метод последовательного уточнения числа существенных факторов состоит в проверке необходимости дальнейшего повышения (понижения) степени искомого уравнения регрессии. Достоинством этого метода является довольно экономичное использование памяти вычислительных машин. Кроме того, он позволяет обойти трудность, связанную с априорным определением числа наиболее существенных факторов. Но метод последовательного уточнения поверхностей регрессии связан с большими затратами машинного времени. Поэтому для сложных систем его удастся реализовать далеко не во всех случаях. В связи с этим при анализе сложных систем большое внимание уделяют вопросам априорного определения числа наиболее существенных факторов.

Когда наиболее существенные факторы определены и метод планирования экспериментов на модели выбран, расчет коэффициентов a_1, a_2, \dots, a_s может быть произведен, если время для нахождения значений R_j^* ($j = 1, 2, \dots, p$) будет допустимым. По этой причине поиск и разработка эффективных вычислительных процедур определения параметров аппроксимирующих поверхностей для сложных систем имеют чрезвычайно важное значение.

Один из возможных методов расчета необходимого числа значений R_j^* ($j = 1, 2, \dots, p$), который в определенной степени решает

проблему вычислений, состоит в организации «взвешенной» обработки результатов статистического моделирования.

Метод взвешенной обработки результатов статистического моделирования. Для определения расчетных формул метода рассмотрим задачу: как рассчитать, зная результаты статистического моделирования

$$\left. \begin{array}{l} x_1 \rightarrow R_1; \\ x_2 \rightarrow R_2; \\ \dots\dots\dots \\ x_N \rightarrow R_N, \end{array} \right\} \quad (10)$$

оценку математического ожидания выходного показателя

$$E \{R(\mathbf{c}, t)\} = \int R(\mathbf{x}) \mu(d\mathbf{x}),$$

если параметры распределения вектора \mathbf{x} имеют численные значения \mathbf{c} , не равные \mathbf{c}^* , при котором осуществлялось моделирование. Задача может быть решена на основании гипотезы: выборка x_1, x_2, \dots, x_N , полученная при фиксированном \mathbf{c}_λ , принадлежит генеральной совокупности выборок, которые можно получить при всевозможных значениях вектора параметров \mathbf{c} . Считая, что эта гипотеза справедлива, можно записать следующее выражение:

$$R(t | \mathbf{c}) = \int R(\mathbf{x}) (\mu/\lambda)(\mathbf{x}) \lambda(d\mathbf{x}). \quad (11)$$

Полученные зависимости показывают, что для определения оценок R_j^* ($j = 1, 2, \dots, p$) необходимо реализовать процесс «взвешенной» обработки результатов статистического моделирования:

$$R(t | \mathbf{c}_j) \approx R_j^* = \frac{1}{N} \sum_{k=1}^N R_k \gamma_{jk}, \quad (12)$$

где

$$\gamma_{jk} = \mu_j(x_k | \mathbf{c}^j) / \lambda(x_k | \mathbf{c}^\lambda).$$

Программная реализуемость такого метода расчета величин R_j^* ($j = 1, 2, \dots, p$) во многом зависит от принятого математического описания законов распределения вектора \mathbf{x} . Если законы распределения гауссовы, то алгоритм расчета γ_{jk} получается достаточно простым и его программная реализация обычно не вызывает серьезных затруднений.

Для гауссовых распределений расчетные формулы могут быть записаны в виде:

$$\gamma_{jk} = \sqrt{\frac{|K_\lambda|}{|K_{\nu_j}|}} \exp \left\{ - \frac{(\mathbf{x}_k - E\{\mathbf{x}_{\nu_j}\})^T K_{\nu_j}^{-1} (\mathbf{x}_k - E\{\mathbf{x}_{\nu_j}\})}{(\mathbf{x}_k - E\{\mathbf{x}_\lambda\})^T K_\lambda^{-1} (\mathbf{x}_k - E\{\mathbf{x}_\lambda\})} \right\}, \quad (13)$$

где \mathbf{K} — ковариационная матрица вектора \mathbf{x} , а $|K|$ ее детерминант.

Эффективность метода взвешенной обработки. Сравнительную оценку метода взвешенной обработки с каким-либо другим методом построения уравнений регрессии произведем для очень простого, но широко распространенного на практике случая, когда необходимо при некотором фиксированном c_0 определить значения

$$R(t|c_0), \frac{\partial R}{\partial c} \Big|_{c_0}$$

на модели, описываемой соотношением вида:

$$R(t|c) = \int_{-\infty}^{\infty} R(x) p(x|c) dx, \quad (14)$$

где $R(x)$ — некоторая известная функция; $p(x|c)$ — плотность распределения x с параметром c .

Будем считать, что оценка значения $R(t|c)$ при каждом фиксированном c находится методом статистических испытаний:

$$R_N^*(c) = \frac{1}{N} \sum_{i=1}^N R_i, \quad (15)$$

для чего рассчитываем последовательности

$$\begin{aligned} x_1 &\rightarrow R_1; \\ x_2 &\rightarrow R_2; \\ &\dots\dots\dots \\ x_N &\rightarrow R_N. \end{aligned}$$

Кроме того, будем предполагать, что закон распределения величин x_i (при $i=1, 2, \dots, N$) гауссов с плотностью распределения

$$p(x|c) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\{-x^2/2\sigma_x^2\}, \quad c = \sigma_x^2, \quad (16)$$

а функция $R(x)$ определяется соотношением:

$$R(x) = R_0 + \frac{\partial R}{\partial x} x + \frac{1}{2} \frac{\partial^2 R}{\partial x^{(2)}} x^2. \quad (17)$$

Чтобы упростить последующие расчеты, искомое уравнение регрессии будем записывать так:

$$R(t|c) = R(t|c_0) + \frac{\partial R(t|c)}{\partial c} \Big|_{c_0} (c - c_0) = a_0 + a_1(c - c_0) \quad (18)$$

и считать, что для нахождения неизвестных коэффициентов a_0, a_1 по результатам моделирования определены оценки $R^*(t|c_0), R^*(t|c_1)$ при $c_0 = \beta c_1$. Для этих условий точность расчета коэффициента

$$a_1^* = \frac{R^*(t|\sigma_{x_1}^2) - R^*(t|\sigma_{x_0}^2)}{\sigma_{x_1}^2 - \sigma_{x_0}^2} \quad (19)$$

можно охарактеризовать значением

$$D\{a_1^*\} = \frac{D\{R^*(t|\sigma_{x_1}^2) - R^*(t|\sigma_{x_0}^2)\}}{(\sigma_{x_1}^2 - \sigma_{x_0}^2)^2}. \quad (20)$$

Величина дисперсии

$$D\{R^*(t|\sigma_x^2)\} = \frac{D\{R(x|\sigma_x^2)\}}{N}, \quad (21)$$

$$D\{R(x|\sigma_x^2)\} = \int_{-\infty}^{\infty} [R(x) - R(t|\sigma_x^2)]^2 p(x|\sigma_x^2) dx = \left(\frac{\partial R}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial^2 R}{\partial x^{(2)}}\right)^2 \frac{\sigma_x^4}{2}.$$

Проведя все необходимые преобразования, получим:

$$D\{a_1^*\} = \frac{\left(\frac{\partial R}{\partial x}\right)^2 \frac{(1+\beta)}{\sigma_{x_0}^2} + \left(\frac{\partial^2 R}{\partial x^{(2)}}\right)^2 \frac{(1+\beta^2)}{2}}{(1-\beta)^2}. \quad (22)$$

С помощью метода взвешенной обработки можно также найти значения $R^*(t|\sigma_{x_0}^2), R^*(t|\sigma_{x_1}^2)$.

Если устремить $\sigma_{x_1}^2$ к $\sigma_{x_0}^2$, то принимая во внимание, что математическое ожидание коэффициента

$$a_1 = \frac{\partial}{\partial \sigma_x^2} \left\{ \int_{-\infty}^{\infty} R(x) p(x|\sigma_x^2) dx \right\} = \frac{1}{2} \frac{\partial^2 R}{\partial x^{(2)}} \quad (23)$$

находим при этом же объеме моделирования

$$D_b\{a_1^*\} = \left[\left(\frac{\partial R}{\partial x}\right)^2 \frac{5}{2\sigma_{x_0}^2} + \left(\frac{\partial^2 R}{\partial x^{(2)}}\right) \frac{37}{8} + R_0 \frac{\partial^2 R}{\partial x^{(2)}} \frac{5}{2\sigma_{x_0}^2} + \frac{R_0^2}{2\sigma_{x_0}^4} \right] 2N. \quad (24)$$

Рассмотрим несколько случаев. Для простоты положим $R_0=1$, то нетрудно достичь соответствующей нормировкой членов уравнения (17).

Случай 1. Пусть выполнены следующие условия:

$$\left| \frac{\partial^2 R}{\partial x^{(2)}} \right| \gg \frac{1}{\sigma_{x_0}^2}; \quad \left(\frac{\partial R}{\partial x} \right)^2 \approx \left| \frac{\partial^2 R}{\partial x^{(2)}} \right|. \quad (25)$$

Тогда для отношения дисперсий $D_b\{a_1^*\}$ и $D\{a_1^*\}$ справедливо соотношение

$$\frac{D_b}{D} = \frac{37(1-\beta)^2}{8(1+\beta^2)}. \quad (26)$$

Из уравнения (26) нетрудно получить, что при $\beta > 0,5$ метод «взвешенной» обработки позволяет найти более точную оценку для коэффициента a_1 .

Случай 2. Для условий

$$\left| \frac{\partial^2 R}{\partial x^{(2)}} \right| \ll \frac{1}{\sigma_{x_0}^2}; \quad \left| \frac{\partial R}{\partial x} \right|^2 \gg \frac{1}{\sigma_{x_0}^2} \quad (27)$$

нетрудно получить

$$\frac{D_b}{D} = \frac{5(1-\beta)^2}{4(1+\beta)}$$

и найти, что при $\beta > 0,1$ дисперсия D_b меньше дисперсии D .

Случай 3. Если выполнены соотношения

$$\left| \frac{\partial^2 R}{\partial x^{(2)}} \right| \gg \left| \frac{\partial R}{\partial x} \right|^2; \quad \left| \frac{\partial^2 R}{\partial x^{(2)}} \right| \sim \frac{1}{\sigma_{x_0}^2},$$

то

$$\frac{D_b}{D} = \frac{15(1-\beta)^2}{2(1+\beta)^2}$$

и эффективность применения метода «взвешенной» обработки при оценке коэффициента a_1 будет гарантирована при $\beta > 0,55$.

Так как сравнительная оценка анализируемых методов расчет коэффициентов a_0, a_1 должна производиться при одинаковом числе статистических испытаний, то нетрудно показать, что в рассматриваемом случае при использовании метода «взвешенной» обработки точность оценки коэффициента a_0 получается вдвое большей.

Итак, метод «взвешенной» обработки позволяет при определенных условиях реализовать более оптимальную обработку результатов статистического моделирования. Причем, область эффективного применения метода зависит от формы и параметров искомого поверхностей регрессии, а также в значительной степени определяется характеристиками законов распределения варьируемых при статистическом моделировании параметров. Для сложных поверхностей нахождение областей применимости этого метода представляет собой довольно сложную задачу.

1.5.6. Оценка методических ошибок моделирования в пределах всей области рассматриваемого факторного пространства

При использовании результатов моделирования в целях оценки параметров и характеристик сложной системы возникает, как уже было отмечено, сложная и очень важная задача обеспечения состоятельности, несмещенности и эффективности получаемых оценок. Успешное решение этой задачи во многом зависит от выбранного порядка отладки и калибровки математических моделей

В связи с ограниченной возможностью проведения экспериментов на сложных системах для калибровки моделей и устранения расхождений между результатами натурных экспериментов и моделирования обычно в факторном пространстве назначают крайне небольшое количество точек из числа возможных. При этом, прежде всего, ставят задачу: провести соответствующие параметрическую и структурную доработки математических моделей и в этих условиях достичь несмещенности получаемых оценок. Однако подобная калибровка математических моделей не может служить основанием для заключения об отсутствии разностной ошибки между результатами натурального эксперимента и моделирования во всей области изменения рассматриваемых факторов. Поэтому методика оценки и учета ошибки моделирования во всей области рассматриваемого факторного пространства важна и необходима. Положим, что методическая ошибка моделирования Δ_R в общем случае может быть представлена в виде случайной функции, зависящей от ряда параметров c_1, c_2, \dots, c_i , где c_i — параметры, характеризующие влияние факторов, связанных с изменением положения выбранной точки для калибровки. Влияние всех остальных факторов, не связанных с изменением положения точки в факторном пространстве, в данном случае не рассматривается, так как оно должно быть учтено в результатах калибровки.

Функцию Δ_R в окрестности точки $(c_1^*, c_2^*, \dots, c_r^*)$, где проводят натурные эксперименты, можно представить степенным рядом

$$\begin{aligned} \Delta_R &= \Delta_R(c_1^* + \delta c_1, c_2^* + \delta c_2, \dots, c_r^* + \delta c_r) = \\ &= \Delta_R(c_1^*, c_2^*, \dots, c_r^*) + \sum_{q=1}^r \frac{\partial \Delta_R^*}{\partial c_q} \delta c_q + \frac{1}{2} \sum_{f,q=1}^r \frac{\partial^2 \Delta_R^*}{\partial c_f \partial c_q} \delta c_f \delta c_q. \quad (1) \end{aligned}$$

Значения частных производных соответствуют значениям производных в выбранной точке рабочего режима, а величины δc_q , $\delta c_f, \dots$ — суть небольшие отклонения параметров от их значений в точке рабочего режима.

Так как отклонения параметров лежат в небольших пределах, то рассматриваемую функцию в первом приближении принимают локально линейной и ограничиваются разложением до частных производных первого порядка. Эти значения производных представляют собой компоненты градиента функции $\Delta_{\mathbf{R}}$. Характер изменения этих компонент позволяет судить о характере поведения ошибки моделирования в рассматриваемой области. Равенство величины градиента функции $\Delta_{\mathbf{R}}$ нулю свидетельствует о том, что в данной точке имеет место экстремум функции, или что рассматриваемая функция по своей величине неизменна. При отличии градиента от нуля, используя метод направленного поиска, можно найти точку (в общем случае ряд точек), в которой функция $\Delta_{\mathbf{R}}$ принимает экстремальное значение.

Таким образом, при исследовании характера поведения величины $\Delta_{\mathbf{R}}$ представляется картина изменения ошибки моделирования по всей области рассматриваемого факторного пространства. Непременным условием осуществления такого исследования является возможность получения экспериментальных данных в окрестностях всех точек, составляющих каждый шаг направленного поиска.

В общем случае при испытаниях сложных систем, когда возможности проведения натуральных экспериментов крайне ограничены, подобное решение задачи неосуществимо. Для некоторых конкретных задач указанный подход может быть использован с учетом ряда упрощающих допущений.

Поэтому всегда ставится задача уточнить совокупность параметров c_1, c_2, \dots, c_7 , оказывающих наибольшее влияние на величину ошибки моделирования. Эта задача сводится к определению величины линейных коэффициентов разложения в выражении (1) и оценке их значимости. Оценку значимости в первом приближении можно получить с помощью метода случайного баланса, который не только позволяет произвести оценку наиболее существенных параметров по результатам эксперимента, но и дает возможность определить примерное направление изменения градиента. Решение проводят в несколько этапов, на каждом из которых сначала качественно выделяют существенные параметры, затем осуществляют статистический анализ, обеспечивающий получение количественной

меры оценки влияния рассматриваемого параметра на величину ошибки моделирования.

Для качественного выделения существенных параметров используют диаграммы рассеяния, представляющие гистограмму величины $\Delta_{\mathbf{R}}$, полученную при случайном изменении всех параметров c_0 , за исключением одного параметра c_k , зафиксированного на определенном уровне. Если c_k оказывает существенное влияние на величину ошибки моделирования, то при построении гистограммы для другого фиксированного уровня этого параметра получают смещение центра распределения величины $\Delta_{\mathbf{R}}$:

$$\Delta L = \Delta_{\mathbf{R}}^1 - \Delta_{\mathbf{R}}^2, \quad (2)$$

где $\Delta_{\mathbf{R}}^1$ — центр распределения $\Delta_{\mathbf{R}}$ при фиксированном значении параметра c_k на первом уровне; $\Delta_{\mathbf{R}}^2$ — центр распределения $\Delta_{\mathbf{R}}$ при c_k на втором уровне.

Для получения оценки истинного значения центра распределения с приемлемой степенью надежности иногда можно ограничиться 5-8 значениями $\Delta_{\mathbf{R}}$. Наиболее значимыми будут те параметры c_k , у которых по диаграмме рассеяния получается наибольшее смещение, если же смещения одинаковые, то те параметры, гистограмма которых сдвигается наиболее сильно.

Проведенный таким образом качественный отбор позволяет исключить из рассмотрения параметры, которые можно отнести к разряду случайных возмущений, вносящих погрешности в измерения величины $\Delta_{\mathbf{R}}$.

После качественного отбора проводят статистический анализ, по результатам которого оценивают значимость коэффициентов регрессии. Значимые коэффициенты признают существенными и учитывают в уравнении (1). С помощью критерия Фишера проверяют адекватность полученного уравнения. Если по результатам обработки это уравнение окажется неадекватным, то следует перейти к более сложной форме уравнения связи между ошибкой моделирования и рассматриваемыми переменными. Полученная на основании обработки экспериментальных данных в рабочей точке зависимость дает возможность определить направление градиента. В случае, если функция $\Delta_{\mathbf{R}}$ в пределах области определения может быть отнесена к классу монотонных функций, то при выборе координат каждой последующей рабочей точки в целях оценки коэффициентов уравнения регрессии и определения направления градиента можно ориентироваться на крайние значения параметров c_i в пределах их изменений. Так, если исследуемыми факторами являются параметры

c_1 и c_2 , то в качестве рабочих точек выбирают точки, соответствующие сочетаниям

$$(c_{1\max}, c_{2\min}), (c_{1\max}, c_{2\max}), (c_{1\min}, c_{2\max}), (c_{1\min}, c_{2\min}).$$

Проведя качественный отбор и статистический анализ в указанных четырех рабочих точках, можно получить представление о диапазоне и характере изменения ошибки моделирования в рассматриваемом факторном пространстве. Если подобные исследования удастся провести на этапе планирования натурных испытаний, то в результате можно определить условия, при которых нужно получить реальную информацию для оценки максимально возможных ошибок моделирования.

1.6. Методы обработки и анализа результатов испытаний сложных систем

1.6.1. Требования к методам обработки результатов измерений

Испытания любого устройства, подсистемы или системы в целом неразрывно связаны с проведением измерений. Именно в результате измерений получают данные, характеризующие способность испытываемой системы решать те задачи, для которых ее создавали. Поэтому успех испытаний зависит прежде всего от организации эксперимента, выбранных условий и метода измерения, который был использован при проведении наблюдений. Рассмотрим задачи, связанные с организацией вычислений при математической обработке результатов измерений, так как только правильное их решение позволит оценить истинность полученного результата в ходе испытаний, даст возможность установить необходимые закономерности и обеспечит безошибочность выводов и заключений.

При организации обработки результатов измерений следует прежде всего учитывать, что любые измерения содержат ошибки. По характеру ошибки измерений разделяют на систематические и случайные. *Систематические ошибки* порождаются определенными закономерностями, существующими при данных условиях эксперимента, или же объективными факторами, присущими данному объекту. *Случайные ошибки* при данных условиях эксперимента характеризуются непостоянностью, отражая суммарное воздействие большого числа различных, но не определяющих факторов.

Подобное деление ошибок измерений в какой-то степени носит условный характер, поскольку ошибки, являющиеся при данных условиях эксперимента систематическими, в других условиях могут быть случайными. Например, если имеет место реализация, полученная для нестационарного процесса, то ошибка, рассматриваемая как систематическая на малом промежутке времени, приобретает характер случайной при увеличении времени наблюдений. Именно отсутствие четко выраженной границы деления между систематическими и случайными ошибками приводит иногда к необходимости выделения *медленно меняющихся ошибок*.

Поэтому одной из основных задач, которую следует решать при организации обработки, является установление способов получения оценок, обеспечивающих наилучшее приближение к истинным значениям параметров по результатам измерений, обладающих соответствующими ошибками. Выбранный метод обработки при этом не только должен обеспечивать достижение наилучшего приближения, но и позволять провести апостериорную оценку достигнутой точности приближения.

Для получения оценок искомых параметров проводят статистическую обработку данных измерений, в результате которой в случае минимальных потребностей находят приближенные значения параметров распределения в виде математического ожидания и среднего квадратического отклонения. В большинстве случаев используют более полный аппарат математической статистики с применением методов дисперсионного и регрессионного анализа, а также методов проверки гипотез.

Оценки параметров распределения, полученные по результатам обработки измерений, иногда называют *статистиками*. Для оценки одного и того же параметра можно воспользоваться различными статистиками. При выборе конкретной статистики обычно учитывают их **состоятельность, несмещенность, эффективность и достаточность**.

Выполнение условий *состоятельности* означает, что полученная оценка сходится по вероятности при большом числе измерений к истинному значению оцениваемого параметра x , т. е.

$$\lim_{n \rightarrow \infty} P(|x_n^* - x| < \varepsilon) \rightarrow 1,$$

где ε — сколь угодно малое положительное число.

Несмещенность оценки говорит об отсутствии систематической погрешности, т. е. выполнении следующего равенства:

$$E\{x_n^*\} = x.$$

Оценки, несмещенные и обладающие наименьшей возможной дисперсией, называют *эффективными*. Условие эффективности оценок можно выразить в виде:

$$D\{x_n^*\}/D_n = 1,$$

где D_n — представляет значение нижней границы дисперсии статистики x_n^* для данного объема выборки.

Оценку называют *достаточной*, если она построена на основании всей информации об искомом параметре, содержащейся в данном объеме наблюдений.

1.6.2. Первичная обработка результатов измерений

Результаты измерений, полученные при испытаниях, представляют собой фактические данные, на основании анализа которых можно прийти к определенному заключению. Но для проведения анализа весь огромный массив наблюдений предварительно подвергают обработке, которую на практике обычно подразделяют на первичную и вторичную. К *первичной обработке* относят такие операции как сортировка и объединение данных, представление их в виде, удобном для дальнейшей работы, отбраковка грубых и аномальных результатов.

Вторичная обработка представляет собой проведение непосредственных вычислений интересующих параметров. Анализ результатов, полученных при вторичной обработке позволяет исследователю прийти к тому или другому заключению о полноте решения задач испытаний.

Совокупность наблюдений, полученная при испытаниях и представляющая собой выборку из генеральной совокупности, может быть записана в виде упорядоченного или вариационного, возрастающего ряда.

Если обозначить рассматриваемую дискретную случайную величину через x , то для вариационного ряда будет справедлива последовательность вида $x_1 < x_2 < \dots < x_n$.

При многократных измерениях величины x легко подсчитать число k появления событий $x \leq x_i$ ($i=1, 2, \dots, n$), на основании чего нетрудно определить эмпирические вероятности $P^*(x \leq x_i) = k/n$. Результаты наблюдений и обработки можно представить в виде табл. 1.

Таблица 1

x_i	x_1	x_2	...	x_n
$P^*(x < x_i)$	$P^*(x \leq x_1)$	$P^*(x < x_2)$...	$P^*(x \leq x_n)$

Часто при большом объеме выборки для облегчения исследований прибегают к упорядочению и уплотнению статистических рядов путем построения интервального вариационного ряда. Весь статистический ряд разбивают на определенное число интервалов, для которых подсчитывают частоту попадания рассматриваемой случайной величины x . В этом случае принимают, что результаты наблюдений, попавшие в один интервал, обладают одним и тем же значением, соответствующим середине интервала. Тогда все данные могут быть сведены в табл. 2 в следующем виде:

Таблица 2

Интервал	От x_0 до x_1	От x_1 до x_2	...	От x_{n-1} до x_n
Частота	p_1^*	p_2^*	...	p_n^*
Эмпирическая вероятность	$P^*(x_0 \leq x < x_1)$	$P^*(x_1 \leq x < x_2)$...	$P^*(x_{n-1} \leq x < x_n)$

В таком случае допускается погрешность, которая, однако, не будет превосходить половины длины выбранного интервала. Тем не менее это может привести к появлению систематических ошибок при определении параметров эмпирического распределения.

По данным, собранным в табл. 2, можно получить эмпирическую функцию распределения, представляющую накопленные частоты, отнесенные к серединам выбранных интервалов. Подсчет значений функции распределения можно произвести по формулам:

$$F^*(\bar{x}_1) = 0; F^*(\bar{x}_2) = p_1^*; F^*(\bar{x}_3) = p_1^* + p_2^*; F^*(\bar{x}_i) = \sum_{k=1}^i p_k^*$$

для каждого $\bar{x}_i = (x_i + x_{i-1})/2$.

График полученной интегральной кривой функции распределения будет иметь вид, показанный на рис. 1.

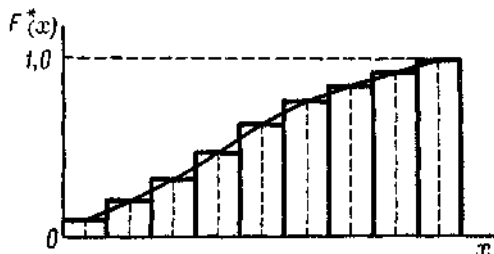


Рис. 1

Для характеристики непрерывных случайных величин часто используют плотность распределения вероятностей, которая представляет собой производную от функции распределения. Подсчет значений эмпирической плотности распределения можно произвести по формуле:

$$f^*(\bar{x}_i) = F^*(\bar{x}_i) / l_i,$$

где l_i — длина i -го интервала.

Графическое представление полученной дифференциальной кривой распределения (рис. 2) называют *гистограммой*.

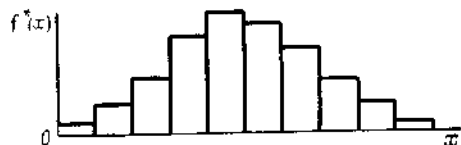


Рис. 2

Ступенчатый график интегральной кривой и гистограмму обычно сглаживают непрерывной функцией, характер и вид которой во многом зависит от выбранной длины интервала и соотношения выбранных масштабов по оси ординат и абсцисс. С увеличением длины интервалов и уменьшением их количества сложнее воссоздать истинный характер кривой распределения, поэтому длину интервалов выбирают такой, чтобы количество последних составляло 10-20. Иногда для расчета длины интервала l используют формулу:

$$l = (x_{\max} - x_{\min}) / (1 + 3 \cdot 2 \log_2 n),$$

где n — число наблюдений.

Соотношение масштабов вдоль осей координат выбирают, руководствуясь правилом «золотого» сечения. При выборе интервалов и

составлении вариационного ряда следует учитывать, что отдельные результаты наблюдений могут резко отличаться от всех остальных результатов вследствие появления грубых ошибок или недопустимого нарушения условий измерений.

При натуральных испытаниях сложных систем обычно привлекают большое количество измерительной техники, обслуживающий персонал которой имеет различный уровень подготовки. Вследствие этого, а также из-за случайной неоднотипности условий проведения натуральных экспериментов порой возникают значительные ошибки в выходных результатах. При ограниченном объеме реальной информации эти ошибки могут привести к неправильным статистическим выводам. Чтобы избежать подобные явления, при анализе сложных систем большое внимание уделяют определению условий, при которых можно исключить из рассмотрения резко выделяющиеся наблюдения.

При выявлении подобных аномальных результатов наблюдений прежде всего необходимо провести тщательный анализ измерений и проверить, не являются ли эти результаты следствием грубого промаха или нарушения условий измерений. Если подобный анализ не устранил появившихся сомнений, то прибегают к помощи статистических методов выявления грубых ошибок, позволяющих производить целесообразную отбраковку аномальных данных. Для этого можно использовать ряд различных критериев: Колмогорова; Пирсона; Шовенэ и др. В частности, если допустить предположение о нормальном распределении погрешностей измерений, то можно прибегнуть к правилу «трех сигма».

Учитывая, что появление значительной по абсолютной величине ошибки маловероятно, так как

$$P \{ -3\sigma_x \leq x - m_x^* \leq 3\sigma_x \} \approx 0,997,$$

где $3\sigma_x$ — предельная ошибка измерений, то все $|x_i - m_x^*|$, превышающие величину $3\sigma_x$, можно отнести к категории грубых и из дальнейшего рассмотрения их исключить. Однако подобный подход к отработке результатов измерений требует особо тщательного анализа, так как при малой выборке оценить правильно предельную погрешность достаточно трудно.

Для отбраковки грубых ошибок можно также воспользоваться критерием Смирнова. При этом необходимо подсчитать среднюю арифметическую величину

$$\bar{x}_n^* = \frac{1}{n} \sum_{i=1}^n x_i$$

и эмпирическое среднее квадратическое отклонение

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - x_n^*)^2}$$

Затем необходимо найти отношение абсолютной величины разности между сомнительным результатом измерений x^* и средним арифметическим значением к величине S_n , т. е. рассчитать статистику:

$$\xi = |x_* - x_n^*| / S_n$$

Если для данного числа измерений n и выбранной надежности $P=1-\beta$ величина ξ превосходит критическое значение $\xi_n(P)$, то с вероятностью, большей P , можно считать сомнительный результат измерения грубой ошибкой. В табл. 3 приведены критические границы для различных β .

Таблица 3

n	Критические значения ξ_β								
	β								
	0,0005	0,001	0,002	0,005	0,01	0,02	0,05	0,10	0,20
3	1,414	1,414	1,414	1,414	1,414	1,414	1,414	1,412	1,406
4	1,732	1,732	1,731	1,730	1,728	1,723	1,710	1,689	1,645
5	1,996	1,994	1,990	1,982	1,972	1,955	1,917	1,869	1,791
6	2,219	2,212	2,203	2,183	2,161	2,130	2,067	1,996	1,894
7	2,408	2,395	2,377	2,344	2,310	2,265	2,182	2,093	1,974
8	2,568	2,547	2,521	2,476	2,431	2,374	2,273	2,172	2,041
9	2,704	2,677	2,643	2,586	2,532	2,464	2,349	2,238	2,097
10	2,822	2,788	2,747	2,680	2,616	2,540	2,414	2,294	2,146
11	2,925	2,884	2,837	2,760	2,689	2,606	2,470	2,343	2,190
12	3,015	2,969	2,915	2,830	2,753	2,663	2,519	2,387	2,229
13	3,098	3,044	2,984	2,892	2,809	2,713	2,563	2,426	2,264
14	3,167	3,111	3,046	2,947	2,859	2,759	2,602	2,461	2,297
15	3,232	3,171	3,102	2,997	2,905	2,800	2,638	2,494	2,327
16	3,290	3,225	3,152	3,042	2,946	2,837	2,670	2,523	2,354
17	3,343	3,274	3,198	3,083	2,983	2,871	2,701	2,551	2,380
18	3,392	3,320	3,240	3,120	3,017	2,903	2,728	2,577	2,404
19	3,437	3,361	3,278	3,155	3,049	2,932	2,754	2,601	2,426
20	3,478	3,400	3,314	3,187	3,079	2,959	2,779	2,623	2,447
22	3,552	3,469	3,378	3,245	3,132	3,008	2,823	2,664	2,486
24	3,616	3,529	3,434	3,295	3,179	3,051	2,862	2,701	2,521
26	3,673	3,582	3,483	3,340	3,220	3,089	2,897	2,734	2,553
28	3,724	3,629	3,528	3,380	3,258	3,124	2,929	2,764	2,582
30	3,769	3,672	3,567	3,416	3,291	3,156	2,958	2,792	2,609
35	3,866	3,762	3,652	3,494	3,364	3,224	3,022	2,853	2,668
40	3,943	3,835	3,720	3,557	3,424	3,281	3,075	2,904	2,718
45	4,007	3,896	3,778	3,610	3,474	3,329	3,120	2,948	2,762
50	4,062	3,948	3,827	3,656	3,518	3,370	3,160	2,987	2,800

Если $n > 50$, то верхние β -процентные точки статистики $\xi(m_{x^*}, \sigma_{x^*})$ можно рассчитывать при $\beta \leq 0,2$ по формулам:

$$\xi_\beta = u \sqrt{\frac{2(n-1)}{2n-5+u^2+(3+u^2+2u^4)(1/6(2n-5))}};$$

$$u = \Psi(1 - \beta/2n),$$

где $\Psi(\beta, n)$ — функция, обратная функции нормального распределения.

Пример 1. Пусть в результате натурных испытаний получена выборка x_1, x_2, \dots, x_n , численные значения элементов которой совпадают с первой группой нормально распределенных чисел. Указанные числа выписаны и представлены в табл 4.

Таблица 4

i	x_i	i	x_i	i	x_i	i	x_i	i	x_i
1	0,464	11	2,455	21	-0,068	31	-0,288	41	0,241
2	0,060	12	-0,531	22	0,543	32	0,187	42	0,022
3	1,486	13	-0,634	23	0,926	33	0,785	43	-0,853
4	2,022	14	1,279	24	0,571	34	0,0194	44	-0,501
5	1,394	15	0,046	25	2,945	35	-0,258	45	0,439
6	0,137	16	-0,323	26	0,296	36	1,298	46	-0,957
7	-2,526	17	-0,194	27	-1,558	37	-1,190	47	0,525
8	-0,354	18	0,697	28	1,375	38	-0,963	48	-1,865
9	-0,472	19	3,521	29	-1,851	39	1,192	49	-0,273
10	-0,555	20	0,321	30	1,974	40	0,412	50	-0,035

Докажем, что в результатах 19-го эксперимента не допущено грубых ошибок.

Для решения поставленной задачи рассчитаем выборочные значения:

а) математического ожидания

$$m_x^* = 0,211;$$

б) дисперсии

$$\sigma_x^{*2} = 1,348^2.$$

На основании этих значений величина статистики:

$$\xi_{50}(m_x^*, \sigma_x^*) = 2,851.$$

Если принять вероятность ложной отбраковки анализируемых результатов, равной 0,05, то по табл. 3 при заданном $n = 50$ нетрудно найти ей соответствующее значение $\xi_{\beta=0,05} = 3,16$.

Сопоставляя выборочное значение $\xi_{50}(m_x^*, \sigma_x^*)$ с критическим ξ_β , приходим к следующему выводу: так как $\xi_{50} < \xi_{\beta=0,05}$, то нет никаких оснований считать результаты 19-го эксперимента аномальными среди всех остальных результатов.

Нетрудно убедиться в том, что этот вывод справедлив и для $\beta < 0,10$.

Пример 2. Пусть для $n=15$ независимых равнооточных измерений некоторой величины получено среднее арифметическое значение, равное $m_x^* = 10,17$ и эмпирическое значение среднего квадратического отклонения $\sigma_x^* = 0,73$.

Пусть значение $x_{15} = 11,5$ будет сомнительным. Необходимо, используя статистические методы, решить, можно ли отбраковать этот результат?

Находим значение

$$\xi_{15}(m_x^*, \sigma_x^*) = 1,82.$$

Для $n=15$ и $1 - \beta = 0,95$ критическим значением ξ_β является величина 2,638, что свидетельствует о том, что данный результат не следует относить к числу грубых ошибок.

1.6.3. Оценка дисперсий и математических ожиданий

В результате измерений получают некоторую выборку, состоящую из совокупности значений $x_1, x_2, x_3, \dots, x_n$. В общем случае каждый замер отличается от других замеров и их отклонение от истинного значения искомого параметра носит случайный характер (при этом предполагается, что результаты измерений уже свободны от грубых и систематических ошибок).

Чтобы по данной совокупности измерений получить приближенное значение величины x , необходимо найти центр группирования последней и оценить рассеивание всех рассматриваемых результатов относительно этого центра. В качестве числовых характеристик центра группирования обычно принимают математическое ожидание, медиану и моду. Правда, последние две характеристики в этих целях используют сравнительно редко. На практике для характеристики центра группирования определяют эмпирическую числовую характеристику, среднюю арифметическую величину, полученную по совокупности значений x_1, x_2, \dots, x_n . Среднее арифметическое значение x^* в случае равнооточных измерений находят по формуле:

$$x^* = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1)$$

Как следует из обоснований, приводимых в математической статистике, такая оценка является несмещенной и состоятельной. Если ошибки измерения подчинены нормальному закону распределения, то

эта оценка будет к тому же и эффективной. Последнее предположение о нормальном характере распределения ошибок измерения на практике в большинстве случаев оправдано.

Если же распределение ошибок измерений подчиняется другому закону, то эффективной оценкой может быть другая статистика. Так, для равномерного распределения ошибок измерений медиана является более эффективной оценкой, чем арифметическая средняя.

В случае неравнооточных измерений в качестве оценки искомой величины x принимают взвешенное среднее арифметическое значение:

$$x^* = \frac{p_1 x_1 + p_2 x_2 + \dots + p_n x_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}, \quad (2)$$

где p_i — веса измерений, принимаемые обычно обратно пропорциональными значениям дисперсий ошибок, т. е.

$$p_i = 1/\sigma_i^2, \quad i = 1, 2, \dots, n. \quad (3)$$

Вследствие влияния различных случайных факторов данные наблюдений можно охарактеризовать определенным рассеиванием результатов измерений. В качестве характеристики рассеивания случайных величин при обработке результатов испытаний обычно принимают эмпирическую дисперсию:

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - x^*)^2. \quad (4)$$

Приведенное выражение для эмпирической дисперсии позволяет получить несмещенную, состоятельную и асимптотически эффективную оценку. Последнее означает, что при n , стремящемся к бесконечности, отношение дисперсии $D\{S_x^2\}$ к минимально возможной ее величине, неограниченно приближается к 1.

Для равнооточных и независимых измерений дисперсию оценки можно охарактеризовать выражением:

$$D\{x^*\} = D\left\{\sum_{i=1}^n x_i/n\right\} = \frac{1}{n^2} \sum_{i=1}^n D\{x_i\} = \frac{D\{x_i\}}{n}. \quad (5)$$

Таким образом, средняя квадратическая ошибка средней арифметической величины может быть представлена в виде:

$$\sigma_{x^*} = \sigma_{x_i} / \sqrt{n}, \quad (6)$$

т. е. средняя арифметическая величина x^* в \sqrt{n} раз точнее отдельного измерения x_i ($i = 1, 2, \dots, n$). В случае неравнооточных измерений

необходимо учитывать веса последних. Взвешенное эмпирическое значение среднего квадратического отклонения величин x_1, x_2, \dots, x_n от их среднего значения x^* при этом можно определить по формуле:

$$S_x = \sqrt{\left[\sum_{i=1}^n p_i (x_i - x^*)^2 \right] / \sum_{i=1}^k p_i}, \quad (7)$$

где p_i — веса измерений.

Если же имеют k серий наблюдений, в которых было получено соответственно n_1, n_2, \dots, n_k количество измерений, то оценку дисперсии можно получить в виде средней взвешенной величины из эмпирических дисперсий по сериям:

$$S_x^2 = \frac{(n_1 - 1) S_{x_1}^2 + (n_2 - 1) S_{x_2}^2 + \dots + (n_k - 1) S_{x_k}^2}{(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1)}.$$

Для эмпирической оценки среднего квадратического отклонения величины x^* в случае неравноточных измерений можно воспользоваться формулой:

$$S_{x^*} = \sqrt{\left[\sum_{i=1}^n p_i (x_i - x^*)^2 \right] / \left[(n - 1) \sum_{i=1}^n p_i \right]}. \quad (8)$$

Рассуждая аналогично, находят погрешность, допускаемую при оценке среднего квадратического отклонения σ_x . В математической статистике доказывается, что эмпирическая дисперсия распределена по нормальному закону и характеризуется средним квадратическим отклонением:

$$\sigma(S_x^2) = \sigma_{x_i} / \sqrt{2(n-1)}.$$

Рассмотренные эмпирические характеристики представляют собой оценки искомых параметров в виде чисел, характеризующих распределение случайных величин в данной совокупности наблюдений. Такую оценку параметров называют *точечной*. Кроме этого способа оценки параметров на практике прибегают к другому способу, основанному на определении интервала, накрывающего истинное значение искомого параметра с заданной вероятностью. Эту оценку называют *интервальной*, или *доверительной*.

Например, доверительная оценка для математического ожидания в симметричном случае может быть представлена в виде:

$$x^* - \epsilon < x < x^* + \epsilon \text{ или } |x - x^*| < \epsilon. \quad (9)$$

Доверительный интервал $I_\beta = (x^* - \epsilon, x^* + \epsilon)$ определяет область возможных значений полученной при обработке средней арифметической величины x^* для данного параметра x . Причем попадание x^* в

эту область гарантируется с заданной доверительной вероятностью, т. е.

$$P(|x - x^*| < \epsilon_\beta) = \beta,$$

где x^* — среднее арифметическое значение, полученное в результате обработки данных наблюдений; x — истинное значение искомого параметра; ϵ_β — наперед заданная положительная величина, определяемая по доверительной вероятности (надежности оценки), принимаемой равной величине β .

В случае равноточных измерений, когда заранее известна точность измерений σ_x , доверительную оценку математического ожидания можно представить через функцию Лапласа, предположив, что ошибки измерения подчинены нормальному закону. Тогда, принимая во внимание, что

$$P(|x - x^*| < \epsilon_\beta) = \Phi\left(\frac{\epsilon_\beta}{\sqrt{2}\sigma_{x^*}}\right),$$

из уравнения

$$\Phi\left(\frac{\epsilon_\beta}{\sqrt{2}\sigma_{x^*}}\right) = \beta$$

можно найти ϵ_β :

$$\epsilon_\beta = \sqrt{2}\sigma_{x^*} \Phi^{-1}(\beta) = \sqrt{2} \frac{\sigma_x}{\sqrt{n}} \Phi^{-1}(\beta),$$

где $\Phi^{-1}(\beta)$ — функция, обратная функции Лапласа.

Если положить

$$t_\beta = \sqrt{2} \Phi^{-1}(\beta),$$

то можно воспользоваться таблицей, в которой приведены значения t_β в зависимости от выбранной величины β (табл. 1).

Таблица 1

β	t_β	β	t_β	β	t_β	β	t_β
0,80	1,282	0,86	1,475	0,92	1,750	0,98	2,325
0,81	1,310	0,87	1,513	0,93	1,810	0,99	2,576
0,82	1,340	0,88	1,554	0,94	1,880	0,999	3,290
0,83	1,371	0,89	1,597	0,95	1,960		
0,84	1,404	0,90	1,643	0,96	2,053		
0,85	1,439	0,91	1,694	0,97	2,169		

Таким образом, доверительная оценка для математического ожидания x будет иметь вид:

$$|x - x^*| < t_{\beta} (\sigma_x / \sqrt{n}). \quad (10)$$

В случае, если заранее неизвестна точность измерений, но известно, что измерения равноточны, доверительная оценка математического ожидания может быть получена по формуле

$$|x - x^*| < t(\beta, k) (S_x / \sqrt{n}), \quad (11)$$

где

$$S_x = \sqrt{\sum_{i=1}^n (x_i - x^*)^2 / (n-1)}$$

— эмпирическая оценка средней квадратической ошибки; n — число измерений; $k = n-1$ — число степеней свободы; $p=1-\beta$ — заданная доверительная вероятность, $t(\beta, k)$ — величина, определяемая по табл. 2 в зависимости от β и k .

Таблица 2

Значения $t(\beta, k)$											
k	$1-\beta$			k	$1-\beta$			k	$1-\beta$		
	0,9	0,95	0,99		0,9	0,95	0,99		0,9	0,95	0,99
4	2,132	2,776	4,604	15	1,753	2,131	2,947	80	1,664	1,990	2,639
5	2,015	2,571	4,032	20	1,725	2,086	2,845	90	1,662	1,987	2,632
6	1,943	2,447	3,707	30	1,697	2,042	2,750	100	1,660	1,984	2,626
7	1,895	2,365	3,499	40	1,684	2,021	2,704		1,645	1,960	2,576
8	1,860	2,306	3,355	50	1,676	2,008	2,677				
9	1,833	2,262	3,250	60	1,671	2,000	2,660				
10	1,812	2,228	3,169	70	1,667	1,995	2,648				

В случае необходимости для получения величины $t(\beta, k)$ при промежуточных значениях аргумента k можно воспользоваться линейной интерполяцией.

При получении доверительных оценок следует учитывать, что формула (11) не может применяться, если имеет место обработка результатов наблюдений, сгруппированных в интервальный ряд. Доверительные границы при обработке интервального ряда можно приближенно оценить по правилу «трех сигма».

При неравноточных измерениях с неизвестной дисперсией σ_x доверительная оценка математического ожидания искомого параметра x может быть получена по формуле:

$$|x - x^*| < t(\beta, k) S_{x^*}, \quad (12)$$

где $x^* = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}$

$$S_{x^*} = \sqrt{\left[\sum_{i=1}^n p_i (x_i - x^*)^2 \right] / \left[(n-1) \sum_{i=1}^n p_i \right]}$$

S_{x^*} — эмпирическая средняя квадратическая ошибка величины x^* ; $k=n-1$ — число степеней свободы; p_i — веса измерений; $P=1-\beta$ — доверительная вероятность (значение находят из табл. 2).

Для построения доверительной оценки среднего квадратического отклонения нормально распределенной величины можно прибегнуть к определению доверительных границ в долях от полученной эмпирической оценки S_x . Тогда

$$P(z_1 S_x < \sigma_x < z_2 S_x) = P,$$

т. е. с вероятностью P можно ожидать выполнения неравенства

$$z_1 S_x < \sigma_x < z_2 S_x, \quad (13)$$

где $z_1 S_x$ и $z_2 S_x$ — определяют доверительные границы для σ_x .

Значения z_1 и z_2 для заданных величин P и n можно найти из табл. 3.

Таблица 3

n	$P=0,90$		$P=0,95$		$P=0,99$	
	z_1	z_2	z_1	z_2	z_1	z_2
5	0,649	2,429	0,599	2,875	0,519	4,394
6	0,672	2,090	0,624	2,453	0,546	3,484
7	0,690	1,916	0,644	2,202	0,569	2,979
8	0,705	1,797	0,661	2,035	0,588	2,660
9	0,718	1,711	0,675	1,916	0,604	2,440
10	0,729	1,645	0,688	1,826	0,618	2,274
15	0,769	1,460	0,732	1,577	0,669	1,853
20	0,794	1,370	0,760	1,460	0,702	1,666
30	0,826	1,280	0,796	1,344	0,744	1,487
40	0,845	1,232	0,819	1,284	0,722	1,397
50	0,859	1,202	0,835	1,246	0,791	1,341
70	0,879	1,164	0,857	1,200	0,818	1,274
100	0,896	1,134	0,878	1,162	0,844	1,220
200	0,921	1,102	0,911	1,109	0,885	1,147

Интерполяция начений z_1 и z_2 допустима лишь для промежуточных значений n .

Часто встречающаяся на практике задача, связанная с оценкой вероятности по частоте, по существу представляет собой частный случай получения доверительной оценки для математического ожидания. Рассматриваемая случайная величина может принимать лишь значения $x=1$, если событие совершилось, и $x=0$, если событие не совершилось.

Если математическое ожидание величины x равно p а дисперсия равна qp , где $q=1-p$, то по результатам n независимых опытов можно

вычислить частоту появления рассматриваемого события представляющую оценку искомой вероятности, т. е.

$$p^* = \left(\sum_{i=1}^n x_i \right) / n,$$

где $\sum_{i=1}^n x_i$ — число появлений рассматриваемого события; n — число испытаний.

Дисперсия величины p^*

$$D[p^*] = (pq)/n.$$

Для получения доверительной оценки в виде

$$p_2(p^*) < p < p_1(p^*)$$

необходимо определить доверительные границы p_1 и p_2 по формулам:

$$p_1 = \left[p^* + \frac{1}{2} \cdot \frac{t_\beta^2}{n} - t_\beta \sqrt{\frac{p^*(1-p^*)}{n} + \frac{1}{4} \frac{t_\beta^2}{n^2}} \right] / \left(1 + \frac{t_\beta^2}{n^2} \right);$$

$$p_2 = \left[p^* + \frac{1}{2} \frac{t_\beta^2}{n} + t_\beta \sqrt{\frac{p^*(1-p^*)}{n} + \frac{1}{4} \frac{t_\beta^2}{n^2}} \right] / \left(1 + \frac{t_\beta^2}{n^2} \right).$$

Если $n > 100$ и значения np и nq порядка 10 и более, приведенные формулы для p_1 и p_2 можно упростить и применять в следующем

$$p_1 = p^* - t_\beta \sqrt{[p^*(1-p^*)]/n};$$

$$p_2 = p^* + t_\beta \sqrt{[p^*(1-p^*)]/n}.$$

При малом числе опытов указанный подход к получению доверительной оценки неприемлем, так как распределение частоты уже нельзя считать подчиненным нормальному закону. В этом случае определяя доверительные границы, необходимо учитывать то, что частота подчиняется биномиальному распределению. Для приближенной оценки доверительных интервалов на рис. 1 приведены кривые, определяющие область (в зависимости от числа испытаний n и доверительной вероятности β) возможных значений частоты p .

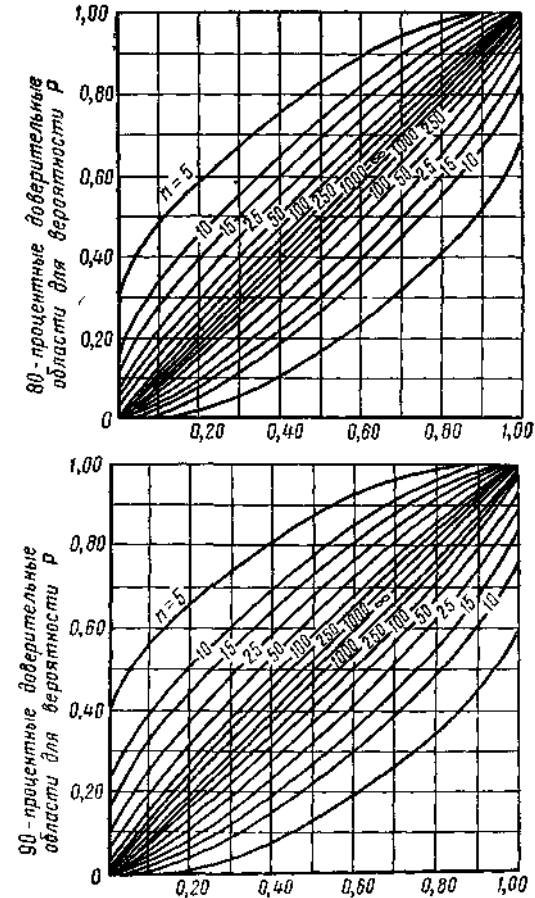


Рис.1

Для нахождения верхней и нижней границы доверительного интервала необходимо значение частоты p^* отложить по оси абсцисс и из полученной точки провести параллельно оси ординат прямую до пересечения с кривыми, соответствующими по параметру n числу испытаний в данном опыте. Значения ординат точек пересечения с этими кривыми и будут верхней и нижней границей доверительного интервала.

Частным случаем задачи, связанной с получением доверительной оценки вероятности по полученной из опыта частоте, является задача, когда $p^* = 0$, что свидетельствует о малости самого значения ве-

роятности. Подобные задачи могут возникнуть при оценке вероятности безотказной работы для высоконадежной аппаратуры. Ранее изложенные методы построения доверительного интервала здесь неприменимы.

Для получения значения p_2 (верхней границы доверительного интервала) можно воспользоваться формулой:

$$p_2 = 1 - \sqrt[n]{1 - \beta};$$

где n — число испытаний; β — заданная доверительная вероятность.

Нижняя граница доверительного интервала p_1 очевидно равна нулю.

Метод вычисления средней арифметической величины и эмпирической дисперсии с помощью «ложного нуля». Этот метод удобно применять в тех случаях, когда приходится иметь дело с большими числами. Введение «ложного нуля» намного упрощает и облегчает проведение вычислительных операций, способствует уменьшению ошибок и просчетов при вычислениях.

Формулу (1) можно представить в виде

$$x^* = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n x_i}{n} - c + c = \frac{\sum_{i=1}^n (x_i - c)}{n} + c = b_1 + c,$$

где b_1 — статистический момент первого порядка; c — постоянная («ложный нуль»).

При обработке результатов наблюдений значения величины c выбирают таким образом, чтобы разности $(x_i - c)$ имели относительно небольшие значения.

Аналогично формулу для эмпирической дисперсии можно представить в виде:

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - x^*)^2}{n} = \frac{\sum_{i=1}^n (x_i - c)^2}{n} - (x^* - c)^2 = b_2 - b_1^2,$$

где b_2 и b_1 соответственно моменты второго и первого порядков.

Пример. Пусть в результате двадцати одного замера ($n=21$), проведенных в одинаковых условиях, были получены следующие значения расстояния x_i (в километрах) между двумя пунктами: 844,4; 852,2; 848,8; 844,4; 841,7; 846,8; 847,2; 849,8; 845,8; 848,7; 854,2; 840,4; 848,7; 846,2; 847,3; 844,8; 848,8; 849,6; 844,9; 844,0.

Рассчитать дисперсию и математическое ожидание.

Если принять в качестве «ложного нуля» величину $c=844$ км, тогда средняя арифметическая величина

$$x^* = \frac{\sum_{i=1}^n (x_i - c)}{n} + c = 846,7$$

$$n = 21.$$

Эмпирическая дисперсия для данной выборки

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - c)^2}{n} - (x^* - c)^2 = 11,5.$$

Найдем доверительные оценки для искомого параметра и его дисперсии с надежностью $P=0,9$. Чтобы найти доверительные границы по формуле (11), необходимо с помощью табл. 2 определить множитель $t(\beta, k)$, где P — заданная величина надежности ($P=1-\beta=0,9$), а $k=n-1$ — число степеней свободы ($k=20$). Получим $t(\beta, k)=1,725$, а доверительная оценка

$$|x - 846,7| < 1,725 \cdot 11,5 / \sqrt{21} = 4,1.$$

Таким образом, с надежностью 0,9 можно считать, что истинное значение расстояния x заключено в интервале:

$$842,6 < x < 850,8.$$

Для нахождения доверительных границ по формуле (13) при заданных величинах P и n найдем из табл. 3 значения z_1 и z_2 . Тогда, среднее квадратическое отклонение с надежностью 0,9 заключено в интервале

$$0,829 \sqrt{11,5} < \sigma_x < 1,361 \sqrt{11,5},$$

$$2,7 < \sigma_x < 4,5.$$

1.6.4. Анализ качества измерений от различных источников

В процессе испытаний сложных систем для регистрации процессов обычно привлекают большое количество разнотипной измерительной техники. В связи с этим при обработке результатов наблюдений возникает необходимость в доказательстве возможности совместного использования всех выборок при оценке показателей реальной системы. Комплекс задач, решаемых при рассмотрении подобных

вопросов, в большинстве практических случаев сводится к задачам, связанным с определением характера распределения случайных величин в каждой выборке и с проверкой некоторой статистической эквивалентности двух выборок.

Рассмотрим на конкретных примерах методы решения задач, возникающих при анализе качества измерений от различных источников.

Задача проверки гипотезы о равенстве средних арифметических величин. Предположим, что имеются две выборки результатов наблюдений над одним объектом, по которым получены оценки двух центров рассеяния измерений в виде средних арифметических величин x^* и z^* . Естественно то, что возникает задача сравнения найденных числовых характеристик с целью проверки гипотезы о равенстве полученных средних величин. В случае утвердительного ответа на последний вопрос можно считать, что различие в значениях x^* и z^* обусловлено случайными ошибками. Если же разность между x^* и z^* оказывается достаточно большой, то различие в их значениях не имеет случайного характера и свидетельствует о существенном отклонении в условиях экспериментов.

Проверку гипотезы осуществляют с помощью различных критериев. Для малых значений вероятностей β , называемых уровнем значимости, можно определить критическую область значений критериев проверки, попадание в которую означает, что гипотезу следует отвергнуть. При этом обычно рассматривают две возможные ошибки: *первого рода* (бракует правильную гипотезу); *второго рода* (принимают ложную гипотезу). Чтобы избежать появления ошибки первого рода, надо выбрать достаточно малый уровень значимости β . С уменьшением значения β критическая область критерия проверки сужается и вероятность отвергнуть правильную гипотезу вполне естественно уменьшается. Но при этом возрастает вероятность принять неверную гипотезу за правильную, т. е. возрастает вероятность допустить ошибку второго рода.

При анализе результатов моделирования важно убедиться в том, что оценки выходных показателей, рассчитываемых с помощью математических моделей, являются несмещенными. Практически для сложных систем такая проверка сводится к определению условий, при которых оценки x^* , z^* , рассчитанные по результатам натуральных испытаний x_1, x_2, \dots, x_n и по результатам моделирования z_1, z_2, \dots, z_m , следует считать значимо отличными друг от друга. Конечная цель подобной проверки состоит в выработке заключения, позволяющего утверждать, что $E\{x\} \neq E\{z\}$, если только модуль разности выборочных средних $D = x^* - z^*$ превосходит некоторую заранее расчи-

танную границу. Для широко распространенного случая, когда выборки $x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_m$ независимы и получены из некоторых нормальных генеральных совокупностей, проверка нулевой гипотезы $H_0: E\{x\} = E\{z\}$ относительно альтернативы $H_1: E\{x\} \neq E\{z\}$ при неизвестных, но равных друг другу дисперсиях $\sigma_x^2 = \sigma_z^2$, может быть проведена на основании двустороннего критерия Стьюдента.

Критерий Стьюдента основан на распределении статистики

$$t = D/S, \tag{1}$$

где D — разность выборочных средних арифметических величин x^* и z^* :

$$x^* = \frac{1}{n} \sum_{i=1}^n x_i; \quad z^* = \frac{1}{m} \sum_{i=1}^m z_i;$$

S^2 — случайная величина, равная

$$S^2 = \frac{(n-1)S_x^2 + (m-1)S_z^2}{n+m-2}; \tag{2}$$

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - x_n^*)^2;$$

$$S_z^2 = \frac{1}{m-1} \sum_{i=1}^m (z_i - z_m^*)^2.$$

Случайная величина S^2 имеет χ^2 — распределение с числом степеней свободы $f = n + m - 2$, что следует из (2), если принять во внимание, что для нормальных генеральных совокупностей D и S^2 зависимы соответственно, как функции среднего арифметического и выборочной дисперсии.

Предполагая, что $\sigma_x^2 = \sigma_z^2$, уравнение (1) можно привести к виду

$$t = \frac{x^* - z^*}{\sqrt{(n-1)S_x^2 + (m-1)S_z^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}. \tag{3}$$

Из (3) следует, что случайная величина t при условии справедливости нулевой гипотезы $H_0: E\{x\} = E\{z\}$ имеет распределение Стьюдента с числом степеней свободы $f = n + m - 2$ и параметрами, которые определяются объемами выборок и полученными численными значениями x_i ($i = 1, 2, \dots, n$); z_j ($j = 1, 2, \dots, m$).

Практическое применение критерия Стьюдента в рассматриваемом случае заключается в последовательном выполнении следующих операций:

1. Расчет величины статистики t с использованием формулы (3);
2. Выбор уровня значимости 2β и определение по таблицам распределения Стьюдента (см. табл. 1) соответствующих границ критической области $t_{2\beta, f}$ (при определении границ нужно помнить, что число степеней свободы равно $f = n + m - 2$);
3. Сравнение значений t с величинами $t_{2\beta, f}$ (это позволяет выработать заключение о справедливости проверяемой нулевой гипотезы: нулевая гипотеза $H_0: E\{x\} = E\{z\}$ справедлива, когда $|t| < t_{2\beta, f}$; в противном случае, гипотеза отвергается).

Если по условиям задачи нужно убедиться в справедливости соотношения $E\{x\} \geq E\{z\}$ ($E\{x\} \leq E\{z\}$), то нулевую гипотезу записывают в виде $H_0: E\{x\} \geq E\{z\}$ ($H_0: E\{x\} \leq E\{z\}$) и ее проверку осуществляют на основании одностороннего критерия Стьюдента. В последнем случае вероятность ошибочно отвергнуть правильную гипотезу не превышает β .

Задача сравнения точности измерений. Полученные в результате обработки эмпирические дисперсии для двух выборок могут отличаться друг от друга, так как по сути они являются лишь приближенными оценками. В связи с этим может возникнуть задача проверки гипотезы $H_0: \sigma_x^2 = \sigma_z^2$ о равнозначности измерений в рассматриваемых выборках, несмотря на имеющиеся различия в эмпирических дисперсиях.

Доверительные границы для двустороннего критерия Стьюдента представлены в табл. 1.

Для проверки гипотезы $H_0: \sigma_x^2 = \sigma_z^2$ можно воспользоваться критерием, основанным на распределении статистики:

$$F = \max [S_x^2/S_z^2, S_z^2/S_x^2]; \tag{4}$$

где

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - x^*)^2;$$

$$S_z^2 = \frac{1}{m-1} \sum_{i=1}^m (z_i - z^*)^2.$$

Статистики S_x^2, S_z^2 имеют χ^2 — распределение с числами степеней свободы, равными соответственно $k_1 = n - 1; k_2 = m - 1$.

Таблица 1

f	Критические значения $t_{2\beta}$				
	2β				
	0,10	0,05	0,02	0,01	0,005
1	0,314	12,706	31,821	63,657	123,321
2	2,320	4,303	6,965	9,925	14,089
3	2,333	3,182	4,541	5,841	7,453
4	2,192	2,776	3,474	4,604	5,597
5	2,015	2,571	3,365	4,032	4,773
6	1,943	2,447	3,143	3,707	4,317
7	1,895	2,365	2,998	3,499	4,029
8	1,860	2,306	2,896	3,553	3,833
9	1,833	2,262	2,821	3,250	3,690
10	1,812	2,228	2,764	3,169	3,581
12	1,782	2,179	2,681	3,055	3,428
14	1,761	2,145	2,624	2,977	3,326
16	1,746	2,120	2,583	2,921	3,252
18	1,734	2,101	2,552	2,878	3,193
20	1,725	2,086	2,528	2,845	3,153
22	1,717	2,074	2,508	2,819	3,119
24	1,711	2,064	2,492	2,797	3,032
26	1,706	2,056	2,479	2,779	3,067
28	1,701	2,048	2,467	2,763	3,047
30	1,697	2,042	2,457	2,750	3,030
40	1,684	2,021	2,423	2,704	2,971
50	1,676	2,009	2,403	2,678	2,937
60	1,671	2,000	2,390	2,660	2,915
80	1,664	1,990	2,374	2,639	2,887
100	1,660	1,984	3,365	2,626	2,871
00	1,645	1,960	2,326	2,576	2,807

Эти результаты показывают, что для F-критерия числа степеней свободы нужно определять по формулам

$$\left. \begin{matrix} k_1 = n - 1 \\ k_2 = m - 1 \end{matrix} \right\} \text{ при } S_x^2 > S_z^2; \tag{5a}$$

$$\left. \begin{matrix} k_1 = m - 1; \\ k_2 = n - 1 \end{matrix} \right\} \text{ при } S_z^2 > S_x^2. \tag{5б}$$

Вероятности наступления событий $P(F > F_{\beta}^+) = \beta$ затабулированы в виде таблиц для различных k_1, k_2, β . Эти таблицы позволяют по заданному уровню значимости и найденным числам степеней свободы k_1, k_2 рассчитать правые критические точки F_{β}^+ для одностороннего F-критерия. Уровень значимости одностороннего F-критерия равен β . Такой критерий применяют при проверке гипотез вида $H_0: \sigma_x^2 = \sigma_z^2$ относительно альтернатив $H_1: \sigma_x^2 > \sigma_z^2$ или $H_1: \sigma_x^2 < \sigma_z^2$.

Если организовать проверку гипотезы $H_0: \sigma_x^2 = \sigma_z^2$ дважды при альтернативах $H_1: \sigma_x^2 > \sigma_z^2$ и $H_1: \sigma_x^2 < \sigma_z^2$, то получим двусторонний F -критерий с уровнем значимости 2β . Так как вероятности наступления событий $P(F > F_{\beta}^+)$ и $P(\frac{1}{F} < \frac{1}{F_{\beta}^+} = F^-)$ равны между собой, то проверку гипотезы $H_0: \sigma_x^2 = \sigma_z^2$ относительно альтернативы

$$H_1: \sigma_x^2 \neq \sigma_z^2$$

можно осуществлять на основании двустороннего критерия по выполнению соотношения $F < F_{\beta}^+$. Если окажется, что рассчитанное значение F меньше правой границы F_{β}^+ , то гипотезу $H_0: \sigma_x^2 = \sigma_z^2$

следует считать справедливой. При таком правиле проверки гипотезы H_0 вероятность ее отвергнуть, когда она правильная, не превышает 2β .

Пример 1. Пусть в результате натуральных испытаний и моделирования получены две выборки, относительно которых известно, что они независимы и принадлежат некоторым нормально распределенным генеральным совокупностям. Случайные значения, реализовавшиеся в указанных экспериментах, приведены в табл. 2.

Таблица 2

i	1	2	3	4	5	6	7	8	9	10
x_i	-1,324	-0,726	-1,618	1,695	0,790	1,792	1,771	-1,438	-0,294	-1,966
z_i	-1,621	0,163	1,786	2,140	0,064	0,789	-0,011	-0,463	-1,210	1,157

На основании полученных результатов опыта нужно установить справедливость предположения о том, что математические ожидания генеральных совокупностей, из которых получены выборки x_i, z_i ($i=1, 2, \dots, 10$), равны между собой.

Так как априори неизвестен знак разности $E\{x\} - E\{z\}$, то проверку нулевой гипотезы $H_0: E\{x\} = E\{z\}$ при альтернативе $H_1: E\{x\} \neq E\{z\}$ будем производить на основании двустороннего критерия t .

Чтобы воспользоваться указанным критерием, сначала проверим гипотезу о равенстве дисперсий $\sigma_x^2 = \sigma_z^2$. Для этого по (4) находим

$$x^* = \frac{\sum_{i=1}^n x_i}{n} = -0,232;$$

$$z^* = \frac{\sum_{i=1}^m z_i}{m} = +0,279$$

и далее рассчитываем

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - x^*)^2}{(n-1)} = 1,965;$$

$$S_z^2 = \frac{\sum_{i=1}^m (z_i - z^*)^2}{(m-1)} = 1,476.$$

В рассматриваемом случае получаем:

$$F = \max [S_x^* / S_z^2, S_z^* / S_x^2] = S_x^2 / S_z^2 = 1,331.$$

Задаваясь довольно высоким уровнем значимости $2\beta=0,1$, по табл. 5 находим верхнее критическое значение $F_{\beta}^+ = 2,440$. Так как $F < F_{\beta}^+$,

то нет основания считать, что дисперсии σ_x^2, σ_z^2 отличны друг от друга.

Далее определяем

$$\{x_n^* - z_m^*\} = 0,511;$$

$$|t| = 0,871;$$

$$f = 10 + 10 - 2 = 18.$$

Из табл. 1 для уровня значимости $2\beta = 0,10$, соответствующего проверяемой гипотезе $H_0: E\{x\} = E\{z\}$, находим $t_{2\beta, f} = 1,734$.

Ввиду того, что опытное значение критерия $|t| = 0,871$ получилось меньше $t_{2\beta, f} = 1,734$, то проверяемую нулевую гипотезу H_0 о равенстве математических ожиданий $E\{x\} = E\{z\}$ следует считать справедливой.

Если точно известно, что $\sigma_x^2 = \sigma_z^2$, вероятность отвергнуть правильную гипотезу $H_0: E\{x\} = E\{z\}$ не превосходит $2\beta=0,10$. Когда равенство дисперсий $\sigma_x^2 = \sigma_z^2$ устанавливается по тем же наблюдавшимся выборкам, то указать истинный уровень значимости критерия t очень трудно.

Пример 2. Измерение расстояния было проведено при экспериментах. В результате обработки первой выборки ($n=16$) получили $x_n^* = 855,1$ км и $S_x^2 = 12,5$ км², а по данным второй выборки ($m=21$) получили $z_m^* = 846,7$ км и $S_z^2 = 11$ км².

Проверить гипотезу о равенстве средних и эмпирических дисперсий. Подсчитаем значения $S=3,4; t=2,1; f=35$ и из табл. 1 найдем для $P=1-2\beta=0,99$ ($2\beta=0,01$), $t_{2\beta, f} = 2,725$. Подсчитанное значение t

меньше критического значения, что с надежностью $P=0,99$ не дает основания отвергать гипотезу о равенстве средних.

Значение F -критерия будет: $F=1,12$.

Пределы уклонений величин F можно найти по табл. 3, 4, 5, в которых соответственно даны доверительные границы для F -распределения с уровнями значимости 0,01; 0,05; 0,10. Откуда однопроцентный предел уклонения величин F_{β}^+ для значений $k_1 = 15$ и $k_2=20$ равен 3,09.

Следовательно, полученное значение F -критерия меньше критического значения, поэтому нет оснований отвергнуть гипотезу о равнозначности измерений.

Задача проверки гипотезы о тождественности законов распределения двух выборок. В процессе разработки математических моделей сложных систем часто анализу подвергают несколько схем их структурного построения. Тогда при достаточно высокой стоимости проигрыша одной реализации не удастся достичь такого объема выборок на модели, чтобы для каждого допустимого вариант;) ее построения можно было бы рассчитать с требуемой точность законы распределения оцениваемых показателей.

Таблица 3

k_2	Критические значения F_{β}^+										∞			
	1	2	3	4	5	6	7	8	9	10		15	20	30
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6157	6209	6260	936,6
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,43	99,45	99,47	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	26,87	26,69	26,50	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	15,20	14,02	13,84	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,888	9,553	9,379	9,020
6	13,74	10,92	9,779	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,559	7,396	7,228	6,880
7	12,25	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,314	6,155	5,992	5,649
8	11,26	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,515	5,359	5,198	4,859
9	10,56	8,021	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,256	4,962	4,808	4,649	4,310
10	10,04	7,559	6,532	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,558	4,405	4,247	3,909
15	8,683	6,359	5,417	5,893	4,556	4,318	4,141	4,004	3,895	3,805	3,522	3,372	3,214	2,868
20	8,096	5,843	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,088	2,938	2,778	2,421
30	7,562	5,390	4,510	4,018	3,699	3,473	3,304	3,173	3,066	2,979	2,700	2,549	2,386	2,006
	6,635	4,605	3,782	3,319	3,017	2,802	2,639	2,511	2,407	2,321	2,038	1,878	1,696	1,000

Таблица 4

k_2		Критические значения F_{β}^+													∞
		k_1													
		1	2	3	4	5	6	7	8	9	10	15	20	30	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	245,95	248,01	250,09	254,33	
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,429	19,446	19,462	19,496	
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	8,703	8,660	8,617	8,526	
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,965	5,858	5,802	5,846	5,628	
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,619	4,558	4,496	4,365	
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,089	4,060	3,938	3,874	3,808	3,669	
7	5,591	4,737	4,347	4,120	3,971	3,866	3,787	3,726	3,677	3,636	3,511	3,444	3,376	3,230	
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,218	3,150	3,079	2,928	
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,006	2,936	2,864	2,707	
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,845	2,774	2,700	2,538	
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,403	2,327	2,247	2,066	
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,203	2,124	2,039	1,843	
30	4,171	3,316	2,922	2,690	2,534	2,420	2,334	2,266	2,211	2,165	2,015	1,932	1,842	1,622	
	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,880	1,831	1,666	1,570	1,459	1,000	

Таблица 5

k_1		Критические значения F_{β}^+													∞
		k_2													
		1	2	3	4	5	6	7	8	9	10	15	20	30	∞
1	39,864	49,500	53,593	55,833	57,241	58,204	58,906	59,439	59,858	60,195	61,220	61,740	62,265	63,328	
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,425	9,441	9,458	9,491	
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,200	5,184	5,168	5,134	
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,870	3,843	3,817	3,761	
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,238	3,207	3,174	3,105	
6	3,776	3,463	3,289	3,181	3,108	3,055	3,015	2,983	2,958	2,937	2,871	2,876	2,800	2,722	
7	3,589	3,257	3,074	2,960	2,883	2,827	2,785	2,752	2,725	2,703	2,632	2,595	2,556	2,471	
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,464	2,425	2,383	2,293	
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,340	2,298	2,255	2,159	
10	3,285	2,924	2,732	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,244	2,201	2,155	2,055	
15	3,073	2,695	2,489	2,361	2,273	2,208	2,158	2,118	2,086	2,059	1,972	1,924	1,873	1,755	
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,845	1,794	1,738	1,607	
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819	1,722	1,667	1,606	1,456	
	2,705	2,303	2,084	1,945	1,847	1,774	1,717	1,630	1,631	1,599	1,487	1,421	1,342	1,000	

Указанные причины приводят к тому, что заключения о статистической совместимости результатов моделирования и натуральных испытаний нужно принимать по выборкам ограниченного объема.

На практике указанную задачу статистической проверки гипотез чаще всего рассматривают как задачу проверки адекватности законов распределения выборок, полученных при натуральных испытаниях x_1, x_2, \dots, x_n и моделировании z_1, z_2, \dots, z_m . В процессе решения подобных задач обычно предполагают, что выборки x_1, x_2, \dots, x_n и z_1, z_2, \dots, z_m независимы, а законы распределения их совокупностей $F_1(x), F_2(z)$, из которых получены анализируемые выборки, являются непрерывными функциями от своих аргументов x, z .

При такой интерпретации условий задачи естественно обозначить множество всех пар $(F_1(x), F_2(z))$ через \mathcal{F} , а подмножестве пар $(F_1(x), F_2(z))$ множества \mathcal{F} , для которых $F_1(x) \equiv F_2(z)$, через \mathcal{F}_0 . Тогда задачу проверки гипотезы о тождественности распределений $F_1(x) \equiv F_2(z)$ можно свести к выбору критерия, состоятельного для проверки гипотезы $(F_1(x) \equiv F_2(z)) \in \mathcal{F}_0$ относительно любой альтернативы $(F_1(x), F_2(z)) \in \mathcal{F} \setminus \mathcal{F}_0$, где символ $\mathcal{F} \setminus \mathcal{F}_0$ определяет множество всех пар, для которых $F_1(x) \neq F_2(z)$.

При проверке гипотез подобного класса широко применяют критерий Смирнова, который для рассматриваемых условий состоятелен и использует при определении количественной меры статистической совместимости результатов моделирования и натуральных испытаний распределение статистики

$$D_{n,m} = \sup_{x=z} |F_{1n}(x) - F_{2m}(z)|, \quad (6)$$

где $F_{1n}(x), F_{2m}(z)$ эмпирические функции распределения, рассчитанные по выборкам $x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_m$.

Н. В. Смирнов доказал, что вероятность появления событий $D_{n,m} > D_\beta = \lambda_\beta \sqrt{1/n + 1/m}$ определяется соотношением

$$\lim_{n, m \rightarrow \infty} P(D_{n,m} > D_\beta = \lambda_\beta \sqrt{1/n + 1/m}) = 1 - 2 \sum_{i=1}^{\infty} (-1)^i e^{-2i^2 \lambda_\beta^2}. \quad (7)$$

Для того чтобы уровень значимости критерия Смирнова был равен некоторому заданному числу 2β , т. е.

$$\lim_{n, m \rightarrow \infty} P(D_{n,m} > \lambda_\beta \sqrt{1/n + 1/m}) = 2\beta, \quad (8)$$

необходимо значения λ_β находить из уравнения

$$2 \sum_{i=1}^{\infty} (-1)^i e^{-2i^2 \lambda_\beta^2} = 1 - 2\beta. \quad (9)$$

Практическое применение критерия Смирнова заключается в последовательном выполнении следующих операций;

1. Расчет эмпирических функций распределения

$$F_{1n}(x), F_{2m}(z).$$

Для определения эмпирических функций распределения $F_{1n}(x), F_{2m}(z)$ необходимо упорядочить выборки $x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_m$ в порядке возрастания случайных величин, т. е. получить последовательности $x^{(1)}, x^{(2)}, \dots, x^{(n)}; z^{(1)}, z^{(2)}, \dots, z^{(m)}$. После этого используя полученные порядковые статистики $x^{(i)}$ и $z^{(j)}$ ($i = 1, 2, \dots, n, j = 1, 2, \dots, m$), рассчитать эмпирические функции распределения, например для выборки x_1, x_2, \dots, x_n по формулам:

$$F_{1n}(x) = \begin{cases} 0 & x \leq x^{(1)}; \\ i/n & \text{при } x^{(i)} \leq x < x^{(i+1)}; \\ 1 & x \geq x^{(n)}; \end{cases} \quad (10)$$

2. Определение верхней грани $D_{n,m}$ модуля разности эмпирических функций $F_{1n}(x), F_{2m}(z)$, т. е.

$$D_{n,m} = \sup_{x=z} |F_{1n}(x) - F_{2m}(z)|; \quad (11)$$

3. Определение при заданном уровне значимости 2β допустимых отклонений значений верхней грани D_β .

Высокая скорость сходимости ряда (7) при практических оценках позволяет ограничиться первым членом разложения и использовать для расчета D_β формулу

$$D_\beta = \sqrt{-\ln \beta / (1/n + 1/m) / 2}. \quad (12)$$

Значения D_β , найденные по (12), отличаются от истинных точных значений в сторону увеличения степени доверия о статистической совместимости результатов моделирования с результатами натуральных испытаний (увеличивается надежность критерия);

4. Сравнение значений D_β и $D_{n,m}$:

а) если $D_{n,m} > D_\beta$, то гипотеза H_0 о тождественности законов распределения $F_{1n}(x), F_{2m}(z)$ с доверительной вероятностью $1 - 2\beta$ отвергается;

б) если $D_{n,m} \leq D_\beta$, то гипотеза H_0 справедлива.

Иногда, исходя из целевого назначения разрабатываемой математической модели, необходимо и достаточно проверить гипотезу H_0 о

тождественности законов распределения результатов моделирования и натуральных испытаний на основании односторонних критериев, использующих распределения статистик:

$$D_{n,m}^+ = \sup_{x=z} (F_{1n}(x) - F_{2m}(z)); \quad (13)$$

$$D_{n,m}^- = \inf_{x=z} (F_{1n}(x) - F_{2m}(z)). \quad (14)$$

Случайные величины $D_{m,n}^+$, $D_{m,n}^-$, если гипотеза H_0 верна, имеют одинаковое распределение.

При проверке указанных гипотез необходимо проделать те же вычисления, что и для двустороннего критерия, основанного на статистике $D_{m,n}$, с той лишь разницей, что для одностороннего критерия уровень значимости при одних и тех же значениях $D_\beta = D_\beta^+$ равен не 2β , а β . Такое положение с достаточной степенью точности справедливо и для $2\beta \leq 0,2$. Если гипотеза H_0 верна и объемы выборок неограниченно увеличиваются, то

$$\lim_{n \rightarrow \infty} P(D_{n,m}^+ > D_\beta^+) = \lambda_\beta \sqrt{1/n + 1/m} = e^{-2\lambda_\beta^2}, \quad m \geq n, \quad (15)$$

и для расчета D_β^+ можно при заданном β пользоваться формулой (12). Когда $\min[n, m] > 50$, то замена истинных границ на асимптотические для широко используемых уровней значимости $\beta = (0,01-0,05)$ приводит к увеличению надежности критерия приблизительно на 2%, а при $\min[n, m] \geq 10$ не более чем на 5%.

При практических расчетах статистик целесообразно определять по следующим формулам:

$$\left. \begin{aligned} D_{m,n}^+ &= \max_{1 < i < m} \left(\frac{i}{m} - F_n(x_i) \right) = \max_{1 < s < n} \left(F_m(z_s) - \frac{(s-1)}{n} \right); \\ D_{m,n}^- &= \max_{1 < i < m} \left(F_n(x_i) - \frac{(i-1)}{m} \right) = \max_{1 < s < n} \left(\frac{s}{n} - F_m(z_s) \right); \\ D_{m,n} &= \max(D_{m,n}^+, D_{m,n}^-). \end{aligned} \right\} \quad (16)$$

На практике зависимостью (12) пользуются при определении объема моделирования и числа натуральных испытаний, необходимых для обнаружения с любым наперед заданным уровнем β значимых отклонений между функциями распределения $F_1(x)$ и $F_2(z)$. Но при такой оценке объемы планируемых выборок получают несколько завышенными.

При анализе сложных систем, когда стоимость проведения одного натурального испытания очень велика, точность получаемых значений n , m по формуле (12) может оказаться недостаточной. В этих случаях для

определения n , m нужно применять более точные соотношения, приведенные в работах, или находить значения n , m на основании различного рода интерполяционных формул, использующих в качестве опорных данных табличные значения истинных границ $D_{n,m}$, рассчитанных для некоторого фиксированного набора уровней значимости.

Если объемы выборок на модели и при натуральных испытаниях предполагают сделать одинаковыми, что имеет место при очень большом числе вариантов построения и большой стоимости экспериментов на модели, то зависимости $D_{n,m}^+(\beta)$ и $D_{n,m}(2\beta)$ как функции β и $n = m$ можно для $\beta \leq 5\%$ охарактеризовать простыми и в то же время достаточно точными соотношениями:

$$D_{n,m}^+(\beta) = D_{n,m}(2\beta) = r^+(\beta)m, \quad m = n; \quad (17)$$

$$r^+(\beta) = 1 + \sqrt{-m \ln \beta - \frac{\ln \beta (3 + \ln \beta)}{6}}. \quad (18)$$

В последнем случае истинные значения вероятностей появления событий $D_{n,m}^+ \leq (k/m)$ и $D_{n,m} \leq (k/m)$ можно рассчитать на основании формул:

$$P\{D_{m,m}^+ \leq (k/m)\} = 1 - C_{2m}^{m+k-1} / C_{2m}^m; \quad (19)$$

$$P\{D_{m,m} \leq (k/m)\} = \sum_{i=-l}^l (-1)^i (C_{2m}^{m+i(k+1)} / C_{2m}^m), \quad (20)$$

где l — целая часть числа $m/(k+1)$; k — целое число, определяемое так, чтобы выполнялись соотношения:

$$P\{D_{m,m}^+ > D_\beta = k/m\} = \beta \quad (21)$$

$$P\{D_{m,m} > D_\beta = k/m\} = 2\beta. \quad (22)$$

Особым преимуществом критерия Смирнова является его состоятельность, которая позволяет при неограниченном увеличении объема выборок на модели и при натуральных испытаниях обнаружить практически любое отличие между функциями распределения $F_1(x)$ и $F_2(z)$. Однако при больших объемах выборок нужно считаться с тем, что проверка гипотез на основании критерия Смирнова связана с довольно значительным объемом вычислений, обусловленным необходимостью построения вариационных рядов $x^{(1)}, x^{(2)}, \dots, x^{(m)}$; $z^{(1)}, z^{(2)}, \dots, z^{(m)}$. При использовании известных методов упорядочивания выборок x_1, x_2, \dots, x_n ; z_1, z_2, \dots, z_m в порядке возрастания значения x_i, z_j нужно выполнить количество алгоритмических операций, пропорциональное квадрату их объемов.

Пример. На математической модели планируют провести такое же количество экспериментов, как и при натуральных испытаниях на реальной системе.

На основании полученных выборок необходимо проверить гипотезу о тождественности функций их распределения.

Для этих условий взаимосвязь критических значений $r(2\beta)$ и уровней значимости 2β для различных значений $n = m$ протабулирована и представлена в табл. 6.

Таблица 6

n=m	Критические значения r(2β)							
	2β=0,10		2β=0,05		2β=0,02		2β=0,01	
	r(2β)	2β _{ист}	r(2β)	2β _{ист}	r(2β)	2β _{ист}	r(2β)	2β _{ист}
5	4	7,9	5	0,8	5	0,8	5	0,8
6	5	2,6	5	2,6	6	0,2	6	0,2
7	5	5,3	6	0,8	6	0,8	6	0,8
8	5	8,7	6	1,9	6	1,9	7	0,2
9	6	3,4	6	3,4	7	0,6	7	0,6
10	6	5,2	7	1,2	7	1,2	8	0,2
11	6	7,5	7	2,1	8	0,4	8	0,4
12	6	10	7	3,1	8	0,8	8	0,8
13	7	4,4	7	4,4	8	1,3	9	0,3
14	7	5,9	8	1,9	8	1,9	9	0,5
15	7	7,5	8	2,6	9	0,8	9	0,8
16	7	9,3	8	3,5	9	1,1	10	0,3
17	8	4,5	8	4,5	9	1,6	10	0,5
18	8	5,6	9	2,1	10	0,7	10	0,7
19	8	6,8	9	2,7	10	0,9	10	0,9
20	8	8,1	9	3,4	10	1,2	11	0,4

Пусть случайные величины $x_i, z_i (i=1, 2, \dots, 10)$, полученные на модели и при испытаниях реальной системы, имеют значения, приведенные в табл. 2.

На основании указанных выборок составляем им соответствующие вариационные ряды $x^{(1)}, x^{(2)}, \dots, x^{(10)}; z^{(1)}, z^{(2)}, \dots, z^{(10)}$ (табл. 7)

Таблица 7

i	1	2	3	4	5	6	7	8	9	10
$x^{(i)}$	-1,966	-1,618	-1,438	-1,324	-0,726	-0,294	0,771	0,790	1,695	1,792
$z^{(i)}$	-1,621	-1,210	-0,463	-0,011	-0,064	0,163	0,789	1,157	1,786	2,140

Используя формулы (10), рассчитываем эмпирические функции распределения $F_{1n}(x), F_{2m}(z)$ (рис. 1) и находим, что в рассматриваемом случае случайная величина статистики Смирнова равна

$$D_{10;10} = \sup_{x=z} |F_{1,10}(x) - F_{2,10}(z)| = 0,3.$$

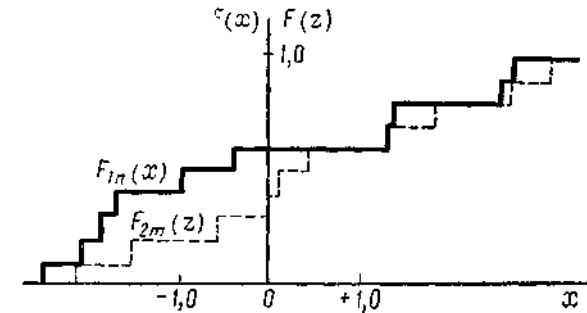


Рис. 1

Из табл. 6 для $n = m = 10$ находим $r(0,10) = 6, r(0,5) = 7, r(0,02) = 7$ и $r(0,01) = 8$, что позволяет рассчитать искомые критические значения: $D_{10;10}(2\beta) = 0,6$ (при $2\beta = 0,10$); $D_{10;10}(2\beta) = 0,7$ (при $2\beta = 0,05$), $D_{10;10}(2\beta) = 0,7$ (при $2\beta = 0,02$); $D_{10;10}(2\beta) = 0,8$ (при $2\beta = 0,01$). Предположим, что по условиям задачи выбран уровень значимости $2\beta = 0,10$. Тогда, сравнивая $D_{10;10}$ с найденным критическим значением $D_{10;10}(2\beta) = 0,6$ (при $2\beta = 0,10$), получаем, что $D_{10;10} < D_{10;10}(2\beta)$ и, следовательно, гипотезу о тождественности распределения $F_1(x)$ и $F_2(z)$ можно считать справедливой. В данном случае при выбранном критическом значении $D_{10;10}(2\beta) = 0,6$ истинный уровень значимости равен не 0,10, а 0,052, что непосредственно следует из табл. 6, в которой рядом с каждым критическим значением $r(2\beta)$ указан истинный уровень значимости, выраженный в процентах.

Если необходимо проверить гипотезу о тождественности распределений $F_1(x) \equiv F_2(z)$ на основании статистики $D_{10;10}^+(\beta)$, используя в качестве критического значения $r(\beta) = 6$, то в рассматриваемом случае имеем $D_{10;10}^+ = D_{10;10}^+(\beta)$; $D_{10;10}^+(\beta) < 0,6$ и поэтому получаем, что проверяемая гипотеза оказывается справедливой, но уже с истинным уровнем значимости, равным $\beta_{ист} = 0,026$.

Задача проверки гипотезы о тождественности эмпирических и теоретических функций распределения.

Когда стоимость и время проигрыша одной реализации на модели сравнительно невелики, то на модели можно получить выборку z_1, z_2, \dots, z_m достаточно большого объема и по этой выборке построить закон распределения $F(z)$.

Для объемов выборок $m > 10\,000$ параметры закона распределения $F(z)$ можно считать найденными абсолютно точно. В этих условиях, если есть необходимость в проверке гипотезы H_0 о том, что выборка результатов натуральных испытаний x_i ($i=1, 2, \dots, n$) принадлежит некоторой совокупности, имеющей точно известное распределение $F(z)$, то она может быть осуществлена на основании критерия Колмогорова.

Статистику Колмогорова записывают в виде:

$$D_n = \sup_x |F_n(x) - F(x)|, \quad (23)$$

где $F_n(x)$ — эмпирическая функция распределения, построенная по выборке x_1, x_2, \dots, x_n .

Согласно критерию Колмогорова гипотеза H_0 считается справедливой, если вероятность появления событий $D_n > D_\beta$ при правильной гипотезе H_0 не превосходит 2β . Асимптотические свойства критерия Колмогорова можно охарактеризовать соотношениями

$$\lim_{n \rightarrow \infty} P \left\{ \sup_x |F_n(x) - F(x)| < D_\beta = \frac{\lambda_\beta}{\sqrt{n}} \right\} = K(\lambda_\beta); \quad (24)$$

$$K(\lambda_\beta) = 1 - 2 \sum_{i=1}^{\infty} (-1)^i e^{-2i\lambda_\beta^2}, \quad \lambda_\beta \geq 0. \quad (25)$$

Если $n > 50$ расчет асимптотических границ D_β для $2\beta = 0,05$ и $2\beta = 0,01$ производят по формулам:

$$D_{\beta=0,05} = 1,36/\sqrt{n}; \quad D_{\beta=0,01} = 1,63/\sqrt{n}. \quad (26)$$

и при этом получают некоторое увеличение надежности критерия, что в конечном счете приводит к более обоснованным выводам о неслучайности расхождений между $F_n(x)$ и $F(x)$. На практике в качестве рабочего интервала значений $\sqrt{n}D_\beta$ используют промежуток $\lambda_\beta \in [0,7 \div 1]$, в который при правильной гипотезе H_0 попадает свыше 50% ожидаемых значений $D_n \sqrt{n}$. Часто при $n \geq 20$ применяют асимптотическую формулу Смирнова

$$2\beta \sim 2e^{-2n\lambda_\beta^2},$$

которая, как нетрудно убедиться, эквивалентна соотношению (12) и позволяет рассчитать критические значения D_β при заданных β и n .

Критерий Колмогорова нетрудно получить из критерия Смирнова, если объем выборки z_1, z_2, \dots, z_m , получаемой при моделировании, устремить к бесконечности ($m \rightarrow \infty$). Подобная общность приводит к тому, что последовательность выполняемых операций и соответственно трудности реализации критерия Колмогорова при проверке гипотезы H_0 будут такими же, как и для критерия Смирнова.

Односторонний критерий Колмогорова (Δ -критерий) основывается на распределении одной из статистик:

$$D_n^+ = \max_x (F(x) - F_n(x)); \quad (27)$$

$$D_n^- = \max_x (F_n(x) - F(x)). \quad (28)$$

Уровень значимости Δ -критерия по сравнению с двусторонним критерием Колмогорова для одних и тех же критических значений получают равным уже не 2β , а β . С помощью статистик D_n^+, D_n^- выражение для статистики D_n можно записать в виде:

$$D_n = \max(D_n^+, D_n^-). \quad (29)$$

При практической проверке справедливости гипотезы H_0 на основании двустороннего критерия Колмогорова использование соотношений (27)-(29) дает возможность в определенной степени упорядочить процедуру проводимых расчетов.

Взаимосвязь между критическими значениями λ_β/\sqrt{n} статистик D_n^+, D_n^-, D_n и истинными уровнями значимости позволяет с помощью эмпирической функции распределения $F_n(x)$ строить область определения функции распределения генеральной совокупности $F(x) = F_n(x) \pm \lambda_\beta/\sqrt{n}$. При анализе сложных систем, когда функция распределения $F(x)$ неизвестна, решение указанной выше задачи иногда позволяет получить практически полезные выводы относительно вида функции распределения $F(x)$ и объема планируемых экспериментов.

Пример. Пусть случайные величины x_i ($i=1, 2, \dots, 10$), полученные при испытаниях реальной системы, имеют численные значения, которые приведены в табл. 2. Предполагается также, что выборка x_1, x_2, \dots, x_{10} принадлежит генеральной совокупности с законом распределения $F(x) = N\{m_x, \sigma_x^2\}$.

Проверить указанное предположение при условии, что параметры m_x, σ_x^2 рассчитаны по результатам моделирования абсолютно точно. Для рассматриваемых условий примем значения, соответствующие истине: $m_x = 0; \sigma_x^2 = 1,0$.

Проверку будем осуществлять на основании критерия Колмогорова, для чего рассчитаем эмпирическую функцию $F_n(x)$ и по формуле (29) найдем $D_n = 0,3072485$. Если в качестве критического значения взять $D_\beta = 0,32260$, которое соответствует $2\beta = 0,20$ при $n=10$ (табл. 8), то поскольку $D_n < D_\beta$, получаем, что проверяемая гипотеза H_0 в рассматриваемом случае справедлива.

Таблица 8

n	Критические значения D_n				
	$2\beta=0,20$	$2\beta=0,10$	$2\beta=0,05$	$2\beta=0,02$	$2\beta=0,01$
5	0,44698	0,50945	0,56328	0,62718	0,66853
10	0,32260	0,36866	0,40925	0,45662	0,48893
15	0,26558	0,30397	0,33760	0,37713	0,40420
20	0,23156	0,26473	0,29408	0,32866	0,35241
25	0,20790	0,23768	0,26404	0,28516	0,31657
30	0,19032	0,21756	0,24170	0,27023	0,28987
35	0,17695	0,20185	0,22425	0,25073	0,26897
40	0,16547	0,18913	0,21012	0,23494	0,25205
45	0,15623	0,17856	0,19837	0,2218	0,23798
50	0,14840	0,16959	0,18841	0,21068	0,22604
55	0,14164	0,16186	0,17981	0,20107	0,21574
60	0,13573	0,15511	0,17231	0,19267	0,20673
65	0,13052	0,14913	0,16576	0,18525	0,19877
70	0,12586	0,14381	0,15975	0,17863	0,19167
75	0,12167	0,13901	0,15442	0,17268	0,18528
80	0,11787	0,13467	0,14960	0,16728	0,17949
85	0,11442	0,13072	0,14520	0,16236	0,17421
90	0,11125	0,12709	0,14117	0,15786	0,16938
95	0,10833	0,12375	0,13746	0,15371	0,16493
100	0,10563	0,12067	0,13403	0,14987	0,16081

Критические значения для наибольшего отклонения эмпирического распределения $F_n(x)$ от теоретического $F(x)$ протабулированы и сведены в таблицы. Выборочная таблица 8, широко используемая на практике, приведена для значений $n \leq 100$ и различных уровней значимости $2\beta = 0,2; 0,1; 0,05; 0,02; 0,01$

Задача проверки гипотезы о нормальности распределения результатов измерений (критерий Саркади). Если нет достаточной уверенности в однородной точности результатов измерений, то для проверки гипотезы о нормальном характере ошибок измерений целесообразно воспользоваться критерием Саркади, который не требует знания параметров распределения. Кратко суть этого критерия сводится к следующему.

По результатам наблюдений для рассматриваемой выборки вычисляют случайные величины

$$\eta_i = x_i - \frac{1}{\sqrt{n+1}} (\sqrt{n}x^* + x_i), \quad i=1, 2, \dots, n, \quad (30)$$

по которым рассчитывают новые случайные величины:

$$\xi_i = \eta_i \sqrt{n-i-1} / \sqrt{\sum_{k=i+1}^{n-1} \eta_k^2} \quad (31)$$

Далее, с помощью таблиц распределения Стьюдента находят случайные величины

$$\delta_{v_i} = S_{v_i}(\xi_i), \quad i=1, 2, \dots, n-2, \quad (32)$$

где $S_{v_i}(\xi_i)$ — функция распределения Стьюдента с числом степеней свободы $v_i = n - i - 1$, ($i=1, 2, \dots, n-2$).

Если распределение результатов измерений в рассматриваемой выборке согласуется с гипотезой о нормальном характере их распределения, то случайные величины δ_{v_i} должны подчиняться равномерному распределению на отрезке (0,1).

Таким образом, чтобы проверить гипотезу о нормальном характере распределения величин x_i , ($i=1, 2, \dots, n$) необходимо установить, подчиняются ли случайные величины δ_{v_i} равномерному распределению. Это можно сделать с помощью критерия ω^2 , рассчитав величину $(n-2)\omega_{n-2}^2$ по формуле

$$(n-2)\omega_{n-2}^2 = \frac{1}{12(n-2)} + \sum_{i=1}^{n-2} \left\{ \delta_{v_i} - \frac{2i-1}{2(n-2)} \right\}^2 \quad (33)$$

и сравнив полученное значение с граничной величиной критической области. В частном случае (надежность $P = 0,95$) критическая область для этого критерия удовлетворяет неравенству

$$(n-2)\omega_{n-2}^2 > 0,4614. \quad (34)$$

Если рассчитанное значение $(n-2)\omega_{n-2}^2$ будет меньше граничной величины, то можно с заданной вероятностью считать распределение ошибок измерений нормальным.

Задача проверки гипотез на основании χ^2 -критерия. В случае, если обработка опытных данных проводится с использованием интервального вариационного ряда, то проверку гипотез о виде эмпирической функции распределения можно осуществить с помощью критерия соответствия χ^2 . Практическое применение этого критерия предопределяется двумя причинами.

Во-первых, критерий χ^2 может быть применен тогда, когда часть или все параметры предполагаемого закона распределения априори неизвестны и их нужно определить по результатам испытаний.

Во-вторых, критерий χ^2 не требует построения упорядоченных вариационных рядов для наблюдаемых выборок и поэтому при моделировании, когда можно получить выборки достаточно большого объема, является достаточно эффективной вычислительной процедурой.

Критерий χ^2 основан на распределении статистики

$$\chi^2 = \sum_{i=1}^e (m_i - mp_i^*) / (mp_i^*) \quad (35)$$

в предположении, что частота m_i/m попадания значений случайной величины z в некоторый интервал $[L_{i-1}, L_i]$ определяется непосредственно по выборке z_1, z_2, \dots, z_m , а оценки вероятностей появления событий $p_i^* = P(L_{i-1} \leq z \leq L_i)$ могут быть рассчитаны с помощью закона распределения $F(z, a_1, a_2, \dots, a_s)$, известного с точностью до некоторой группы параметров a_1, a_2, \dots, a_s . Если в качестве оценок параметров a_1, a_2, \dots, a_s использовались асимптотически эффективные оценки, определенные по результатам моделирования z_1, z_2, \dots, z_m , то при $m \rightarrow \infty$ статистика (35) имеет χ^2 — распределение с числом степеней свободы

$$k = l - s - 1.$$

Из множества различных критериев χ^2 — критерий Пирсона отличается наибольшей мощностью, т. е. характеризуется минимальными вероятностями ошибок второго рода. Минимально допустимое количество наблюдений в группе составляет примерно пять измерений, а число групп следует выбирать порядка восьми, если приходится пользоваться эмпирическими параметрами распределения.

Когда относительно выборки z_1, z_2, \dots, z_m предполагают, что она распределена по нормальному закону, то для проверки выдвинутой гипотезы $H_0: F(z) = N\{m_z, \sigma_z^2\}$ нужно:

1) определить минимальное и максимальное значения z_i ($i=1, 2, \dots, m$) и затем найденный диапазон изменения результатов моделирования разбить на заранее выбранное число l интервалов. При выборе числа интервалов нужно руководствоваться общим положением, чтобы в каждом интервале было не менее пяти значений случайной величины z . Если это условие не выполняется, то целесообразно скорректировать априори выбранное число разбиений l путем слияния соседних интервалов так, чтобы число попавших точек в каждый интервал было больше или равно пяти;

2) рассчитать оценки математического ожидания m_z^* дисперсии σ_z^{*2}

$$m_z^* = \frac{1}{m} \sum_{i=1}^m z_i; \quad (36)$$

$$\sigma_z^{*2} = \frac{1}{m-1} \sum_{i=1}^m (z_i - m_z^*)^2 \quad (37)$$

и с помощью их по таблицам распределения функции Лапласа $\Phi(x)$ найти вероятности p_i появления значений случайной величины z в каждом i -м интервале ($i=1, 2, \dots, l$). Если гипотеза $H_0: F(z) = N\{m_z, \sigma_z^2\}$ верна, то оценки (36) и (37) являются асимптотически эффективными оценками;

3) по данному уровню значимости β и найденному числу степеней свободы $k = l - 3$ определить с помощью таблиц χ^2 -распределения (табл. 9) границу χ_{β}^2 критической области;

4) сравнить найденную границу χ_{β}^2 со значением статистики, рассчитанной по формуле (35). Если окажется, что вычисленное значение статистики χ^{*2} меньше границы χ_{β}^2 , то проверяемую гипотезу следует считать справедливой.

Вероятность ошибочно отвергнуть правильную гипотезу H_0 не превосходит β . Эту вероятность обычно называют уровнем значимости χ^2 -критерия. Для χ^2 -критерия характерна некоторая неопределенность выбора числа и способа разбиения диапазона изменения случайной величины z на интервалы. Поэтому при малых объемах выборки к результатам проверки нужно относиться критически и рассматривать их с учетом тех априорных сведений, которые обычно известны об изучаемых явлениях до моделирования.

В сомнительных случаях, когда χ^2 -критерий принимает гипотезу $H_0: F(z) = N\{m_z, \sigma_z^2\}$ с невысоким уровнем значимости, а экспериментатор убежден в обратном, целесообразно воспользоваться критериями, позволяющими оценить степень отклонения реального распределения от нормального. Эти критерии основаны на том, что если случайная величина z распределена по нормальному закону $N\{m_z, \sigma_z^2\}$,

то справедливы следующие соотношения:

для нормированного среднего абсолютного отклонения

$$\delta = \frac{E\{|z - m_z|\}}{\sigma_z} = \sqrt{\frac{2}{\pi}} = 0,79788; \quad (38)$$

Таблица 9

		Критические значения χ^2_{β}															
		0,999	0,99	0,975	0,95	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01
β	k																
1	1	0,00000157	0,000157	0,000982	0,00393	0,0158	0,0642	0,148	0,275	0,455	0,708	1,07	1,64	2,71	3,84	5,02	6,63
2	1	0,00200	0,0201	0,0506	0,103	0,211	0,446	0,713	1,02	1,39	1,83	2,41	3,22	4,61	5,99	7,38	9,21
3	1	0,0243	0,115	0,216	0,352	0,584	1,00	1,42	1,87	2,37	2,95	3,67	4,64	6,25	7,81	9,35	11,3
4	1	0,0908	0,297	0,484	0,711	1,06	1,65	2,19	2,75	3,36	4,04	4,88	5,99	7,78	9,49	11,1	13,3
5	1	0,210	0,334	0,531	0,711	1,06	1,65	2,19	2,75	3,36	4,04	4,88	5,99	7,78	9,49	11,1	13,3
6	1	0,381	0,531	0,711	1,06	1,65	2,19	2,75	3,36	4,04	4,88	5,99	7,78	9,49	11,1	13,3	16,8
7	1	0,598	0,872	1,24	1,64	2,20	3,07	3,83	4,57	5,35	6,21	7,23	8,56	10,6	12,6	14,4	16,8
8	1	0,857	1,24	1,69	2,17	2,83	3,82	4,67	5,49	6,35	7,28	8,38	9,80	12,0	14,0	16,0	18,5
9	1	1,15	1,65	2,18	2,73	3,49	4,59	5,53	6,42	7,34	8,35	9,52	11,0	13,4	15,5	17,5	20,1
10	1	1,48	2,09	2,70	3,33	4,17	5,38	6,39	7,36	8,34	9,41	10,7	12,2	14,7	16,9	19,0	21,7
12	1	2,21	3,57	4,40	5,23	6,30	7,81	9,03	10,2	11,3	12,6	14,0	15,8	18,5	21,0	23,3	26,2
14	1	3,04	4,66	5,63	6,57	7,79	9,47	10,8	12,1	13,3	14,7	16,2	18,2	21,1	23,7	26,1	29,1
16	1	3,94	5,81	6,91	7,96	9,31	11,2	12,6	14,0	15,3	16,8	18,4	20,5	23,5	26,3	28,8	32,0
18	1	4,90	7,01	8,23	9,39	10,9	12,9	14,4	15,9	17,3	18,9	20,6	22,8	26,0	28,9	31,5	34,8
20	1	5,92	8,26	9,59	10,9	12,4	14,6	16,3	17,8	19,3	21,0	22,8	25,0	28,4	31,4	34,2	37,6
25	1	8,65	11,5	13,1	14,6	16,5	18,9	20,9	22,6	24,3	26,1	28,2	30,7	34,4	37,7	40,6	44,3
30	1	11,6	15,0	16,8	18,5	20,6	23,4	25,5	27,4	29,3	31,3	33,5	36,3	40,3	43,8	47,0	50,9
35	1	14,7	18,5	20,6	22,5	24,8	27,8	30,2	32,3	34,3	36,5	38,9	41,8	46,1	49,8	53,2	57,3
40	1	17,9	22,2	24,4	26,5	29,1	32,2	34,9	37,1	39,3	41,6	44,2	47,3	51,8	55,8	59,3	63,7
45	1	21,3	25,9	28,4	30,6	33,4	36,9	39,6	42,0	44,3	46,8	49,5	52,7	57,5	61,7	65,4	70,0
50	1	24,7	29,7	32,4	34,8	37,7	41,4	44,3	46,9	49,3	51,9	54,7	58,2	63,2	67,5	71,4	76,2
60	1	31,7	37,5	40,5	43,2	46,5	50,6	53,8	56,6	59,3	62,1	65,2	69,0	74,4	79,1	83,3	88,4
70	1	39,0	45,4	48,8	51,7	55,3	59,9	63,3	66,4	69,3	72,4	75,7	79,7	85,5	90,5	95,0	100,4
80	1	46,5	53,5	57,2	60,4	64,3	69,2	72,9	76,2	79,3	82,6	86,1	90,4	96,6	101,9	106,6	112,3
90	1	54,2	61,0	65,6	69,1	73,3	78,6	82,5	86,0	89,3	92,8	96,5	101,1	107,6	113,1	118,1	124,1
100	1	61,9	70,1	74,2	77,9	82,4	87,9	92,1	96,8	99,3	102,9	106,9	111,7	118,5	124,3	129,6	135,8

для коэффициента асимметрии

$$\gamma_1 = \frac{E\{|z - m_z|^3\}}{\sigma_z^3} = 0; \tag{39}$$

для коэффициента эксцесса

$$\beta_2 - 3 = \frac{E\{(z - m_z)^4\}}{\sigma_z^4} - 3 = 0. \tag{40}$$

Чтобы узнать степень невыполнения равенств (38)-(40), нужно рассчитать статистики:

$$\delta^* = \frac{1}{m\sigma_z^3} \sum_{i=1}^m |z_i - m_z^*|;$$

$$\gamma_1^* = \frac{1}{m\sigma_z^{*3}} \sum_{i=1}^m |z_i - m_z^*|^3;$$

$$\beta_2^* = \frac{1}{m\sigma_z^{*4}} \sum_{i=1}^m |z_i - m_z^*|^4;$$

$$m_z^* = \frac{1}{m} \sum_{i=1}^m z_i;$$

$$\sigma_z^{*2} = \frac{1}{m-1} \sum_{i=1}^m |z_i - m_z^*|^2.$$

Когда статистики δ^* , γ_1^* , β_2^* найдены, то с помощью таблиц, нетрудно рассчитать количественные оценки, позволяющие определить, по каким параметрам выборочное распределение наиболее сильно отличается от нормального.

Пример. Чтобы проиллюстрировать методику проверки нормальности распределения результатов моделирования по некоторой выборке z_1, z_2, \dots, z_m , примем в качестве ее элементов случайные значения x_i , которые приведены в табл. 4 (т. е. примем $z_i = x_i$ при $i=1, 2, \dots, 50$).

Следуя изложенной выше процедуре проверки:

1) определяем $z_{\max} = 3,521$; $z_{\min} = -2,526$ и для $l=6$ с учетом округления находим границы интервалов деления $(-\infty, -2]$, $(-2, -1]$, $(-1, 0]$, $(0, 1]$; $(1, 2]$, $(2, +\infty)$. Чтобы сумма вероятностей попадания в каждый интервал была равна единице, примем левой границей первого интервала $-\infty$, а правой границей шестого интервала $+\infty$;

2) находим частоты попадания случайной величины z_i в каждый интервал ($m_1'=1; m_2'=4; m_3'=16; m_4'=18; m_5'=8; m_6'=3$) Так как числа m_1' и m_6' меньше пяти, то объединим первый интервал со вторым, а шестой интервал с пятым. Таким образом, границами уточненного разбиения интервала $[z_{\max}, z_{\min}]$ будут $(-\infty, -1], (-1, 0], (0, 1], (1, +\infty)$, т. е. $l=4$.

Для нового разбиения получим $m_1=5; m_2=16; m_3=18; m_4=11$;

3) по формулам (736) и (37) находим, что $m_2^*=0,21116; \sigma_2^*=1,160886$ и с помощью таблицы распределения $\Phi(x)$ определяем вероятности попадания в каждый интервал ($p_1^*=0,14917; p_2^*=0,27941; p_3^*=0,32317; p_4^*=0,24825$);

4) вычислим выборочное значение χ^2 — критерия по формуле (35) : $\chi^{*2}=1,47582$;

5) сравнивая значения χ^{*2} с критическими значениями χ_{β}^2 для $k=l-3=4-3=1$ и для $\beta \leq 0,2$, получим, что для всех рассматриваемых β справедливо неравенство: $\chi^{*2} < \chi_{\beta}^2$

Следовательно, анализируемую выборку результатов моделирования можно считать нормально распределенной, если априори выбираемый уровень значимости меньше 0,2. Рассчитывая значения статистик $\delta^*=0,63233675; \gamma_1^*=0,3797265; \beta_2^*=3,6929147$ и сравнивая их с критическими значениями, приведенными в табл. 10, нетрудно убедиться в несущественности отклонения выборочного распределения от нормального по крайней мере для общего уровня значимости, равного 0,05. Для указанного уровня значимости критические значения соответственно равны $\delta_{\beta=0,05} = 0,8644; \gamma_{1,\beta=0,05} = 0,533; \beta_{2,\beta=0,05} = 4,01$.

Задача проверки гипотезы об однородности распределения двух выборок (критерий Уилкоксона). При решении задачи, связанной с проверкой гипотезы о принадлежности двух случайных выборок одной генеральной совокупности, можно воспользоваться критерием Уилкоксона, который основан на применении теории инверсий. Суть этого критерия сводится к следующему.

Пусть выборка результатов x_1, x_2, \dots, x_n статистически независима с результатами z_1, z_2, \dots, z_m . Для этих условий проверка однородности распределения двух выборок состоит в проверке адекватности законов распределения $F_1(x), F_2(z)$, характеризующих их генеральные совокупности. Законы распределения $F_1(x), F_2(z)$ априори неизвестны. При общей альтернативе $H_1: F_1 \neq F_2$ решение поставленной задачи может быть найдено на основании критерия Смирнова,

Таблица 10

m	Критические значения δ_{β}					m	Критические значения $\gamma_{1,\beta}$		m	Критические значения $\beta_{2,\beta}$			
	β						0,01	0,05		0,95	0,99		
	0	0,05	0,10	0,90	0,95							0,99	
11	0,936	0,907	0,890	0,741	0,715	0,668	0,711	0,061	50	4,92	4,01	2,13	1,95
16	0,914	0,888	0,873	0,745	0,724	0,683	0,661	0,982	100	4,40	3,77	2,35	2,18
21	0,900	0,877	0,863	0,750	0,730	0,695	0,621	0,921	150	4,14	3,66	2,45	2,30
26	0,890	0,869	0,857	0,753	0,736	0,704	0,587	0,869	200	3,98	3,57	2,51	2,37
31	0,883	0,863	0,851	0,756	0,740	0,711	0,558	0,825	250	3,87	3,51	2,56	2,42
36	0,887	0,857	0,847	0,758	0,744	0,717	0,533	0,787	300	3,79	3,47	2,59	2,46
41	0,872	0,854	0,844	0,760	0,747	0,722	0,492	0,723	350	3,72	3,44	2,62	2,50
46	0,868	0,851	0,841	0,762	0,750	0,726	0,459	0,673	400	3,67	3,41	2,64	2,52
51	0,865	0,848	0,839	0,764	0,752	0,729	0,432	0,631	450	3,63	3,39	2,66	2,55
51	0,859	0,843	0,835	0,766	0,755	0,735	0,409	0,596	500	3,60	3,37	2,67	2,57
61	0,855	0,840	0,832	0,768	0,758	0,739	0,389	0,567	550	3,57	3,35	2,69	2,5
71	0,852	0,838	0,830	0,770	0,761	0,743	0,350	0,508	600	3,54	3,34	2,70	2,5
81	0,848	0,835	0,828	0,771	0,763	0,746	0,321	0,464	650	3,52	3,33	2,71	2,6
91	0,846	0,834	0,826	0,783	0,774	0,764	0,298	0,430	700	3,50	3,31	2,72	2,6
101	0,842	0,832	0,824	0,780	0,774	0,763	0,280	0,403	750	3,48	3,30	2,73	2,6
201	0,832	0,823	0,818	0,783	0,778	0,769	0,230	0,329	800	3,46	3,29	2,74	2,6
301	0,826	0,818	0,814	0,783	0,778	0,773	0,200	0,285	850	3,45	3,28	2,75	2,6
401	0,822	0,815	0,812	0,786	0,783	0,776	0,173	0,255	900	3,43	3,28	2,75	2,6
501	0,820	0,814	0,810	0,788	0,785	0,779	0,151	0,215	950	3,42	3,27	2,76	2,6
701	0,816	0,811	0,808	0,788	0,785	0,779	0,127	0,180	1000	3,41	3,26	2,76	2,6
1001	0,813	0,809	0,807	0,789	0,787	0,782							

если для каждой выборки рассчитать эмпирические функции распределения. В частных случаях процедуры расчета эмпирических функций распределения можно избежать, воспользовавшись при проверке гипотезы $H_0 : F_1(x) = F_2(z)$ порядковым критерием, каким и является критерий Уилкоксона. Состоятельность критерия Уилкоксона доказана, когда в качестве альтернативы выступает односторонняя гипотеза о том, что величины x_i стохастически больше или меньше величин z_j . Чтобы воспользоваться критерием Уилкоксона, нужно:

- 1) построить общий вариационный ряд в порядке совместного возрастания случайных величин x_i, z_j (при $i=1, 2, \dots, n; j=1, 2, \dots, m$);
- 2) определить число инверсий. Количество инверсий для данного x_i определяется как число тех z_j , которые удовлетворяют условию $x_i > z_j$. Если для каждой пары наблюдений x_i, z_i определить функцию y_{ij}

$$y_{ij} = \begin{cases} 1, & \text{если } x_i > z_j; \\ 0, & \text{если } x_i \leq z_j, \end{cases} \quad (41)$$

то общее число инверсий можно найти следующим образом:

$$U = \sum_{i=1}^n \sum_{j=1}^m y_{ij}, \quad (42)$$

где U — статистика, носящая название статистики Уилкоксона;

- 3) рассчитать по заданному уровню значимости β нижнюю допустимую границу U_β .

Манном и Уитни получено, что при $n, m \rightarrow \infty$ имеют место зависимости:

$$P \{U \leq u_\beta\} = \Phi \left\{ \frac{u_\beta - m_U}{D_U} \right\}; \quad (43)$$

$$m_U = \frac{nm}{2}; \quad D_U = \frac{nm(n+m+1)}{12}. \quad (44)$$

На основании этих асимптотических соотношений нетрудно рассчитать приближенное значение U_β .

Однако при практических расчетах для нахождения U_β по заданному β рекомендуется использовать более точную аппроксимацию, приводящую к уравнению

$$u_\beta = [m_U - 0,05 + \Psi(\beta) D_U], \quad (45)$$

где $\Psi(\beta)$ — значение функции нормального распределения с параметрами (0,1).

При $\min(n, m) \geq 4, n+m \geq 20$ точность аппроксимации истинного распределения статистики U асимптотическим получается вполне достаточной.

С помощью статистики U и значений U_β , соответствующих заданному уровню значимости β , можно сформулировать односторонний критерий Уилкоксона. Согласно этому критерию, нулевую гипотезу $H_0 : F_1(x) = F_2(z)$ нужно отвергнуть, когда на самом деле гипотеза H_0 верна, в том случае, если количество инверсий U больше рассчитанной допустимой границы U_β .

Сравнивая асимптотическую эффективность одностороннего критерия Уилкоксона с t -критерием, можно показать, что критерий Уилкоксона требует по сравнению с t -критерием несколько больших объемов выборок. Например, если $F_1(x)$ и $F_2(z)$ распределены по нормальному закону с одинаковой дисперсией и критерий Уилкоксона применяют к выборкам объема n и m , а t -критерий Стьюдента — к выборкам объема n' и m' , то равные мощности рангового критерия Уилкоксона и t -критерия реализуются в случае, когда $n' \approx \left(\frac{3}{\pi}\right)n, m' \approx \left(\frac{3}{\pi}\right)m$, т. е. асимптотическая эффективность критерия Уилкоксона равна $3/\pi$.

Если выполнимы соотношения $U < u_\beta, nm - U < U_\beta = 2m_U - u_\beta$ (U_β — верхнее критическое значение), то говорят, что нулевая гипотеза об однородности распределения выборок $x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_m$ справедлива при ее проверке на основании двустороннего критерия Уилкоксона.

Уровень значимости двустороннего критерия Уилкоксона равен 2β .

Проверка гипотез об однородности распределения двух выборок достаточно проста и может быть реализована при полном отсутствии каких-либо априорных сведений о законах распределения $F_1(x), F_2(z)$. Такая универсальность критерия является весомым обоснованием его широкого применения при анализе результатов испытаний и моделирования сложных систем. Недостатки критерия Уилкоксона проявляются при больших объемах выборок и связаны с трудностями определения общего числа инверсий. В этих случаях целесообразно использовать быстродействующие ЭВМ.

Пример 1. Пусть в результате моделирования и натуральных испытаний получены две выборки $x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_m$, относительно которых сформулировано предположение, что они имеют одинаковые функции распределения. Необходимо проверить это предположение.

Проверку предположения будем осуществлять на основании двустороннего критерия Уилкоксона.

Следуя процедуре проверки, описанной в данном параграфе, для случайных величин x_i, z_j ($i=1, 2, \dots, n=10; j=1, 2, \dots, m=10$), которые приведены в табл. 2, определим, что общее количество инверсий $U=38$.

Задаваясь уровнем значимости $\beta=0,05$, по формулам (43) и (44) находим нижнюю допустимую границу: $u_\beta=82$.

Так как количество прямых инверсий удовлетворяет соотношению $U < u_\beta$, то проверяемую гипотезу об однородности выборок на основании одностороннего критерия Уилкоксона нужно считать справедливой с уровнем значимости, равным $\beta=0,05$.

Рассчитав обратное количество инверсий $n \cdot m - U = 100 - 38 = 62$ и верхнее критическое значение $U_\beta = 2m_{U-u_\beta} = 210 - 82 = 128$, убеждаемся в том, что $nm - U < U_\beta$, а следовательно, проверяемая нулевая гипотеза $H_0: F_1(x) \equiv F_2(z)$ верна на основании двустороннего критерия Уилкоксона, но уже с удвоенным уровнем значимости $2\beta=0,1$.

Итак, в рассматриваемом примере критерий Смирнова и критерий Уилкоксона приводят к одному и тому же заключению: выборка результатов моделирования однородна выборке, полученной в процессе натурных испытаний.

Пример 2. В результате двух проведенных экспериментов получили следующие выборки: 1) —6,5; 6,6; 6,9; 7,0; $n = 4$; 2) —6,0; 6,2; 6,9; 7,3; 7,4; $m = 5$.

Необходимо проверить нулевую гипотезу $H_0: F_1(x) \equiv F_2(z)$.

Составляем вариационный ряд: 6,0; 6,2; 6,5; 6,6; 6,9; 6,9; 7,0; 7,3; 7,4 и для него получаем: $U = 9$; $nm - U = 11$; $u_{\beta=0,05} = 12$; $U_{\beta=0,05} = 40 - 12 = 28$. Таким образом, $U < u_\beta$ и $nm - U < U_\beta$ и нет оснований отвергать гипотезу о принадлежности этих выборок одной генеральной совокупности.

1.6.5. Оценка корреляционных моментов и коэффициентов регрессии

В ходе проведения измерений при испытаниях можно получить статистическую совокупность, состоящую из ряда соответствующих друг другу значений двух и большего числа случайных величин. При этом между рассматриваемыми случайными величинами может быть зафиксирована не однозначная функциональная зависимость, а сложная стохастическая связь, проявляющаяся в том, что измерение одной из случайных величин приводит к изменению закона распределения второй случайной величины. Наиболее простым и важным для практики случаем стохастической связи является корреляционная зависимость.

Коэффициент корреляции представляет собой второй корреляционный момент для нормированных случайных величин. Так, коэф-

фициент корреляции для двух случайных величин x и y может быть представлен в виде

$$\rho_{xy} = \frac{E \{ (x - m_x)(y - m_y) \}}{\sigma_x \sigma_y} = \frac{K_{xy}}{\sigma_x \sigma_y}.$$

При этом ρ_{xy} достигает значения ± 1 лишь тогда, когда имеет место точная функциональная зависимость. Таким образом, с помощью коэффициента корреляции можно оценить степень близости корреляционной зависимости к функциональной.

Второй корреляционный момент K_{xy} (смешанный момент, случайных величин x и y) характеризует рассеивание величин x и y и связь между ними. Значение K_{xy} можно получить по данным совокупности наблюдений, применив формулу:

$$K_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x^*)(y_i - m_y^*).$$

В общем случае, если имеет место система случайных величин x_1, x_2, \dots, x_m и проводится обработка данных, состоящих из результатов измерений при n независимых наблюдениях, то для вычисления элементов корреляционной матрицы можно воспользоваться формулой:

$$K_{kl}^* = \frac{1}{n-1} \sum_{i=1}^n (x_{ki} - m_{xk}^*)(x_{li} - m_{xl}^*).$$

Полученные таким образом корреляционные моменты можно представить в виде корреляционной матрицы:

$$\begin{pmatrix} K_{11}^* & K_{12}^* & \dots & K_{1m}^* \\ K_{21}^* & K_{22}^* & \dots & K_{2m}^* \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ K_{m1}^* & K_{m2}^* & \dots & K_{mm}^* \end{pmatrix}$$

В этой матрице по главной диагонали расположены дисперсии каждой из случайных величин. Действительно,

$$K_{jj}^* = \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - m_{x_j}^*)(x_{ji} - m_{x_j}^*) =$$

$$= \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - m_{x_j}^*)^2 = D^*(x_j).$$

Оценки элементов нормированной корреляционной матрицы можно подсчитать по формуле:

$$\rho_{kl}^* = K_{kl}^* / (S_k S_l),$$

где S_k и S_l — оценки среднего квадратического отклонения, полученные по данным наблюдений.

Стохастическая связь между средними значениями случайных величин x и y может быть охарактеризована в виде линейных уравнений регрессии, которые обычно записывают в виде:

$$y - m_y^* = \frac{\mu_{11}}{\sigma_x^{*2}} (x - m_x^*);$$

$$x - m_x^* = \frac{\mu_{11}}{\sigma_y^{*2}} (y - m_y^*),$$

где

$$m_x^* = \frac{1}{n} \sum_{i=1}^n x_i;$$

$$m_y^* = \frac{1}{n} \sum_{i=1}^n y_i;$$

$$\sigma_x^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x^*)^2;$$

$$\sigma_y^{*2} = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_y^*)^2;$$

$$\mu_{11} = \frac{1}{n} \sum_{i=1}^n x_i y_i - m_x^* m_y^*.$$

Величины

$$\mu_{11}/\sigma_x^{*2} = b_{x|y} \text{ и } \mu_{11}/\sigma_y^{*2} = b_{y|x}$$

называют *коэффициентами регрессии*.

Коэффициент регрессии выражают через коэффициент корреляции так:

$$b_{y|x}^* = \rho^* (\sigma_y^*/\sigma_x^*); \quad b_{x|y}^* = \rho^* (\sigma_x^*/\sigma_y^*).$$

Эмпирическое значение коэффициента регрессии

$$b_{y|x}^* = \sqrt{\frac{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}.$$

Среднее квадратическое отклонение эмпирического коэффициента регрессии

$$\sigma_{b_{y|x}^*} \approx \frac{\sigma_y^* (1 - \rho^{*2})}{\sigma_x^* \sqrt{n}}.$$

Используя приближенное выражение для среднего квадратического отклонения эмпирического коэффициента регрессии, можно представить доверительный интервал для $b_{y|x}^*$:

$$b_{y|x}^* - \alpha \frac{\sigma_y^*}{\sigma_x^*} \frac{1 - \rho^{*2}}{\sqrt{n}} < b_{y|x} < b_{y|x}^* + \alpha \frac{\sigma_y^*}{\sigma_x^*} \frac{(1 - \rho^{*2})}{\sqrt{n}},$$

где α — коэффициент, зависящий от уровня доверительной вероятности P и определяемый из соотношения: $P = 2\Phi(\alpha)$.

С целью построения доверительного интервала для коэффициента корреляции также можно воспользоваться выражением

$$\rho^* - \alpha \frac{1 - \rho^{*2}}{\sqrt{n}} < \rho < \rho^* + \alpha \frac{1 - \rho^{*2}}{\sqrt{n}}.$$

При анализе полученных результатов возникает необходимость проверки значимости вычисленных по данным наблюдений коэффициентов корреляции.

Пусть надо проверить гипотезу о равенстве нулю коэффициента корреляции, что в случае нормальной корреляции равносильно проверке утверждения о независимости рассматриваемых случайных величин.

Для решения этой задачи полученный эмпирический коэффициент корреляции сравнивают с критическим значением, определяемым для заданного уровня надежности.

Если полученное эмпирическое значение ρ^* больше граничного значения, т. е.

$$|\rho^*| > t\sigma_{\rho^*},$$

то гипотезу о равенстве нулю коэффициента корреляции следует отбросить.

1.6.6. Байесова процедура оценки

Байесов подход возможен при наличии априорной информации о законах распределения оцениваемых показателей качества или эффективности. Методы определения стратегий оценки параметров для этого подхода хорошо изучены и сводятся к минимизации среднего риска:

$$\rho[R^*(\vec{y}_n)] = \int w(R) \rho_R[R^*(\vec{y}_n), R] dR, \quad (1)$$

где $w(R)$ — априорное распределение оцениваемого показателя качества; \vec{y}_n — наблюдаемая выборка, полученная либо при статистическом моделировании, либо в процессе проведения натурных испытаний (временной вектор $\vec{y}_n = (y_1, y_2, \dots, y_n)$).

Выражение для условного риска ρ_R , как среднего значения функции потерь $W[R^*(\vec{y}_n), R]$, можно определить соотношением

$$\rho_R[R^*(\vec{y}_n), R] = \int W[R^*(\vec{y}_n), R] l(\vec{y}_n | R) d\vec{y}_n, \quad (2)$$

где $l(\vec{y}_n | R)$ — функция правдоподобия или совместная функция плотности вероятности временного вектора \vec{y}_n .

Переоценку априорных распределений показателя R в апостериорные осуществляют на основании формулы Байеса

$$p(R | \vec{y}_n) = \frac{w(R) l(\vec{y}_n | R)}{p(\vec{y}_n)}; \quad (3)$$

$$p(\vec{y}_n) = \int w(R) l(\vec{y}_n | R) dR. \quad (4)$$

Если ввести условный риск относительно наблюдаемой выборки \vec{y}_n

$$\rho_y[R^*(\vec{y}_n), \vec{y}_n] = \int W[R^*(\vec{y}_n), R] p(R | \vec{y}_n) d\vec{y}_n, \quad (5)$$

то средний риск можно записать в виде:

$$\rho[R^*(\vec{y}_n)] = \int \rho_y[R^*(\vec{y}_n), R] p(\vec{y}_n) d\vec{y}_n.$$

Так как функция плотности вероятности $p(\vec{y}_n)$ положительна во

всей области определения \vec{y}_n , то минимизация среднего риска эквивалентна минимизации условного риска $\rho_y[R^*(\vec{y}_n), R]$ для каждой фиксированной выборки \vec{y}_n . С математической точки зрения решение возникающих функциональных уравнений для всевозможные выпуклых положительных функций потерь $W[R^*(\vec{y}_n), R]$ всегда существует и может быть найдено, например, с помощью градиентного метода поиска экстремумов.

Вычислительная реализуемость байесовой процедуры оценки в значительной степени зависит от сложности математического описания выбираемой функции потерь. Если функция потерь равна

$$W[R^*(\vec{y}_n), R] = [R^*(\vec{y}_n) - R]^2,$$

то уравнения оптимальности оценок получают наиболее простыми, а при гауссовых распределениях их решение может быть найдено в аналитической форме. Однако в некоторых случаях указать конкретный вид функции потерь очень трудно. Поэтому, когда функция потерь неизвестна, на практике в качестве оценок искомых параметров принимают такие, которые максимизируют значение апостериорной функции плотности вероятности $p(R | \vec{y}_n)$.

При отсутствии априорных сведений, оставаясь в рамках байесова подхода, можно рассматривать задачи, связанные с поисками таких наилучших априорных распределений $w(R)$, которые гарантируют определенную оптимальность получаемых оценок в самых неблагоприятных случаях. Если решение подобных задач удастся найти, то говорят, что полученный алгоритм расчета оценок реализует стратегию минимаксного типа в том смысле, что удовлетворяет уравнению

$$\rho^*[R^*(\vec{y}_n)] = \max_{w(R)} \rho[R^*(\vec{y}_n)], \quad (6)$$

где ρ^* — минимаксный риск.

Постулат Байеса гласит, что в случае отсутствия априорных данных, чтобы сохранить замкнутость и логическую обоснованность байесова подхода, нужно принять гипотезу, что все значения истинного показателя R равновероятны, т. е. $w(R) = \text{const}$.

При оценке характеристик реальных систем, когда все физические переменные ограничены размерами области своего изменения, практическое использование постулата Байеса приводит порой к противоречивым заключениям. В этом нетрудно убедиться, если рассмотреть задачу оценки дисперсии нормально распределенной

случайной величины двумя различными способами, которые по своему смыслу должны привести к одному и тому же результату.

Первый способ заключается в том, что сначала рассчитывают оценку среднеквадратического отклонения σ , а потом с помощью известных преобразований находят оценку дисперсии σ^2 . Реализация второго способа оценивания состоит в непосредственном расчете величины дисперсии по наблюдаемой выборке

Чтобы найти искомые оценки, используя постулат Байеса, записывают априорные плотности вероятности для σ

$$w(\sigma) d\sigma = \begin{cases} \frac{d\sigma}{\sigma_{\max} - \sigma_{\min}}, & \text{если } \sigma_{\min} \leq \sigma \leq \sigma_{\max}; \\ 0, & \text{в противном случае,} \end{cases} \quad (7)$$

для σ^2

$$w(\sigma^2) d\sigma^2 = \begin{cases} \frac{d\sigma^2}{\sigma_{\max}^2 - \sigma_{\min}^2}, & \text{если } \sigma_{\min}^2 \leq \sigma^2 \leq \sigma_{\max}^2; \\ 0, & \text{в противном случае.} \end{cases} \quad (8)$$

Рассматривая одну и ту же выборку, например выборку единичного объема, получают, что функции правдоподобия для оцениваемых параметров должны быть равными и записываться в виде:

$$l_1(y|\sigma) = l_2(y|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}. \quad (9)$$

Тогда на основании формул (3), (7) и (8) можно определить распределение апостериорных вероятностей:

$$p_1(\sigma|y) = \frac{c_1}{\sigma} e^{-y^2/2\sigma^2}, \quad \sigma_{\min} \leq \sigma \leq \sigma_{\max};$$

$$p_2(\sigma^2|y) = \frac{c_2}{\sigma^2} e^{-y^2/2\sigma^2}, \quad \sigma_{\min}^2 \leq \sigma^2 \leq \sigma_{\max}^2,$$

где c_1, c_2 — некоторые константы.

Принимая во внимание, что $d\sigma = d\sigma^2/2\sqrt{\sigma^2}$, нетрудно получить распределение $p_1(\sigma^2|y)$, соответствующее распределению $p_1(\sigma|y)$:

$$p_1(\sigma^2|y) = (c_3/\sigma^2) e^{-y^2/2\sigma^2}, \quad \sigma_{\min}^2 \leq \sigma^2 \leq \sigma_{\max}^2.$$

Сравнивая $p_1(\sigma^2|y)$ с $p_2(\sigma^2|y)$, можно убедиться в том, что полученные апостериорные распределения отличны друг от друга, а следовательно, будут разными и оценки величины дисперсии, хотя исходная выборка была одной и той же. Этот факт обусловлен не-

однозначностью, которую допускает постулат Байеса при выборе априорных распределений.

Пример. Пусть задана выборка результатов наблюдений y_1, y_2, \dots, y_n . Относительно выборки известно, что закон распределения гауссов с плотностью вероятности, равной

$$p(y_i | m_y) = \frac{1}{\sqrt{2\pi\sigma_y}} e^{-(y_i - m_y)^2/2\sigma_y^2}, \quad i = 1, 2, \dots, n.$$

Требуется определить оценку математического ожидания при условии, что величина дисперсии σ_y^2 известна, а априорное распределение оцениваемого параметра

$$w(m_y) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{(m_y - m_0)^2}{2\sigma_0^2}\right\},$$

где σ_0, m_0 — заданные параметры.

В качестве функции потерь выберем функцию

$$W\{m_y^*, m_y\} = (m_y^* - m_y)^2.$$

Далее находим совместную функцию плотности вероятности значений $y_i, i = 1, 2, \dots, n$

$$l(\vec{y}_n | m_y) = \prod_{i=1}^n p(y_i | m_y) = (\sqrt{2\pi}\sigma_y)^{-n} \exp\left\{-\sum_{i=1}^n (y_i - m_y)^2/2\sigma_y^2\right\},$$

$$\vec{y}_n = (y_1, y_2, \dots, y_n).$$

Применяя формулу Байеса, получаем:

$$p(m_y | \vec{y}_n) = \frac{1}{\sqrt{2\pi}\sigma_{m_y}^*} \exp\left\{-\frac{[m_y - \sigma_{m_y}^* (ny^*/\sigma_y^2 + m_0/\sigma_0^2)]^2}{2\sigma_{m_y}^{*2}}\right\}; \quad (10)$$

$$\frac{1}{\sigma_{m_y}^{*2}} = \frac{n}{\sigma_y^2} + \frac{1}{\sigma_0^2}; \quad y^* = \frac{1}{n} \sum_{i=1}^n y_i.$$

Для выбранной функции потерь на основании (10) нетрудно записать алгоритм построения искомых оценок:

$$m_y^* = y^* [1 + \sigma_y^2/\sigma_0^2 n]^{-1} + m_0 [1 + n\sigma_0^2/\sigma_y^2]^{-1}.$$

Оценки, найденные на основании этого алгоритма, получают смещенными; величина этого смещения

$$E\{m_y^*\} - m_y = \frac{m_0 - m_y}{1 + \sigma_0^2 n / \sigma_y^2}.$$

Подставив выражения для m_y^* , $w(m_y)$ в (2), а затем в (1) и преобразовав полученные выражения, находим зависимость среднего риска $\rho[m_y^*]$ от объема выборки и точности априорных сведений:

$$\rho = \frac{\sigma_y^2}{n} [1 + \sigma_0^2 / \sigma_y^2 n]^{-1}.$$

Если устремить $\sigma_0^2 \rightarrow \infty$, что соответствует полному отсутствию априорных сведений, то можно убедиться в том, что в этом случае риск как функция σ_0^2 достигает своего максимума.

Стратегия оценки, соответствующая этой ситуации, сводится к тому, что в качестве оценки математического ожидания следует принять величину y^* , т. е.

$$m_y^* = y^* = \frac{1}{n} \sum_{i=1}^n y_i.$$

Так как риск $\rho[m_y^*]$ максимален при $m_y^* = y^*$, то такая стратегия оценки является минимаксной и в рассматриваемом случае приводит к несмещенным оценкам.

Рассмотрим более сложный случай байесовой оценки параметров, характеризующих качество управляемых систем, в которых процессы смены состояний описываются многомерными марковскими цепями.

Пусть исследуемая система вместе с измерительными устройствами описывается разностными уравнениями:

$$\mathbf{x}_{i+1} = \mathbf{F}_i(\mathbf{x}_i, \mathbf{u}_i, \xi_i), \quad i=0, 1, 2, \dots; \quad (11)$$

$$\mathbf{y}_i = \mathbf{G}_i(\mathbf{x}_i, \eta_i), \quad i=0, 1, 2, \dots, \quad (12)$$

где \mathbf{x}_i — n -мерный вектор состояния в i -й момент; \mathbf{u}_i — r -мерный вектор управления в i -й момент; \mathbf{y}_i — m -мерный вектор наблюдений в i -й момент; ξ_i, η_i — взаимно независимые случайные шумы измерительных устройств и отклонения параметров в i -й момент.

Чтобы проиллюстрировать байесову процедуру оценки вектора \mathbf{x}_i , будем предполагать, что априорное распределение вектора \mathbf{x}_0 задано и закон распределения $p(\mathbf{y}_{i+1} | \mathbf{x}_{i+1})$, определяющий точность измерений, известен.

Для нахождения искомого апостериорного распределения $p(\mathbf{x}_{i+1} | \mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{i+1})$ рассмотрим равенство:

$$p(\mathbf{x}_{i+1}, \mathbf{y}_{i+1} | \vec{\mathbf{y}}_i) = p(\mathbf{y}_{i+1} | \vec{\mathbf{y}}_i) p(\mathbf{x}_{i+1} | \vec{\mathbf{y}}_{i+1}), \quad (13)$$

$$\vec{\mathbf{y}}_{i+1} = (\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{i+1}).$$

Выражая $p(\mathbf{y}_{i+1} | \vec{\mathbf{y}}_i)$ и $p(\mathbf{x}_{i+1}, \mathbf{y}_{i+1} | \vec{\mathbf{y}}_i)$ через известные распределения, получают:

$$p(\mathbf{y}_{i+1} | \vec{\mathbf{y}}_i) = \int p(\mathbf{x}_i | \vec{\mathbf{y}}_i) p(\mathbf{x}_{i+1} | \mathbf{x}_i, \mathbf{u}_i) p(\mathbf{y}_{i+1} | \mathbf{x}_{i+1}) d(\mathbf{x}_i, \mathbf{x}_{i+1}), \quad (14)$$

$$p(\mathbf{x}_{i+1}, \mathbf{y}_{i+1} | \vec{\mathbf{y}}_i) = \int p(\mathbf{x}_i | \vec{\mathbf{y}}_i) p(\mathbf{x}_{i+1} | \mathbf{x}_i, \mathbf{u}_i) p(\mathbf{y}_{i+1} | \mathbf{x}_{i+1}) d\mathbf{x}_i. \quad (15)$$

Распределение $p(\mathbf{x}_0 | \mathbf{y}_0)$ может быть найдено на основании исходных распределений $p(\mathbf{x}_0), p(\mathbf{y}_0 | \mathbf{x}_0)$:

$$p(\mathbf{x}_0 | \mathbf{y}_0) = \frac{p(\mathbf{x}_0) p(\mathbf{y}_0 | \mathbf{x}_0)}{\int p(\mathbf{x}_0) p(\mathbf{y}_0 | \mathbf{x}_0) d\mathbf{x}_0}, \quad (16)$$

Соотношение (13) позволяет записать рекуррентный алгоритм переоценки апостериорных распределений:

$$p(\mathbf{x}_{i+1} | \vec{\mathbf{y}}_{i+1}) = \frac{p(\mathbf{x}_{i+1}, \mathbf{y}_{i+1} | \vec{\mathbf{y}}_i)}{p(\mathbf{y}_{i+1} | \vec{\mathbf{y}}_i)} \quad (17)$$

при условии, что

$$p(\mathbf{x}_i | \vec{\mathbf{y}}_i) \text{ и } p(\mathbf{y}_{i+1} | \vec{\mathbf{y}}_i)$$

ранее уже получены.

В частных случаях для $p(\mathbf{x}_{i+1} | \vec{\mathbf{y}}_i)$ удастся получить компактные аналитические соотношения.

Пример. Если система и процессы измерений описываются уравнениями

$$\mathbf{x}_{i+1} = \Phi \mathbf{x}_i + \mathbf{H} u_i;$$

$$y_{i+1} = \mathbf{h}^T \mathbf{x}_{i+1} + \eta_{i+1},$$

где Φ — матрица размерности $n \times n$; u_i — скалярная переменная; \mathbf{H} — n -мерный вектор; \mathbf{h} — постоянный n -мерный вектор $\mathbf{h}^T = (1, 0, 0, \dots)$; η_{i+1} — шумы измерительного устройства (скалярная нормально распределенная переменная с дисперсией σ_η^2), и стоит задача уточнения начального вектора состояния по результатам последующих наблюдений y_0, y_1, \dots, y_b , то алгоритм переоценки может быть записан в виде:

$$\begin{aligned} \mathbf{x}_{0,i+1}^{*T} &= \left[\mathbf{x}_{0,i}^{*T} \mathbf{K}_i + \frac{1}{2\sigma_\eta^2} \mathbf{h}^T \Phi^{i+1} (y_{i+1} - \mathbf{h}^T \mathbf{L}_{i+1}) \right] \times \\ &\times \left[\mathbf{K}_i + \frac{1}{2\sigma_\eta^2} \mathbf{h}^T \Phi^{i+1} \mathbf{h}^T \Phi^{i+1} \right]^{-1}; \\ \mathbf{K}_{i+1} &= \mathbf{K}_i + \frac{1}{2\sigma_\eta^2} \mathbf{h}^T \Phi^{i+1} \mathbf{h}^T \Phi^{i+1}; \\ \mathbf{L}_{i+1} &= \sum_{l=1}^{i+1} \Phi^{i-l} \mathbf{H} \mathbf{u}_{l-\varepsilon}; \\ \Phi^i &= \underbrace{\Phi \times \Phi \times \dots \times \Phi}_{i \text{ раз}}. \end{aligned} \quad (18)$$

При выводе (18) предполагалось, что априорное распределение $p(\mathbf{x}_0)$ гауссово с плотностью вероятности

$$p(\mathbf{x}_0) = \frac{\sqrt{|\mathbf{K}_0|}}{(2\pi)^{\frac{1}{2}n}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_0 - \mathbf{m}_{\mathbf{x}_0})^T \mathbf{K}_0^{-1} (\mathbf{x}_0 - \mathbf{m}_{\mathbf{x}_0}) \right\},$$

где $\mathbf{m}_{\mathbf{x}_0}$ — заданный n -мерный вектор; \mathbf{K}_0 — заданная ковариационная матрица размерности $n \times n$.

Если положить $\mathbf{H} = 0$ и

$$\Phi = \begin{pmatrix} 1 & \tau & \tau^2/2! & \tau^3/3! & \dots & \tau^n/n! \\ 0 & 1 & \tau & \tau^2/2! & \dots & \tau^{n-1}/(n-1)! \\ 0 & 0 & 1 & \tau & \dots & \tau^{n-2}/(n-2)! \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

то алгоритм (18) будет соответствовать режиму последовательного уточнения коэффициентов сглаживающего полинома n степени, когда интервал квантования входной информации равен τ и привязка коэффициентов полинома осуществляется к левому концу временного интервала наблюдения.

Прием, посредством которого можно избежать довольно трудоемкие вычисления при нахождении интегралов видов (14) и (15), заключается в следующем; если удастся установить взаимнооднозначное соответствие между текущим вектором состояния системы и его начальным значением, то задачи переоценки апостериорных распределений $p(\mathbf{x}_{i+1} | \mathbf{y}_{i+1})$ могут быть трансформированы в задачи,

решение которых должно определить алгоритм последовательного уточнения распределений начального вектора состояния $p(\mathbf{x}_{0,i+1} | \mathbf{y}_{i+1})$

Достоинство такого перехода состоит в том, что распределение $p(\mathbf{x}_{i+1} | \mathbf{x}_i, u_i)$ вырождается в $\delta(\mathbf{x}_{i+1} - \mathbf{x}_i)$ распределение и операции интегрирования уравнений (14) и (15) по \mathbf{x}_i сводятся к простой замене в подинтегральных функциях \mathbf{x}_i на \mathbf{x}_{i+1} .

Если система описывается линейными разностными или дифференциальными уравнениями, а шумы измерений подчиняются гауссовому распределению, то этот прием позволяет сравнительно просто получить правило нахождения $p(\mathbf{x}_{0,i+1} | \mathbf{y}_{i+1})$, а затем, используя законы линейного преобразования, найти алгоритм расчета апостериорных распределений для текущего вектора состояния \mathbf{x}_{i+1} .

1.6.7. Принцип максимального правдоподобия

При отсутствии априорных сведений широко используемым способом оценки выходных показателей является метод максимального правдоподобия. Для определения оценок по этому методу составляют уравнения:

$$\frac{\partial l(\vec{\mathbf{y}}_n | R)}{\partial R} = 0, \quad \vec{\mathbf{y}}_n = (y_1, y_2, \dots, y_n) \quad (1)$$

решают их и в качестве оценок выходных показателей R берут максимально правдоподобные значения. Следовательно, оценка максимального правдоподобия удовлетворяет уравнению

$$\max_R l(\vec{\mathbf{y}}_n | R) = l(\vec{\mathbf{y}}_n | R^*).$$

При использовании метода максимального правдоподобия необходимо знать:

1. Существуют ли корни уравнения правдоподобия (1)?
2. Обращает ли найденный корень в абсолютный максимум

функцию правдоподобия $l(\vec{\mathbf{y}}_n | R)$?

Если известны ответы на поставленные вопросы, то можно охарактеризовать свойства оценок максимального правдоподобия. Например, если найденный корень уравнения обращает неравенство Крамера — Рао в равенство, то можно утверждать, что полученная оценка будет эффективной и более точной оценки найти нельзя. Доказательство этого утверждения основывают на том, что в данном случае просто реализуется верхняя грань точности, которая в принципе может быть достигнута при идентификации параметров по результатам проведенных наблюдений.

Обычно неравенство Крамера — Рао записывают в виде

$$E \{ (R^* - R)^2 \} \geq \frac{\gamma^2(R)}{n},$$

где n — объем выборки; $\gamma^2(R)$ — величина, численно равная

$$\gamma^2(R) = \frac{[1 + d\beta(R)dR]^2}{\int \left[\frac{\partial \ln l(\vec{y} | R)}{\partial R} \right]^2 l(\vec{y} | R) d\vec{y}},$$

где $\beta(R)$ — величина смещения, рассчитываемая по формуле

$$\beta(R) = E \{ R^* \} - R.$$

При достаточно общих предположениях, какими являются следующие условия:

- 1) если $R_1 = R_2$, то $l(\vec{y} | R_1) = l(\vec{y} | R_2)$ почти всюду в области определения R_1 и R_2 (однозначность функции правдоподобия);
- 2) если R_0 является истинным значением оцениваемого параметра, то

$$\lim_{\Delta R \rightarrow 0} E \left\{ \frac{l(\vec{y} | R_0 + \Delta R)}{l(\vec{y} | R_0)} \right\} = E \left\{ \ln \frac{l(\vec{y} | R_0)}{l(\vec{y} | R_0)} \right\}$$

(непрерывность логарифмической функции правдоподобия в точке $R=R_0$);

- 3) для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что для всех R_1 и R_2 , для которых $|R_1 - R_2| < \delta$, справедливо неравенство $|l(\vec{y}_n | R_1) - l(\vec{y}_n | R_2)| < \varepsilon$ для всех \vec{y}_n ,

- 4) если функция правдоподобия $l(\vec{y} | R)$ имеет единственный максимум, что соответствует одному корню уравнения (1), то можно показать состоятельность оценок максимального правдоподобия, т. е. $P\{|R^* - R| \leq \varepsilon\} \rightarrow 1$ для любого $\varepsilon > 0$.

Если от функции правдоподобия потребовать также, чтобы при каждом \vec{y}_n функция $l(\vec{y}_n | R)$ была дважды непрерывно дифференцируема по параметру R , то можно доказать асимптотическую нормальность распределения получаемых оценок с параметрами R и дисперсией:

$$\sigma_{R^*}^2 = \frac{1}{-E \left\{ \frac{\partial^2 L}{\partial R^2} \right\}} \approx \frac{1}{-\left[\frac{\partial^2 L}{\partial R^2} \right]_{R=R^*}},$$

где L — логарифмическая функция правдоподобия

$$L(\vec{y}_n | R) = \ln l(\vec{y}_n | R).$$

Используя асимптотическую нормальность оценок максимального правдоподобия, можно доказать их асимптотическую эффективность.

Эти выводы получают для рассматриваемых условий с использованием неравенства Крамера — Рао при $n \rightarrow \infty$.

Для малых объемов выборок, когда известно, что эффективная оценка существует, принцип максимального правдоподобия гарантирует получение именно этой оценки и при сделанных предположениях регулярности функции правдоподобия найти более точную оценку нельзя.

Первые результаты по определению условий состоятельности и асимптотической эффективности оценок максимального правдоподобия были получены Г. Крамером.

Пример. Пусть в процессе испытаний получена выборка y_1, y_2, \dots, y_n являющаяся элементом нормальной совокупности $N\{m_y, \sigma_y^2\}$. Причем, параметры m_y, σ_y^2 неизвестны и априорные распределения $p(m_y), p(\sigma_y^2)$ не заданы.

Необходимо найти максимально правдоподобные оценки параметров m_y и σ_y^2 .

Чтобы рассчитать оценки m_y, σ_y^2 , следуя принципу максимального правдоподобия, составим функцию правдоподобия, которая в рассматриваемом случае имеет вид

$$l(\vec{y}_n | m_y, \sigma_y^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp \left\{ -\frac{(y_i - m_y)^2}{2\sigma_y^2} \right\} = (1/\sigma_y)^n (2\pi)^{-n/2} \exp \left\{ -\sum_{i=1}^n \frac{(y_i - m_y)^2}{2\sigma_y^2} \right\}.$$

Записывая (без учета множителя $(2\pi)^{-n/2}$)

$$\frac{\partial \ln l(\vec{y}_n | \sigma_y^2)}{\partial \sigma_y^2} = -\frac{n}{\sigma_y^2} + \frac{1}{\sigma_y^4} \sum_{i=1}^n (y_i - m_y)^2 = 0,$$

$$\frac{\partial \ln l(\vec{y}_n | m_y)}{\partial m_y} = \frac{1}{\sigma_y^2} \sum_{i=1}^n (y_i - m_y) = 0;$$

нетрудно получить алгоритмы построения оценок

$$m_y' = \frac{1}{n} \sum_{i=1}^n y_i;$$

$$\sigma_y^{*2} = \frac{1}{n} \sum_{i=1}^n (y_i - m_y^*)^2.$$

Относительно этих оценок можно сказать, что m_y^* несмещенная оценка, σ_y^{*2} смещена, так как

$$E \{ \sigma_y^{*2} \} = \frac{n-1}{n} \sigma_y^2.$$

Естественным обобщением рассмотренного примера является многомерный случай, когда элементами выборки y_1, y_2, \dots, y_n являются r -мерные нормально распределенные векторы. Стоит задача по результатам наблюдений необходимо оценить вектор математического ожидания \mathbf{m}_y и ковариационную матрицу \mathbf{K}_y .

Функция правдоподобия

$$l(\vec{y}_n | \mathbf{m}_y, \mathbf{K}_y) = \prod_{i=1}^n \frac{(2\pi)^{-r/2}}{|\mathbf{K}_y|^{1/2}} \exp \left\{ -\frac{1}{2} (y_i - \mathbf{m}_y)^T \mathbf{K}_y^{-1} (y_i - \mathbf{m}_y) \right\} =$$

$$= (2\pi)^{-nr/2} |\mathbf{K}_y|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \mathbf{m}_y)^T \mathbf{K}_y^{-1} (y_i - \mathbf{m}_y) \right\},$$

$$\vec{y}_n = (y_1, y_2, \dots, y_n).$$

Оценки максимального правдоподобия в данном случае

$$\mathbf{m}_y^* = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\mathbf{K}_y^* = \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{m}_y^*) (y_i - \mathbf{m}_y^*)^T.$$

Когда функции правдоподобия не допускают аналитического исследования корней уравнения правдоподобия, для определения оценок максимального правдоподобия необходимо привлекать итерационные методы поиска экстремумов. В одномерном случае решение подобных задач может быть найдено графическим способом. Для этого нужно функцию правдоподобия нанести на график и визуально определить искомые оценки. При этом очень просто указать доверительный интервал найденной оценки.

1.6.8. Алгоритмы метода наименьших квадратов

При обработке и анализе экспериментальных данных широкое применение находит метод наименьших квадратов. Вычислительная процедура, соответствующая методу наименьших квадратов, позволяет найти искомые оценки параметров и может быть получена как результат решения следующих уравнений:

$$\sum_{i=1}^n [y(t_i) - \varphi(t_i, a_1^*, a_2^*, \dots, a_j^*, \dots, a_r^*)]^2 = \min_{a_j} \quad (1)$$

где $y(t_i)$ — i -й элемент выборки результатов наблюдений y_1, y_2, \dots, y_n ; $\varphi(t)$ — заданная функция; a_j^* — оценки неизвестных коэффициентов a_j , t_i — момент времени, соответствующий y_i наблюдению.

Физический смысл уравнений (1) состоит в том, что в качестве оценок параметров a_j необходимо принять такие значения a_j^* , которые минимизируют сумму квадратов отклонений результатов наблюдений от аппроксимирующей кривой.

В общем случае задача определения оценок параметров a_j^* сводится к разработке методов поиска корней следующей системы уравнений:

$$\sum_{i=1}^n [y(t_i) - \varphi(t_i, \mathbf{a}^*)] \frac{\partial \varphi(t_i, \mathbf{a}^*)}{\partial a_j^*} = 0;$$

$$\mathbf{a}^* = (a_1^*, a_2^*, \dots, a_j^*, \dots, a_r^*), \quad j = 1, 2, \dots, r. \quad (2)$$

В частном случае, когда

$$\varphi(t_i, \mathbf{a}^*) = \sum_{j=1}^r a_j^* f_j(t_i), \quad (3)$$

где $f_j(t_i)$ — заданная система линейно независимых функций, получаем систему

$$\sum_{i=1}^n \sum_{j=1}^r a_j^* f_j(t_i) f_k(t_i) = \sum_{i=1}^n y(t_i) f_k(t_i), \quad (4)$$

$$(k = 1, 2, \dots, r)$$

которая линейна относительно оцениваемых параметров. Если обозначить:

$$\mathbf{B} = \begin{pmatrix} \sum_{i=1}^n f_1(t_i) f_1(t_i); & \sum_{i=1}^n f_2(t_i) f_1(t_i); & \dots; & \sum_{i=1}^n f_r(t_i) f_1(t_i) \\ \sum_{i=1}^n f_1(t_i) f_2(t_i); & \sum_{i=1}^n f_2(t_i) f_2(t_i); & \dots; & \sum_{i=1}^n f_r(t_i) f_2(t_i) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n f_1(t_i) f_r(t_i); & \sum_{i=1}^n f_2(t_i) f_r(t_i); & \dots; & \sum_{i=1}^n f_r(t_i) f_r(t_i) \end{pmatrix} \quad (5)$$

$$\mathbf{y} = \begin{pmatrix} \sum_{i=1}^n y(t_i) f_1(t_i) \\ \sum_{i=1}^n y(t_i) f_2(t_i) \\ \vdots \\ \sum_{i=1}^n y(t_i) f_r(t_i) \end{pmatrix}, \quad (6)$$

то систему уравнений (4) можно записать в матричном виде:

$$\mathbf{B}\mathbf{a}^* = \mathbf{y}, \quad (7)$$

а ее решение представить в следующей форме

$$\mathbf{a}^* = \mathbf{B}^{-1} \mathbf{y}, \quad (8)$$

если существует обратная матрица \mathbf{B}^{-1} .

Из полученного соотношения следует, что в рассматриваемом случае оценки параметров a_j^* являются линейными функциями результатов наблюдений, так как элементы вектора \mathbf{y} линейно зависят от элементов выборки $y(t_1), y(t_2), \dots, y(t_n)$.

Поскольку выражение (1) представляет собой неотрицательный квадратный многочлен относительно переменных $a_1^*, a_2^*, \dots, a_r^*$, то минимум (1) всегда существует и бывает единственным при $r \leq n$.

Если матрица \mathbf{B} вырожденная, то система уравнений (4) будет иметь несколько линейно зависимых решений.

Метод наименьших квадратов позволяет в классе линейных оценок находить такие оценки, которые являются эффективными среди всех линейных несмещенных оценок. Причем, это свойство не зависит от вида распределения элементов выборки y_1, y_2, \dots, y_n .

Пример. Пусть задана выборка результатов наблюдений $y(t_1), \dots, y(t_n)$ и стоит задача оценки точности расчета коэффициентов аппроксимирующего полинома

$$\varphi(t, a_1^*, a_2^*, \dots, a_r^*) = \sum_{j=1}^r a_j^* t^j \quad (9)$$

на основании метода наименьших квадратов.

В рассматриваемом случае

$$\mathbf{B} = \begin{pmatrix} n; & \sum_{i=1}^n t_i; & \sum_{i=1}^n t_i^2; & \dots; & \sum_{i=1}^n t_i^r \\ \sum_{i=1}^n t_i; & \sum_{i=1}^n t_i^2; & \sum_{i=1}^n t_i^3; & \dots; & \sum_{i=1}^n t_i^{r+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n t_i^r; & \sum_{i=1}^n t_i^{r+1}; & \sum_{i=1}^n t_i^{r+2}; & \dots; & \sum_{i=1}^n t_i^{2r} \end{pmatrix} \quad (10)$$

$$\mathbf{y} = \begin{pmatrix} \sum_{i=1}^n y(t_i) \\ \sum_{i=1}^n y(t_i) t_i \\ \vdots \\ \sum_{i=1}^n y(t_i) t_i^r \end{pmatrix} \quad (11)$$

и для равноточных наблюдений

$$\mathbf{K}(\mathbf{a}^*) = \mathbf{B}^{-1} \sigma_y^2, \quad (12)$$

где σ_y^2 — дисперсия наблюдений y_i ; $\mathbf{K}(\mathbf{a}^*)$ — ковариационная матрица, характеризующая точность расчета оценок a_j^* .

Приведенное соотношение определяет как бы предельно достижимую точность оценок, получаемых на основании метода наименьших квадратов. Это объясняется тем, что выражение (12) не учитывает ошибок, возникающих в процессе расчета оценок на вычислительных машинах. Количественно охарактеризовать эту группу ошибок очень трудно. Однако общие закономерности указать можно, состоят они в том, что при увеличении объема выборки n так же, как и при возрастании степени аппроксимирующего полинома r , матрица \mathbf{B} становится менее обусловленной. В свою очередь этот факт приводит к возрастанию роли ошибок округления при расчете оценок a_j^* по формулам (10).

Влияние ошибок округления на точность получаемых оценок прямо пропорционально квадратному корню из отношения наибольшего собственного значения симметрической матрицы \mathbf{B} к ее наименьшему собственному значению. Если указанное отношение превышает 10^4 , то полученные оценки не будут приемлемой точности. Для малых степеней аппроксимирующего полинома можно преобразовать уравнения (12) и записать непосредственно весовые функции сглаживающих фильтров:

при $r=1$

$$K\{i\} = \frac{6}{(n+1)(n+2)} \left(\frac{2n+1}{3} - i \right), \quad 0 \leq i \leq n;$$

при $r=2$

$$K\{i\} = \frac{3}{(n+1)(n+2)(n+3)} [3n^2 + 3n + 2 - 6(1+2n)i + 10i^2],$$

$$0 \leq i \leq n.$$

При такой записи для формирования оценок $y^*(t_n)$ нет необходимости в расчете коэффициентов a_i^* с последующим раскрытием полинома (9) по $t=t_n$. В этом случае определение $y^*(t_n)$ можно произвести по формуле

$$y^*(t_n) = \sum_{i=0}^n K[n-i] y(t_i).$$

Точность подобных оценок, если шумы измерений $y(t_i)$ гауссовы, для рассматриваемых примеров равны:

при $r=1$

$$D\{y^*(t_n)\} = \frac{2\sigma_y^2(2n+1)}{(n+1)(n+2)};$$

при $r=2$

$$D\{y^*(t_n)\} = \frac{3(3n^2 + 3n + 2)\sigma_y^2}{(n+1)(n+2)(n+3)}.$$

Для достаточно больших n

$$D\{y^*(t_n)\} \cong \frac{4\sigma_y^2}{n} \quad \text{при } r=1;$$

$$D\{y^*(t_n)\} \cong \frac{9\sigma_y^2}{n} \quad \text{при } r=2.$$

В методе наименьших квадратов, так же как и при использовании методов обработки экспериментальных данных, изложенных в предыдущих параграфах, предполагают, что математическая зависимость $\varphi(t, \mathbf{a})$, описывающая процесс изменения полезного сигнала $x(t)$, известна с точностью до некоторой группы параметров

$\mathbf{a} = (a_1, a_2, \dots, a_r)$. Однако в реальных условиях таких сведений может и не быть. В этом случае возникает задача разработки процедуры, позволяющей в заданном классе функции \mathbf{H} выбирать такую функцию $\varphi_{\mathbf{H}}^*(t, \mathbf{a})$, чтобы близость значений к $\varphi_{\mathbf{H}}^*(t, \mathbf{a})$ результатам наблюдений $\vec{y}_n = (y_1, y_2, \dots, y_n)$ была в некотором смысле наилучшей. Близость $\varphi_{\mathbf{H}}(t, \mathbf{a})$ к $x(t)$ можно оценивать в точках разностной сетки $t_i, i=1, 2, \dots, n$. Для сравнения различных решений выбирают критерий оптимальности $W_{\mathbf{H}}(x(t), \varphi_{\mathbf{H}}(t, \mathbf{a}))|_{t_i} = W_{\mathbf{H}, t_i}$. В результате получают, что наилучшее приближение $\varphi^*(t, \mathbf{a}^*)$ и найденные значения параметров \mathbf{a}^* должны удовлетворять уравнению

$$W_{\mathbf{H}, t_i}^* = \min_{\mathbf{a}, \varphi_{\mathbf{H}}} W_{\mathbf{H}}(x(t), \varphi_{\mathbf{H}}(t, \mathbf{a}))|_{t_i}. \quad (13)$$

В зависимости от качества априорной информации, характера распределений y_i , а также сложности аппроксимирующих функций $\varphi_{\mathbf{H}}(t, \mathbf{a}) \in \mathbf{H}$, конкретные выражения для $W_{\mathbf{H}, t_i}$ будут различными. Однако для понимания существа рассматриваемых явлений важно знать, что $W_{\mathbf{H}, t_i} = W_{\mathbf{H}, t_i}(\delta_{\varphi}, \delta_a)$ характеризует как динамическую точность δ_a приближения $\varphi_{\mathbf{H}}^*(t, \mathbf{a})$ к $x(t)$, так и случайные ошибки δ_{φ} описания $x(t)$ с помощью $\varphi_{\mathbf{H}}(t, \mathbf{a}^*)$, когда оценки \mathbf{a}^* найдены по выборке \vec{y}_n ограниченного объема.

В методе наименьших квадратов в качестве $W_{\mathbf{H}, t_i}$ обычно принимают квадрат суммарной ошибки, равный сумме квадратов динамической δ_a и случайной δ_{φ} ошибок.

Для определения δ_a нужно знать способ конструирования $x(t)$ с помощью функций $\varphi_{\mathbf{H}}(t, \mathbf{a})$. Конкретное выражение δ_a можно получить только на основании теоретических исследований.

Чтобы найти δ_{φ} нужно располагать сведениями о точности расчета оценок \mathbf{a}^* при условии, что рассматриваемая функция $\varphi_{\mathbf{H}}(t, \mathbf{a})$ точно описывает полезный сигнал $x(t)$.

В условиях полной информации (известен закон распределения $p(y_i)$ и $x(t)$ точно описывается некоторой функцией $\varphi_{\mathbf{H}}(t, \mathbf{a}) \in \mathbf{H}$) влияние флуктуационных ошибок измерений на $W_{\mathbf{H}, t_i}$ можно описать аналитическим способом. Когда априорная информация недостоверна или ее недостаточно, то оценку влияния δ_{φ} на $W_{\mathbf{H}, t_i}$ можно осуществить непосредственно по результатам наблюдений y_n .

При известных значениях $W_{\mathbf{H}}(x(t), \varphi_{\mathbf{H}}(t, \mathbf{a}^*))$ для определения

наилучшего приближения $\varphi^*(t, \mathbf{a}^*)$ нужно сравнить значения $W_{\mathbf{H}}(x(t), \varphi_{\mathbf{H}}(t, \mathbf{a}^*))$ и в качестве решения (13) использовать $\varphi_{\mathbf{H}}(t, \mathbf{a}^*)$, для которой $W_{\mathbf{H}, t_i}$ минимально.

Таким образом, когда известно, что полезный сигнал $x(t)$ принадлежит множеству \mathbf{H} , то при известном способе расчета $W_{\mathbf{H}, t_i}$ минимизация (13) позволяет выбрать наилучшее приближение $\varphi^*(t, \mathbf{a}^*)$ к $x(t)$.

На практике наиболее характерен случай, когда наблюдаемый сигнал $y(t)$ состоит из аддитивной смеси полезного сигнала $x(t)$ и гауссовой помехи $\delta(t)$ с нулевым средним и известной дисперсией σ^2 . Аппроксимация полезного сигнала

$$x(t) = \sum_{i=1}^r a_i Q_i(t) \quad (14)$$

осуществляется с помощью ортонормированных полиномов $Q_i(t)$. При этом рассматривается задача выбора наименьшей степени аппроксимирующего полинома r , при которой следующее $r+1$ приближение не приводит к существенному уточнению оценки

$$x_r^*(t) = x_{r-1}^*(t) + Q_r(t) \sum_{i=1}^n y_i Q_r(t_i).$$

Решение этой задачи нетрудно получить, если принять во внимание, что при $\lambda_r = Q_r(t) \sum_{i=1}^n y_i Q_r(t_i) = 0$ отношение λ_r^2/σ^{2*} имеет $\chi^2(v)$ распределение с числом степеней свободы $v = 1$.

Отсюда следует правило выбора степени аппроксимирующего полинома: до тех пор пока величина λ_r значительно отличается от нуля, то в разложение (14) необходимо ввести полином $Q_r(t)$ и осуществить аналогичные операции для полинома более высокой $r+1$ степени.

При практическом применении этого правила нужно знать критическое значение χ_{β}^2 , соответствующее принятому или заданному уровню значимости β .

Если дисперсия σ^2 неизвестна и рассчитывается по результатам наблюдений y_n , то отношение λ_r^2/σ^{2*} имеет $F_{1, v}$ распределение с $v = n - r - 1$. Расчет σ^{2*} можно осуществить по формуле

$$\sigma^{2*} = \left(\sum_{i=1}^n y_i^2 - \sum_{j=1}^r \lambda_j^2 \right) / (n - r - 1).$$

Критическое значение $F_{1, v}^+(\beta)$ рассчитывают по заданному уровню значимости β . Гипотезу $H_0: \lambda_{r+1} = 0$ принимают, если реализо-

ванное значение $F_{1, v}$ меньше критического значения $F_{1, v}^+(\beta)$.

Если для аппроксимации $x(t)$ используют обычные полиномы:

$$x(t) = \sum_{k=1}^r a_k t^k,$$

то процедуру выбора r можно сократить, воспользовавшись статистикой:

$$\frac{R_r - R_{r+1}}{R_{r+1}} (n - r - 1) \sim F_{1, n-r-1},$$

где $R_r = \sum_{i=1}^n (y_i - y_i^*)^2$; $y_i^* = \sum_{k=1}^r a_k^* t^k$.

Процедура проверки гипотез на основании F -критерия Фишера изложена ранее.

Достаточность приведенного правила выбора наименьшей степени аппроксимирующего полинома основана на предложении, что если для некоторой степени r гипотеза $H_0: \lambda_r = 0$ оказалась справедливой, то по индукции справедлива и гипотеза $H_0: \lambda_{r+1} = 0$. Однако при практических оценках такое индуктивное распространение результатов проверки гипотез $H_0: \lambda_r = 0$ может и не выполняться. В этих случаях возможны ошибки при выборе значений наилучшей степени аппроксимирующего полинома. Чтобы исключить эти ошибки, правило выбора r нужно скорректировать и сформулировать так: если по результатам наблюдений \mathbf{y}_n гипотезы $H_0: \lambda_{r+1} = 0$ и $H_0: \lambda_{r+2} = 0$ оказываются справедливыми, то полиномы r степени отвечают требованиям, предъявляемым выбранным критерием оптимальности к наилучшему приближению. Если одна из гипотез не подтверждается, то степень аппроксимирующего полинома нужно увеличить на 1 и для нового приближения выполнить все указанные выше процедуры.

Для сокращения объема вычислительных работ целесообразно сначала проверить гипотезу $H_0: \lambda_{r+2} = 0$ и в случае ее выполнения решить вопрос о возможности понижения степени аппроксимирующего полинома на 1 (проверить гипотезу $H_0: \lambda_{r+1} = 0$).

Если не подтверждается гипотеза $H_0: \lambda_{r+2} = 0$, то можно сразу степень полинома увеличить на 2 и снова приступить к проверке, но уже гипотезы $H_0: \lambda_{r+4} = 0$. При такой организации вычислительного процесса удается сократить вдвое объем расчетов.

1.7. Методы статистической обработки информации

1.7.1. Ощая характеристика и формализация задач обработки информации

Автоматизированные системы обработки информации (АСОИ) входят в качестве подсистем в различные автоматизированные системы управления производственными процессами (АСУТП), автоматизации научных исследований (АСНИ), сбора и обработки метеоинформации, поиска полезных ископаемых, управления транспортом, медицинской диагностики и др. АСОИ реализуются часто в виде так называемых информационно-измерительных систем (ИИС), функционирующих в реальном масштабе времени. Другая форма реализации АСОИ предполагает ее программную (иногда машинно-независимую) реализацию в виде пакета прикладных программ (ППП), имеющего определенный входной язык представления заданий и данных и реализуемого в рамках определенной операционной системы не обязательно в реальном времени. В ИИС сигналы от датчиков после предварительной обработки в приемных устройствах вводятся для последующей математической (статистической) обработки в ЭВМ или в специализированные вычислители (спецпроцессоры). Информация в ЭВМ (спецпроцессоры) вводится автоматически путем использования соответствующих интерфейсов.

Эффективность статистической обработки сигналов определяется многими факторами (параметрами датчиков, отношением сигнал/помеха на выходе приемника, разрядностью АЦП и т. п.), однако важнейшую роль при этом играет качество специального алгоритмического обеспечения (САО) задач обработки. Для качественной характеристики задач статистической обработки информации (сигналов, данных) рассмотрим обобщенную структурную схему и классификацию ИИС, уточним базовые понятия и дадим классификацию САО ИИС.

Структурная схема автоматизированной ИИС показана на рис. 1.

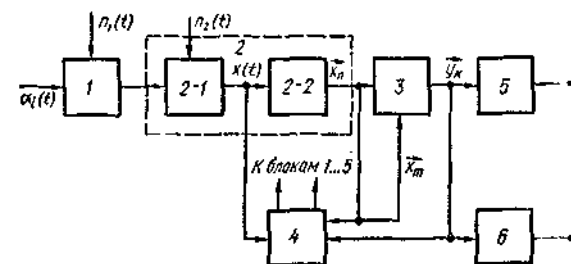


Рис. 1. Структурная схема ИИС:

1 — датчики; 2 — приемник (2-1 — каналы аналоговой обработки; 2-2 — дискретизатор процесса $x(t)$); 3 — ЭВМ или спецпроцессор; 4 — блок управления параметрами и алгоритмами блоков 1...5 в процессе адаптации к текущей сигнально-помеховой обстановке; 5 — устройства регистрации, отображения или дополнительной обработки вектора оценок информационных параметров; 6 — исполнительное устройство; \vec{x}_n — сигнальная (рабочая), \vec{x}_{nm} — обучающая выборка (n, m — объемы выборок)

Блок 1 (датчиков) выполняет функции в общем случае нелинейного инерционного преобразования входных сообщений $\alpha_i(t)$ или сигналов $f_c(\alpha_i, \beta, t)$, несущих сообщения, в выходные сигналы датчиков $S_1(\alpha_i, \beta, t)$, где β — вектор неинформационных параметров сигнала. Природа входных воздействий при этом может быть любой (химической, механической, акустической или электромагнитной). Часть физических параметров входных сигналов игнорируется датчиками, сам датчик вносит линейные, а иногда и нелинейные искажения в формируемый сигнал.

Воздействие $\alpha_i(t)$ от i -го источника, $i = \overline{1, M}$, может фиксироваться одновременно конечным числом различных по параметрам датчиков, поэтому процессы $S_1(\cdot)$ могут быть векторными.

Приемники 2 реализуют функции предварительной аналоговой обработки смеси сигналов и помех $\xi(t) = S_1(\alpha, \beta, t) + n_1(t)$, обеспечивая усиление слабых сигналов на фоне собственных шумов; полосовую, низкочастотную и режекторную фильтрацию как внешних $n_1(t)$, так и внутренних $n_2(t)$ помех; нормировку выходных процессов $x(t)$ по интенсивности (с помощью различных схем автоматической регулировки усиления и нелинейного сжатия динамического диапазона), а иногда и нормировку по средней частоте спектра (например, используя системы автоматической подстройки частоты); аналого-цифровое преобразование процесса $x(t)$.

Будем рассматривать операции по обработке сигналов в спецпроцессоре 3 над дискретной выборкой $x_n = \{x_1, \dots, x_n\}$ объема n из генеральной совокупности Γ , где $x_i = x(t_i)$ — отсчеты, сформированные из, в общем случае, неаддитивной смеси сигналов и помех $x(t) = F[S(\alpha, \beta, t), n_n(t)] = F(S, n_n)^*$. (Символ «п» в обозначении реализации помехи $n_n(t)$ или n_n будем опускать, если нет опасности перепутывания с обозначением объема выборки n .)

Специфика датчиков 1 и приемников 2 учитывается в n -мерной плотности распределения вероятностей (ПРВ) $\omega(\vec{x}_n/\Theta_n, \Theta)$ (обозначение параметра ПРВ \vec{x}_n и случайной величины \vec{x}_n одним символом, удобно, но требует ясного понимания их смысла: в ПРВ \vec{x}_n — неслучайный аргумент функции; выборка \vec{x}_n — вектор случайных величин), где Θ_n и Θ — соответственно p - и q -мерные векторы параметров сигнала и помех, совокупность которых (Θ_n, Θ) в ряде случаев будем обозначать одним вектором Θ , вводя по необходимости нижние индексы. Информационный параметр Θ_n является функцией (функционалом) параметров Θ_α ПРВ сообщения α , параметров Θ_β ПРВ β и вида сигнала $S(\alpha, \beta, t)$, вида в общем случае нелинейного инерционного (с «памятью») оператора взаимодействия сигналов и помех $F(S, n_n)$.

Обычно стремятся выбором режима работы приемника 2 упростить вид функции $F(S, n_n)$, избегая нелинейных эффектов взаимодействия сигналов и помех и добиваясь их аддитивного взаимодействия: $x(t) = S(\alpha, \beta, t) + n_n(t)$ или в частном случае $x(t) = \alpha(t) + n_n(t)$. В некоторых простейших случаях параметр ПРВ Θ_n совпадает с моментами ПРВ процесса $\alpha(t)$ или тождественен отсчетам самого процесса $\alpha(t_i)$, $i = \overline{1, p}$.

В ИИС, где применяются узкополосные акустические или радиосигналы и для обработки используются их огибающая или фаза, функция $F(S, n_n)$ нелинейная. Для огибающей $F(\cdot)$ определяется, например, преобразованием Гильберта, для фазы — нелинейным безынерционным преобразованием квадратур процесса $x(t)$.

Типовыми для спецпроцессоров 3 являются следующие операции статистической обработки выборок \vec{x}_n :

проверки статистической гипотезы о значении параметра Θ_n плотности распределения $\omega(\vec{x}_n/\Theta_n, \Theta)$, если известно, что Θ_n принимает одно из двух допустимых значений $\Theta_n = 0, \Theta_n = 1$ ($p = 1$, индекс «1» в обозначении $\Theta_n = 1$ опущен);

измерения («оценки») параметров $\Theta_n = \Theta_{np}, \Theta = \Theta_q$, если за время формирования наблюдаемой выборки n эти параметры остаются неизменными (обычно $p + q \ll n$);

фильтрации параметров Θ_{np} или Θ_q , если некоторые (или все) $p + q$ параметров изменяют свое значение за время формирования выборки n (в частных случаях говорят об операциях «сглаживания», интерполяции, прогнозирования или экстраполяции поведения параметров Θ_n, Θ в общепринятом смысле этих понятий);

некоторые другие операции, обобщающие вышеуказанные, в частности, составные операции совместной проверки гипотез и оценивания и т. п.

Операцию проверки гипотез часто называют операцией «обнаружения сигналов», так как значению параметра $\Theta_n = 1$ соответствует наличие на входе датчика информационного сигнала $f_c(\alpha \neq 0, \beta, t)$, а значению $\Theta_n = 0 - f_0(\alpha = 0, \beta, t) = 0$ и выборка \vec{x}_n формируется из реализации помехи. Если $f_0(\alpha = 0, \beta, t) \neq 0$, то чаще говорят о «различении» или «распознавании» сигналов.

Типовым операциям математической обработки сигналов соответствует специфичная структура «пространства статистических решений» y_k на выходе спецпроцессора 3 (рис. 1). Следуя общепринятым в математической статистике обозначениям,

будем полагать $\vec{y}_k = \vec{\hat{\Theta}}_k$, где компоненты вектора $\vec{\hat{\Theta}}_k = \{\hat{\Theta}_1, \dots, \dots, \hat{\Theta}_k\}$ называют «оценками» неизвестных параметров ПРВ $\omega(\vec{x}_n/\Theta_n, \Theta)$. Для задач проверки гипотез $k = 1$, опуская индекс «1», будем полагать $y = \hat{\Theta}_n$. Здесь он принимает одно из двух возможных значений: $\hat{\Theta}_n = 1$ («решение»: сигнал присутствует в выборке \vec{x}_n) или $\hat{\Theta}_n = 0$ (сигнал отсутствует). При измерении p параметров Θ_{np} размерность вектора $\vec{y}_k = \vec{\hat{\Theta}}_k$ равна $k = p$ и т. д.

Четко разделить функции блоков 2 и 3 на рис. 1 трудно, так как они в совокупности решают общую задачу выделения сообщений. Условно можно считать, что обработка в блоке 2 выполняется на интервалах времени $T_0 \approx (\Delta f)^{-1}$, где Δf — ширина энергетического спектра сигнала. В блоке 3 обработка выполняется на временных интервалах $T = nT_0$, где n — целое число (объем выборки). Оба блока аппаратурно (программно) реализуют единые линейные или нелинейные операторы обработки случайных процессов (последовательностей),

декомпозиция которых на два блока обуславливается только техническими (экономическими) ограничениями.

Блок 4 реализует функции управления процессом обработки сигналов в приемниках и процессоре. В некоторых разновидностях ИИС, называемых адаптивными, блоку 4 поручаются и задачи математической обработки случайных процессов, в частности, задачи измерения текущих статистических характеристик помех по «обучающей» выборке \vec{x}_m объемом m .

Оценки этих характеристик, например, вектора $\vec{\Theta}$ используются затем для оптимизации процессов обработки полезных сигналов в блоках 2, 3.

Для классификации ИИС по особенностям процессов обработки сигналов можно использовать следующие информационные признаки.

1. Тип и количество сообщений $\alpha_i(t)$, $i = \overline{1, M}$; α — непрерывная или дискретная функция времени; при $M > 1$ принято называть ИИС многоканальными. Число M в ИИС может составлять величину до $10^5 \dots 10^6$.
2. Тип переносчика сообщения $\alpha(t)$, определяющего структуру информационного сигнала $S(\alpha, \beta, t)$; различают переносчики гармонические (когерентные и некогерентные), импульсные (радио- или видеопульсная последовательности), шумоподобные (их называют псевдослучайными когерентными последовательностями). В частности, переносчик может отсутствовать, и сообщение $\alpha(t)$ тогда поступает непосредственно на датчик 1 (рис. 1).
3. Тип модуляции переносчика информации $S(\alpha = 0, \beta, t)$ сообщениями $\alpha(t)$.
4. Диапазон частот переносчика сигналов (или просто сообщения, если переносчик отсутствует): низкочастотные, радиочастотные, оптические, рентгеновские и другие ИИС.
5. Тип датчика.
6. Структурные особенности обрабатывающих блоков 2, 3, зависящие от типа решаемой задачи (обнаружение сообщений, измерения, фильтрация сообщений и др.).
7. Особенности технической реализации ИИС (программная, аналоговая, цифровая и т. п.). В ИИС применяют и пакеты прикладных программ (ППП) обработки сигналов, в частности, библиотечные программы по статистической обработке данных.

Для классификации алгоритмов обработки сигналов в ИИС существуют следующие признаки: тип задачи обработки (признак 6

классификации ИИС); особенности одномерной ПРВ, а с учетом статистической связи x_i и x_j ($i \neq j$) и многомерной ПРВ $\omega(\vec{x}_{n+m} / \Theta_n, \Theta)$, определяемой признаками 1—5 классификации ИИС; количество неизвестных, «мешающих» компонентов (параметров) помеховой обстановки в составе вектора Θ , а также диапазон их варьирования, наличие (отсутствие) статистических сведений о ПРВ $\omega(\vec{x}_{n+m} / \Theta_n, \Theta)$ (говорят об априорной неопределенности в задании модели сигнально-помеховой обстановки); отношение b сигнал/помеха по мощности на входе спецпроцессора и рабочий диапазон $b_{\min} \leq b \leq b_{\max}$.

Возможности технической реализации спецпроцессора ИИС (признак 7 классификации ИИС) выступают в ряде случаев в качестве жестких ограничений при выборе численных методов для САО. Векторную запись $\vec{y}_k = f(\vec{x}_{n+m})$ алгоритмов обработки сигналов будем понимать в обычном смысле;

$$\begin{cases} y_1 = f_1(x_1, \dots, x_n, x_{n+1}, \dots, x_{n+m}); \\ y_k = f_k(x_1, \dots, x_n, x_{n+1}, \dots, x_{n+m}); \end{cases} \quad (1)$$

где, согласно рис. 1, $\vec{y}_k = \vec{\Theta}_k$ — выделенное сообщение («статистика» $\vec{\Theta}_k$), как отмечалось выше, для задач обнаружения или измерения (оценивания) одного параметра (Θ_n или Θ — скаляры), $k = 1$, для задач фильтрации $k = n$. Другие случаи $k \neq 1$ и $k \neq n$ оговариваются в каждом конкретном случае обработки сигналов. При $k = 1$ будем полагать $y_1 = y$. Аналогично при $n = 1$ — $x_1 = x$.

Для \vec{y}_k используются также понятия «теста» и другие, ряд из которых имеет устоявшиеся обозначения и наименования (тесты Стьюдента, Хоттеллинга, критерии χ^2 , r -статистики Диксона и т. п.).

В качестве одной из основных моделей сигнально-помеховой обстановки рассмотрим ПРВ вектора \vec{x}_n , сформированного путем случайного выбора из составной генеральной совокупности $\Gamma = \Gamma_1 \cup \Gamma_2$ с одномерными плотностями $\omega(x/\Theta)$, $\omega(x/\Theta_1)$, $\omega(x/\Theta_2)$ соответственно.

Процедура отбора отсчетов x_i определяется стохастическим правилом:

$$x_i = (1 - \xi_i) x_{1i} + \xi_i x_{2i}, \quad i = \overline{1, n}, \quad (2)$$

где ξ_i — случайная величина, принимающая значения 0 или 1 с вероятностями $1 - p$ и p соответственно; x_{1i} , x_{2i} — случайные величины с ПРВ $\omega(x/\Theta_1)$, $\omega(x/\Theta_2)$.

Таким образом, при $p \neq 0$ к отсчетам x_{1i} из основной генеральной совокупности Γ_1 подмешиваются «загрязняющие» выборку отсчеты x_{2i} из инородной генеральной совокупности Γ_2 , которые назовем аномальными. На практике аномальными являются выбросы (или обнуления), обусловленные действием импульсных помех индустриального или атмосферного происхождения, сбоем синхронизации в коммутаторах приемников, кратковременно действующие помехи, создаваемые излучениями радиоэлектронных систем, в том числе преднамеренно.

Одномерную ПРВ отсчетов x_i , соответствующую процедуре (2), запишем в виде

$$\omega(x/\Theta) = (1 - p)\omega_1(x/\Theta_1) + p\omega_2(x/\Theta_2)$$

или, вводя информационный параметр Θ_n ,

$$\omega(x/\Theta_n, \Theta) = (1 - p)\omega_1(x/\Theta_n, \Theta_1) + p\omega_2(x/\Theta_n, \Theta_2), \quad (3)$$

где смысл векторов параметров $\Theta_n, \Theta_1, \Theta_2$ будет уточняться в конкретных задачах; $0 < p < 1$ (в литературе часто рассматриваются только случаи малых p). В качестве $\omega_i(x/\Theta_i)$ в примерах рассмотрим не только гауссовы ПРВ (в сокращенной записи $N(m, \sigma^2)$), но и другие, в частности, типичные для огибающих $x(t)$ узкополосных процессов релеевские (райсовские) ПРВ. Вид ПРВ $\omega_1(x/\Theta_n, \Theta)$ всегда будем считать известным, p — неизвестным, относительно $\omega_2(x/\Theta_n, \Theta_2)$ будут рассматриваться различные предположения. Основная задача — синтез алгоритмов, обладающих свойствами эффективности (высокого качества выделения сообщений) и устойчивости, т. е. независимости эффективности обработки от параметров $\Theta = \{p, \Theta_2\}$.

1.7.2. Показатели эффективности алгоритмов обработки сигналов

Постановка задачи и основные функционалы эффективности. Формируемые на выходе блока 3 (рис. 1) измеренные значения («оценки») $\vec{y}_k = \vec{\hat{\Theta}}_k$ параметров сигнальной (или помеховой) обстановки отличаются от истинных значений $\vec{\Theta}_k$ из-за искажающего воздействия помех на входе ИИС, наличия ошибок в канале обработки (блоки 1—3), в частности, из-за переходных процессов, обусловленных блоком 4. Информация об эффективности САО ИИС заключена в условной ПРВ $\omega(\vec{\hat{\Theta}}_k/U)$, или $\omega(\vec{\epsilon}_k/U)$, где $\vec{\epsilon}_k = \vec{\hat{\Theta}}_k - \vec{\Theta}_k$ — вектор

ошибок; U — комплекс условий. Используя в дальнейшем скалярные обозначения для всех встречающихся векторов (например, ϵ вместо $\vec{\epsilon}_k$; α вместо $\alpha(t)$) и полагая, что непрерывную реализацию можно представить вектором в результате разложения типа Карунена — Лоева, Котельникова и других, рассмотрим два предельных случая при формализации задачи отыскания $\omega(\epsilon/U)$. Считаем фиксированной сигнально-помеховую обстановку и оператор обработки в блоках 2—4 ИИС. В частности, полагаем известными операторы обработки блоков 2, 3, а при учете работы блока 4 — оператор $F(S, n, \vec{x}_n)$ (см. рис. 1).

Случай 1. Полностью известна (в результате расчетов, эксперимента) ПРВ $\omega(\vec{x}_{n+m}/\Theta_n, \Theta)$. Тогда комплекс условий определим вектором $U = \{\Theta_n, \Theta, b, f, n+m\}$. При таком подходе сложная задача анализа декомпозируется на две самостоятельные и более простые задачи, что часто используется на практике. В дальнейших рассуждениях такой подход будет принят в качестве основного.

Случай 2 (анализ «в целом»). Даны ПРВ α, β, n_n и операторы обработки сигналов в ИИС (см. рис. 1). Тогда комплекс условий U принимает вид $U = \{\Theta_\alpha, \Theta_\beta, \Theta_{n_n}; F(\cdot); \text{оператор АЦП}; b, f, n+m\}$. Искомую ПРВ ошибки обработки $\omega(\epsilon/U)$ можно записать аналитически через кратные интегралы, используя формулу полной вероятности

$$\omega(\epsilon/U) = \int_{(\alpha)} \omega(\epsilon/\alpha) \omega(\alpha/\Theta_\alpha) d\alpha, \quad (4)$$

где

$$\omega(\epsilon/\alpha) = \int_{(\beta)} \omega(\epsilon/\alpha, \beta) \omega(\beta/\Theta_\beta) d\beta; \quad (5)$$

$$\omega(\epsilon/\alpha, \beta) = \int_{(n_n)} \omega(\epsilon/\alpha, \beta, n_n) \omega(n_n/\Theta_{n_n}) dn_n, \quad (6)$$

где условная ПРВ $\omega(\epsilon/\alpha, \beta, n_n)$ определяется видом взаимодействия сигналов и помех $F(\bullet)$ и алгоритмом процессора $f(\bullet)$; пределы (\bullet) кратных интегралов определяются диапазоном варьирования векторных переменных α, β, n_n ; дифференциалы $d\alpha = \prod_{k=1}^c d\alpha_k$;

$d\beta = \prod_{k=1}^d d\beta_k$; $dn_n = \prod_{k=1}^l dn_{nk}$; c, d, l -мерности соответствующих векторов, совпадающие с мерностью их ПРВ. Интегралы (4) — (6) редко удается вычислить аналитически, поэтому для отыскания ПРВ ошибки (4) используют метод полунатурного моделирования. Однако

корректные результаты при моделировании можно получить только в случае ясного понимания и точной фиксации комплекса условий U . Как показывает опыт, несоблюдение этого правила приводит к снижению эксплуатационной эффективности САО относительно ожидаемой.

Кроме двух рассмотренных предельных случаев формализации задачи анализа эффективности обработки сигналов, в ИИС применяют и различные промежуточные случаи декомпозиции сложной задачи к двум, трем и более простым. Результатом их решения, естественно,

должна являться одна и та же плотность $\omega(\mathbf{e}/U)$ или $\omega(\vec{\Theta}/U)$, а конкретное содержание вектора условий U определяется самим разработчиком в соответствии с решаемой задачей. Далее всюду $U = \{\Theta_n, \Theta, b, f, n + m\}$, что соответствует заданию ПРВ на входе спецпроцессора.

Вместо многомерной ПРВ $\omega(\mathbf{e}/U)$, работать с которой неудобно из-за отсутствия наглядности, целесообразно перейти к вектору \vec{K}_h функционалов, достаточно полно характеризующих эффективность обработки сигналов САО ИИС. В качестве таких функционалов используют:

начальные моменты порядка S ПРВ (4) вектора ошибок $\mathbf{e} = \vec{\mathbf{e}}_k = \{e_1, \dots, e_k\}$

$$m_S \{e_i/U\} = \int_{(\mathbf{e})} e_i^S \omega(\mathbf{e}/U) d\mathbf{e}, \quad (7)$$

где $d\mathbf{e} = \prod_{i=1}^k de_i$;

центральные моменты

$$\mu_S \{e_i/U\} = \int_{(\mathbf{e})} (e_i - m_1 \{e_i/U\})^S \omega(\mathbf{e}/U) d\mathbf{e}; \quad (8)$$

матрицу ковариаций ошибок

$$\begin{aligned} \text{cov} \{e_i, e_j/U\} &= \mu_{11} \{e_i, e_j/U\} = \\ &= \int_{(\mathbf{e})} (e_i - m_1 \{e_i/U\})(e_j - m_1 \{e_j/U\}) \omega(\mathbf{e}/U) d\mathbf{e}; \end{aligned} \quad (9)$$

вероятности пребывания j -й компоненты ошибки в пределах допуска $(e_j^{\text{доп min}} - e_j^{\text{доп max}})$

$$P(e_j^{\text{доп min}} \leq e_j \leq e_j^{\text{доп max}}/U) = \int_{(\mathbf{e}_1)} \dots \int_{e_j^{\text{доп min}}}^{e_j^{\text{доп max}}} \dots \int_{(\mathbf{e}_k)} \omega(\mathbf{e}/U) d\mathbf{e}; \quad (10)$$

функционалы безусловного

$$R(\vec{U}) = \int_{(\mathbf{e})} \int_{(\vec{\Theta}_n)} \Pi(\mathbf{e}) \omega(\Theta_n) \omega(\mathbf{e}/\Theta_n, \Theta, b, f, n + m) d\mathbf{e} d\vec{\Theta}_n, \quad (11)$$

где $d\vec{\Theta}_n = \prod_{i=1}^n d\Theta_{ni}$; $\vec{U} = \{\omega(\Theta_n), \Theta, b, f, n + m\}$,

и условного при $\Theta_n = \text{const}$

$$R(\Theta_n, \vec{U}) = \int_{(\mathbf{e})} \Pi(\mathbf{e}) \omega(\mathbf{e}/\Theta_n, U) d\mathbf{e}, \quad (12)$$

где $i = \overline{1, k}$; $j = \overline{1, k}$; $\Pi(\mathbf{e})$ — функция потерь, учитывающая «вес» ошибки \mathbf{e} ; $\vec{U} = \{\Theta, b, f, n + m\}$.

Функционалы типа (7) — (12) находят применение в задачах синтеза и анализа алгоритмов обработки сигналов. Отметим, что соотношения (7) — (10) можно рассматривать как частные случаи формулы (12) при соответствующем выборе $\Pi(\mathbf{e})$. Функционалы (7) — (12) содержат практически всю доступную разработчику САО информацию об эффективности конкретного алгоритма $f(\vec{x}_{n+m})$ в сигнально-помеховой обстановке, определяемой, например, ПРВ (3). На стадии синтеза решается задача поиска «оптимального» $f(\cdot)$ (поиска экстремума выбранного функционала при ограничениях на другие функционалы. На стадии анализа для выбранного допустимого набора алгоритмов $f_i(\cdot)$, $i = \overline{1, L}$ и конкретной сигнально-помеховой обстановки решается задача отыскания числовых значений функционалов типа (7) — (12) и сравнения их с требованиями технического задания на разработку САО ИИС.

Показатели устойчивости алгоритмов. Качественно понятие устойчивости алгоритма $f(\vec{x}_{n+m})$ конкретизируется количественной зависимостью вектора \vec{K}_h от компоненты Θ вектора условий U . В условиях эксплуатации САО ИИС вместо ожидаемой в соответствии, например, с формулой (3) «основной», априорной модели $\omega_1(x/\Theta_n, \Theta_1)$, в которой параметры Θ_1 обычно предполагают известными, часто реализуется другая модель $\omega_1(x/\Theta_n, \Theta)$ с другими параметрами Θ .

В этих условиях некоторый алгоритм $f(\cdot)$, возможно оптимальный при сигнально-помеховой ситуации $\omega_1(x/\Theta_n, \Theta)$, может резко ухудшить свои показатели качества или стать полностью неработоспособным. Такой алгоритм целесообразно назвать неустойчивым.

Другую причину неустойчивости алгоритмов можно связать с существенным изменением класса ПРВ наблюдаемых отсчетов \vec{x}_{n+m} . Например, как это следует из записи ПРВ (3), вместо ожидаемой плотности $\omega_1(x/\theta_n, \theta)$ в условиях эксплуатации может реализоваться совершенно другая плотность $\omega(x/\theta_n, \theta)$. Последствия такой «подмены» ПРВ для эффективности алгоритма также могут быть катастрофическими. Эти два различных случая рассмотрим с общих позиций, если в качестве основной модели принять ПРВ $\omega(x/\theta_n, \theta)$ вида (3), т. е. в параметр θ включить параметры θ_1, p, θ_2 и рассматривать далее частные безусловные

$$\tilde{K}_i = K_i(\bar{U}) = K_i(\theta), \quad i = \overline{1, h} \quad (13)$$

или частные условные

$$\bar{K}_i = K_i(\theta_n, \bar{U}) = \bar{K}_i(\theta), \quad i = \overline{1, h} \quad (14)$$

функционалы эффективности вектора \vec{K}_h как функции обобщенного вектора помех Θ .

Для анализа устойчивости алгоритма $f(\vec{x}_{n+m})$ по формулам (13), (14) достаточно задать диапазон $(\theta_{\min}, \theta_{\max})$ варьирования θ , выполнить расчеты показателей с шагом $\Delta\theta$ и сравнить $\Delta\tilde{K}_i = \tilde{K}_{i \max} - \tilde{K}_{i \min}$ (или $\Delta\bar{K}_i = \bar{K}_{i \max} - \bar{K}_{i \min}$) с допустимой по ТЗ неустойчивостью показателя в условиях эксплуатации $\Delta\tilde{K}_{i \text{ТЗ}}$ ($\Delta\bar{K}_{i \text{ТЗ}}$). При случайных вариациях θ числа \tilde{K}_i и \bar{K}_i в формулах (13), (14) также будут случайными. Если задана ПРВ $\omega(\theta)$ этих параметров, решение об устойчивости алгоритма $f(\vec{x}_{n+m})$ целесообразно выносить сравнением математических ожиданий

$$m_1\{\tilde{K}_i(\theta)\} \text{ (или } m_1\{\bar{K}_i(\theta)\}) \text{ и дисперсий } \mu_2\{\tilde{K}_i(\theta)\} = \sigma_{\tilde{K}_i}^2$$

(или $\sigma_{\bar{K}_i}^2$) с заданными по ТЗ. Возможен учет и моментов более высокого порядка, включая смешанные, или отыскание h -мерных ПРВ вектора критериев \vec{K}_h , содержащих всю доступную информацию об эффективности САО. В качестве первого приближения при поиске, например, $\Delta\tilde{K}_i(\theta)$, можно использовать дифференциальные критерии устойчивости вида

$$\delta\tilde{K}_i(\theta)|_{\theta=\theta_0} = \sum_{j=1}^q \left[\frac{\partial \tilde{K}_i(\theta)}{\partial \theta_j} \Big|_{\theta=\theta_0} \right]^2 \Delta\theta_j, \quad (16)$$

где $i = \overline{1, h}$; $\theta_0 = (\theta_{01}, p_0, \theta_{02}) = (\theta_{01}, \dots, \theta_{0q})$ — вектор средних значений параметров помех; $\delta\tilde{K}_i(\theta)$ и $\Delta\theta_j$ — приращения i -го показателя и j -й независимой переменной.

Отметим, что использовать критерий (15) надо осторожно: варьирование энергетических параметров помех меняет одновременно параметр b , а в этом случае условие $\delta\tilde{K}_i(\theta)|_{\theta=\theta_0} = 0$ свидетельствует о том, что устойчивость обработки сигналов при использовании алгоритма $f(\vec{x}_{n+m})$ достигнута снижением чувствительности САО ИИС к полезным сигналам.

Рассмотрим примеры показателей эффективности некоторых алгоритмов обработки сигналов.

Пример 1. В задачах обнаружения сигналов (проверки статистических гипотез H_1 , соответствующей $\Theta_n = 1$ или H_0 — $\Theta_n = 0$) алгоритм обработки отсчетов $\vec{x}_n = \{x_1, \dots, x_n\}$ ($m = 0$) записывается в виде $z = f(\vec{x}_n)$ и

$$y = \hat{\theta}_n = \begin{cases} 1, & \text{если } z \geq V; \\ 0, & \text{если } z < V, \end{cases} \quad (16)$$

где V — порог принятия решения.

Так как z и $y = \hat{\theta}_n$ (или $e = \hat{\theta}_n - \theta_n$) функционально связаны при $\theta_n = \text{const}$, то, используя в формуле (10) под интегралом условную ПРВ $\omega(z/U)$ вместо $\omega(e/U)$ и полагая $e_{1\min}^{\text{доп}} = -1$, $e_{1\max}^{\text{доп}} = 1$, запишем

условные вероятности $D(\theta, b) = P(\hat{\theta}_n = 1/\theta_n = 1, b, f, n) = 1 - P(e = e_{1\min}^{\text{доп}})$ правильного обнаружения и $F(\theta) = P(\hat{\theta}_n = 1/\theta_n = 0, b = 0, f, n) = P(e = e_{1\max}^{\text{доп}})$ — ложной тревоги как

$$D(\theta, b) = \int_V^\infty \omega(z/\theta_n = 1, \theta, b, f, n) dz; \quad (17)$$

$$F(\theta) = \int_V^\infty \omega(z/\theta_n = 0, \theta, b = 0, f, n) dz. \quad (18)$$

При этом $D(\theta, b = 0) = F(\theta)$ и, таким образом, определена пара условных показателей эффективности ($h = 1$) $\bar{K}_1\{\theta_n = 1, \theta\} = D(\theta, b)$; $\bar{K}_2\{\theta_n = 0, \theta\} = F(\theta)$ в соответствии с формулой (14).

При расчетах (эксперименте, моделировании на ЭВМ) обычно фиксируют $f(\cdot)$, n , требуемый по ТЗ уровень $F_{\text{ТЗ}} \leq F(\theta)$ для всех $\theta_{\min} < \theta < \theta_{\max}$, определяют V из уравнения (18) для наименее

благоприятного $\Theta_{И}$ и переходят к построению мощностной характеристики $D = f_D(b/\Theta_{И})$. При $\Theta_{И} = \text{const}$ из функции $b = D^{-1}(\Theta_{И}, D)$, обратной к мощностной характеристике, определяют пороговое отношение сигнал/помеха $b_{п} = D(\Theta_{И}, D_{ТЗ})$. Если корень уравнения $D_{ТЗ} = f_D(b/\Theta_{И})$ единственный, то условие $b > b_{п}$ обеспечит при некоторых дополнительных ограничениях на вид функции $f(\vec{x}_n)$ выполнение неравенства $D > D_{ТЗ}$. Параметр $b_{п}$ играет важную роль, разделяя всю область возможных соотношений сигнал/помеха на две подобласти: рабочую для ИИС, если $b > b_{п}$ и «отказа», «сбоя», если $b < b_{п}$. Наличие «сбоя», вызванного усложнением сигнально-помеховой обстановки, наступает также в случае, если нарушается условие $F_{ТЗ} \leq F$. Задача проектирования устойчивых САО и сводится к отысканию таких алгоритмов $f(\bullet)$, чтобы при вариациях параметров сигнально-помеховой обстановки Θ и b обеспечить минимизацию относительного времени пребывания ИИС в состоянии «отказа». Отметим также, что b является важнейшим тактическим параметром ряда ИИС, определяя, например, в радиосвязи и эхолокации максимальные дальности действия R_{max} по формулам

$$R_{\text{max}}^{\text{р.с}} = K_{\text{р.с}} b_{п}^{-0,5}, \quad R_{\text{max}}^{\text{эхол}} = K_{\text{эхол}} b_{п}^{-0,25}, \quad (19)$$

где константы зависят от внешних по отношению к САО ИИС параметров: условий распространения сигнала, параметров датчиков l и приемников 2 (см. рис. 1), мощности передатчика и др.

При неизвестном $\Theta_{И}$ воспользоваться формулой (19) нельзя, так как от $\Theta_{И}$ зависят $F, D, b_{п}$. Алгоритмы обнаружения, обеспечивающие практическую независимость (устойчивость) $F, D, b_{п}$ для ПРВ, задаваемых формулой (3), будут рассмотрены далее, где в качестве мешающих приняты параметры $\Theta = (p, \Theta_2)$ формулы (3).

Если задана ПРВ $\omega(\Theta)$, $\Theta_{\text{min}} < \Theta < \Theta_{\text{max}}$, то вычисляется математическое ожидание вероятности ложной тревоги

$$m_1 \{F(\Theta)\} = \int_{\Theta} f(\Theta) \omega(\Theta) d\Theta, \quad (20)$$

где

$$d\Theta = \prod_{i=1}^q d\Theta_i,$$

ищется допустимое значение порога решения V_{ω} из уравнения $F_{ТЗ} = m_1 \{F(\Theta)\}$; найденное значение порога подставляется в формулу (17), усреднение которой по Θ дает уравнение

$$m_1 \{D(\Theta, b)\} = D_{ТЗ}, \quad (21)$$

из которого определяют $b_{п\omega}$. Найденные этим способом $b_{п\omega}$ не

совпадают с $b_{п}$ для $\Theta_{И}$ и при подстановке в формулу (19) дадут разные результаты. Для ИИС с высокой ответственностью и избыточными ресурсами (энергетикой сигналов, n) можно рекомендовать минимаксный по отношению к F способ расчета $V, F, D, b_{п}$ и других косвенных показателей, гарантирующий достоверное выполнение условия $R \geq R_{\text{max}}$. Однако при ограниченных ресурсах ИИС более реалистичной оказывается оценка $b_{п\omega}$, полученная методом усреднения.

Иногда для фиксированных F, D, n пороговое отношение сигнал/помеха выступает в качестве меры сравнительной помехоустойчивости различных алгоритмов обнаружения.

Пример 2. В задачах измерения (оценивания) $\Theta_{и}$ при $p=1$ пользуются формулами (7), (8) для $k=1$, определяя смещенность оценки $m_1 \{e/U\} (S=1)$ и дисперсию $\mu^2 \{e/U\} = \sigma^2(U)$ ошибки оценки $e = \hat{\Theta}_{и} - \Theta_{и}$ как функции $\Theta_{и}, \Theta, b, f, n$. Фиксируя $\Theta_{и}, \Theta_{п}, l, n$, определяют два значения пороговых сигнала $b_{п1} = m_1^{-1} \{\Theta_{и}, f, n, m_{1ТЗ}\}$ и $b_{п2} = m_2^{-1} \{\Theta_{и}, f, n, m_{2ТЗ}\}$, которые в общем случае различны. Для расчета тактических параметров ИИС по формуле (19) необходимо полагать

$$b_{п} = \max_i \{b_{пi}, i = 1, 2\}. \quad (22)$$

Анализ устойчивости алгоритма $f(\vec{r}_n)$ оценивания $\hat{\Theta}_{и}$ при случайных вариациях параметров помех в выполняется по методике примера 1.

Методы получения числовых значений показателей: аналитический; экспериментальный; имитационного моделирования на ЭВМ, называемый также вычислительным экспериментом, статистических испытаний, Монте-Карло.

Аналитический метод расчета \vec{K}_h эффективен в относительно редких случаях: гауссовы сигналы и помехи, линейные операции обработки; процессы, происходящие из гауссовых при специальных нелинейных преобразованиях; несложные нелинейные функциональные преобразования. Известны отдельные результаты в математической статистике типа отыскания ПРВ Стьюдента, Уишарта и др. Задача анализа облегчается при больших выборках, т. е. при $n \gg 1, m \gg 1$, когда возможно использование асимптотических результатов, например, центральной предельной теоремы. Метод требует определенной математической культуры разработчика алгоритмического обеспечения ИИС, однако даже и в этом случае не всегда приводит к решению задачи.

Экспериментальный метод предполагает макетирование спецпроцессора или эмуляцию САО ИИС, использование имитаторов

сигналов и помех, аппаратуры (или программы для ЭВМ) сбора и обработки результатов N -кратного (10^2 — 10^4) «прогона» реализаций смеси сигналов и помех через макет ИИС. В случае натуральных испытаний постановка эксперимента осложняется организационными трудностями.

Имитационный метод моделирования на ЭВМ реализуется следующим образом. Имитаторы (реальные источники) сигналов и помех, макеты блоков 1 — 4 ИИС (см. рис. 1) и устройства сбора и обработки реализации y_k заменяют соответствующими подпрограммами генерации случайных процессов \vec{x}_{n+m} , обработки $f(\vec{x}_{n+m})$, вычисления функционалов эффективности \vec{K} путем N -кратного «прогона» реализаций \vec{x}_{n+m} смеси сигналов и помех. В частности, ведется оценка \hat{F} , \hat{D} , $\hat{m}_1\{e/U\}$ и т. п. при переборе параметров сигнально-помеховой обстановки $\Theta_\alpha, \Theta_\beta, \Theta_{np}, b, n, m$ (или $\Theta_n, \Theta_1, \Theta_{2..p}$ в формуле (3)) и алгоритмов обработки $f_i(\cdot), i = \overline{1, L}$. Экономически этот метод выгоден, оперативен, точен, однако требует высокой статистической, прикладной математической и инженерной квалификации постановщика задачи имитационного моделирования, так как результаты моделирования на неадекватных моделях не имеют практической значимости.

Применяют смешанные (комбинированные) методы оценки эффективности ИИС, в частности, **аналитико-имитационный**. При этом методе вводится замена операторов части обрабатывающих блоков ИИС (например, блоков 1, 2 на рис. 1) их статистическими эквивалентами (датчиками чисел с ПРВ $\omega(x/\Theta_n, \Theta)$ типа (3)), полученными путем аналитических расчетов. Применяют другие комбинации рассмотренных основных методов.

1.7.3. Классификация методов синтеза алгоритмов статистической обработки сигналов

При классификации методов синтеза алгоритмов обработки сигналов используют такой существенный признак как объем априорной информации: о статистических характеристиках сообщений, сигналов, помех, взаимодействии сигналов и помех в датчике и приемнике. Выделим четыре основные градации этого признака и рассмотрим типовые задачи.

Задача 1. В теории статистических решений А. Вальда предполагается полностью, т. е. до числовых значений параметров Θ , известной условная ПРВ выборки $\omega(\vec{x}_n/\Theta_n, \Theta)$, ПРВ информационного параметра $\omega(\Theta_n)$, функция потерь в формуле (11) и для фиксированных Θ, b, n ищется неизвестная решающая функция $\hat{\Theta} = f_{\text{опт}}(\vec{x}_n)$, при которой достигается экстремум (обычно минимум) безусловного риска $R(\vec{U})$ в формуле (11). Искомая функция (оператор) $f(\cdot)$ входит в качестве переменной в комплекс условий $\vec{U} = \{\omega(\Theta_n), \Theta, b, n\}$. Оптимальная решающая функция $f_{\text{опт}}(\cdot)$ называется байесовской, а соответствующее ей минимальное значение риска

$$R(\vec{U})|_{f=f_{\text{опт}}}$$

— байесовским риском.

Задача 2. При неизвестной ПРВ информационных параметров $\omega(\Theta_n)$ ищется наименее благоприятное распределение $\omega_n(\Theta_n)$ и соответствующая ему специальная байесовская минимаксная решающая функция $f_{\text{мм}}$, при которой минимум байесовского риска достигает своего максимума при наиболее неблагоприятном ПРВ.

Обе задачи являются вариационными. Постановке и решению их для различных частных случаев посвящена обширная литература по синтезу алгоритмов обнаружения, измерения и фильтрации сигналов. Получены существенные результаты в задачах обнаружения, измерения, теории нелинейной фильтрации марковских процессов. В частности, для произвольных непрерывных ПРВ $\omega(\vec{x}_n/\Theta_n, \Theta)$ и $\Theta = \text{const}$ Нейманом и Пирсоном доказано, что оптимальный алгоритм обнаружения определяется статистикой $l(\vec{x}_n)$ отношения функций правдоподобия $L(\Theta_n = 1) = \omega(\vec{x}_n/\Theta_n = 1)$ и $L(\Theta_n = 0) = \omega(\vec{x}_n/\Theta_n = 0)$

$$l(\vec{x}_n) = L(\Theta_n = 1)L^{-1}(\Theta_n = 0)$$

или статистикой

$$z = f(\vec{x}_n) = \varphi[l(\vec{x}_n)], \quad (23)$$

где $\varphi(\cdot)$ — произвольная монотонно возрастающая функция, с последующим сравнением $l(\vec{x}_n)$ или z с порогом решения в соответствии с правилом (16).

Доказана универсальность статистики $l(\vec{x}_n)$: для различных функций потерь; при неизвестной функции потерь; в случае, когда пользуются формулой Байеса для расчета апостериорной ПРВ

$\omega(\Theta_u/\vec{x}_n) = K\omega(\Theta_u)\omega(\vec{x}_n/\Theta_u) = K\omega(\Theta_u)L(\Theta_u)$, $K=\text{const}$, а также при неизвестной ПРВ $\omega(\Theta_u)$, когда для заданной функции потерь ищут наиболее благоприятное распределение $\omega_u(\Theta_u)$.

Задачи 1 и 2 решаются методами математического программирования в простых случаях аналитически и, как правило, численно. Отметим, что такая постановка задач обработки сигналов не учитывает затрат ресурсов на получение всей необходимой информации о сигнально-помеховой обстановке, а иногда все вариации сигнально-помеховой обстановки не могут быть ни изучены, ни предсказаны в принципе. Алгоритмы, синтезированные исходя из допущений априорной (допытной) вероятностной известности и стабильности сигнально-помеховой обстановки, оказываются совершенно неработоспособными (неустойчивыми) при отклонениях реальных характеристик сигналов и помех от модельных. Поэтому рассмотрим другие постановки задач.

Задача 3. Отличается от задачи 2 отсутствием априорных сведений о параметрах в ПРВ $\omega(\vec{x}_{n+m}/\Theta)$. Ищется алгоритм $y = f(\vec{x}_{n+m})$ обработки вектора x_{n+m} , который обеспечит удовлетворительное качество функционирования ИИС при вариациях Θ . В частности, неопределенность знания обусловлена неполностью известной ПРВ $\omega(\vec{n}_n/\Theta_n)$.

Постановка задачи 3 значительно ближе к реальности эксплуатации ИИС. Решается задача поиска алгоритма аналитическими или численными методами, образующими в совокупности группу так называемых параметрических методов прикладной статистики. В той или иной степени все они основаны на идее сведения (редукции) сложной задачи 3 к более простым задачам 1 или 2. При конечных объемах выборки n, m применяют методы редукции в смысле А. Вальда, подобия и инвариантности, прямой оценки $\hat{\Theta} = \Theta(\vec{x}_m)$ или $\hat{\Theta} = \Theta(\vec{x}_n)$, где в первом случае процесс формирования оценки называют обучением с учителем, а во втором — самообучением.

Метод редукции А. Вальда заключается в усреднении $\omega(\vec{x}_n/\Theta_u, \Theta)$ по ПРВ $\omega(\Theta/\Theta_0)$, полученной экспериментально или выбранной из класса наиболее благоприятных:

$$\omega(\vec{x}_n/\Theta_u, \Theta_0) = \int_{(\Theta)} \omega(\Theta/\Theta_0) \omega(\vec{x}_n/\Theta_u, \Theta) d\Theta, \quad (24)$$

где Θ_0 — известный теперь вектор постоянных параметров ПРВ $\omega(\Theta/\Theta_0)$, компоненты которого имеют смысл, например, для

гауссовых $\omega(\Theta/\Theta_0)$, математического ожидания и дисперсии дрейфующих параметров помех $\Theta(t)$.

Методы подобия и инвариантности, использованные в работах для решения задач проверки гипотез, когда $\omega(\vec{x}_{n+m}/\Theta_u, \Theta)$ принадлежат экспоненциальному семейству ПРВ, позволили получить строгие результаты в области синтеза алгоритмов, обеспечивающих постоянство $F(\Theta)$ при вариациях Θ , выполнение условия несмещенности $D(\Theta, b=0) \geq F(\Theta)$, высокую эффективность, оцениваемую пороговым соотношением b_n или мощностью $D(b)$ алгоритма обнаружения. Принцип подобия использует понятие достаточных статистик для мешающих параметров Θ и позволяет реализовать процедуру самообучения, извлекая информацию о Θ по единственной выборке $\vec{x}_n (m=0)$. Принцип инвариантности основан на отыскании так называемых максимальных инвариантов, т. е. специального вида функций $f(\vec{x}_{n+m})$, ПРВ которых не зависит от Θ при нулевой гипотезе.

Экспоненциальное семейство включает в себя гауссовы ПРВ, но само является лишь частным случаем используемых в практике моделей сигнально-помеховых ситуаций. ПРВ вида (3), рассматриваемые в задачах устойчивости алгоритмов обработки сигналов, не относятся к типу экспоненциальных.

Метод прямой оценки в режимах обучения или самообучения иногда называют эмпирическим байесовым (ЭВМ) или адаптивным. Строгого обоснования метод не имеет. Алгоритмы обработки сигналов при конечных $n+m$, найденные с его помощью, не являются оптимальными, а учитывая свободу выбора численных методов формирования оценок $\hat{\Theta}$, их следует отнести к классу эвристических.

При использовании для формирования $\hat{\Theta} = \Theta(\vec{x}_n)$ метода максимального правдоподобия (МП) для гауссовых ПРВ $\omega(\vec{x}_n/\Theta_u, \Theta)$ синтезированы тесты Стьюдента, Хотеллинга и другие, оптимальность которых доказывается и методами подобия. В общем случае поиск алгоритма оценивания ведется перебором известных методов: МП, моментов (ММ), квантилей (МК), λ^2 -квадрат и их комбинаций. При этом реализуют, как правило, режим обучения, т. е. формируют $\hat{\Theta} = \Theta(\vec{x}_m)$. Сложность решаемых здесь задач растет с размерностью вектора $\Theta = (\Theta_1, \dots, \Theta_q)$, а качество алгоритмов обработки снижается. Введение одного неизвестного параметра увеличивает порядок системы дифференциальных уравнений не-

линейной фильтрации, заданных в форме Коши, со второго до пятого, а двух — до девятого.

Задача 4. Априорная неопределенность относительно $\omega(\vec{x}_n/\Theta_u, \Theta)$ такова, что вид ПРВ задан с точностью до указания класса функций $\omega(\cdot) \in \Omega$. Класс функций определяется качественно, например, одномерные ПРВ $\omega(x/\Theta_u, \Theta)$ симметричны; ПРВ стохастически упорядочены по интегральной $\omega(x/\Theta_u, \Theta)$ функции распределения вероятностей (ИФРВ), т. е.

$$\int_{-\infty}^x \omega(x/\Theta'_u, \Theta) dx > \int_{-\infty}^x \omega(x/\Theta''_u, \Theta) dx$$

при $\Theta'_u < \Theta''_u$ для всех Θ и всех x из интервала $-\infty < x < \infty$.

Методы и алгоритмы обработки сигналов, применяемых в этом случае, называются непараметрическими. Разработаны непараметрические методы проверки статистических гипотез, позволяющие получать достаточно эффективные алгоритмы обнаружения сигналов на фоне помех с неизвестной ПРВ. Классификацию непараметрических методов можно выполнить на основе различных наборов признаков. По частным задачам проверки непараметрических гипотез $H_0(\vec{x}_n \in \Gamma_0)$, определяемой классом ПРВ $\omega(\vec{x}_n/\Theta_u = 0, \Theta)$, против гипотезы $H_1(\vec{x}_n \in \Gamma_1)$, определяемой $\omega(\vec{x}_n/\Theta_u = 1, \Theta)$, их удобно классифицировать на методы для проверки нулевых гипотез о симметрии ПРВ $\omega(\vec{x}_n/\Theta_u = 0, \Theta)$ по рабочей выборке \vec{x}_n (обучение не используется); методы проверки нулевой гипотезы о принадлежности обучающей \vec{x}_m и рабочей \vec{x}_n выборок к одной и той же генеральной совокупности с ПРВ $\omega(\vec{x}_{n+m}/\Theta_u, \Theta)$.

Первая из задач называется задачей о симметрии ПРВ $\omega(\vec{x}_n/\Theta_u = 0, \Theta)$, вторая — задачей об однородности двух выборок (в общем случае k выборок).

Задача о симметрии ПРВ $\omega(x/\Theta_u = 0, \Theta)$ для независимых выборок и $m_1 \{x\} = 0$ приводит к известному алгоритму знаков

$$y = f(\vec{x}_n) = \left| \sum_{i=1}^n \text{sign } x_i \right| > V, \quad (25)$$

где

$$\text{sign } x = \begin{cases} 1, & x \geq 0; \\ -1, & x < 0. \end{cases}$$

Эффективность алгоритма зависит от конкретного вида ПРВ $\omega(\vec{x}_n/\Theta_u, \Theta)$ и в ряде случаев низка, если вносимая сигналом асимметрия мала.

Задача однородности двух выборок \vec{x}_m и \vec{x}_n , т. е. задача о совпадении из ПРВ или ИФРВ $F_m(x)$ и $F_n(x)$ решается группой тестов (и соответственно алгоритмов) типа Смирнова — Колмогорова, двухвыборочного Мизеса, Вилкоксона и ряда других, которые являются частными случаями ранговых критериев. Понятие «ранга» R_i выборочного отсчета x_j из выборки $\vec{x}_n = \{x_1, \dots, x_n\}$ определяется как порядковый номер x_j в общем вариационном ряду $\{x_{(1)}, \dots, x_{(i)} = x_j, x_{(i+1)}, \dots, x_{(n+m)}\} = \vec{x}_{(n+m)}$ объединенной выборки \vec{x}_{n+m} объемом $n + m$. Заметим, что статистику $x_{(k)}$ называют k -й порядковой статистикой, в частности, статистики $x_{(1)}$, $x_{(n+m)}$ называют крайними, экстремальными.

Алгоритмы обработки сигналов, использующие функции $f(\vec{x}_{(n+m)})$, $f(\vec{R}_{(n+m)})$ порядковых и ранговых статистик, обладают устойчивостью при вариациях ПРВ в широких пределах, что привлекает к ним особое внимание разработчиков САО ИИС. Функции $f(\vec{x}_{(n+m)})$ и $f(\vec{R}_{(n+m)})$ для краткости будем называть L - и R -статистиками соответственно. Рассмотрим, например, некоторые свойства R -статистик.

Вектор рангов $\vec{R}_n = \{R_1, \dots, R_n\}$ не изменяется при любых монотонных преобразованиях отсчетов выборки \vec{x}_{n+m} , в частности, при умножении на масштабный коэффициент, и является в этом смысле максимальным инвариантом. Для независимых x_i , x_i , $i = \overline{1, n+m}$; $j = \overline{1, n+m}$, названные алгоритмы проверки гипотез об однородности двух выборок являются частными случаями функций вида

$$y = f(\vec{R}_n, \vec{R}_m) \geq V. \quad (6)$$

Так, алгоритм обнаружения сигналов, реализующий двух выборочный тест Мизеса, дается известным в статистике выражением

$$y = n \sum_{i=1}^n (R_i - i)^2 + m \sum_{j=1}^m (R'_j - j)^2 \geq V, \quad (7)$$

где m — объем обучающей выборки; n — объем рабочей выборки; R_i , R'_j — ранги отсчетов рабочей и обучающей выборок в общем вариационном ряду.

Развитая теория ранговых критериев используется в задачах обработки сигналов, однако, гарантируя устойчивость $F = \text{const}$, ранговые алгоритмы при небольших $n + m$ не всегда обеспечивают требуемую мощность D .

Другой метод решения задачи однородности двух выборок использует понятие квантиля X_p неизвестного распределения $F(x)$, определяемого как решение уравнения $F(X_p) = p$. Для выборки \vec{x}_k оценкой \hat{X}_{pk} квантиля является скаляр, ниже которого лежит $p \cdot 100\%$ членов вариационного ряда $\vec{x}_{(k)} = \{x_{(1)}, \dots, x_{(k)}\}$. По вариационным рядам выборок \vec{x}_m и \vec{x}_n находят оценки квантилей \hat{X}_{pm} и \hat{X}_{pn} и строят L -статистику вида $y = f(\hat{X}_{pm}, \hat{X}_{pn})$.

Например, L -статистика

$$y = |\hat{X}_{pn} - X_{pm}| \quad (28)$$

может послужить основой построения непараметрического алгоритма обнаружения появления полезного сигнала на фоне помех с повышенной стабильностью F (в отличие от ранговых обнаружителей здесь остается, хотя и ослаблена, зависимость F от Θ). Алгоритмы типа (28) могут превосходить ранговые алгоритмы по мощности (вероятности правильного обнаружения), в частности, на малых $n + m$.

1.7.4. Специальные методы устойчивой обработки сигналов

Практический интерес представляют смешанные параметрически-непараметрические постановки задач обработки сигналов для существенно негауссовых, в частности, мультимодальных ПРВ типа (3). Такие задачи не тривиальны и привлекают особое внимание. Мультимодальные ПРВ не относятся к классу экспоненциальных, поэтому строгие методы теорий подобия и инвариантности к ним неприменимы. Эмпирический байесов (адаптивный) метод не всегда эффективен при низком качестве оценок $\hat{\Theta}$. Непараметрические методы недостаточно разработаны, кроме задач обнаружения, и малоэффективны при небольших выборках. На практике априорные сведения о сигнально-помеховой обстановке в ряде случаев позволяют сформулировать, параметрические ограничения на основную модель $\omega_1(x/\Theta_n, \Theta_1)$ в формуле (3), указав тип ПРВ и диапазон вариации компонент вектора Θ_1 . Относительно второго слагаемого в формуле (3) возможны две постановки задач: параметрическая и непараметрическая (см. задачу 4, разд. 1.7.3).

Для решения задач синтеза устойчивых алгоритмов обработки сигналов применим следующие параметрические и непараметрические методы.

В параметрической постановке задач обработки сигналов с ПРВ (3) для оценки неизвестных параметров p, Θ_2 применяются методы моментов и квантилей, используемые либо непосредственно, либо как оценки первого приближения для применения метода максимального правдоподобия.

При непараметрической гипотезе относительно $\omega(x/\Theta_n, \Theta_2)$ в формуле (3) применяют метод M -оценок Хубера, являющийся минимаксным по отношению к ПРВ вида (3) и дающий возможность найти вид функционального преобразования $y_i = \Psi(x_i)$, при котором гарантируется подавление аномальных ошибок (выбросов) в выборке \vec{x}_n , если они имеются. Одновременно сохраняются высокие показатели эффективности обработки сигналов при отсутствии аномальных ошибок. При конечных p в формуле (3) оптимальность метода M -оценок не гарантируется (при $p=1$ алгоритмы, синтезированные по методу M -оценок, могут быть полностью неработоспособными).

В любой из рассмотренных постановок задач обработки сигналов с ПРВ вида (3) перспективным представляется применение метода расслоения неоднородной выборки объемом \vec{x}_k , определяемой формулой (2), на две подвыборки объемом K_1 и K_2 , где $K_1 + K_2 = K$ в некотором смысле более однородные. В алгоритмах с обучением расслоению классификации сначала предлагается подвергнуть обучающую выборку \vec{x}_m ($k = m$). По выборкам $\vec{x}_{m1}, \vec{x}_{m2}$ объемом m_1 и m_2 , где $m_1 + m_2 = m$, необходимо найти оценки $\hat{\Theta}_1, \hat{\Theta}_2$ параметров помех и, если качество оценок $\hat{p} = m_2/m, \Theta_1, \Theta_2$ удовлетворительное, можно считать задачу решенной в первом приближении, т. е. редуцированной к известным задачам 1 — 3 разд. 1.7.3.

Обработка сигнальной выборки \vec{x}_n проводится по двухканальной схеме (с расслоением) или без расслоения. Основная специфика возникающих здесь задач классификации и оценивания p, Θ_1, Θ_2 связана с жестким требованием устойчивости показателей качества алгоритма классификации к вариациям этих параметров, а также с устойчивостью алгоритма оценивания $\hat{\Theta}_1, \hat{\Theta}_2$ к аномальным выбросам в подвыборках m_1, m_2 из-за ошибок классификации. Тре-

бование устойчивости остается основным и при выборе алгоритма обработки рабочей выборки.

Так как при $p \approx 0$ в формуле (3), а также при некоторых Θ_2 бимодальность ПРВ теряется и классификация становится неэффективной, алгоритм обработки должен быть дополнен соответствующими алгоритмами проверки гипотез относительно бимодальности ПРВ (3). При малых $n + m$ и $p = 0$ алгоритм обработки должен быть одноканальным параметрическим; при выраженной бимодальности ПРВ алгоритм классификации реализуется поканально параметрическими методами; наконец, в случае $n + m \ll 1$, $p \neq 0$ и при невыраженной бимодальности ПРВ алгоритм обработки в каждом канале должен быть непараметрическим.

Задачи расслоения неоднородных выборок изучались в статистике для случая малых p . В литературе рассмотрены алгоритмы («критерии») для выделения аномальных наблюдений в выборке, которая в основном описывается нормальным распределением

$$m_1(x/\Theta, \Theta) = N(m_1, \sigma_1^2).$$

Выбросы или аномальные значения принадлежат также нормальному распределению, но с другим математическим ожиданием — $N(m_2, \sigma_2^2)$. Классификаторы П. В. Смирнова и некоторые другие, использующие порядковые статистики, рассмотрены далее.

В заключение отметим, что методы проектирования устойчивых алгоритмов обработки сигналов при конечных объемах выборки в целом плохо формализуются и остаются в своей основе эвристическими. Имеем обширную патентную литературу, накоплен опыт решения частных задач.

1.7.5. Особенности реализации устойчивых алгоритмов обработки сигналов

Задача реализации алгоритмов обработки сигналов решается, как правило, в два этапа. На первом этапе (этапе аппроксимации) на основе численных методов разрабатывается структурная схема и определяются параметры вычислительной процедуры, реализующей выбранный алгоритм обработки. На втором этапе определяются способ реализации (программный, аппаратный, смешанный) и необходимые технические средства. Оба этапа взаимосвязаны, а желание оптимизировать технические решения приводит к необходимости их многократной итерации.

Отличительной особенностью устойчивых алгоритмов обработки сигналов является использование следующих основных операций,

выполнение которых, как правило, должно обеспечиваться по мере поступления отсчетов x_i ; выборки $\vec{x}_{(n+m)}$; решения линейных, алгебраических и трансцендентных уравнений (систем уравнений) в задачах оценивания неизвестных параметров сигналов и помех $\Theta_n, \Theta_1, p, \Theta_2$ в ПРВ (3); ранжировки отсчетов рабочей выборки \vec{x}_n по вариационному ряду составной обучающей и рабочей выборки $\vec{x}_{(n+m)}$, т. е. построение вектора рангов \vec{R}_n ; вычисления функций выборочных отсчетов вида $f(\vec{x}_{(n+m)})$, порядковых L - и ранговых R -статистик в задачах оценивания, обнаружения и фильтрации. В устойчивых алгоритмах применяется также обычная операция вычисления линейной свертки двух функций, описывающая в интегральной форме реакцию линейной системы на входное воздействие, заданное во временной или частотной области. В ряде случаев экономия памяти достигается заданием линейных операторов в виде дифференциальных (разностных) уравнений.

Используемые для реализации алгоритмов обработки сигналов технические средства принято классифицировать по следующим признакам: формы представления входных сигналов $x(t)$ — аналоговые, цифровые и дискретные (непрерывные по состоянию, дискретные по времени); степени универсальности вычислительных средств (универсальные и специализированные); архитектуры ЭВМ (специализированные); используемой элементной базы.

Рассматривая специализированные и универсальные средства цифровой обработки сигналов, внимание обращают на особые, иногда труднореализуемые требования обработки в реальном времени: обеспечение выдачи результатов с задержкой, не превышающей заданную и, как правило, небольшую по сравнению с интервалом поступления данных; высокую надежность аппаратных и программных средств обработки; контролируемую погрешность вычислений.

В памяти вычислительной системы (ВС), работающей в реальном времени, должно храниться не более $n+m$ отсчетов \vec{x}_m, \vec{x}_n . При поступлении очередных $(m+1)$ -го и $(n+1)$ -го отсчетов обучающей и рабочей выборок в памяти стираются отсчеты x_1 обеих выборок и состав \vec{x}_m обучающей выборки обновляется: $\vec{x}_m = \{x_2, \dots, x_{m+1}\}$; $\vec{x}_n = \{x_2, \dots, x_{n+1}\}$ и т. д. Такой режим называют режимом скользящего окна (PCO). При M источниках сообщений ОЗУ ВС должно иметь не менее $(n + m) M$ разрядных ячеек памяти. Если учитывать необходимость хранения промежуточных и конечных результатов вычислений $y_k = f(\vec{x}_{(n+m)})$, R - и L -статистик,

решений уравнений правдоподобия и т. д., то уже при $n + m \approx 10^2$, $M \approx 10^2 - 10^5$, $K \approx 16 - 8$ несложно оценить требуемую емкость ОЗУ ($10^4 - 10^7$ байт) для хранения данных, емкость ПЗУ для табличного вычисления функций $y_k = f(\cdot)$ и сформулировать требования к производительности спецпроцессора, определяемой в основном стандартным численным методом решения уравнений, заданной шириной энергетического спектра процесса и темпом поступления выборочных отсчетов.

Отметим одну особенность обработки сигналов в РСО: после заполнения ОЗУ и проведения необходимых вычислений на этапе «запуска» дальнейшее функционирование алгоритма происходит при относительно медленном темпе смены массива данных. Поэтому результаты обработки предыдущего i -го шага можно использовать в качестве первого приближения для формирования результата $(i + 1)$ -го шага. Знание первого приближения часто упрощает алгоритм и сокращает время поиска решений уравнений методами простых итераций Ньютона, поиска глобального минимума. Совершенно очевидно, что на этапе «запуска» проблема «нулевого» приближения часто является основной и предполагает значительные затраты вычислительных ресурсов.

1.8. Устойчивые алгоритмы измерения (оценки) параметров сигналов на фоне помех

1.8.1. Основные понятия и определения

Рассмотрим задачи измерения (оценки) информационных параметров полезных сигналов, помех или сигналов на фоне помех с позиций устойчивости алгоритмов оценки.

Модель наблюдения в заданном сечении оси времени t_i определим одномерной составной плотностью распределения вероятностей (ПРВ) независимых отчетов $x(t_i) = x_i$ типа (3 п.1.7):

$$\omega(x_i/\Theta_n, \Theta) = (1 - p)\omega_1(x_i/\Theta_n, \Theta_1) + p\omega_2(x_i/\Theta_n, \Theta_2), \quad i = \overline{1, n}, \quad (1)$$

где Θ_n — вектор информационных параметров, подлежащий оценке; $\Theta = (\Theta_1, \Theta_2, p)$ — вектор неинформационных (мешающих) параметров сигналов и помех, существенно влияющих на качество оценки Θ_n ; $\omega_1(x_i/\Theta_n, \Theta_1)$ — «основная модель» сигнально-

помеховой ситуации; $\omega_2(x_i/\Theta_n, \Theta_2)$ — известная или неизвестная ПРВ аномальных ошибок, загрязняющих выборку наблюдений \vec{x}_n выбросами или обнулениями, не учитываемыми в основной ПРВ $\omega_1(x_i/\Theta_n, \Theta_1)$.

В соответствии с общепринятым в технике делением случайных процессов на широкополосные и узкополосные, ширина энергетического спектра которых значительно меньше его средней частоты, будем рассматривать следующие частные случаи условных ПРВ в формуле (1): гауссовские, релеевские, Релея — Раиса.

Составные ПРВ для этих случаев запишем, используя буквенные обозначения ПРВ:

$$N(m, \sigma^2), R(\sigma)^2 = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right],$$

$$RR(\sigma^2, u) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2 + u^2}{2\sigma^2}\right] I_0\left(\frac{xy}{\sigma^2}\right);$$

$$R(\sigma^2) = RR(\sigma^2, 0)$$

В примерах синтеза и выбора алгоритмов оценивания будем использовать следующие частные виды формулы (1):

$$\omega(x_i/m_1, m_2, \sigma_1^2, \sigma_2^2, p) = (1 - p)N(m_1, \sigma_1^2) + pN(m_2, \sigma_2^2); \quad (2)$$

$$\omega(x_i/\sigma_1^2, \sigma_2^2, p) = (1 - p)R(\sigma_1^2) + pR(\sigma_2^2); \quad (3)$$

$$\omega(x_i/\sigma_1^2, u, p) = (1 - p)R(\sigma_1^2) + pRR(\sigma_1^2, u) \quad (4)$$

и др.

Для коррелированных отчетов x_i и x_j ($i \neq j$) вид многомерной ПРВ будет конкретизироваться в примерах, однако одномерная ПРВ всегда полагается относящейся к типу (1), т. е. существенно негауссовой и, кроме того, в ряде случаев удовлетворяющей требованиям бимодальности.

Содержательная трактовка параметра Θ_n может быть любой в зависимости от рассматриваемой задачи оценивания и частного вида наблюдаемого процесса $x(t)$:

$$x(t) = \alpha(t); \quad x(t) = S[\alpha(t), \beta, t];$$

$$x(t) = S[\alpha(t), \beta, t] + n(t)$$

и т. п. (см. рис. 1).

Для повторной выборки Θ_n имеет смысл: моментов ПРВ $\omega_1(\cdot)$, $\omega_2(\cdot)$, если $\omega_i(\cdot) = N(m_i, \sigma_i^2)$, $i = 1, 2$; некоторых неслучайных, но неизвестных физических параметров случайных процессов $x(t)$, в частности, таких, как неизвестная постоянная составляющая процесса $x(t)$, называемая в статистике параметром

сдвига $\Theta_n = \Theta_c$; неизвестного среднеквадратического отклонения $\Theta_n = \sigma$, где σ — коэффициент масштаба. Если Θ_n — известная функция сообщения α , например, $\Theta_n = S(\alpha)$, то, оценивая Θ_n , получаем возможность формирования оценки $\hat{\alpha} = S^{-1}(\hat{\Theta}_n)$. Это случай так называемых косвенных измерений. Поскольку устойчивая и эффективная оценка $\hat{\Theta}_n$ гарантирует при однозначных и достаточно гладких функциях $S^{-1}(\cdot)$ устойчивость и качество $\hat{\alpha}$, то будем рассматривать только задачи устойчивости «прямых» измерений Θ_n .

Следующий пример характеризует количественно понятие устойчивости оценок. Рассмотрим для модели (2) задачу оценки $\Theta_n = m_1$, полагая, что в основной модели, определяемой ПРВ $N_1(m_1, \sigma_1^2)$, параметр σ_1^2 известен. Если не учитывать возможности появления аномальных ошибок и полагать $p=0$, то алгоритм оценки $\hat{\Theta}_n = f_1(\vec{x}_n)$ синтезируется элементарно по методу максимального правдоподобия и дается статистикой выборочного среднего

$$\hat{\Theta}_n = \hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i = f_1(\vec{x}_n). \quad (5)$$

Качество оценок $\hat{\Theta}_n$ определяется ПРВ $\omega(\varepsilon/m_1, \sigma_1, f_1, n)$, где $\varepsilon = \hat{\Theta}_n - \Theta_n$ — ошибка измерения. Эта ПРВ гауссова с нулевым математическим ожиданием и дисперсией $\mu_2\{\varepsilon/m_1, \sigma_1^2, f_1, n\} = \sigma_1^2/n$, т. е. алгоритм $f_1(\vec{x}_n)$ обеспечивает несмещенность оценки Θ_n , а ее дисперсия соответствует минимальной границе дисперсии, определяемой неравенством Рао — Крамера.

Если при эксплуатации устройства, реализующего алгоритм (5), условие $p = 0$ будет нарушено и в гауссову выборку будут аддитивно подмешаны выбросы из другой генеральной совокупности $N(u_b, \sigma_b^2)$, то реализуется совершенно отличная от основной модели плотность выборочных отсчетов \vec{x}_n вида (2), где $m_b = m_1 + u_b$; $\sigma_b^2 = \sigma_1^2 + \sigma_b^2$.

Нетрудно вычислить математическое ожидание и дисперсию «загрязненной» выборки, полагая все параметры ПРВ (2) известными,

$$m_1\{x_i/\Theta_n, \Theta_n + u_b, \sigma_1^2, \sigma_1^2 + \sigma_b^2, f_1, n, p\} = \Theta_n + p u_b; \quad (6)$$

$$\begin{aligned} \mu_2\{x_i/\Theta_n, \Theta_n + u_b, \sigma_1^2, \sigma_1^2 + \sigma_b^2, f_1, n, p\} = \\ = \sigma_1^2 + p[u_b^2(1-p) + \sigma_b^2] \end{aligned} \quad (7)$$

и в итоге оценить смещенность $\varepsilon = f_1(\vec{x}_n) - \Theta_n$ оценки (5)

$$\begin{aligned} m_1\{\varepsilon\} &= p u_b; \\ \mu_2\{\varepsilon\} &= \sigma_1^2/n + \Delta\mu_2, \end{aligned} \quad (8)$$

где

$$\Delta\mu_2 = \frac{p}{n} [u_b^2(1-p) + \sigma_b^2]. \quad (9)$$

Например, при $n = 10$, $m_2 = 1$, $u_b = 5$, $\sigma_1^2 = \sigma_b^2 = 1$ и $p = 0,2$ смещенность оценки $\hat{\Theta}$ составит примерно единицу, т. е. величину, равную самому оцениваемому параметру, а приращение дисперсии оценки $\Delta\mu_2 = 0,42$, что составит 42 % к ожидаемой величине 0,1. Именно в этом смысле будем понимать неустойчивость алгоритмов оценивая, обращая внимание на чувствительность алгоритмов оценивания к текущим значениям параметров выбросов. В частности, алгоритм (5) монотонно ухудшает свои показатели с ростом u_b , при $p \rightarrow 1$ растет смещенность, поведение $\Delta\mu_2$ относительно p более сложно: при $\sigma_b = 0$ наименее благоприятным является случай $p_n = 0,5$; при $\sigma_b \leq u_b$, например, наименее благоприятна вероятность выбросов, определяемая условием

$$p_n = 0,5 \left(1 + \frac{\sigma_b^2}{u_b^2} \right).$$

В задачах с обучением формулу (1) будем использовать в виде (3 п.1.7), т. е. будем полагать, что $\Theta_n = 0$, по выборке \vec{x}_m будем формировать оценки мешающих параметров Θ , а затем, используя рабочую выборку \vec{x}_n , оценивать Θ_n . Такая процедура часто значительно упрощает алгоритм оценки в вычислительном плане, хотя в общем случае не является оптимальной в смысле каких-то строгих критериев эффективности оценивания. В целом последовательность изложения методов синтеза и эвристического выбора устойчивых алгоритмов оценивания информационных параметров процессов $x(t)$ аналогична методике, принятой в разд. 1.7.4.

Сначала рассматриваются параметрические алгоритмы, затем последовательно алгоритмы M -, L - и R -оценивания и, наконец, алгоритмы, использующие расслоение выборки. M -оценками называются алгоритмы, позволяющие получать оценки, максимизирующие некоторую функцию выборочных отсчетов $\Lambda(\vec{x}_n, \Theta)$, выполняющую роль функции правдоподобия:

$$\hat{\Theta} = \arg \max_{\Theta \in \Omega} \{\Lambda(\vec{x}_n, \Theta)\}. \quad (10)$$

L -оценками называются алгоритмы получения устойчивых оценок в виде взвешенных сумм порядковых статистик:

$$\hat{\theta} = \sum_{i=1}^n d_{in} x_{(i)}. \quad (11)$$

R-оценками называются алгоритмы получения устойчивых оценок на базе статистик, зависящих лишь от рангов наблюдений. Эти алгоритмы предполагают наличие двух выборок x_1, \dots, x_m и y_1, \dots, y_n . Используются ранги наблюдений $x_i, i = 1, \dots, m$, в объединенной выборке. R-оценки позволяют оценить относительный сдвиг, масштаб двух выборок.

1.8.2. Параметрические алгоритмы оценки

Рассмотрим повторную выборку с ПРВ (1), где будем считать заданными плотности основной модели $\omega_1(x_i/\theta_u, \theta)$ и аномальных ошибок $\omega_2(x_i/\theta_u, \theta_2)$.

Основное внимание уделим особенностям построения простых и достаточно эффективных и устойчивых алгоритмов оценивания параметров бимодальных ПРВ (1).

Принцип совместного оценивания информационных θ_u и мешающих θ, ρ параметров в ПРВ вида (1) освобождает ПРВ оценки $\hat{\theta}_u$ или $\epsilon = \hat{\theta}_u - \theta_u$ от непосредственной зависимости от мешающих параметров θ_2, ρ . Однако расширение размерности вектора одновременно оцениваемых по обучающей x_m и рабочей \vec{x}_n выборкам параметров (θ_u, θ, ρ) снижает точность оценок $\hat{\theta}$ и усложняет алгоритм формирования оценок. Поскольку установившейся общей теории устойчивого параметрического оценивания, использующей принцип расширения размерности оцениваемых векторов мешающих параметров, не существует, дадим краткую характеристику используемых методов синтеза устойчивых алгоритмов оценивания и рассмотрим примеры для типовых ПРВ вида (2) — (4).

Общий подход в теории параметрического оценивания, как известно, основан на методе максимального правдоподобия и сводится к поиску максимума функции правдоподобия

$$L(\theta_u, \theta) = \prod_{i=1}^k \omega(x_i/\theta_u, \theta) \quad (12)$$

по параметрам θ_u, θ для фиксированной по результату наблюдения выборки \vec{x}_k .

При большой размерности векторов θ_u, θ , учитывая, что логарифмирование (12), обычно применяемое для упрощения исходной системы уравнений

$$\left. \begin{aligned} \frac{\partial L(\theta_u, \theta)}{\partial \theta_{u_j}} &= 0; \quad j = \overline{1, q_u}; \\ \frac{\partial L(\theta_u, \theta)}{\partial \theta_j} &= 0; \quad j = \overline{1, q}, \end{aligned} \right\} \quad (13)$$

неэффективно для ПРВ вида (1), очевидна сложность алгоритмов получения оценок $\hat{\theta}_u, \hat{\theta}$ по методу максимального правдоподобия. При этом принципиальной является необходимость получения оценок нулевого приближения $\hat{\theta}_{u0}, \hat{\theta}_0$ для обеспечения сходимости вычислительных алгоритмов к глобальному экстремуму функции (12). При отсутствии таких оценок метод максимального правдоподобия в ряде случаев неработоспособен. Более простые алгоритмы можно получить эвристически на основе поиска набора некоторых функционалов ПРВ (1) приравниванием этих функционалов их выборочным эквивалентам и решением систем алгебраических или трансцендентных уравнений относительно оцениваемых параметров θ_u, θ, ρ . Сущность метода заключается в поиске системы уравнений, имеющей аналитическое решение. Тогда алгоритм оценивания найдем в виде явной функции выборки, т. е.

$$\hat{\theta}_u = f(\vec{x}_{n+m}). \quad (14)$$

Эффективность этих алгоритмов, оцениваемая по критериям достоверности (7) — (12) из п.1.7, может быть, как показывает опыт, достаточно высокой. Частными случаями такого подхода являются известные методы моментов (ММ), минимума χ -квадрат (МХК), квантилей (МК).

Рассмотрим набор функционалов от ПРВ (1) следующего вида:

$$\Phi_j(\theta_u, \theta, \rho) = \int_{(x)} \rho_j(x) \omega(x/\theta_u, \theta) dx, \quad j = \overline{1, l} \quad (15)$$

и систему l уравнений

$$\hat{\Phi}_j = \Phi_j(\theta_u, \theta, \rho), \quad (15a)$$

где $\rho_j(x)$ — весовая функция; $\hat{\Phi}_j$ — выборочное значение функционала: $\hat{\Phi}_j = f_j(\vec{x}_{n+m})$.

Существуют такие $\rho_j(x)$, что значения функционалов $\Phi_j(\theta_u, \theta, \rho)$ допускают несмещенное оценивание.

Действительно, если полагать $\rho_j(x) = x^j$, то $\Phi_j(\theta_n, \theta, \rho) = m_j\{x\}$, т. е. начальному моменту j -го порядка. Известно, что выборочный начальный момент

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^n x_i^j \quad (16)$$

имеет математическое ожидание

$$m_1\{\hat{m}_j\} = \frac{1}{n} \sum_{i=1}^n m_1\{x_i^j\} = m_j\{x\}.$$

Для центральных моментов, когда в формуле (15) $\rho_j(x) = (x - m_1\{x\})^j$, смещенность выборочных центральных моментов

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{m}_1)^j \quad (17)$$

устраняется. Несмещенными оценками нескольких первых центральных моментов являются следующие выборочные функции:

$$\begin{aligned} \hat{\mu}_{2\text{несм}} &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m}_1)^2; \\ \hat{\mu}_{3\text{несм}} &= \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \hat{m}_1)^3; \\ \hat{\mu}_{4\text{несм}} &= \frac{(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n (x_i - \hat{m}_1)^4 - \\ &\quad - \frac{3(2n-3)}{n(n-1)(n-2)(n-3)} \left[\sum_{i=1}^n (x_i - \hat{m}_1)^2 \right]^2. \end{aligned} \quad (18)$$

Применение несмещенных оценок (16), (18) в формуле (15) приводит к методу моментов Пирсона. К методу квантилей в формуле (15а) приводит специальный вид весовой функции

$$\rho_j(x) = \begin{cases} 1, & \text{если } x < V_j; \\ 0, & \text{если } x > V_j, \end{cases} \quad j = \overline{1, l}; \quad (19)$$

где V_j — набор известных не совпадающих между собой чисел из области значений случайной величины x . Выборочным эквивалентом функционала $\Phi_j(\theta_n, \theta, \rho)$ в формуле (15) является значение эмпирической функции распределения $\hat{F}_j = \hat{F}(V_j)$. Очевидны возможности использования других модификаций метода выборочных функционалов (15), соответствующих различным комбинациям уравнений метода моментов, когда часть функций $\rho_j(x)$ определяется

формулами (16), (18), а другая часть — функциями (19). Применение других видов $\rho_j(x)$, в частности,

$$\rho_j(x) = \begin{cases} 0, & \text{если } x < V_j; \\ x_j^k, & \text{если } V_j \leq x \leq V_{j+1}; \\ 0, & \text{если } x > V_{j+1}, \end{cases} \quad j = \overline{1, l}, \quad k_j = 0, 1, 2, \dots \quad (20)$$

обобщает известные методы моментов и квантилей. Выборочный функционал в формуле (15) имеет смысл выборочного момента k_j -го порядка не всей выборки \bar{x}_n , а ее части, т. е. только тех выборочных отсчетов, значения которых находятся в «щели» V_j, V_{j+1} . Такие алгоритмы в ряде случаев эффективны, так как по критерию помехоустойчивости к одиночным аномальным ошибкам, не учитываемым параметрической моделью (1), они эквивалентны квантильным алгоритмам (вкладом одиночной ошибки в оценку $\hat{F}(V_j)$ или $\hat{F}(V_j) - \hat{F}(V_{j+1})$ можно пренебречь) и оставляют

возможность варьирования аналитической структурой уравнений в системе (15).

К известному методу χ -квадрат и его обобщения — можно прийти, если в формуле (15) составить невязки для отдельных уравнений

$$\Delta\Phi_j(\theta_n, \theta, \rho) = \Phi_j(\theta_n, \theta, \rho) - \hat{\Phi}_j$$

и перейти к новому функционалу

$$\Phi(\theta_n, \theta, \rho) = \sum_{j=1}^l \Delta\Phi_j^2 F[\theta_n, \theta, \rho, \hat{\Phi}_j], \quad (21)$$

численный алгоритм поиска минимума которого и определит искомый алгоритм оценивания θ_n . Для метода χ -квадрат

$$F(\cdot) = \frac{1}{\Phi_j(\theta_n, \theta, \rho)},$$

весовая функция функционала определяется формулой (20). Для так называемого модифицированного критерия χ -квадрат — формулой $F(\cdot) = \hat{\Phi}_j^{-1}$, что упрощает вид уравнений без существенной потери в эффективности.

Перейдем к рассмотрению методов расчета точностных характеристик алгоритмов оценивания в случае, если систему уравнений (15а) удастся решить аналитически и записать алгоритм оценивания информационного параметра в виде явной функции

$$\hat{\theta}_n = f_1(x_1, \dots, x_{n+m}) \quad (22)$$

выборочных векторов \vec{x}_n, \vec{x}_m или функций

$$\hat{\theta}_n = f_2(\hat{\Phi}_1, \dots, \hat{\Phi}_l) \quad (23)$$

выборочных функционалов (моментов, вероятностей и др.).

Отметим, что в формуле (22) аргументы являются независимыми случайными величинами и для расчета математического ожидания и дисперсии $f_1(\cdot)$ можно воспользоваться формулами

$$m_1\{\hat{\Theta}_n\} = f_1[m_1\{x_1\}, \dots, m_1\{x_{n+m}\}] + \frac{1}{2} \sum_{i=1}^{n+m} \frac{\partial^2 f_1}{\partial x_i^2} \mu_2\{x_i\}; \quad (24)$$

$$\mu_2\{\hat{\Theta}_n\} = \sum_{i=1}^{n+m} \left(\frac{\partial \bar{f}_1}{\partial x_i} \right)^2 \mu_2\{x_i\} + \sum_{i=1}^{n+m} \left(\frac{\partial \bar{f}_1}{\partial x_i} \right) \left(\frac{\partial^2 \bar{f}_1}{\partial x_i^2} \right) \mu_3\{x_i\}, \quad (25)$$

где черта над символом функции означает, что соответствующие производные от этой функции взяты при значениях аргументов, равных математическому ожиданию.

Формулы (24), (25) получены при аппроксимации функции $f_1(\cdot)$ полиномом второго порядка. Второй член в формуле (24) определяет смещение оценки. Для бимодальных ПРВ типа (1) существенно влияет на дисперсию оценки, как это следует из формулы (25), коэффициент асимметрии $\mu_3\{x_i\}$, который может быть большим. Так как число оцениваемых параметров $l \ll n + m$, то иногда полезно искать моменты оценки (23), учитывая, однако, что аргументы являются коррелированными величинами. Точность аппроксимации многомерной параболой в этом случае выше, так как дисперсии выборочных функционалов меньше дисперсий выборочных отсчетов, однако вычисления смешанных ковариаций до третьего порядка включительно громоздко. Рассмотрим примеры использования метода выборочных функционалов.

Пример 3. ПРВ (1) имеет вид

$$\omega(x/u_1, p) = (1-p)\delta(x-u_1) + p\delta(x).$$

Пусть u_1 — амплитуда потока импульсных сигналов и ставится задача оценки параметров p, u_1 по выборке объемом n . Воспользуемся методом моментов, полагая в формуле (15) $\rho_j(x) = x^j$ и записывая систему уравнений для конкретного набора параметров (Θ_n, Θ, p) в виде

$$\hat{\Phi}_j(u_1, p) = \hat{m}_j = (1-p)u_1^j, \quad j = 1, 2. \quad (26)$$

Из первого уравнения системы (26) следует искомый алгоритм оценки

$$\hat{p} = 1 - \frac{1}{nu_1} \sum_{i=1}^n x_i, \quad (27)$$

где $\sum_{i=1}^n x_i = \sum_{i=1}^n \xi_i u_1$, ξ_i — случайная величина, принимающая значения 1 или 0 с вероятностями соответственно $1-p$ и p .

Таким образом, формулу (27) запишем в виде

$$\hat{p} = 1 - \frac{\hat{k}}{n}, \quad (28)$$

где $\hat{k} = \sum_{i=1}^n \xi_i$ — случайная величина, имеющая распределение Бернулли,

$$P(k) = C_n^k (1-p)^k p^{n-k} \quad (29)$$

с моментами $m_1\{k\} = n(1-p)$; $\mu_2\{k\} = np(1-p)$.

Качество оценки \hat{p} поэтому определится выражениями;

$$m_1\{\hat{p}\} = 1 - \frac{n(1-p)}{n} p,$$

$$\mu_2\{\hat{p}\} = \frac{p(1-p)}{n}.$$

Как и следовало ожидать, из самой физической постановки задачи при известной или неизвестной амплитуде импульсов u_1 , следующих с пропусками, синтезированный по методу моментов алгоритм измерения вероятности появления импульсов (28) сводится к подсчету числа k импульсов в выборке. Этот алгоритм дает несмещенную оценку вероятности и обладает дисперсией, совпадающей с дисперсией оценки максимального правдоподобия (в данном случае с минимальной границей дисперсии, определяемой неравенством Рао — Крамера).

Если далее усложнить задачу и интересоваться оценкой u_1 , то необходимо воспользоваться вторым уравнением в системе (26) и, поделив второе уравнение на первое, получим:

$$\hat{u}_1 = \begin{cases} \frac{\hat{m}_2}{\hat{m}_1}, & \text{если } k \geq 1; \\ 0, & \text{если } k = 0. \end{cases} \quad (30)$$

Случай $k = 0$ соответствует событию ненаблюдения импульсов в выборке объема n . Вероятность такого события, как это следует из формулы (29), $P(0) = p^n$. Вероятность противоположного события $P(k \geq 0) = 1 - p^n$, что позволяет записать моменты оценки (30) в виде

$$\begin{aligned} m_1 \{\hat{u}_1\} &= u_1 (1 - p^n); \\ m_2 \{\hat{u}_1\} &= u_1^2 (1 - p^n); \\ \mu_2 \{\hat{u}_1\} &= u_1^2 (1 - p^n) p^n. \end{aligned} \quad (31)$$

При $n = 10, p = 0,2$ смещенностью и дисперсией оценки можно пренебречь. Однако при $n \approx 3-5$ эти ошибки могут быть существенными.

Пример 4. Случай выборки из широкополосного гауссова процесса с аномальными ошибками и виде сигналов с постоянной неизвестной амплитудой u_b , аддитивно подмешиваемых с вероятностью p , изучался Джоном, и соответствует составной гауссовой модели (2) с неизвестными параметрами: m_1 — параметр сдвига; $\sigma_1^2 = \sigma_2^2$ — параметр масштаба основной модели $N(m_1, \sigma_1^2)$, $m_2 = m_1 + u_b$; p . Алгоритм Джона исследован методом статистического моделирования для $n = 40$ и 120 . При $m_1 = 0,375$; $m_2 = 0,625$; $\sigma_1^2 = \sigma_2^2 = 1$; $p = 0,5$ и $n=40$ результаты моделирования свидетельствуют о работоспособности алгоритма оценивания, синтезированного по методу моментов: смещение оценок не превышало 6%, отношение

$\mu_2 \{\hat{\Theta}_i\}^2 / m_1 \{\Theta_i\}$, где символом $\Theta_i, i = \overline{1,4}$, обозначены оцениваемые параметры, составляет от 7 до 30 %. Если, в частности, информационным параметром считать $m_1 = \Theta_{11}$, то при отказе от оценки мешающих параметров σ_1^2, u_b, p , следуя формуле (2.8), смещенность оценки \hat{m}_1 увеличивалась бы на порядок по сравнению с результатами работы Джона. Это свидетельствует о целесообразности метода момента для получения работоспособных алгоритмов устойчивого оценивания.

Пример 5. Случай выборки с выхода квадратичного детектора из огибающей узкополосного гауссова шума с дисперсией σ^2 при наличии импульсных синусоидальных помех с неизвестной амплитудой u_b и вероятностью p .

Исходная плотность огибающей имеет вид (4). В результате квадратичного преобразования огибающей $\rho_i = x_i^2$ плотность изменяется:

$$\begin{aligned} \omega(\rho/\sigma^2, u, p) &= (1 - p) \exp\left(-\frac{\rho}{2\sigma^2}\right) + \\ &+ p \exp\left[-\frac{\rho + u^2}{2\sigma^2}\right] J_0\left(\frac{u \sqrt{\rho}}{\sigma^2}\right). \end{aligned} \quad (32)$$

Первые два начальных момента величин ρ_i вычисляются точно:

$$m_1 \{\rho\} = 2\sigma^2 + pu^2, \quad m_2 \{\rho\} = 8\sigma^4 + 8p\sigma^2u^2 + pu^4, \quad (33)$$

где σ^2 считаем известным.

Система уравнений метода моментов

$$\begin{aligned} \hat{m}_1 &= \frac{1}{n} \sum_{i=1}^n \rho_i = m_1 \{\rho\}; \\ \hat{m}_2 &= \frac{1}{n} \sum_{i=1}^n \rho_i^2 = m_2 \{\rho\} \end{aligned} \quad (34)$$

решается аналитически относительно параметров u и p , определяя алгоритм оценки

$$\begin{aligned} \hat{u} &= \left[\frac{\hat{m}_2 - 8\sigma^2\hat{m}_1 + 8\sigma^4}{\hat{m}_1 - 2\sigma^2} \right]^{0,5}; \\ \hat{p} &= \frac{(\hat{m}_1 - 2\sigma^2)^2}{\hat{m}_2 - 8\sigma^2\hat{m}_1 + 8\sigma^4}. \end{aligned} \quad (35)$$

Вычисления по известным формулам показали, что смещенность оценок (35) не существенна, а дисперсии, например, оценки амплитуды \hat{u} определяются соотношением:

$$\begin{aligned} \mu_2 \{\hat{u}\} &= \frac{0,25}{p^2 u^6 n} \{ (8\sigma^2 + u^2)^2 (m_2 \{\rho\} - m_1^2 \{\rho\}) + m_4 \{\rho\} - \\ &- m_2^2 \{\rho\} - 2(8\sigma^2 + u^2) (m_3 \{\rho\} - m_1 \{\rho\} m_2 \{\rho\}) \}, \end{aligned} \quad (36)$$

где $m_3 \{\rho\} = 48\sigma^6 + 72p\sigma^4u^2 + 18\sigma^2u^4 + pu^6$;

$$m_4 \{\rho\} = 384\sigma^8 + 768p\sigma^6u^2 + 288p\sigma^4u^4 + 32p\sigma^2u^2 + pu^8.$$

Результаты статистического моделирования подтверждают расчетные формулы для дисперсий оценок \hat{u} и \hat{p} .

Пример 6. Рассмотрим задачу оценки дисперсии гауссова узкополосного шума по результатам обработки отсчетов огибающей при наличии интенсивных радиоимпульсных помех с малой вероятностью появления (модель определяется формулой (4))

Воспользуемся методом квантилей, принимая для второго члена в формуле (4) гауссову аппроксимацию. При $p = 0$ из формулы (4)

$$p_V = \int_V^\infty \omega(x/\sigma^2) dx = e^{-\frac{V^2}{2\sigma^2}}. \quad (37)$$

Формула (37) является частным случаем формулы (15) для $f_j(x)$ типа (19) и ПРВ, определяемой соотношением (3) при $p = 0$. Заменяя вероятность p_V ее эмпирическим эквивалентом, получаем уравнение относительно σ^2 , которое и дает алгоритм оценки

$$\hat{\sigma}^2 = -\frac{V^2}{2 \ln \hat{p}_V^2}. \quad (38)$$

Точность оценки (2.38) определяется формулами

$$m_1 \{\hat{\sigma}^2\} = -\frac{V^2}{2 \ln [p_V (1 + \delta)]};$$

$$m_2 \{\hat{\sigma}^2\} = \frac{V^4 (1 - p_V) (1 - p)}{4n \{ \ln [p_V (1 - \delta)] \}^4 p_V (1 + \delta)}, \quad (39)$$

где

$$\delta = \frac{1 - p_V}{p_V} p$$

— смещение оценки p_V из-за импульсных помех, вероятность которых в выборочном отсчете равна p .

Область работоспособности алгоритма (38) ограничивается условием $p_V (1 + \delta) < 1$, так как в противном случае $m_1 \{\hat{\sigma}^2\} < 0$, что не имеет смысла.

Расчеты показывают, что при $p \rightarrow 0$ оценка (38) несмещенная, но уступающая оценке максимального правдоподобия по эффективности на 50%. Однако алгоритм оценки устойчив при воздействии импульсных помех. Так, для случая $\sigma^2 = 1$, $u/\sigma = \nu$, $p = 0,2$, $V = 1,8$ смещение оценки составляет 54 % (против 350 % в случае оценки максимального правдоподобия), а дисперсия возрастает в 2,7 раза (против увеличения ее в 31 раз для максимального правдоподобия оценки).

Если получить алгоритмы оценки в виде явных функций не удастся, для решения систем уравнений (15) необходимы известные численные методы: простых итераций, Ньютона, спуска. В методе простых итераций систему уравнений (15) необходимо представить в эквивалентном виде

$$\vec{\Theta} = \varphi(\vec{\Theta}),$$

где вектор $\vec{\Theta} = (\Theta_n, \Theta, p)$, выбрать в качестве нулевого приближения некоторый вектор компонент $\vec{\Theta}^0$ и организовать итеративную процедуру

$$\vec{\Theta}^{(s+1)} = \varphi(\vec{\Theta}^s), \quad s = 0, 1, \dots$$

Построение спецпроцессора по алгоритму $\varphi(\cdot)$ технически реализуемо, однако метод не гарантирует сходимости процедуры.

Известный метод Ньютона сложнее метода простых итераций, так как на каждом шаге необходимо вычислять матрицу производных $\left[\frac{\partial f_k}{\partial \Theta_i} \right]$ исходной системы уравнений (15), записанной в виде $f(\Theta) = 0$, где f_k, Θ_i — компоненты вектор-функции $f(\cdot)$ вектора аргументов $\vec{\Theta}$.

Кроме того, на этом же шаге должна решаться система l линейных уравнений. Для сокращения вычислительных ресурсов (временных или аппаратных) можно пытаться обратную матрицу $\left[\frac{\partial f_k}{\partial \Theta_i} \right]^{-1}$ вычислить только на первом шаге и использовать ее на всех итерациях. Скорость сходимости итерационной процедуры при этом, естественно, снижается. Одним из способов запуска в методе Ньютона является вычисление матрицы производных в окрестности $\vec{\Theta}_0$, получаемой путем статистической линеаризации исходной нелинейной системы уравнений (т. е. путем замены встречающихся нелинейностей типа $f(x_i)$ линейными функциями $ax_i + b$, где a, b удовлетворяют условиям приближенного совпадения математических ожиданий и дисперсий исходной функции и ее линейной аппроксимации).

Рассмотрим стохастические итерационные методы спуска, т. е. методы поиска минимума функционалов типа (21), использующие совместно известную процедуру слепого случайного поиска и метода наискорейшего спуска. Поскольку функционалы вида (21) многоэкстремальны, то вероятность случайного попадания пробной l -мерной точки в область глобального экстремума мала даже при большом числе испытаний. Действительно, пусть линейным преобразованием координат Θ_i область вариации переменных функционала (21) сведена к l -мерному единичному кубу. Для диаметра дна глобального минимума $\varepsilon \ll 1$ вероятность попадания набора l равномерных случайных чисел в окрестность дна $p = \varepsilon^l$. Нетрудно подсчитать, что при $\varepsilon = 0,1$ и $l = 5$ $p = 10^{-5}$, т. е. для попадания в среднем одного набора случайных чисел на «дно» необходимо число бросаний не менее $k=10^5$. Эффективность процедуры резко повышается, если слепой случайный поиск дополнить градиентным спуском в область минимума (возможно, локального) с последующим выбором глобального минимума. Действительно, для отыскания локального минимума теперь необходимо попасть всего лишь в область «горловины» минимума и если горловина в q раз шире дна, то выигрыш в числе испытаний составит q^l (при $q = 2, l = 5$ это соответствует снижению числа испытаний примерно в 10^3 раз). Для систем обработки сигналов в реальном времени такой метод самый перспективный для реализации на мультипроцессорных ЭВМ с применением табличной формы записи функционалов $\Phi_j(\Theta_n, \Theta, p)$ в формуле (21). Основное его преимущество — контролируемая вероятность сходимости процедуры поиска к глобальному экстремуму, доступность экспериментальной оценки этой вероятности методом имитационного моделирования на этапе разработки алгоритма.

1.8.3. Алгоритмы, использующие М-оценки Хьюбера

Параметрические алгоритмы оценивания, основанные на методах максимального правдоподобия, моментов и других, в ряде случаев при отклонениях реальных распределений от принятой модели не удовлетворяют требованию устойчивости.

Хьюбером предложены процедуры оценивания неизвестных параметров, основанные не на поиске максимума функции правдоподобия выборки, а на поиске максимума некоторой функции выборочных значений и параметров распределений — $\Lambda(\bar{x}, \Theta)$, удовлетворяющей весьма общим условиям. Алгоритм оценивания неизвестных параметров Θ по независимым выборочным значениям определяется решением задачи на поиск максимума функции выборочных отсчетов вида:

$$\Lambda(\bar{x}, \Theta) = \sum_{i=1}^n \rho(x_i, \Theta), \quad (40)$$

где ρ — в общем случае произвольная функция, удовлетворяющая условиям непрерывности и ограниченности.

Под Θ здесь будем понимать объединенный вектор информационных и неинформационных параметров основной модели (см. формулу (1)):

$$\Theta = (\Theta_{ин}, \Theta_{1}).$$

Такой подход по сути игнорирует параметрическую информацию о «выбросах», учитывая лишь весьма общие различия основной модели и мешающих воздействий.

Оценка вектора параметров Θ^* может быть в неявном виде определена как решение системы уравнений

$$\sum_{i=1}^n \Psi_{\theta_j}(x_i, \Theta) = 0, \quad j = 1, \dots, k, \quad (41)$$

где

$$\Psi_{\theta_j}(x, \Theta) = \frac{\partial \rho(x, \Theta)}{\partial \theta_j}; \quad \Theta = (\theta_1, \dots, \theta_k)$$

— вектор параметров. Если $\Theta = (m, \sigma)$ — вектор параметров сдвига (m) и масштаба (σ), то уравнения (41) приобретают следующий вид:

$$\sum_{i=1}^n \Psi_m\left(\frac{x_i - m}{\sigma}\right) = 0; \quad \sum_{i=1}^n \Psi_{\sigma}\left(\frac{x_i - m}{\sigma}\right) = 0. \quad (42)$$

Если в качестве $\rho(x, \Theta)$ выбрать логарифм функции правдоподобия $\rho(x, \Theta) = \ln f(x, \Theta)$ выборки, формулы (42) дадут обычные оценки максимального правдоподобия.

Выбор функции $\rho(\cdot)$ играет определяющую роль в теории устойчивого (робастного) оценивания. Робастными свойствами обладают М-оценки, полученные как оценки максимального правдоподобия для наименее благоприятного распределения в p -окрестности основного распределения выборки $\omega_1(x)$, т. е. $\rho(\cdot) = \log \omega_0(x)$, где $\omega_0(\cdot)$ — наименее благоприятное в определенном смысле распределение, p -окрестность соответствует множеству распределений вида (1).

Таким образом М-оценки параметров сдвига и масштаба являются асимптотически эффективными минимаксными оценками, минимизирующими максимальное асимптотическое смещение или максимальную асимптотическую дисперсию.

Оценки сдвига. Рассмотрим отдельно задачу оценивания параметра сдвига. В этом случае надо решать одно уравнение системы (41):

$$\sum_{i=1}^n \Psi_m(x_i - m) = 0. \quad (43)$$

Для существования решения уравнения (43) функция $\Psi(x, m)$ должна удовлетворять следующим условиям:

1. $\Psi(x, m)$ — измерима как функция x и монотонно убывает по m .
2. Имеется хотя бы одно значение m_0 , при котором

$$\lambda(m_0) = \int_{-\infty}^{\infty} \Psi(x, m_0) f(x) dx = 0. \quad (44)$$

Для обеспечения робастности оценок функция $\Psi(x)$ должна быть ограниченной и условие (44) должно выполняться только в одной точке.

Приведем примеры нескольких устойчивых алгоритмов оценивания параметров сдвига, полученных минимаксными методами.

Пример 7. $\Psi(x) = \text{sign } x$.

Оценкой параметра сдвига, получаемой в этом случае, является выборочная медиана

$$\hat{m} = \text{med}\{x_i\}, \quad i = \overline{1, n}. \quad (45)$$

Выборочная медиана минимизирует максимальное смещение оценки и окрестности вида (1) симметричного распределения $\omega_1(x)$.

Таким образом, устройством, реализующим алгоритм (45), является медианный фильтр (рис. 2).

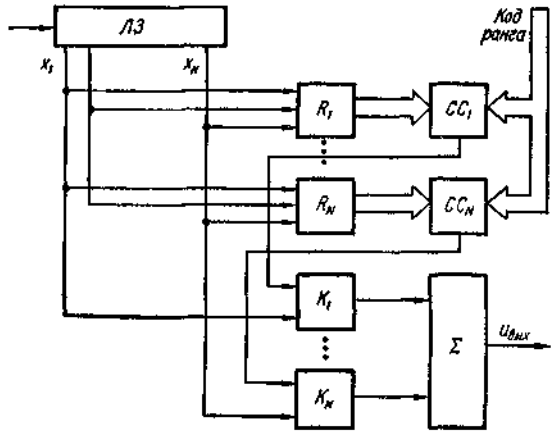


Рис. 2. Структурная схема фильтра для выбора заданной порядковой статистики

Фильтр для выбора заданной порядковой статистики работает следующим образом. Реализация входного процесса $x(t)$ запоминается в линии задержки ЛЗ. N отводов линии задержки подключены к N устройствам вычисления ранга $R_i, i = \overline{1, n}$, каждое из которых вычисляет ранг i -го выборочного отсчета. Код ранга i -го выборочного отсчета подается на вход i -й схемы совпадения CC_i , на другой вход которой подается код нужного номера порядковой статистики. Ранг только одного выборочного отсчета x_j совпадает с требуемым номером порядковой статистики и на выходе только одной схемы совпадения CC_j вырабатывается импульс, открывающий j -ю ключевую схему, на другой вход которой подается сигнал с j -го отвода линии задержки. Этот сигнал проходит через ключевую схему k_j на сумматор 2 и далее на выход устройства.

Рассмотрим M -оценки, минимизирующие максимальную асимптотическую дисперсию в p -окрестности основного распределения выборки $\omega_1(x, \theta)$. Минимаксное решение задачи дает распределение $\omega_0(\cdot)$, обладающее минимальной информацией Фишера в заданной окрестности. Это распределение имеет вид:

$$\omega_0(x) = \begin{cases} (1-p)\omega_1(x_0)e^{k(x-m-x_0)} & \text{для } x \leq x_0; \\ (1-p)\omega_1(x) & \text{для } x_0 < x < x_1; \\ (1-p)\omega_1(x_1)e^{-k(x-m-x_1)} & \text{для } x \geq x_1. \end{cases} \quad (46)$$

Здесь x_0 и x_1 — концы интервала, на котором выполняется неравенство $|\omega'/\omega| \leq k$, а величина k связана со значением p соотношением

$$\int_{x_0}^{x_1} \omega_1(x) dx + [\omega_1(x_0) + \omega_1(x_1)]/k = 1/(1-p). \quad (47)$$

Таким образом, искомое распределение $\omega_0(x)$ в центральной части на интервале $[x_0, x_1]$ совпадает по форме с основным распределением $\omega_1(x)$, а вне этого интервала имеет «длинные хвосты» экспоненциального типа. Искомая Ψ -функция, которая определяет алгоритм робастного оценивания, имеет следующий вид:

$$\Psi(x) = \frac{\partial \ln \omega_0}{\partial m} \Big|_{m=0} = \begin{cases} -k, & x \leq x_0, \\ -\omega_1(x)/\omega_1(x), & x_0 < x < x_1; \\ k, & x \geq x_1. \end{cases} \quad (48)$$

Пример 8. Если основное распределение $\omega_1(x)$ является нормальным, наименее благоприятная в p -окрестности плотность, согласно формуле (46) имеет вид

$$\omega_0(x) = \begin{cases} \frac{1-p}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, & |x| \leq k; \\ \frac{(1-p)}{\sqrt{2\pi}} e^{-\frac{k^2}{2} - k|x|}, & |x| > k, \end{cases}$$

а k и p связаны соотношением

$$2\omega_1(k)/k - 2 \int_{-\infty}^{-k} \omega_1(x) dx = p/(1-p).$$

Таким образом

$$\Psi(x) = -[\ln \omega_0(x)]'_{m=0} = \max[-k, \min(k, x)]. \quad (49)$$

Эта оценка называется оценкой благоприятной плотности Хьюбера.

Графики $\omega_0(x)$ и $\Psi(x)$ изображены на рис. 3.

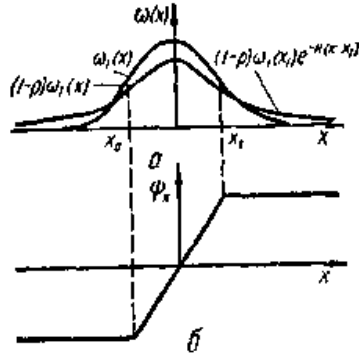


Рис. 3. Графики наименее благоприятной плотности распределения (а); характеристика нелинейного преобразователя (б)

Оценки параметра сдвига

$$\sum_{i=1}^n \Psi(x_i - \hat{m}) = 0. \tag{50}$$

Корень этого уравнения можно найти одним из методов последовательных приближений (метод наискорейшего спуска, процедура Ньютона — Рафсона и др). Возможна рекуррентная процедура поиска корня уравнения (50):

$$\hat{m}_{i+1} = \hat{m}_i + k(i) \Psi(\hat{m}_i - x_{i+1}), \tag{51}$$

где $k(i) |_{i=1, \dots, \infty}$ — последовательность коэффициентов, удовлетворяющая условиям

$$\sum_{i=1}^{\infty} k(i) = \infty, \quad \sum_{i=1}^{\infty} k^2(i) < \infty.$$

Структурная схема, изображенная на рис. 4, иллюстрирует возможную техническую реализацию алгоритма (51).

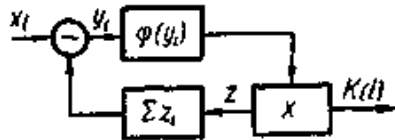


Рис. 4. Структурная схема рекуррентного параметров сдвига

Устройства, работа которых соответствует алгоритму (51), применяют в технике. Это системы автомагической регулировки порога квантования радиолокационных сигналов, различные следящие системы, рекурсивные фильтры и др.

Приведем несколько примеров возможных нелинейных преобразований входных процессов в устройствах оценивания, предложенных эвристически.

Функция Хампеля

$$\Psi(x) = -\Psi(-x) = \begin{cases} x & \text{при } 0 \leq x \leq a; \\ a & \text{при } a \leq x < b; \\ \frac{c-x}{c-b} a & \text{при } b \leq x < c; \\ 0 & \text{при } x \geq c. \end{cases} \tag{52}$$

Синусоида Эндрюса

$$\Psi(x) = \begin{cases} \sin x & \text{при } -\pi \leq x \leq \pi; \\ 0 & \text{при } x \notin (-\pi, \pi). \end{cases} \tag{53}$$

Бивес Тьюки

$$\Psi(x) = \begin{cases} x(1-x^2)^2 & \text{при } |x| \leq 1; \\ 0 & \text{при } |x| > 1. \end{cases} \tag{54}$$

Графическое изображение характеристик нелинейных преобразователей (52), (53), (54) показано на рис. 5.

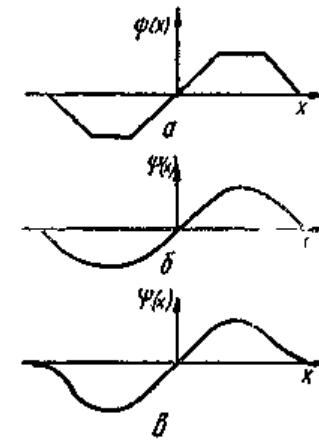


Рис. 5. Примеры характеристик нелинейных преобразователей Хампеля (а), Эндрюса (б), Тьюки (в)

Алгоритмы и устройства, использующие нелинейности такого типа, производят цензурирование особо выделяющихся наблюдений. В литературе отмечается, что качество оценок, полученных в этих алгоритмах, больше зависит от правильности масштабирования, чем в алгоритмах с монотонной функцией.

Оценки масштаба. M -оценки масштаба определяются неявным образом как решение уравнения

$$\sum_{i=1}^n \Psi_{\sigma} \left(\frac{x_i}{\sigma} \right) = 0. \quad (65)$$

Обычно $\Psi(\Theta)$ — четная функция. Для того чтобы решение (55) давало оценку максимального правдоподобия параметра масштаба однопараметрического семейства плотностей $\omega_{\sigma}(x) = \frac{1}{\sigma} \omega \left(\frac{x}{\sigma} \right)$, надо, чтобы выполнялось равенство

$$\Psi_{\sigma}(\tau) = \frac{d \ln \omega_{\sigma}(x)}{d\sigma} \Big|_{\sigma=1} = -x [\omega'(x)/\omega(x)] - 1. \quad (56)$$

Оценки с $\Psi_{\sigma}(x)$ такого вида будут обладать робастными свойствами в p -окрестности основного распределения, если в качестве ω будет выбрано наименее благоприятное в этой окрестности распределение, минимизирующее информацию Фишера для параметра масштаба.

Для поиска этого распределения рассматривают случайную величину $y = \ln x$ и семейство плотностей

$$\omega_{\tau}^*(y) = \omega^*(y - \tau) = 2e^y \omega(e^y), \quad (57)$$

зависящих от параметра сдвига $\tau = \ln \sigma$. При таком преобразовании p -окрестность основного распределения $\omega_1(x)$ преобразуется в p -окрестность для распределения $\omega_1^*(y)$ и задача поиска наименее благоприятного распределения решается так же, как для задачи оценки сдвига в окрестности

$$\omega^*(y) = (1 - p) \omega_1^*(y) + p \omega_2(y), \quad (58)$$

где $\omega_2(y)$ принадлежит множеству всех распределений на действительной прямой. Тогда элемент p -окрестности $\omega_1^*(y)$, на котором информация Фишера достигает минимума, имеет плотность

$$\omega_0^*(y) = \begin{cases} (1 - p) \omega_1^*(y_0) e^{k(y-y_0)} & \text{при } y < y_0; \\ (1 - p) \omega_1^*(y) & \text{при } y_0 < y < y_1; \\ (1 - p) \omega_1^*(y_1) e^{-k(y-y_1)} & \text{при } y > y_1, \end{cases} \quad (59)$$

где $e^{2y_0} - 1 = -k$, $e^{2y_1} - 1 = k$. Величина $k < 1$ связана с p соотношением

$$\int_{y_0}^{y_1} \omega_1^*(y) dy + [\omega_1^*(y_0) + \omega_1^*(y_1)]/k = 1/(1 - p). \quad (61)$$

Если теперь совершить обратный переход к переменной x , то соотношения (59) — (61) будут иметь следующий вид:

$$x_0^2 = (1 - k); \quad x_1^2 = 1 + k;$$

$$2 \int_{x_0}^{x_1} \omega_1(x) dx + [2x_0 \omega_1(x_0) + 2x_1 \omega_1(x_1)]/(x_1^2 - 1) = 1/(1 - p);$$

$$\omega_0(x) = \begin{cases} (1 - p) \omega_1(x_0) \left(\frac{x_0}{|x|} \right)^{x_0^2} & \text{при } |x| < x_0; \\ (1 - p) \omega_1(x) & \text{при } x_0 \leq |x| \leq x_1; \\ (1 - p) \omega_1(x_1) \left(\frac{x_1}{|x|} \right)^{x_1^2} & \text{при } |x| > x_1. \end{cases}$$

Если p — мало, $x_0 = 0$.

Пример 9. $\omega_1(x)$ — нормальное распределение. M — оценка параметра масштаба определяется функцией:

$$\Psi_{\sigma}(x) = -x \frac{\omega_0'(x)}{\omega_0(x)} - 1 = \begin{cases} x_0^2 - 1 & \text{при } |x| < x_0; \\ x^2 - 1 & \text{при } x_0 \leq |x| \leq x_1; \\ x_1^2 - 1 & \text{при } |x| > x_1. \end{cases} \quad (62)$$

Для малых p $x_0 = 0$, $x_1 > \sqrt{2}$. Алгоритм оценки масштаба в рекуррентной форме может быть записан в виде выражения:

$$\sigma_{i+1}^* = \sigma_i + K_{(i)} \Psi_{\sigma}(x_{i+1}/\sigma_i^*), \quad (63)$$

где $K_{(i)}$ — последовательность коэффициентов, удовлетворяющая условиям:

$$\sum_{i=1}^{\infty} K_{(i)} > \infty; \quad \sum_{i=1}^{\infty} K_{(i)}^2 < \infty. \quad (64)$$

Если известны ПРВ, то можно брать $K_{(i)} = 1/\int [(\ln \omega_0^i(x))']^2 \times \omega_0^i(x) dx$.

График функции (62) и структурная схема алгоритма (63) изображены на рис. 6 и 7.

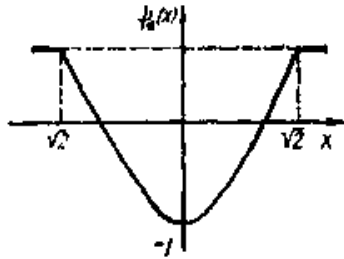


Рис. 6. График функции (62)

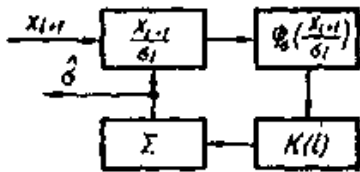


Рис. 7. Структурная схема рекуррентного алгоритма оценки масштаба

Пример 10. В качестве возможной применяют оценку масштаба, получающуюся при выборе $\Psi_{\sigma}(\cdot)$ функции в виде

$$\Psi_{\sigma}(x) = \text{sign}(|x| - 1).$$

График функции изображен на рис. 8.

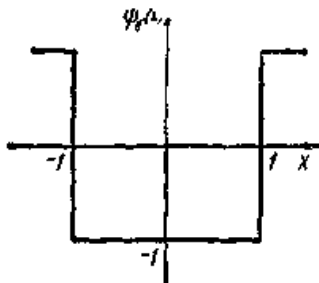


Рис. 8. График функций (65)

Такой выбор приводит к абсолютному медианному отклонению от 0.

$$S = \text{med}\{|x|\}, \tag{65}$$

Более общая модификация функции (65)

$$\Psi_{\sigma}(x) = \text{sign}(|x| - \text{med}\{x\} - 1) \tag{66}$$

приводит к абсолютному медианному отклонению от медианн. График функции (66) выглядит так же, как и график функции (65), но смещен по оси x на расстояние, равное выборочной медиане.

Пара оценок — выборочная медиана и абсолютное медианное отклонение служат примером алгоритма совместного робастного оценивания параметров сдвига и масштаба. Согласно многим источникам эти оценки обладают «наиболее робастными» свойствами.

К выборочной медиане и абсолютному медианному отклонению будем часто обращаться при построении сложных составных алгоритмов обработки сигналов САО ИИС.

1.8.4. Алгоритмы, использующие порядковые статистики

Рассмотрим алгоритмы, в которых оценки неизвестных параметров находят в виде взвешенных сумм порядковых статистик:

$$\hat{\theta} = \sum_{i=1}^n a_{in} x_{(i)}. \tag{67}$$

Алгоритм оценивания считается найденным, если получены значения коэффициентов a_{in} .

Известны различные методы определения весовых коэффициентов. Основной — это обобщенный метод наименьших квадратов. Для того чтобы вычислить значения коэффициентов a_{in} необходимо:

вычислить математические ожидания порядковых статистик нормированной случайной величины $u_{(i)} = (x_{(i)} - m)/\sigma$, распределенной по закону $\omega_1(x, \theta)$.

Вектор математических ожиданий обозначим $E[u]$, координаты вектора — $E[u_{(i)}] = \alpha_i$;

найти их дисперсии и ковариации, т. е. ковариационную матрицу

$$\text{cov}[u_{(i)}; u_{(j)}] = \vartheta_{ij};$$

записать вектор математических ожиданий упорядоченных фактических наблюдений в форме

$$E[X] = A\theta, \tag{68}$$

где A — матрица известных коэффициентов

$$A = \begin{pmatrix} 1 & \alpha_1 \\ 1 & \alpha_2 \\ \dots & \dots \\ 1 & \alpha_n \end{pmatrix},$$

Θ — вектор неизвестных параметров

$$\Theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix};$$

записать ковариационную матрицу наблюдений

$$V(X) = V\sigma^2, \quad (69)$$

где V — матрица, элементы которой Φ_{ij} , $i, j = \overline{1, n}$;

найти $\hat{\Theta}$, минимизирующий квадратичную форму

$$(X - A\Theta)' V^{-1} (X - A\Theta). \quad (70)$$

Вектор $\hat{\Theta}$ находят как решение системы нормальных уравнений

$$A'V^{-1}A\hat{\Theta} = A'V^{-1}X, \quad (71)$$

где X — вектор порядковых статистик наблюдаемой случайной величины. Решение имеет вид:

$$\hat{\Theta} = (A'V^{-1}A)^{-1} A'V^{-1}X. \quad (72)$$

Нормальные уравнения для нахождения оценок параметров \hat{m} и $\hat{\sigma}$ по вектору порядковых статистик имеют вид

$$\left. \begin{aligned} (1'\Omega 1) \hat{m} + (1'\Omega \alpha) \hat{\sigma} &= 1'\Omega X; \\ (1'\Omega \alpha) \hat{m} + (\alpha'\Omega \alpha) \hat{\sigma} &= \alpha'\Omega X, \end{aligned} \right\} \quad (73)$$

где $\Omega = V^{-1}$.

Решение уравнений (73) приводит к оценкам

$$\left. \begin{aligned} \hat{m} &= -\alpha'\Gamma X; \\ \hat{\sigma} &= 1'\Gamma X, \end{aligned} \right\} \quad (74)$$

где

$$\Gamma = \frac{1}{\Delta} \Omega (1\alpha' - \alpha 1') \Omega,$$

$$\Delta = (1'\Omega 1) (\alpha'\Omega \alpha) - (1'\Omega \alpha)^2,$$

где 1 — столбец из единиц; α — столбец α_i ; $i = 1, n$.

Дисперсии и ковариация оценок параметров сдвига и масштаба находятся по элементам обратной матрицы

$$\left. \begin{aligned} D[\hat{m}] &= \frac{\sigma^2}{\Delta} (\alpha'\Omega \alpha); \\ D[\hat{\sigma}] &= \frac{\sigma^2}{\Delta} (1'\Omega 1); \\ \text{cov}[\hat{m}, \hat{\sigma}] &= -\frac{\sigma^2}{\Delta} (1'\Omega \alpha). \end{aligned} \right\} \quad (75)$$

В частности, для построения L -оценок масштаба α релеевского распределения следует знать вектор математических ожиданий и ковариационную матрицу порядковых статистик в выборке из n значений нормированного релеевского распределения. Начальные моменты и ковариации i -й и j -й порядковых статистик вычисляются по формулам:

$$\begin{aligned} \Phi_{ij} &= E\{x_{(i)}x_{(j)}\} = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times \\ &\times \sum_{k=0}^{i-1} \sum_{l=0}^{j-i-1} (-1)^{i+l} \frac{1}{3(n-i+k+1)^3} \times \\ &\times F\left(1; 3; \frac{5}{2}; \frac{j-i+k-l}{n-i+k+1}\right); \\ m_k\{x_{(i)}\} &= \left(\frac{1}{2}\right)^{-k/2} n \binom{n-1}{i-1} \sum_{l=0}^{n-1} (-1)^l \binom{i-1}{l} \times \\ &\times \Gamma\left(1 + \frac{k}{2}\right) (n-i+l+1)^{-\left(\frac{k}{2}+1\right)}; \quad \alpha_i = m_1\{x_{(i)}\}. \end{aligned}$$

Пример 11. Рассмотрим случай $n = 3$, $i = 1$, $j = 2$. Матрица ковариации порядковых статистик

$$V = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix},$$

где Φ_{ij} вычисляются по формулам, приведенным выше. Обратная матрица

$$\Omega = V^{-1} = \frac{1}{D_V} \begin{pmatrix} \Phi_{22} & -\Phi_{12} \\ -\Phi_{21} & \Phi_{11} \end{pmatrix} = \frac{1}{D_V} W;$$

где D_V — определитель ковариационной матрицы.

Подставляя Ω в формулу (74) и учитывая, что $\Phi_{12} = \Phi_{21}$, получаем

$$\Delta = \frac{1}{D_V^2} (\alpha_1 - \alpha_2)^2 (\Phi_{11}\Phi_{22} - \Phi_{12}^2).$$

Матрица Γ в соответствии с формулой (74) имеет вид

$$\begin{aligned} \Gamma &= \frac{1}{\Delta} \left(\frac{1}{D_V^2}\right) (\alpha_2 - \alpha_1) \begin{bmatrix} 0 & \Phi_{11}\Phi_{22} - \Phi_{12}^2 \\ \Phi_{21} - \Phi_{11}\Phi_{22} & 0 \end{bmatrix} = \\ &= \frac{1}{\alpha_2 - \alpha_1} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \end{aligned}$$

Следовательно, оценка параметра σ релеевского распределения в данной задаче определяется выражением

$$\hat{\sigma} = (x_{(2)} - x_{(1)}) / (\alpha_2 - \alpha_1).$$

Дисперсия оценки вычисляется в соответствии с формулой (75):

$$D[\hat{\sigma}] = \frac{\sigma^2}{(\alpha_2 - \alpha_1)^2} (\vartheta_{22} + \vartheta_{11} - 2\vartheta_{12}).$$

Кроме обобщенного метода наименьших квадратов для оценивания параметров сдвига и масштаба используются линейные оценки, описываемые непрерывной весовой функцией

$$z(h) = \frac{1}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right) x_{(i)}, \quad (76)$$

где $h(\cdot)$ — некоторая непрерывная дифференцируемая на интервале $(0,1)$ функция.

Асимптотически оптимальная оценка параметра

$$\Theta = k_1 m + k_2 \sigma$$

$$\hat{\Theta} = z(h_0) = \frac{1}{n} \sum_{i=1}^n h_0\left(\frac{i}{n+1}\right) x_{(i)}, \quad (77)$$

где

$$h_0[x] = -a_1 \gamma_1'(x) - a_2 \gamma_2'(x);$$

$$\gamma_1(x) = -\frac{\omega'(x)}{\omega(x)}; \quad \gamma_2(x) = -x \frac{\omega'(x)}{\omega(x)} - 1.$$

Обозначим

$$a = \begin{vmatrix} a_1 \\ a_2 \end{vmatrix}; \quad k = \begin{vmatrix} k_1 \\ k_2 \end{vmatrix}; \quad d = \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix}; \quad a = d^{-1} k,$$

где

$$d_{r,s} = \int_{-\infty}^{\infty} \gamma_r(x) \gamma_s(x) f(x) dx; \quad r, s = 1, 2.$$

Дисперсия этой оценки

$$D[\hat{\Theta}] = \frac{\sigma^2}{n} k' d^{-1} k + O(n^{-2}). \quad (78)$$

Оценки \hat{m} и $\hat{\sigma}$ асимптотически совместно эффективны и порознь каждая из них эффективна, если другая играет роль мешающего параметра.

Рассмотрим оценку параметров расположения и рассеянии (m и σ) по выборочным квантилям для больших выборок. Этот метод основан на асимптотическом распределении выборочных квантилей. Параметры расположения оцениваются по методу наименьших квадратов.

Минимаксный подход к синтезу L -оценок параметров сдвига дает весовую функцию

$$h(u) = -\left[\frac{\omega'_0(u)}{\omega_0(u)} \right]' \frac{1}{f(\omega_0)}, \quad (79)$$

где $u \in (0, 1)$, $\omega_0(\cdot)$ — наименее благоприятная плотность распределения в p -окрестности основного распределения; $f(\omega_0)$ — информация Фишера.

Вариационными методами определяется наилучшее распределение в p -окрестности и для этого распределения строятся линейные оценки методами, описанными выше (МНК и др.).

Приведем несколько примеров L -оценок.

Пример 12. Оценка параметра сдвига, называемая α -урезанное среднее,

$$\hat{m} = \frac{1}{n - 2\alpha} \sum_{i=\alpha+1}^{n-\alpha} x_{(i)}. \quad (80)$$

Эта оценка получается как минимаксная линейная оценка в p -окрестности нормального распределения.

Пример 13. В качестве статистики, характеризующей параметр масштаба, часто используют t -квантильный размах

$$S = x_{([n(t-1)])} - x_{([nt])}, \quad (81)$$

$0 < t < 1/2$; символ $[\cdot]$ обозначает операцию взятия целой части.

Выборочная медиана и абсолютное медианное отклонение (см. примеры 7 и 8) являются так же робастными L -оценками параметров сдвига и масштаба, соответственно.

Отметим, что t -квантильный размах может быть получен как оптимальная L -оценка масштаба для симметричных распределений при условии, что выбираются две порядковые статистики. При этом оптимизируется значение t .

1.8.5. Алгоритмы, основанные на ранговых статистиках

Если для получения оценок параметров сдвига и масштаба выборки используются ранговые критерии проверки гипотез о равенстве этих параметров в двух выборках: обучающей — x_1, \dots, x_m , рабочей — y_1, \dots, y_n , то такие оценки называют R -оценками.

Оценки параметра сдвига. Пусть обучающая выборка распределена по закону $\omega(x)$, а рабочая выборка — по закону $g(x) = \omega(x - \Delta)$, т.е. распределения выборок различаются параметром сдвига. Для оценки величины сдвига формируют

объединенную выборку и определяют вектор рангов отсчетов рабочей выборки.

R_i — ранг наблюдения y_i в объединенной выборке. Далее формируют статистику рангового критерия для проверки гипотезы о параметре сдвига. Эта статистика имеет вид

$$\delta_n = \sum_{i=1}^n a_i \left(\frac{R_i}{n+m+1} \right), \quad (82)$$

где a_i — весовая функция рангов, вид которой определяется плотностью распределения $\omega(x)$.

Если используется асимптотически оптимальный ранговый критерий, то

$$a(z_i) = -c \frac{\omega_0(x_i)}{\omega_0(x)}, \quad (83)$$

c — некоторый коэффициент; $z_i = R_i/(n+m+1)$;

$x_i = F^{-1}(z_i)$; $F(\cdot)$ — интегральная функция распределения обучающей выборки. Процедура оценивания состоит в подборе величины Δ , преобразовании рабочей выборки $y_1, \dots, y_n \Rightarrow$

$\Rightarrow y_1 - \Delta, \dots, y_n - \Delta$, проверке гипотезы о равенстве параметров сдвига обучающей и преобразованной выборок. Значение $\hat{\Delta}$, при котором гипотеза о равенстве принимается, служит оценкой относительного сдвига. Алгоритм оценивания изображен на рис. 9.

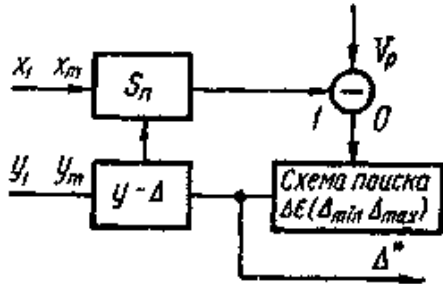


Рис. 9. Алгоритм R -оценивания относительного параметра сдвига

Оценки параметра масштаба. Для получения оценки масштаба одной выборки относительно другой используются ранговые критерии различия в масштабе. Пусть R_i — ранг наблюдения y_i , принадлежащего рабочей выборке, в объединенной выборке объема n . Так же, как и в предыдущем случае, формируется статистика рангового критерия

$$S_n = \sum_{i=1}^n a(R_i/(n+m+1)), \quad (84)$$

где $a(R_i)$ — некоторая функция рангов.

Если используется асимптотически эффективный ранговый критерий масштаба, то функция рангов определяется следующим выражением:

$$a(z_i) = -\frac{\omega'(x_i)}{\omega(x_i)} x_i - 1, \quad (85)$$

где $z_i = R_i/(n+m+1)$; $x_i = F^{-1}(z_i)$; $F(x)$; $\omega(x)$ — интегральная и дифференциальная функции распределения обучающей выборки. Если строятся минимаксные R -оценки, то вместо F и ω следует брать наименее благоприятное распределение F_0 в p -окрестности основного распределения F . Алгоритм оценивания относительного масштаба изображен на рис. 10.

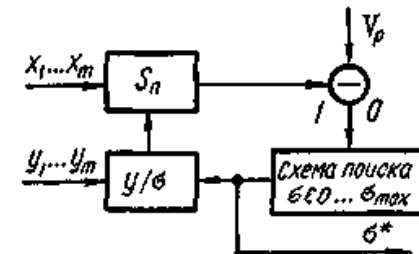


Рис. 10. Алгоритм R -оценивания относительного параметра масштаба

Так же, как и в случае оценивания сдвига, масштаб рабочей выборки ($y_1 \dots y_m$) меняется до тех пор, пока критерий не будет чувствовать различие в масштабе двух выборок.

Для построения R -оценок параметров возможно использование любых известных ранговых критериев (Вилкоксона, Ван-Дер-Вардена и др.), а также минимаксных, синтезированных в соответствии с выражениями (82) — (85).

Пример 14. R -оценка относительного масштаба $\hat{\sigma}$ в задаче с загрязнением нормальных выборок объема n и m . Здесь весовая функция рангов, вычисленная согласно формуле (85),

$$a(z_i) = [\Phi^{-1}(z_i)]^2 - 1,$$

где

$$z_i = R_i/(n+m+1);$$

$\Phi(\cdot)$ — интегральная функция нормального распределения. Построение оценки масштаба состоит в решении уравнения

$$\sum_{i=1}^n a(z_i, \hat{\sigma}) = 0.$$

Для поиска $\hat{\sigma}$, удовлетворяющего этому уравнению, применимы различные процедуры (стохастической аппроксимации, случайного поиска и др.).

Пример 15. R -оценка относительного масштаба σ в задаче с загрязнением выборок из релеевского распределения. Объемы выборок, как и прежде — n и m . Весовая функция рангов, вычисленная по формуле (85), имеет вид

$$a(z_i) = -2 \ln(1 - z_i) - 2; \\ z_i = R_i / (n + m + 1).$$

Построение оценки масштаба так же, как и в предыдущем примере, состоит в решении уравнения

$$\sum_{i=1}^n a(z_i, \hat{\sigma}) = 0.$$

1.8.6. Оценивание с расслоением выборки

Одним из подходов к оцениванию параметров по выборке, описываемой распределением (1), является ее расслоение на две однородные подвыборки или группы, каждая из которых описывается либо плотностью $\omega_1(x, \Theta_1)$, либо плотностью $\omega_2(x, \Theta_2)$ и последующее получение оценок $\hat{\Theta}_1$ и $\hat{\Theta}_2$ одним из методов (максимального правдоподобия, моментов, квантилей, наименьших квадратов) по каждой из групп выборочных отсчетов. Расслоение выборки осуществляется на основании алгоритмов классификации выборочных значений, которые относят i -е выборочное значение либо к генеральной совокупности с распределением ω_1 , либо к генеральной совокупности с распределением ω_2 . Ниже будут рассмотрены некоторые известные и предлагаемые алгоритмы классификации. **Алгоритмы классификации выборочных значений.**

В математической статистике известен ряд подходов к решению задачи классификации при малых значениях p . Это критерии для аномальных наблюдений. Наиболее полная сводка статистик для выделения аномальных наблюдений дана в литературе. Критерии рассчитаны на выделение аномальных наблюдений в выборке, которая в основном описывается нормальным распределением

$$\omega_1(x_1, \Theta) = N(m_1, \sigma_1^2). \tag{86}$$

Выбросы или аномальные значения принадлежат также нормальному распределению, но с другим математическим ожиданием — $N(m_2, \sigma_2^2)$. Рассматриваемые критерии соответствуют таким случаям:

$$1) m_2 > m_1, \sigma_2 = \sigma_1; 2) m_2 < m_1, \sigma_2 = \sigma_1; 3) m_2 > m_1 \vee \\ \vee m_2 < m_1, \sigma_2 = \sigma_1.$$

Рассмотрим случай, когда m_1 и σ_1 неизвестны. Односторонняя статистика, соответствующая обнаружению аномальных значений для случаев 1 и 2, приводится слева; для краткости она записана в форме, подходящей для выявления максимальных аномальных наблюдений. Двухсторонняя статистика, соответствующая случаю 3, приводится справа. Полагаем

$$\bar{x}_n = \sum_{i=1}^{n-1} x_{(i)} / (n - 1), \bar{x}_{n,n-1} = \sum_{i=1}^{n-2} x_{(i)} / (n - 2), S^2 = \\ = \sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1);$$

ω_n — размах.

Таблицы процентных точек статистики приведены в ряде работ;

$$D_1 = (x_{(n)} - \bar{x}_n) / S \quad D_2 = \max |x_i - \bar{x}_n| / S;$$

$$D_3 = \omega_n / S$$

$$D_4 = \frac{n^{1/2} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}}; \quad |D_4|;$$

$$D_5 = \frac{\sum_{i=1}^{n-2} (x_{(i)} - \bar{x}_{n,n-1})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad D_5 = \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2};$$

Для обнаружения аномальных наблюдений применяют r -статистики Диксона, которые представляют собой отношения разностей порядковых статистик, выбранных так, чтобы быть эффективными при следующих условиях.

Проверка одного сомнительного наблюдения $x_{(n)}$

$$r_{10} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(1)}).$$

Проверка одного сомнительного наблюдения $x_{(n)}$, не зависящая от противоположного крайнего наблюдения $x_{(1)}$,

$$r_{11} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(2)}).$$

Проверка одного сомнительного наблюдения $x_{(n)}$, не зависящая от двух крайних наблюдений с противоположного конца,

$$r_{12} = (x_{(n)} - x_{(n-1)}) / (x_{(n)} - x_{(3)}).$$

Проверка сомнительного наблюдения $x_{(n)}$, не зависящая от следующего по величине $x_{(n-1)}$,

$$r_{20} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(1)}).$$

Проверка сомнительного наблюдения $x_{(n)}$, не зависящая от наблюдений $x_{(1)}$ и $x_{(n-1)}$,

$$r_{21} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(2)}).$$

Проверка сомнительного наблюдения $x_{(n)}$, не зависящая от $x_{(n-1)}$, $x_{(1)}$, $x_{(2)}$,

$$r_{22} = (x_{(n)} - x_{(n-2)}) / (x_{(n)} - x_{(3)}).$$

Кроме того, используется статистика Мерфи, которая оптимальна для случая присутствия в выборке k аномальных наблюдений из нормальной генеральной совокупности $N(m + \lambda\sigma_1, \sigma_1^2)$, $x > 0$,

$$D^{(k)} = (x_{(n)} + x_{(n-1)} + \dots + x_{(n-k+1)} - k\bar{x}) / S.$$

Для $k > 1$ никаких таблиц для распределения $D^{(k)}$ нет.

Рассмотрим вопросы оценки качества критериев для классификации выборочных значений.

Для удобства будем считать аномальным значение x_j . Пусть H_1 обозначает соответствующую альтернативную гипотезу (H_0 : в выборке отсутствуют выбросы).

Согласно процедуре классификации выборочное значение x_i , $i = 1, \dots, n$ используется для вычисления статистики критерия V_i . Например, для D_1

$$V_i = (x_i - \bar{x}_n) / S.$$

Обозначая через ϑ_α порог классификации уровня α , укажем следующие возможные характеристики.

1. Вероятность того, что наблюдение x_j классифицировано как выброс:

$$p_1 = p \{V_1 > \vartheta_\alpha / H_1\}.$$

2. Вероятность того, что x_j классифицировано как выброс и при этом является наибольшим в выборке:

$$p_2 = p \{V_1 > \vartheta_\alpha; x_1 > x_2, x_3, \dots, x_n / H_1\}.$$

3. Вероятность того, что только x_j классифицировано как выброс:

$$p_3 = p \{V_1 > \vartheta_\alpha; V_2, V_3, \dots, V_n < \vartheta_\alpha / H_1\}.$$

4. Вероятность того, что x_j классифицировано как выброс; при условии, что x_j — наибольший член выборки:

$$p_4 = p \{V_1 > \vartheta_\alpha / x_1 > x_2, \dots, x_n; H_1\}.$$

б. Мощность критерия

$$p_5 = p \{V > \vartheta_\alpha / H_1\},$$

где

$$V = \max_{1 \leq i \leq n} V_i.$$

Из этих пяти характеристик самая простая с вычислительной точки зрения p_1 ; p_2 и p_3 , наоборот, очень трудны. И хотя p_1 не исключает того, что кроме аномального значения x_j выбросом может быть признано и хорошее наблюдение, эта характеристика оказывается наиболее полезной, так как при малых объемах выборки и высоких уровнях значимости только одна из величин V_1, V_2, \dots, V_n может превышать ϑ_α . В такой ситуации из неравенства $V_1 > \vartheta_\alpha$ вытекает, что $x_1 > x_2 > \dots > x_n$ и $V_2, V_3, \dots, V_n < \vartheta_\alpha$,

$$p_1 = p_2 = p_3. \tag{87}$$

Кроме того, p_1 можно использовать для оценки мощности критерия, так как выполняется следующее неравенство:

$$p_1 \leq p_5 \leq p_1 + (n-1)\beta, \tag{88}$$

где $\beta = p \{V_i > \vartheta_\alpha / H_1\}$.

Следует отметить, что все приведенные критерии применимы для ситуаций, когда в выборке присутствует 1 или 2 аномальных наблюдения, т. е. для малых значений n и p . В случае большой загрязненности выборки сказывается так называемый «маскирующий эффект» и характеристики критериев резко ухудшаются.

Из процедур классификации, заслуживающих внимания, следует отметить критерий Смирнова

$$x_{(n)} > \lambda R, \tag{89}$$

где

$$R = \sqrt{\sum_{i=1}^n x_i^2}, \quad \lambda \in 0 \dots 1.$$

Этот критерий применим для максимальных значений выборки. Критерий (89) допускает модификацию, которая существенно ослабляет «маскирующий» эффект. Модификация состоит в том, что для вычисления R используются первые k порядковых статистик.

Последующие $n - k + 1$ порядковых статистик подвергаются классификации

$$x_{(l)} > \lambda_l R, \quad (90)$$

где $k \leq l \leq n$; $R = \sqrt{\sum_{i=1}^k x_i^2}$.

Значения λ_l , $k \leq l \leq n$ подбираются для обеспечения требуемой вероятности ошибки при классификации l -й порядковой статистики.

Критерий (90) обладает хорошими характеристиками и может успешно применяться в задачах классификации выборочных значений выборки, сформированной согласно модели (1). В случае, когда $\omega_1(x, \Theta)$ — распределение Релея (распределение огибающей нормального процесса) и выполняется гипотеза H_0 , удается получить явные выражения для распределения статистики критерия. Классифицирующая процедура для задачи выделения аномальных значений на фоне выборки с релейским распределением (распределением, имеющим только параметр масштаба) может быть построена на базе вычислителя медианы. Это так называемый медианный классификатор

$$x_{(i)} > \lambda R, \quad \frac{n+1}{2} < i \leq n, \quad (91)$$

где $R = \text{med} \{x_i\}$, $i = \overline{1, n}$.

Заметим, что эта процедура так же, как процедуры (89) и (90), не является инвариантной к параметру сдвига и применима к выборкам, у которых основное распределение — однопараметрическое (параметр масштаба). Следующая классифицирующая процедура использует взвешенную сумму первых m порядковых статистик выборки

$$R = \frac{1}{m} \sum_{i=1}^m a_i x_{(i)}, \quad m < n. \quad (92)$$

Веса a_i вычисляются как величины, обратные квантилям нормированного основного распределения выборки уровня $\frac{n-i+1}{n+1}$, n — объем выборки. Для релейского распределения

$$a_i = \frac{1}{\sqrt{-2 \ln \frac{n-i+1}{n+1}}}. \quad (93)$$

Классификация происходит путем сравнения l -й порядковой статистики выборки с порогом

$$x_{(l)} \geq \lambda R. \quad (94)$$

Значение R вычисляется по формулам (92) и (93).

Заметим, что все, рассмотренные выше классифицирующие статистики, являются робастными оценками параметров основного распределения $\omega_1(x, \Theta_1)$, а процедура классификации представляет собой проверку гипотезы H_0 о принадлежности выборочного значения x_i этому распределению с уровнем значимости, который определяется величиной λ .

Если объем выборки мал, а p достаточно велико ($p > 0,2$), классифицирующие статистики, основанные на робастных оценках параметров основного распределения, становятся неэффективными и приводят к большим ошибкам. В этом случае целесообразно производить расслоение выборки на основе процедур кластерного анализа. Один из возможных алгоритмов заключается в таком разбиении выборки на группы, чтобы сумма дисперсий групп была минимальной. Из выборочных значений образуют вариационный ряд

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

и ведут поиск граничной порядковой статистики, разделяющей две группы.

Обозначим выборочную дисперсию, вычисленную по первым i порядковым статистикам, σ_i^2 , а выборочную дисперсию, вычисленную по следующим $n - i$ порядковым статистикам, $-\sigma_{n-i}^2$. Тогда алгоритм классификации формально можно записать

$$j = \arg \min_{1 \leq i \leq n} \{\sigma_{(i)}^2 + \sigma_{(n-i)}^2\}, \quad (95)$$

где $x_{(l)} \in \Gamma_1$, $l \leq j$; $x_{(l)} \in \Gamma_2$, $l > j$; Γ_1, Γ_2 — генеральные совокупности, описываемые плотностями вероятностей ω_1 и ω_2 соответственно.

Отметим, что при больших значениях $p \in 0,1 \dots 0,9$ алгоритм (95) дает наилучшие результаты среди описанных в данном разделе. При граничных значениях p , близких к нулю или к единице, этот алгоритм уступает другим.

Алгоритмы оценивания. Рассмотрим выборку $\{x_1, \dots, x_n\}$ из составного распределения, описываемого выражением (1). В соответствие вектору выборочных значений поставим вектор параметров $\Xi = (\xi_1, \dots, \xi_n)$ таким образом, что если x_i принадлежит основному распределению с плотностью $\omega_1(x)$, то $\xi_i = 1$, в противном случае $\xi_i = 2$.

Введенные обозначения позволяют при известном векторе параметров записав функцию правдоподобия выборки в виде

$$L(x_1, \dots, x_n) = \prod_{i=1}^n \omega_{\xi_i}(x_i, \Theta_n, \Theta_{\xi_i}). \quad (96)$$

Если сгруппировать члены произведения с $\xi_i = 1$ и $\xi_i = 2$, функция правдоподобия запишется следующим образом:

$$L(x_1, \dots, x_n) = L_1(x_{i1}, \dots, x_{ik}) L_2(x_{i_{k+1}}, \dots, x_{in}) = \prod_{j=1}^k \omega_1(x_{i_j}, \Theta_{1j}, \Theta_1) \prod_{l=k+1}^n \omega_2(x_{i_l}, \Theta_{2l}, \Theta_2), \quad (97)$$

т. е. в виде произведения функций правдоподобия двух подвыборок, одна из которых принадлежит основному распределению, другая — мешающему. Если функцию правдоподобия выборки удастся записать в виде (97), то для оценивания параметров Θ_{1j}, Θ_1 можно пользоваться отсчетами первой подгруппы, отыскивая максимум функции правдоподобия $L_1(\cdot)$, а для оценивания параметров Θ_{2l}, Θ_2 — отсчетами второй подвыборки, отыскивая максимум функции $L_2(\cdot)$. Оценки неизвестных параметров Θ_1 и Θ_2 можно находить по отдельным подвыборкам, пользуясь знанием функций правдоподобия L_1, L_2 и другими методами (моментов, квантилей, M -оценивания, L -оценивания).

В задачах с составной выборкой вектор параметров (ξ_1, \dots, ξ_n) неизвестен. В этом случае требуется оценить его координаты и лишь затем формировать функции правдоподобия L_1 и L_2 .

Таким образом алгоритм оценивания включает две операции.

1. Расслоение выборки на две подгруппы. Для расслоения используются алгоритмы классификации, рассмотренные выше. Эти алгоритмы производят оценку вектора параметров (ξ_1, \dots, ξ_n) .

2. Формирование функций правдоподобия L_1 и L_2 и поиск оценок $\hat{\Theta}_1$ и $\hat{\Theta}_2$ методами разд. 1.8.1—1.8.5.

Пример 16. Плотность распределения выборки (x_1, \dots, x_n) описывается выражением

$$\omega(x) = (1-p) \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} + p \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1-m_2)^2}{2\sigma_1^2}}$$

где σ_1 — известно; m_1, m_2 — неизвестные математические ожидания. Объем выборки $n = 17$. Требуется оценить m_1 при условии

$$m_2 > m_1, \quad p \leq 0,2. \quad (98)$$

Условие (98) позволяет использовать для оценки параметров ξ_1, \dots, ξ_n классифицирующую статистику, основанную на выборочной медиане (91).

Оценим вероятность того, что выборочная медиана не будет принадлежать совокупности выбросов — распределению с математическим ожиданием $m_1 + m_2$. Эта вероятность может быть оценена сверху по асимптотике Пуассона как вероятность появления не более $m = 8$ выбросов в выборке объемом $n = 17$, где вероятность появления выброса в однократном испытании $p = 0,2$:

$$P\{k \leq m\} = P(m, np) \approx \approx 1 - \Phi_1(2\sqrt{np} - 2\sqrt{m+1}) |_{n=17, m=8} \approx 0,99,$$

где k — число выбросов; $\Phi_1(\cdot)$ — интеграл вероятности в форме Лапласа.

Оценка параметра

$$\hat{\xi}_i = \begin{cases} 1, & x_i < \lambda \text{ med } \{x_i\}, \quad i = \overline{1, n}; \\ 2, & x_i > \lambda \text{ med } \{x_i\}, \quad i = \overline{1, n}, \end{cases} \quad (99)$$

где k — некоторое число.

Математическое ожидание m_1 основного распределения оценим по методу максимального правдоподобия, так как вероятность поражения медианы выбросами мала:

$$\hat{m}_1 = \frac{1}{k} \sum_{i:\xi_i=1} x_i, \quad (100)$$

где k — число выборочных значений, отнесенных в результате классификации к распределению ω_1 .

Анализ точности таких составных алгоритмов оценивания представляет сложную задачу. В настоящей работе при анализе будем пользоваться методом статистического моделирования.

1.8.7. Алгоритмы устойчивого оценивания корреляционных функций

Робастное оценивание значений корреляционной функции основано на замене выборочных значений процесса значениями фильтрующей функции

$$\Psi(x, t) = \Psi\left(\frac{x(t) - m_x}{S_x}\right),$$

где $x(t)$ — реализация случайного процесса; m_x — оценка математического ожидания случайного процесса; S_x — оценка масштаба (среднеквадратического отклонения случайного процесса). Оценки m_x и S_x должны быть устойчивыми и могут быть получены одним из способов, описанных в разд. 1.8.6.

Фильтрующая функция $\Psi(x, t)$ удовлетворяет следующим условиям:

$\Psi(x, t)$ сохраняет монотонный порядок значений процесса, т. е. выполнение неравенства $x_1(t_1) > x_2(t_2)$ влечет неравенство $\Psi(x_1, t_1) \geq \Psi(x_2, t_2)$;

$\Psi(x, t)$ — ограниченная функция;

для всякого $a > 0$ и любого b существуют такие $a_1 > 0$ и b_1 , что для всех x выполняется равенство

$$\Psi(ax + b, t) = a_1 \Psi(x, t) + b_1.$$

Значения ковариационной функции вычисляются согласно выражению

$$C(\tau) = \frac{1}{T} \int_0^T \hat{\Psi}(x, t) \hat{\Psi}(x, t - \tau) dt, \quad (101)$$

где $\hat{\Psi}(x, t)$ — центрированная случайная функция.

Робастная оценка корреляционной функции может быть выполнена, если определена дисперсия случайной функции

$$R_c(\tau) = C(\tau)/C(0). \quad (102)$$

М-оценки корреляционных функций. Оценки (102) могут быть оптимизированы для малых значений коэффициентов корреляции. При этом вид фильтрующей функции определяется следующим выражением:

$$\Psi(x) = \omega'(x)/\omega(x), \quad (103)$$

где $\omega'(x)$ — производная по параметру плотности распределения случайного процесса.

Если одномерное распределение вероятностей процесса задано моделью (1), то может быть найдено минимаксное асимптотическое решение задачи оценки корреляционной функции. Это решение дается формулой (103), в которой вместо $\omega(x)$ участвует наименее благоприятное распределение из p -окрестности основного распределения $\omega_1(x)$, минимизирующее информацию Фишера для сдвига (см. разд. 1.8.3). Так, если одномерное распределение случайного процесса является загрязненным, вид фильтрующей функции определяется формулой (49). Отметим, что в общем случае оценки корреляционных функций, получаемые таким образом, являются смещенными.

R-оценки корреляционных функций. Обычно в технических приложениях пользуются отсчетами процесса, взятыми в дискретные моменты времени $x(t_i)$, $i = 1, \dots, n$. Эти отсчеты в совокупности представляют выборку. Если выборочные значения $x_i = x(t_i)$, $i =$

$i = \overline{1, n}$ заменить их рангами R_i в выборке и ввести фильтрующую функцию

$$\Psi(x_i) = a(R_i), \quad (104)$$

где $a(R)$ — монотонная функция, то, используя дискретный аналог формулы (101), в которой $\Psi(x, t)$ определяется выражением (104), получим

$$C(k\tau) = \frac{1}{n-k} \sum_{i=1}^{n-k} \overset{\circ}{a}(R_i) \overset{\circ}{a}(R_{i+k}). \quad (105)$$

Здесь, как и прежде, $\overset{\circ}{a}(\cdot)$ — центрированное значение фильтрующей функции.

Ранговая оценка корреляционной функции вычисляется согласно (51) с учетом выражения (105).

Пример 17. Если положить $a(R_i) = R_i$, получим известный алгоритм оценки ранговом корреляции Спирмена.

Пример 18. Полагая $a(R_i)$ равным математическому ожиданию порядковой статистики с номером R_i выборки объема n , из стандартного нормального распределения $N(0,1)$ получим известный критерий Фишера — Ийтса (критерий нормальных меток).

Отметим, что минимаксная R -оценка коэффициента корреляции в задаче с загрязненным основным распределением (1) получается при выборе $a(R_i)$ в соответствии с формулой (83) для ω_0 , минимизирующей информацию Фишера для сдвига.

1.8.8. Сравнение эффективности алгоритмов оценивания

Эффективность алгоритмов оценивания исследуется в двух задачах.

Задача 1. Рассматривается выборка из загрязненного нормального распределения с известной дисперсией σ_1 и неизвестным математическим ожиданием m_1 . Загрязняющие «выбросы» имеют постоянное и неизвестное значение m_2 .

Одномерное распределение выборочных значений

$$\omega(x) = (1-p) \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} + p \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1-m_2)^2}{2\sigma_1^2}}. \quad (106)$$

Рассмотрим следующие алгоритмы оценивания математического ожидания основного распределения m_1 .

1. Алгоритм, использующий выборочную медиану (пример 7),

$$\hat{m}_1 = \text{med} \{x_i\}, \quad i = \overline{1, n}.$$

2. Итерационная процедура Ньютона — Рафсона для нормального распределения выборки, использующая в качестве начального приближения выборочную медиану

$$\hat{m}_{1i+1} = \hat{m}_{1i} - \left(\frac{\partial \ln L}{\partial m_1} \right)_{m_{1i}} / \left(\frac{\partial^2 \ln L}{\partial m_1^2} \right)_{m_{1i}},$$

которая для гауссовой функции правдоподобия $L(\cdot)$ имеет вид

$$\hat{m}_{1i+1} = \hat{m}_{1i} + \frac{1}{n} \sum_{j=1}^n (x_j - \hat{m}_{1i}),$$

где $\hat{m}_{10} = \text{med} \{x_j\}, j = 1, \dots, n$.

3. Робастная модификация итерационной процедуры Ньютона — Рафсона, синтезированная в предположении наименее благоприятного распределения, описываемого выражением (46). Функция $\Psi(\cdot)$, соответствующая этому распределению, дается формулой (49):

$$\hat{m}_{1i+1} = \hat{m}_{1i} + \frac{1}{l} \sum_{j=1}^l \Psi(x_j - \hat{m}_{1i}),$$

где $\hat{m}_{1i} = \text{med} \{x_j\}; j = 1, \dots, n; l$ — число точек выборки, попавших в интервал $[\hat{m}_{1i} - x_0, \hat{m}_{1i} + x_1]$.

4. Последовательный алгоритм оценивания, использующий процедуру стохастической аппроксимации,

$$\hat{m}_{1i+1} = \hat{m}_{1i} + \frac{1}{l_i} \Psi(x_{i+1} - \hat{m}_{1i}),$$

где l_i — число точек, попавших в интервал $[\hat{m}_{1i} - x_0, \hat{m}_{1i} + x_1]$ на i -м шаге. В качестве первого приближения выбирается первое выборочное значение $\hat{m}_{11} = x_1$.

5. Алгоритм оценивания, использующий расслоение выборки. В качестве классифицирующей статистики выбрана медиана

$$V = \text{med} \{x_i\}, \quad i = \overline{1, n}.$$

Оценка \hat{m}_1 представляет собой среднее по тем выборочным значениям, которые попали в интервал $R = [V - 3\sigma_1, V + 3\sigma_1]$,

$$\hat{m}_1 = \frac{1}{k} \sum_{i: x_i \in R} x_i,$$

где k — число выборочных значений, попавших в интервал R . Для оценки точности моделировалась выборка из распределения (106) с

параметрами $m_1 = 0, m_2 = 10, \sigma_1 = 1$ объемом $n = 16$ производилась оценка параметра m_1 в соответствии с алгоритмами 1—5. Смещение и дисперсия оценок оценивались по числу испытаний $N=200$. Моделировались ситуации с различными значениями p . Результаты приведены в табл. 1.

Таблица 1

1. Результаты моделирования алгоритмов оценивания

p	Алгоритм оценивания*						
	1	2	3	4	5	6	7
<i>Смещение оценки</i>							
0	0,05	0,04	0,04	0,04	0,04	-0,92	0,03
0,1	0,13	1,15	0,45	0,34	0	-0,2	0,68
0,2	0,36	2,01	0,83	1,11	0,05	0	1,51
0,3	1,07	3,05	1,66	1,9	0,2	0,01	2,2
0,4	2,35	3,92	2,8	3,4	0,84	-0,03	2,5
<i>Дисперсия оценки</i>							
0	0,09	0,07	0,07	0,08	0,07	0,46	0,16
0,1	0,12	0,64	0,17	0,17	0,08	0,31	3,97
0,2	0,54	1,08	0,58	1,34	0,08	0,14	8,18
0,3	3,17	1,41	1,88	3,3	1,61	0,1	9,6
0,4	7,99	1,48	5,14	6,1	7,18	0,12	10,35

* Номера алгоритмов оценивания таблицы соответствуют перечню алгоритмов в тексте.

Задача 2. Рассматривается выборка из нормального распределения (106) с неизвестными m_1 и σ_1 загрязненная «выбросами» неизвестной амплитуды m_2 . Априорная информация заключается в том, что $m_2 \gg m_1$.

Рассматриваются следующие алгоритмы оценивания параметра сдвига m_1 основного распределения.

6. Алгоритм оценивания, использующий расслоение выборки по критерию минимума суммы дисперсий (95) и последующее вычисление выборочного среднего по первой выборке.

7. Двухвыборочный алгоритм R оценивания относительного сдвига рабочей выборки $(x_1 \dots x_n)$ и опорной выборки $(y_1 \dots y_n)$. Выборочные отсчеты опорной и рабочей выборок принадлежат загрязненным распределениям (106) с разными $m_1\{y\} \neq m_1\{x\}$. Алгоритм опочиает разность $\Delta = m_1\{y\} - m_1\{x\}$ итерационным способом (см. разд. 1.8.5). Условия моделирования те же, что и в задаче 1.

Результаты оценки точности алгоритмов 6 и 7 содержатся в табл. 1. Данные табл. 1 показывают эффект от использования робастных оценок. Так, при сравнении робастной и неробастной модификаций итерационной процедуры Ньютона — Рафсона (алгоритм 2 и 3) виден существенный выигрыш алгоритма 3 как по смещенности, так и по дисперсии оценки. Еще больший выигрыш в точности достигается при расслоении выборки (алгоритмы 5 и 6), что подтверждает перспективность использования принципа расслоения и задачах статистической обработки данных.

Перспективным при редких выбросах представляется метод оценивания, использующий предварительное обнаружение факта загрязнения выборки на основе алгоритмов (см. разд. 1.8.6). При этом для чистых выборок необходимо использовать классические методы оценивания.

1.9. Устойчивые алгоритмы обнаружения сигналов

1.9.1. Задачи и методы решения устойчивых алгоритмов обнаружения сигналов

В теории синтеза алгоритмов обнаружения особое место занимает лемма Неймана-Пирсона, поскольку параметрические и непараметрические задачи проверки сложных гипотез в ряде случаев удается, как показано ранее, свести (редуцировать) к задачам проверки простой гипотезы против простой альтернативы. Модель наблюдений в данном разделе дается обобщенной ПРВ повторной выборки (1) п. 1.8, а в примерах использованы ПРВ (2) — (4) п. 1.8. Алгоритм («критерий») Неймана — Пирсона (23) п. 1.7 запишем более конкретно для ПРВ (1) п. 1.8 в виде

$$l(\vec{x}_n) = \frac{L(\alpha_0 = 1)}{L(\alpha_0 = 0)} = \frac{\omega(\vec{x}_n/H_1)}{\omega(\vec{x}_n/H_0)} = \frac{\prod_{i=1}^n [(1-p)\omega_1(x_i/\Theta_n, \Theta_1) + p\omega_2(x_i/\Theta_n, \Theta_2)]}{\prod_{i=1}^n [(1-p)\omega_1(x_i/\Theta_1) + p\omega_2(x_i/\Theta_2)]}; \quad (1)$$

$$l(\vec{x}_n) > c, \quad (2)$$

где Θ_n — ожидаемое значение информационного параметра n -мерной ПРВ процесса $x(t)$ при наличии сигнала ($\alpha_0 = 1$).

Алгоритм упрощается, если неравенство (3.2) прологарифмировать

$$\ln [l(\vec{x}_n)] = \sum_{i=1}^n \ln \frac{(1-p)\omega_1(x_i/\Theta_n, \Theta_1) + p\omega_2(x_i/\Theta_n, \Theta_2)}{(1-p)\omega_1(x_i/\Theta_1) + p\omega_2(x_i/\Theta_2)} = \sum_{i=1}^n f_i(x_i/\Theta_n, \Theta_1, \Theta_2, p) \geq \ln c = V_p. \quad (3)$$

Эта формула применяется на практике именно в таком аддитивном виде. Заметим, однако, что алгоритм (3) при неизвестных $\Theta_n, \Theta_1, \Theta_2, p$ не определен, а произвольная подстановка каких-то «ожидаемых» значений параметров при эксплуатации будет иметь следствием ошибку в вероятности ложной тревоги на несколько порядков и, возможно, малую вероятность правильного обнаружения. Рассмотрим это на примере алгоритма (3) при $p = 0$. Пусть

$$\omega_1(x_i/\Theta_n, \Theta_1) = \frac{x_i}{\Theta_1 + \Theta_n} \exp\left[-\frac{x_i^2}{2(\Theta_1 + \Theta_n)}\right],$$

что соответствует случаю обнаружения гауссова сигнала с известной дисперсией $\sigma_w^2 = \Theta_n$ на фоне гауссова шума с известной дисперсией $\sigma_w^2 = \Theta_1$ по наблюдаемым отсчетам \vec{x}_n огибающей смеси сигналов и помех. Алгоритм (3) после эквивалентных преобразований принимает вид:

$$z = f(\vec{x}_n) = \sum_{i=1}^n x_i^2 \geq V_n,$$

где $V_n = -2n(\sigma_w^2 + \sigma_c^2) \ln \frac{\sigma_c^2}{\sigma_w^2 + \sigma_c^2}$ — порог решения, при превышении которого суммой квадратов отсчетов выборки фиксируется $\hat{\alpha}_0 = 1$, т. е. принимается решение о наличии в реализации $x(t)$, наблюдаемой на интервале $(0, T) = \Delta t n$, полезного сигнала $S(t)$, где Δt — интервал дискретизации по времени, а отсчеты $x_i, x_j (i \neq j)$ — независимы (по условию). Устойчивость алгоритма оценим для $n = 1$. Тогда, учитывая, что неравенства $z = x_1^2 > V_{n=1}$ и $x_1 > (V_{n=1})^{0.5}$ эквивалентны, а ПРВ x_1 — релеевская, запишем вероятности ложной тревоги и правильного обнаружения, используя их определения по формулам (17) п.1.7, (18) п.1.7:

$$D = D(\Theta = \sigma_w^2, b = \sigma_c^2/\sigma_w^2, f(x_i) = x_i, n = 1) = \exp\left[-\frac{V^2}{2\sigma_w^2(1+b)}\right]; \quad (4)$$

$$F = D(\Theta = \sigma_{ш}^2, b = 0, f(x_i) = x_i, m = 1) = \exp\left[-\frac{V^2}{2\sigma_{ш}^2}\right]. \quad (5)$$

Пусть при ожидаемой дисперсии шума $\sigma_{ш}^2$ требуется обеспечить вероятность ложной тревоги $F_{\text{треб}} = 10^{-8}$. Из уравнения (5) найдем соответствующее этим условиям значение порога обнаружения $V_{\text{треб}} = (2\sigma_{ш}^2 \ln F_{\text{треб}}^{-1})^{0,5} = 5,678$. Из уравнения (4) по заданному $D_{\text{треб}}$, равному, например, 0,99, можно найти пороговое соотношение сигнал/шум $b_{\text{п}}$. Таким образом, фиксируется набор показателей эффективности алгоритма обнаружения $F_{\text{треб}}$, $D_{\text{треб}}$, $b_{\text{п}}$, соответствующий известному параметру $\sigma_{ш}^2 = 1$. Если, однако, $\sigma_{ш}^2$ изменяется, например, из-за нестабильности коэффициента усиления приемника или внешних причин и достигает значения $\sigma_{ш}^2 = 2$, то по формуле (5) легко оценить новое значение вероятности ложной тревоги:

$$F(\sigma_{ш}^2 = 2) = \exp\left[-\frac{(5,678)^2}{2 \cdot 4}\right] = 1,78 \cdot 10^{-2},$$

т. е. вероятность ложной тревоги увеличилась более, чем на 6 порядков. Такой перепад вероятности ложных решений эквивалентен такому же увеличению интенсивности потока ложных решений, перегрузке памяти, устройств регистрации и отображения информации и полной потере работоспособности ИИС.

Особенно неустойчивы алгоритмы Неймана — Пирсона к аномальным ошибкам, определяемым при $p \neq 0$ вторым членом составных ПРВ в формуле (3).

Применяют двухвыборочные алгоритмы обнаружения, использующие обучающую \vec{x}_m и рабочую \vec{x}_n повторные выборки. Рассмотрим две структурно отличающиеся модификации алгоритмов. Первая подобна структуре алгоритма Неймана — Пирсона (3):

$$z = \sum_{i=1}^n \tilde{\varphi}_i(x_i/\hat{\Theta}_u, \hat{\Theta}) \geq V, \quad (6a)$$

где $\hat{\Theta} = (\hat{\Theta}_1, \hat{\Theta}_2, \hat{\rho})$, или, учитывая, что общий вектор помех $\hat{\Theta} = \hat{\Theta}(\vec{x}_{n+m})$, в другой более общей записи:

$$z = \sum_{i=1}^n \varphi_i(x_i/\vec{x}_{n+m}) \geq V. \quad (6b)$$

Функции $\tilde{\varphi}_i(\cdot)$, $\varphi_i(\cdot)$ назовем предварительными проверочными статистиками, синтез (выбор) которых по критериям стабильности F и максимума D и является основной задачей в параметрических, смешанных параметрически-непараметрических (в формуле (1) п.1.8 p

мало, а $\omega_2(\cdot)$ — неизвестно) и полностью непараметрических сигнально-помеховых ситуациях. Вторая модификация алгоритмов отличается от первой введением подалгоритма обнаружения аномальных ошибок в \vec{x}_n и \vec{x}_m с выбором одного из S подалгоритмов проверки гипотез и объединением S решений по схеме логического ИЛИ

$$z = \sum_{k=1}^s \eta_k \left[\sum_{i=1}^n \tilde{\varphi}_{ki}(x_i/\hat{\Theta}_u, \hat{\Theta}) \right] \geq V, \quad (7a)$$

или по аналогии с записью (6б)

$$z = \sum_{k=1}^s \eta_k \left[\sum_{i=1}^n \varphi_{ki}(x_i/\vec{x}_{n+m}) \right] \geq V, \quad (7b)$$

где $\eta_k = \eta_k(\vec{x}_m, \vec{x}_n)$ — индикаторная функция классификатора помеховых ситуаций, $\eta_k = \begin{cases} 1; \\ 0; \end{cases} \sum_{k=1}^s \eta_k = 1$, обнуляющая вклад

всех k -х «каналов» $\varphi_k = \sum_{i=1}^n \varphi_{ki}(x_i/\vec{x}_{n+m})$, кроме канала, который условно эффективен для реализовавшейся помеховой ситуации. Для простейшего случая $S = 4$, что соответствует следующим помеховым ситуациям $S_1 - S_4$: в \vec{x}_m и \vec{x}_n отсутствуют аномальные ошибки (S_1); ошибки есть или только в \vec{x}_m , или только в \vec{x}_n (ситуации S_2, S_3); ошибки есть как в \vec{x}_m , так и в \vec{x}_n (ситуация S_4). Вероятности этих ситуаций легко вычислить

$$\left. \begin{aligned} p(S_1) &= (1-p)^{m+n}; \\ p(S_2) &= [1 - (1-p)^m](1-p)^n; \\ p(S_3) &= [1 - (1-p)^n](1-p)^m; \\ p(S_4) &= [1 - (1-p)^m] - (1-p)^n - (1-p)^{n+m}. \end{aligned} \right\} \quad (8)$$

Для оценки показателей качества функционирования алгоритма (7) для $S=4$ рассмотрим логику принятия решений о значении индикаторной функции η_j ($j = 1, 4$) по результатам обнаружения выбросов отдельно в обучающей и рабочей выборках. Если эти решения обозначить двоичными символами η_{m1} и η_{n1} , полагая, что $\eta_{m1} = \eta_{n1} = 1$ при обнаружении выбросов в соответствующих выборках, то, очевидно, что

$$\eta_1 = \bar{\eta}_m \bar{\eta}_n; \quad \eta_2 = \eta_m \bar{\eta}_n; \quad \eta_3 = \bar{\eta}_m \eta_n; \quad \eta_4 = \eta_m \eta_n. \quad (9)$$

В качестве решающей функции для проверки гипотез о наличии аномальных ошибок иногда используют выборочные коэффициенты асимметрии и эксцесса, применяют критерий Аббе и др. (см. разд. 1.8.6). Представляют интерес составные алгоритмы, использующие групповое принятие решений по нескольким из указанных критериев с объединением индивидуальных решений по схеме голосования типа «хотя бы два из трех». Целесообразность использования, например, выборочного коэффициента асимметрии \hat{C}_3 для обнаружения аномальных ошибок в выборках из нормальных генеральных совокупностей следует из доказанной Р. Фишером несмещенности \hat{C}_3 при любом объеме выборки K и зависимости дисперсии оценки только от объема выборки:

$$\mu_2 \{\hat{C}_3\} = 6(K-2)(K-1)^{-1}(K-3)^{-1}. \quad (10)$$

Хотя ПРВ \hat{C}_3 негауссово даже асимптотически, формулой (10) можно воспользоваться для вычисления первого приближения порогов решения обнаружителей аномальных ошибок. Для вероятности ложной тревоги порядка 10^{-2} порог решения $V = 2,33 \times \mu_2 \{\hat{C}_3\}^{1/2}$. При таком V действительная вероятность окажется несколько больше ожидаемого значения 10^{-2} и потребуются корректировка порога в сторону увеличения.

Вводя обозначения для вероятностей ложной и правильной классификации F_m, F_n, D_m, D_n выборки \vec{x}_m и \vec{x}_n по признаку наличия в них аномальных ошибок, рассчитываем условные вероятности $p_{kj} = p\{\eta_j | S_k\}$ индикации j -й ситуации (что соответствует включению и отработке j -го подалгоритма $\varphi_j = \sum_{i=1}^n \varphi_{ij}(x_i | \vec{x}_{m+n})$ и формуле (7)) при реализации k -й ситуации S_k :

$$\begin{aligned} p_{11} &= (1 - F_m)(1 - F_n); & p_{12} &= F_m(1 - F_n); \\ p_{13} &= (1 - F_m)F_n; & p_{14} &= F_mF_n; \\ p_{21} &= (1 - D_m)(1 - F_n); & p_{22} &= D_m(1 - F_n); \\ p_{23} &= (1 - D_m)F_n; & p_{24} &= D_mF_n; \\ p_{31} &= (1 - D_m)(1 - D_n); & p_{32} &= F_m(1 - D_n); \end{aligned} \quad (11)$$

$$\begin{aligned} p_{33} &= (1 - F_m)D_n; & p_{34} &= F_mD_n; \\ p_{41} &= (1 - D_m)(1 - D_n); & p_{42} &= D_m(1 - D_n); \\ D_{43} &= (1 - D_m)D_n; & p_{44} &= D_mD_n. \end{aligned}$$

Элементы в формуле (11) образуют стохастическую матрицу, так как $\sum_{i=1}^4 p_{ki} = 1$. Соотношения (8), (11) позволяют на основании формулы полной вероятности записать искомые показатели качества функционирования алгоритма обнаружителя сигналов $f^{(3,7)}$ (7а, 7б) для модели входных данных с аномальными ошибками вида (1) п.1.8:

$$\left. \begin{aligned} F(\Theta, f^{(3,7)}, n+m) &= \\ &= \sum_{k=1}^4 p(S_k) \sum_{j=1}^4 p_{kj} F_{kj}(\Theta_k^{(n+m)}, \varphi_j, n+m, V_k); \\ D(\Theta, b, f^{(3,7)}, n+m) &= \\ &= \sum_{k=1}^4 p(S_k) \sum_{j=1}^4 p_{kj} D_{kj}(\Theta_k^{(n+m)}, b, \varphi_j, n+m, V_k), \end{aligned} \right\} \quad (12)$$

где $\Theta_k^{(n+m)}$ — составной вектор параметров помех обучающей и рабочей выборки:

$$\begin{aligned} \Theta_1^{(n+m)} &= (\Theta_1^{(m)}, \Theta_1^{(n)}); & \Theta_2^{(n+m)} &= (\Theta_1^{(m)}, \Theta_2^{(m)}, p^{(m)}, \Theta_1^{(n)}); \\ \Theta_3^{(n+m)} &= (\Theta_2^{(m)}, \Theta_1^{(n)}, \Theta_2^{(n)}, p^{(n)}); \\ \Theta_4^{(n+m)} &= (\Theta_1^{(m)}, \Theta_2^{(m)}, p^{(m)}; \Theta_1^{(n)}, \Theta_2^{(n)}, p^{(n)}); \end{aligned}$$

V_k — пороги обнаружения, с которыми сравнивается статистика Φ_k , в общем случае различные; b — варьируемый сдвиг информационного параметра сигнала или отношения сигнал/помеха при гипотезе H_1 (иногда b — вектор).

Алгоритм (6) более прост, так как не требует классификатора ситуаций, но в ряде случаев менее эффективен. В параметрических задачах сходство алгоритмов (6) и (3) не только структурное, если в алгоритме (3) заменить в соответствии с эмпирическим байесовым подходом неизвестные параметры сигнала и помех их выборочными оценками. Такая замена определяет вид $\varphi_i(\cdot) = f_i(\cdot)$, если m, n — велики, а оценки Θ_1, Θ_n несмещены и состоятельны.

В более общем случае конечных m, n к структуре (6а... 7б) удается привести широкий класс известных и новых алгоритмов. Используя модификации и обобщения ЭВМ, синтезируем предварительные проверочные статистики $\varphi_i(\cdot)$ в формулах (6а), (7а). Для систематизации методов проверки статистических гипотез непараметрической

статистики удобно использовать запись алгоритмов обнаружения в виде (6б), (7б). Отметим, что в ряде прикладных задач удастся следующим образом организовать процедуру обучения: на каждый отсчет x_i рабочей выборки формируется вектор \vec{x}_{mi} обучающей выборки объемом m_i . Тогда

$$m = \sum_{i=1}^n m_i.$$

1.9.2. Классификация и методы синтеза параметрических (адаптивных) алгоритмов

Структура параметрических алгоритмов обнаружения унифицируется соотношениями (6а), (7а), а их многообразие и соответственно классификация определяется допустимым множеством алгоритмов формирования оценок $\hat{\Theta}, \hat{\Theta}_n$, регулярным методом синтеза или выбором алгоритмов формирования (вычисления)

предварительных проверочных статистик (ППС) $\varphi_i(x_i/\hat{\Theta}_n, \hat{\Theta})$ в формуле (6а) или $\tilde{\varphi}_i(x_i/\hat{\Theta}_n, \Theta)$ в формуле (7а), особенностями аппроксимаций точных алгоритмов. Пример типовой структурной схемы параметрического алгоритма обнаружения, реализующей структуру (7а), показан на рис. 11.

В соответствии с первым признаком классификации выделяют алгоритмы обнаружения, использующие метод моментов, квантилей, максимального правдоподобия, M -оценок, L -оценок, включая в общем случае и непараметрические алгоритмы оценивания. Целесообразность применения последних в задачах проверки сложных параметрических гипотез может обуславливаться, например, простотой алгоритма, несмещенностью оценок в большом диапазоне изменения параметров и т. п.

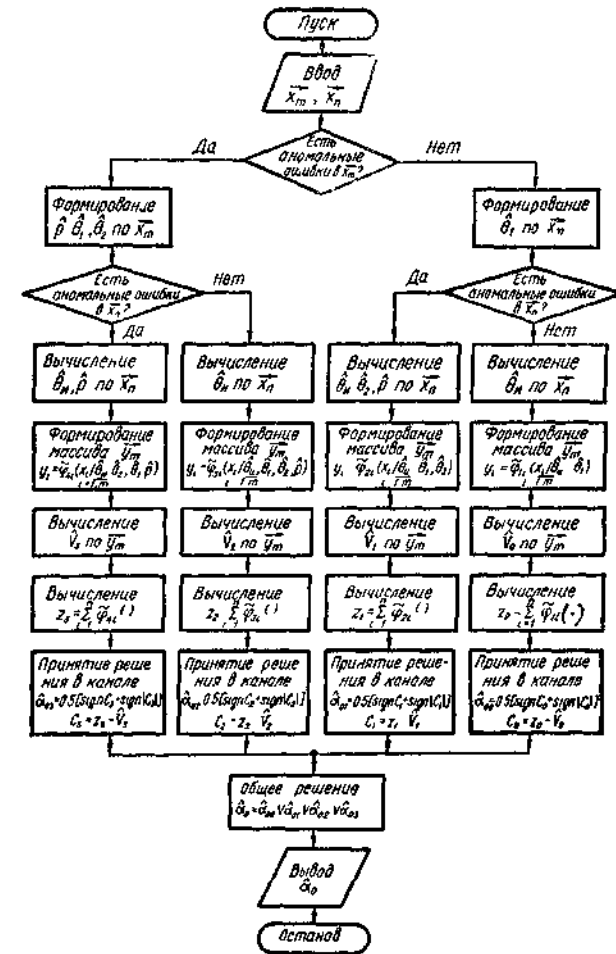


Рис. 11. Структурная схема параметрического алгоритма обнаружения сигналов, использующая подалгоритмы обнаружения аномальных отсчетов в обучающей и рабочей выборках

Перспективно применение комбинированных алгоритмов оценивания $\hat{\Theta}, \hat{\Theta}_n$. В соответствии со вторым признаком классификации алгоритмов указывают метод получения или конкретный вид функции ППС $\varphi_i(x_i/\hat{\Theta}_n, \hat{\Theta})$ в формуле (6а) или

$\bar{\varphi}(x_i/\hat{\Theta}_n, \hat{\Theta})$ в формуле (7а), в существенной степени определяющей эффективность алгоритма.

Операция быстрого арифметического суммирования в «скользящем окне» в соотношениях (6а), (7а) при скорости поступления данных более $10^6 \dots 10^7$ байт/с и больших n ($2^5 \dots 2^7$) требует для реализации больших аппаратных затрат. Поэтому в ряде случаев приходится использовать алгоритмы приближенного суммирования с контролем допустимой величины потерь в пороговом соотношении сигнал/помеха. Иногда особенности этих алгоритмов переносятся на названия алгоритмов обнаружения в целом и их можно отнести к третьему признаку классификации.

Алгоритмы формирования ППС можно получить следующими методами:

эвристик, как например, в случае классического критерия согласия χ^2 в его адаптивной модификации на основе применения метода максимального правдоподобия оценок неизвестных параметров по рабочей выборке в режиме самообучения;

на основе известного в статистике теста «обобщенного отношения правдоподобия» (в ряде работ просто «отношение правдоподобия» (ОП)), использующего оценки максимального правдоподобия и являющегося фактически разновидностью ЭБМ; применяют ЭБМ на основе формулы (3) при использовании других видов оценок (моментов, квантилей, M -оценок, включая робастные и другие непараметрические);

модифицированного ЭБМ (МЭБМ) на основе поиска таких ПП $S\varphi_i(\cdot)$, $\varphi_{ki}(\cdot)$, ПРВ которых не содержит «утяжеленных хвостов», обусловленных ошибками в оценке параметров $\hat{\Theta}$, $\hat{\Theta}_n$;

обобщенного ЭБМ (ОЭБМ), в частности, на основе эмпирического обобщения принципа редукции Вальда (24) п.1.7.

Рассмотрим особенности некоторых из этих методов.

Эмпирический байесовский метод. В эвристических критериях, например, χ^2 и при использовании ЭБМ вид ППС задан однозначно выбором или статистикой (3) и задача синтеза сводится лишь к аппроксимации l в статистике (3). Отмеченные алгоритмы χ^2 , ОП и другие алгоритмы ЭБМ имеют тот общий недостаток, что при низком качестве оценок параметров $\hat{\Theta}_n$ и $\hat{\Theta}$ в задачах с аномальными ошибками их эффективность низка.

Недостаток критерия согласия типа χ^2 — малая мощность (малы D при заданных F), что является следствием его универсальности, т. е. работоспособности в задачах обнаружения любых деформаций $\omega(x_i/H_1)$ по сравнению с $\omega(x_i/H_0)$. Метод ОП для гауссовых ПРВ

позволил получить замечательные результаты (тесты Стьюдента, Хотеллинга и др.). Для рассматриваемых задач с аномальными ошибками в случае ПРВ (2) п.1.8 следует ожидать достаточной эффективности алгоритмов ОП, если применить численный метод решения систем уравнения максимального правдоподобия. Однако для конечных объемов выборки n , m и произвольных ПРВ вида (1) п.1.8 свойства статистики ОП не изучены, достаточная мощность D не гарантируется какими-то общими утверждениями, а в вычислительном плане алгоритмы оценки $\hat{\Theta}_n$, $\hat{\Theta}$ по максимальному правдоподобию ресурсоемки. Асимптотически или уже при n , $m > (1-2) \cdot 10^2$ применение ЭБМ оправдано, если гарантируется высокое качество оценок, по любым из изложенных ранее методов, включая комбинированный. Предположим, что известен алгоритм обработки сигналов, полученный согласно лемме Неймана — Пирсона и определяемый отношением функций правдоподобия

$$l(x, \Theta) = \frac{L(x, \Theta/H_1)}{L(x, \Theta/H_0)}, \quad (13)$$

где Θ — вектор параметров, значения компонент которого, в общем случае, неизвестны. ЭБМ использует структуру алгоритма обработки (13), заменяя неизвестные параметры в их оценках, полученными одним из методов, изложенных ранее.

В качестве примера применения ЭБМ для синтеза алгоритма рассмотрим задачу обнаружения шумоподобного сигнала на выходе линейного детектора огибающей при воздействии внутриприемного шума с неизвестной дисперсией $\sigma^2 = \Psi$ и хаотической импульсной помехи (ХИП) с неизвестными амплитудой u и вероятностью появления p .

Плотность распределения отсчетов огибающей смеси сигнала, внутриприемного шума и ХИП имеет вид (гипотеза H_1):

$$\omega(x/\Psi_c, u, p, \Psi) = \frac{(1-p)x}{\Psi + \Psi_c} e^{-\frac{x^2}{2(\Psi + \Psi_c)}} + \frac{px}{\Psi + \Psi_c} e^{-\frac{x^2+u^2}{2(\Psi + \Psi_c)}} I_0\left(\frac{xu}{\Psi + \Psi_c}\right), \quad (14)$$

где Ψ_c — дисперсия сигнала. Плотность распределения огибающей смеси для случая отсутствия сигнала (гипотеза H_0) определяется формулой (14) при $\Psi_c = 0$. Алгоритм обработки при известных параметрах Ψ , Ψ_c , u , p определяется логарифмом отношения функций правдоподобия для конкурирующих гипотез H_1 и H_0 :

$$\ln l(x/\Psi_c, u, p, \Psi) = \frac{\Psi_c x^2}{2\Psi(\Psi + \Psi_c)} + \ln \frac{1 - p + p \exp\left[-\frac{u^2}{2(\Psi + \Psi_c)} + \ln I_0\left(\frac{xu}{\Psi}\right)\right]}{1 - p \exp\left[-\frac{u^2}{2\Psi} + \ln I_0\left(\frac{xu}{\Psi}\right)\right]} > V. \quad (15)$$

Аппроксимация алгоритма (15):

$$\ln l(x/\Psi_c, u, p, \Psi) \approx y = \begin{cases} cx^2, & x \leq b, & c = \frac{\Psi_c}{2\Psi(\Psi + \Psi_c)}; \\ cx^2 - b(x - b), & x > b, & b = u/2. \end{cases} \quad (16)$$

Вид нелинейного функционального преобразования изображен на рис. 12.

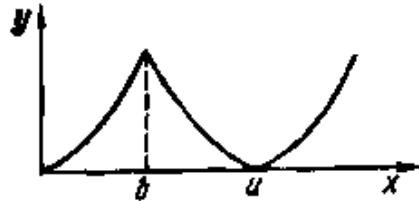


Рис. 12. Характеристика нелинейного преобразователя алгоритма обнаружителя сигналов на фоне гауссова шума и хаотической импульсной помехи

При большой амплитуде ХИП ($u \gg \sqrt{\Psi_c + \Psi}$) вид алгоритма (15) практически не зависит от величины p на всем интервале изменения кроме окрестности крайних точек $p \approx 0$ и $p \approx 1$ и аппроксимация (16) является вполне рабочей.

Для использования аппроксимации алгоритма (16) в ситуации о неизвестной амплитудой помехи u надо, согласно ЭБМ, в отношении правдоподобия (15) либо его аппроксимацию (16) ввести оценки амплитуды u , дисперсий Ψ_c, Ψ .

Для синтеза алгоритма оценивания u воспользуемся методом моментов. Рассмотрим систему трех уравнений относительно трех параметров смеси внутриприемного шума и импульсной помехи с плотностью вероятностей (14) при аппроксимации распределения Раиса нормальным законом с математическим ожиданием u и дисперсией $\Psi_p = \Psi + \Psi_c$:

$$\begin{aligned} f_1(p, u, \Psi_p) &= \hat{m}_1\{x\} - (1 - p) \sqrt{\frac{\pi\Psi_p}{2}} - pu = 0; \\ f_2(p, u, \Psi_p) &= \hat{m}_2\{x\} - (1 - p) 2\Psi_p - p(u^2 + \Psi_p) = 0; \\ f_3(p, u, \Psi_p) &= \hat{m}_4\{x\} - 8\Psi_p^2 - p(u^4 - 5\Psi_p^2 + 6\Psi_p u^2) = 0. \end{aligned} \quad (17)$$

Здесь $\hat{m}_1\{x\}, \hat{m}_2\{x\}, \hat{m}_4\{x\}$ — 1-й, 2-й и 4-й выборочные начальные моменты.

Систему (17) можно решить одним из численных методов, например методом Ньютона. Начальное приближение можно получить исходя из априорных сведений о значении p . При заданном начальном значении $p=p_0$ из первых двух уравнений легко определяются начальные значения параметров u_0 и Ψ_0 .

Другой метод заключается в решении системы из двух первых уравнений для нескольких значений $p_{01}, p_{02}, \dots, p_{0k}$. В качестве начального приближения выбираются такие $p_{0j}, u_{0j}, \Psi_{0j}$, при которых выполняется неравенство

$$\begin{aligned} f_1^2(p_{0j}, u_{0j}, \Psi_{0j}) + f_2^2(p_{0j}, u_{0j}, \Psi_{0j}) + f_3^2(p_{0j}, u_{0j}, \Psi_{0j}) &\leq \\ &\leq f_1^2(p_{0l}, u_{0l}, \Psi_{0l}) + f_2^2(p_{0l}, u_{0l}, \Psi_{0l}) + \\ &+ f_3^2(p_{0l}, u_{0l}, \Psi_{0l}), \quad l \in 1, \dots, k. \end{aligned} \quad (18)$$

В табл. 2 приведены результаты моделирования алгоритма оценки трех параметров (17) при заданных $u = 5; p = 0,2; 0,4; \Psi = 1$ при двух объемах выборки $n = 32, 100$.

Таблица 2

2. Результаты моделирования алгоритма оценки трех параметров

p	$\hat{m}_1(p)$	$m_1(\hat{u})$	$m_1(\hat{\Psi})$	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(\hat{u})$	$\hat{\sigma}^2(\hat{\Psi})$
$n = 32, \Psi = 1, u = 5$						
0,2	0,187	5,28	1,05	0,006	0,304	0,057
0,4	0,473	4,85	0,877	0,027	0,437	0,258
$n = 100, \Psi = 1, u = 5$						
0,2	0,179	5,38	1,05	0,0014	0,09	0,057
0,4	0,523	4,67	0,701	0,0244	0,375	0,196

Использование алгоритма (16) совместно с алгоритмом оценки неизвестных параметров помех определяет адаптивный алгоритм обнаружения шумоподобного сигнала на фоне внутриприемного шума и ХИП с неизвестной вероятностью воздействия и неизвестной амплитудой. Эффективность составного алгоритма (16) — (18) оценивалась методом статистического моделирования. Характеристики обнаружения и условия моделирования показаны на рис. 13.

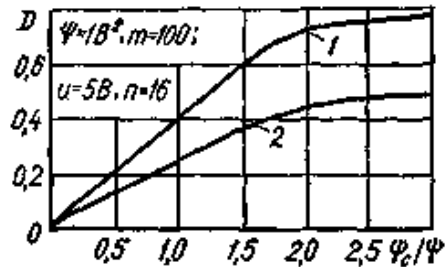


Рис 13. Характеристики адаптивного обнаружителя сигналов на фоне гауссова шума и импульсных помех для $p = 0,4$ (1), $p = 0,2$ (2)

Здесь же приведены потенциально достижимые характеристики обнаружения для алгоритма (16) в условиях априорной определенности параметров помех. Результаты анализа свидетельствуют о работоспособности рассмотренного алгоритма обнаружения. Следует отметить, что в ситуации когда средне-квадратическое отклонение смеси сигнала и шума соизмеримо с амплитудой импульсной помехи (большие соотношения сигнал/шум), эффективность алгоритма снижается. Это объясняется тем, что аппроксимация (15) в этих условиях не является удовлетворительной и структура алгоритма далека от оптимальной.

Модифицированный ЭБМ. Модификация ЭБМ состоит в том, что при синтезе используются сведения о плотности распределения оценок неизвестных параметров. Структура алгоритма обработки определяется при этом следующим образом.

Ищутся плотности распределения статистики (13) с учетом распределения оценок параметров $\hat{\Theta}$ в области гипотез H_0 и $H_1 \rightarrow \omega[l(x, \hat{\Theta})/H_0]$ и $\omega[l(x, \hat{\Theta})/H_1]$ и вычисляется отношение

$$\varphi(l) = \frac{\omega[l(x, \hat{\Theta})/H_1]}{\omega[l(x, \hat{\Theta})/H_0]}, \quad (19)$$

которое определяет структуру дополнительного функционального преобразователя. В случае когда используются оценки с «плохой» функцией распределения, имеющей «тяжелые хвосты», обусловленные ошибками оценивания, функциональный преобразователь (19) имеет нелинейную, а часто и немонотонную характеристику и его применение необходимо для обеспечения устойчивости характеристик алгоритма обработки сигналов. Решение выносится по сравнению статистики (19) с порогом решения.

Заметим, что принцип инвариантности, широко применяемый в задачах с априорной неопределенностью, использует этот подход для отыскания оптимальных решающих правил в классе статистик, основанных на максимальных инвариантах. В качестве $l(\cdot)$ берется максимальный инвариант к группе преобразований выборочного пространства, описывающей априорную неопределенность. На этапе выбора максимального инварианта присутствует большая доля эвристики, что лишает задачу однозначности. Кроме того, не всегда априорную неопределенность можно описать в терминах групповых преобразований выборочного пространства либо пространства параметров.

Например, в задаче обнаружения при воздействии импульсной помехи с неизвестной интенсивностью, рассмотренной выше, неопределенность амплитуды помехи не описывается групповыми преобразованиями.

Специфику использования модифицированного эмпирического байесовского подхода рассмотрим на примере обнаружения детерминированного сигнала с амплитудой u на фоне гауссовской помехи с дисперсией Ψ и неизвестной постоянной составляющей, которую обозначим u_n . Будем решать задачу в условиях, когда амплитуда помехи оценивается и в процессе измерений возможны грубые систематические ошибки, т. е. с конечной вероятностью p оценка амплитуды помехи \hat{u}_n смещена на величину c .

Плотности распределения отсчетов, соответствующих отсутствию и наличию сигнала, для гипотез H_0 и H_1 следующие:

$$\left. \begin{aligned} \omega(x) = \omega_n(x) &= \frac{1}{\sqrt{2\pi\Psi}} e^{-\frac{(x-u_n)^2}{2\Psi}}; \\ \omega(x) = \omega_{c+n}(x) &= \frac{1}{\sqrt{2\pi\Psi}} e^{-\frac{(x-u_n-c)^2}{2\Psi}}. \end{aligned} \right\} \quad (20)$$

Плотность распределения оценки неизвестной амплитуды помехи имеет вид

$$\omega_{\hat{u}_n}(\hat{u}_n/u_n) = \frac{1-p}{\sqrt{2\pi\Psi_{\hat{u}}}} e^{-\frac{(\hat{u}_n-u_n)^2}{2\Psi_{\hat{u}}}} + \frac{p}{\sqrt{2\pi\Psi_{\hat{u}}}} e^{-\frac{(\hat{u}_n-u_n-c)^2}{2\Psi_{\hat{u}}}}, \quad (21)$$

где $\Psi_{\hat{u}}$ — дисперсия оценки.

Алгоритм проверки гипотезы H_1 против альтернативы H_0 при известных параметрах Ψ, u, u_n , как известно, определяется выражением

$$\ln l(x, \Psi, u, u_n) = x - u_n > V(\Psi, u), \quad (22)$$

где $V(\cdot)$ — порог решения, зависящий от известных параметров Ψ и u .

Согласно ЭБМ, алгоритм обнаружения, использующий оценку амплитуды \hat{u}_n , определяется формулой (22) с заменой u_n на \hat{u}_n , т. е.

$$l' = x - \hat{u}_n. \quad (23)$$

На основании МЭБМ найдем другой алгоритм. Плотности распределения вероятностей статистики (23) для гипотез H_0 и H_1 запишутся соответственно

$$\omega(l'/H_0) = \int_{\hat{u}_n} \omega_{\hat{u}_n}(y) \omega_n(y+l') dy, \quad (24)$$

$$\omega(l'/H_1) = \int \omega_{\hat{u}_n}(y) \omega_{c+n}(y+l') dy.$$

Подставляя выражения (20) и (21) в (24), получаем

$$\omega(l'/H_0) = (1-p) \frac{\exp\left[-\frac{l'^2}{2(\Psi_{\hat{u}}+\Psi)}\right]}{\sqrt{2\pi(\Psi_{\hat{u}}+\Psi)}} + p \frac{\exp\left[-\frac{(l'+C)^2}{2(\Psi_{\hat{u}}+\Psi)}\right]}{\sqrt{2\pi(\Psi_{\hat{u}}+\Psi)}}; \quad (25)$$

$$\omega(l'/H_1) = (1-p) \frac{\exp\left[-\frac{(l'-u)^2}{2(\Psi_{\hat{u}}+\Psi)}\right]}{\sqrt{2\pi(\Psi_{\hat{u}}+\Psi)}} + p \frac{\exp\left[-\frac{(l'+C-u)^2}{2(\Psi_{\hat{u}}+\Psi)}\right]}{\sqrt{2\pi(\Psi_{\hat{u}}+\Psi)}}. \quad (26)$$

Алгоритм обработки $\varphi(l')$ определяется логарифмом отношения функций правдоподобия (25) и (26). Графическое изображение характеристики нелинейного преобразователя $q>(l')$ для $c \gg u + \sqrt{\Psi_{\hat{u}} + \Psi}$ показано на рис. 14.

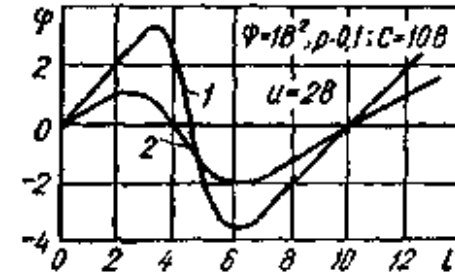


Рис. 14. Характеристика нелинейного преобразователя адаптивного алгоритма обнаружения детерминированного сигнала на фоне гауссовой помехи с неизвестными средним и дисперсией при аномальных ошибках в канале оценки (синтез по МЭБМ) для $\Psi_{\hat{u}} = 1B^2$ (1); $\Psi_{\hat{u}} = 3B^2$ (2)

Как видно из рис. 14, нелинейный элемент компенсирует ошибки измерения амплитуды помехи, увеличивая тем самым устойчивость характеристик качества обнаружителя.

Характеристики обнаружения алгоритмов изображены на рис. 15.

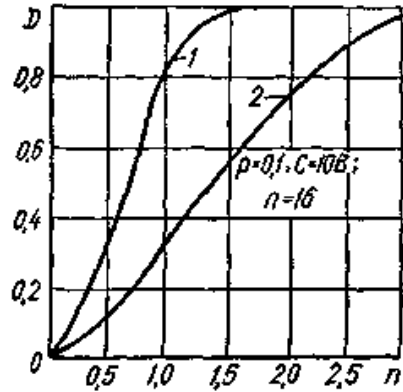


Рис. 15. Характеристики обнаружения:
 1 — алгоритма с нелинейным преобразователем $\Phi (I')$;
 2 — без нелинейного преобразователя

Незащищенный алгоритм (23) существенно уступает модифицированному обнаружителю, использующему нелинейный преобразователь.

Обобщенный ЭБМ. Обобщенный ЭБМ базируется на принципе редукции Вальда, который широко применяется в задачах с параметрической априорной неопределенностью.

Согласно этому принципу решающее правило определяется выражением

$$I = \frac{\int_{(\Theta \in \Omega_{c+n})} \omega_{c+n}(\vec{x}/\Theta) \omega_{\Theta}(\Theta) d\Theta}{\int_{(\Theta \in \Omega_n)} \omega_n(\vec{x}/\Theta) \omega_{\Theta}(\Theta) d\Theta}, \quad (27)$$

где $\omega_{c+n}(\vec{x}/\Theta)$, $\omega_n(\cdot)$ — условные плотности распределения выборки \vec{x} в области смеси сигнала и помехи (только помехи); Θ — вектор неизвестных параметров; $\omega_{\Theta}(\cdot)$ — плотность распределения вероятности параметров. Алгоритм (27) не является адаптивным.

Обобщение принципа редукции Вальда состоит в том, что мы рассматриваем совместное условное распределение вектора выборочных значений \vec{x} и некоторой оценки $\hat{\Theta}$ параметра в

распределения выборки — $\omega(\vec{x}, \hat{\Theta}/\Theta)$ и применяем принцип редукции Вальда (27) для условных совместных распределений \vec{x} и $\hat{\Theta}$ в области гипотезы H_0 и альтернативы H_1 .

Адаптивный алгоритм определяется выражением

$$I = \frac{\int_{(\Theta \in \Omega_{c+n})} \omega_{c+n}(\vec{x}, \hat{\Theta}/\Theta) \omega_{\Theta}(\Theta) d\Theta}{\int_{(\Theta \in \Omega_n)} \omega_n(\vec{x}, \hat{\Theta}/\Theta) \omega_{\Theta}(\Theta) d\Theta} \quad (28)$$

и может быть получен в конечном виде при известной плотности распределения параметров ω_{Θ} . Обычно вид $\omega_{\Theta}(\cdot)$ определяется априорной информацией либо пользуются минимаксным подходом, выбирая наименее благоприятное распределение параметра Θ . Рассмотрим пример использования ОЭБМ в задаче последетекторного обнаружения гауссовского сигнала на фоне внутримемного шума и ХИП.

Пример 19. Дано: условная плотность распределения вероятности отсчетов огибающей смеси сигнал + шум + помеха на выходе линейного детектора

$$\omega(x/u) = (1-p) \frac{x}{\Psi(1+b)} e^{-\frac{x^2}{2\Psi(1+b)}} + \frac{pc}{\sqrt{2\pi\Psi(1+b)}} e^{-\frac{(x-u)^2}{2\Psi(1+b)}}, \quad (29)$$

где p — вероятность воздействия импульсной помехи; b — соотношение сигнал/шум по мощности; u — амплитуда импульсной помехи; c — нормирующий множитель; Ψ — дисперсия шума.

Условная плотность распределения вероятности оценки амплитуды импульсной помехи

$$\omega_u^{\wedge}(\hat{u}/u) = \frac{1}{\sqrt{2\pi\Psi_u^{\wedge}}} e^{-\frac{(u-\hat{u})^2}{2\Psi_u^{\wedge}}}$$

где Ψ_u^{\wedge} — дисперсия оценки амплитуды.

Требуется найти адаптивный алгоритм обнаружения сигнала.

Запишем совместную условную плотность распределения вероятностей выборочных отсчетов и оценки параметров

$$\omega(x, \hat{u}/u) = \left[(1-p) \frac{x}{\Psi(1+b)} e^{-\frac{x^2}{2\Psi(1+b)}} + \frac{pc}{\sqrt{2\pi\Psi(1+b)}} e^{-\frac{(x-u)^2}{2\Psi(1+b)}} \right] \omega_u^{\wedge}(\hat{u}/u). \quad (30)$$

Безусловную плотность распределения $\omega(x, \hat{u})$ можно найти

интегрированием (30) по априорной плотности распределения вероятностей неизвестного параметра $\omega(u)$. В качестве такого примем несобственное равномерное распределение с элементом вероятности du .

Тогда

$$\omega(x, \hat{u}) = \lim_{\substack{b \rightarrow \infty \\ a \rightarrow -\infty}} \int_a^b \omega(x, \hat{u}/u) du, \quad (31)$$

Подставляя (30) в (31), получаем

$$\omega(x, \hat{u}) = (1-p) \frac{x}{\Psi(1+b)} e^{-\frac{x^2}{2\Psi(1+b)}} + \frac{pc}{\sqrt{2\pi[\Psi(1+b) + \Psi_{\hat{u}}]}} e^{-\frac{(x-\hat{u})^2}{2[\Psi(1+b) + \Psi_{\hat{u}}]}}. \quad (32)$$

Алгоритм обнаружения, согласно выражению (28), дается отношением правдоподобия

$$l(x, \hat{u}, b)_{b=b_1} = \frac{\omega(x, \hat{u})_{b=b_1}}{\omega(x, \hat{u})_{b=0}} > V_p, \quad (33)$$

где b_1 — ожидаемое соотношение сигнал/шум либо логарифм выражения (33).

Графическое изображение характеристики логарифма нелинейного преобразователя (33) при фиксированных значениях \hat{u} и b_1 показано на рис. 16.

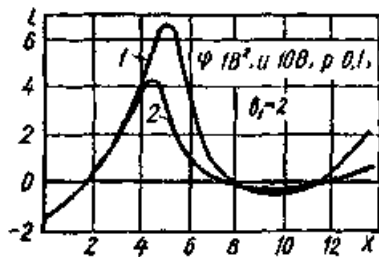


Рис. 16. Характеристики нелинейного преобразователя адаптивную алгоритма обнаружения гауссова сигнала на фоне гауссова шума и импульсной помехи (синтез по ОЭБМ) для $\Psi_{\hat{u}} = 1B^2$ (1); $\Psi_{\hat{u}} = 3B^2$ (2)

Как видно из кривых (рис. 16), адаптивный алгоритм подавляет выборочные отсчеты огибающей смеси сигнала и помехи в окрестности оценки амплитуды импульсной помехи. Размер окрестности зависит от точности оценки и увеличивается с увеличением ее дисперсии $\Psi_{\hat{u}}$.

Характеристики обнаружения алгоритма (33) получены методом статистического моделирования и показаны на рис. 17.

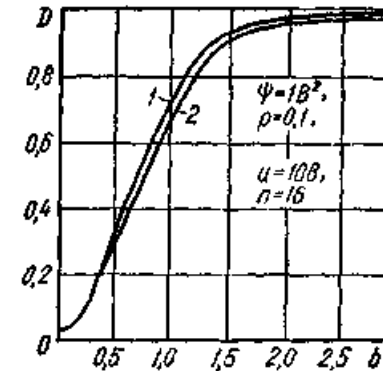


Рис. 17. Характеристики адаптивного алгоритма обнаружения, определяемого формулой (33), для $\Psi_{\hat{u}} = 1B^2$ (1); $\Psi_{\hat{u}} = 3B^2$ (2)

Рассмотренные примеры иллюстрируют полезность и работоспособность ЭБМ и его модификаций в задачах синтеза алгоритмов обнаружения в сложных помеховых ситуациях с параметрической априорной неопределенностью. Использование изложенных методов позволяет в существенной степени увеличить устойчивость характеристик качества алгоритмов к воздействию помех.

1.9.3. Параметрические алгоритмы, использующие L-оценки

Идея использования L-оценок параметров в задачах проверки параметрических гипотез была высказана рядом авторов (Хартли Дэвид, Джонсон, Грабе, Уолш и др.) в начале пятидесятых годов. Ими рассмотрены задачи проверки гипотез относительно нормального распределения выборочных значений, исследована мощность критериев, устойчивость их характеристик при отклонениях от нормального распределения в исходной совокупности. В настоящем разделе рассматривается пример синтеза на основе ОЭБМ решающего

правила при использовании робастных L -оценок в задаче последетекторного обнаружения шумоподобного и детерминированного сигналов на фоне негауссовых помех (см. формулу (2) п.1.8).

Сформулируем задачу обнаружения в терминах теории проверки статистических гипотез относительно рабочей выборки x_1, \dots, x_n , отсчеты которой независимы и одинаково распределены. Проверяются гипотезы H_0 и H_1 соответственно:

$$\omega_0(x/p, u, \Psi) = (1 - p) \frac{1}{2\Psi_0} e^{-\frac{x}{2\Psi_0}} + p\omega_{20}(x); \quad (34)$$

$$\omega_1(x/p, u, \Psi, b) = (1 - p) \frac{1}{2\Psi_0(1+b)} e^{-\frac{x}{2\Psi_0(1+b)}} + p\omega_{21}(x).$$

Модель (34) соответствует случаю обнаружения шумоподобного сигнала на выходе квадратичного детектора огибающей при воздействии внутриприемного шума и импульсной помехи со скважностью $1/p$ ($p \ll 1$), и неизвестной плотностью распределения $\omega_{2i}(x)$, $i = 0, 1$. Параметр масштаба внутриприемного шума Ψ_0 считаем неизвестным. Как видно из формул (34), присутствие сигнала в выборке сказывается на изменении параметра b , который становится отличным от нуля. Предполагается, что кроме рабочей выборки, мы располагаем опорной (обучающей) выборкой y_1, \dots, y_m соответствующей распределению (34) при $b = 0$.

Задача обнаружения сигнала в выборке x_1, \dots, x_n эквивалентна задаче проверки гипотез: $H_0 - b = 0; H_1 - b = b_0 \neq 0$. Введем обозначение $\Psi_1 = \Psi_0(1+b)$. Тогда $\Psi = \Psi_0$ при $b = 0$ и $\Psi_1 > \Psi_0$ при $b > 0$, что позволяет сформулировать задачу в пространстве параметров: $H_0 - \Psi_1 = \Psi_0; H_1 - \Psi_1 > \Psi_0$. Так как распределения (34) не принадлежат к семейству экспоненциальных, достаточных статистик для параметров этих ПРВ не существует и задачу нельзя решать методами подобия и инвариантности. Для ее решения используем «защищенные» от выбросов L -оценки параметров масштаба Ψ_0 и Ψ_1 рабочей и опорной выборок (см. разд. 1.8.3). Пользуясь тем, что $p \ll 1$, произведем цензурирование рабочей и опорной выборок и по цензурированным выборкам вычислим нелинейные оценки параметров Ψ_0 и $\Psi_1 = \Psi_0(1+b)$ экспоненциального распределения по усеченной выборке.

Воспользуемся известными выражениями

$$\hat{\Psi} = \sum_{i=1}^r a_{(i)} y_{(i)} = \frac{1}{r} \left(\sum_{i=1}^r y_{(i)} + (m-r) y_{(r)} \right), \quad r < m; \quad (35)$$

$$\hat{\Psi} = \sum_{i=1}^s a_{(i)} x_{(i)} = \frac{1}{s} \left(\sum_{i=1}^s x_{(i)} + (n-s) x_{(s)} \right), \quad s < n. \quad (36)$$

Распределение статистик

$$\frac{2r\hat{\Psi}_0}{\Psi_0} \text{ и } \frac{2s\hat{\Psi}_1}{\Psi_1}$$

является распределением χ^2 с $2r$ ($2s$) степенями свободы или распределением Эрланга $r(s)$ -го порядка.

Найдем условное распределение оценок $\hat{\Psi}_0$ и $\hat{\Psi}_1$ при

$$\omega(\hat{\Psi}_0/\Psi_0, b = 0) = \frac{(r\mu_0)^r}{\Gamma(r)} \Psi_0^{r-1} e^{-r\mu_0\hat{\Psi}_0}, \quad \mu_0 = \frac{1}{\Psi_0};$$

$$\omega(\hat{\Psi}_1/\Psi_0, b = 0) = \frac{(s\mu_0)^s}{\Gamma(s)} \hat{\Psi}_1^{s-1} e^{-s\mu_0\hat{\Psi}_1}.$$

Безусловное распределение

$$\omega(\hat{\Psi}_0, \hat{\Psi}_1/b = 0) =$$

$$= \frac{(s)^s r^r}{\Gamma(s) \Gamma(r)} \int_0^\infty \hat{\Psi}_0^{r-1} \hat{\Psi}_1^{s-1} \mu_0^{s+r} e^{-\mu_0(r\hat{\Psi}_0 + s\hat{\Psi}_1)} d\mu_0 =$$

$$= \frac{(s+r)! s^s r^r}{(r\hat{\Psi}_0 + s\hat{\Psi}_1)^{s+r+1} \Gamma(s) \Gamma(r)} \hat{\Psi}_0^{r-1} \hat{\Psi}_1^{s-1}.$$

Распределение оценок при H_1

$$\omega(\hat{\Psi}_0/\Psi_0, b = 0) = \frac{(r\mu_0)^r}{\Gamma(r)} \hat{\Psi}_0^{r-1} e^{-r\mu_0\hat{\Psi}_0};$$

$$\omega(\hat{\Psi}_1/\Psi_0(1+b), b \neq 0) = \frac{(s\mu_0)^s}{(1+b)^s \Gamma(s)} \hat{\Psi}_1^{s-1} e^{-\frac{s\mu_0\hat{\Psi}_1}{(1+b)}}.$$

Совместное распределение

$$\omega(\Psi_0, \hat{\Psi}_1) = \frac{s^s r^r}{(1+b)^s \Gamma(r) \Gamma(s)} \hat{\Psi}_0^{r-1} \hat{\Psi}_1^{s-1} \times$$

$$\times \int_0^\infty \mu_0^{r+s} e^{-\mu_0(r\hat{\Psi}_0 + \frac{s\hat{\Psi}_1}{1+b})} d\mu_0 =$$

$$= \Psi_0^{r-1} \Psi_0^{s-1} \frac{s^s r^r (s+r)! (1+b)^{s+r+1}}{(1+b)^s \Gamma(r) \Gamma(s) (r\hat{\Psi}_0(1+b) + s\hat{\Psi}_1)^{s+r+1}}.$$

Отношение правдоподобия

$$l(\hat{\Psi}_0, \hat{\Psi}_1) = \frac{\omega(\hat{\Psi}_0, \hat{\Psi}_1/H_1)}{\omega(\hat{\Psi}_0, \hat{\Psi}_1/H_0)} = (1+b)^{r+1} \left[\frac{r\hat{\Psi}_0 + s\hat{\Psi}_1}{r\hat{\Psi}_0(1+b) + s\hat{\Psi}_1} \right]^{s+r+1}.$$

Введя обозначение $z = \hat{\Psi}_1/\hat{\Psi}_0$, получим

$$l(z) = (1+b)^{r+1} \left[\frac{r+sz}{r(1+b)+sz} \right]^{s+r+1}. \quad (37)$$

Этот результат можно получить и другим способом, если применить принцип инвариантности (z — максимальный инвариант на пространстве параметров масштаба относительно группы масштабных преобразований выборочного пространства) и построить равномерно наиболее мощное (РНМ) инвариантное решающее правило.

Так как отношение правдоподобия монотонно по z , эквивалентное решающее правило может быть получено путем сравнения z с порогом решения V_p :

$$z(x_{(1)} \dots x_{(r)}, y_{(1)} \dots y_{(r)}) > V_p. \quad (38)$$

Исследуем основные характеристики — вероятность ложной тревоги и вероятность правильного обнаружения цели.

Найдем распределение статистики z при отсутствии сигнала (гипотеза H_0). Обозначим $\hat{\Psi}_1 = x$, $\hat{\Psi}_0 = y$. Тогда

$$\omega(z/H_0) = \int_0^\infty \frac{(r\mu_0)^r}{\Gamma(r)} y^{r-1} e^{-r\mu_0 y} \frac{(s\mu_0)^s}{\Gamma(s)} (yz)^{s-1} e^{-s\mu_0 yz} y dy = \frac{s^s r^r \mu_0^{s+r} \Gamma(r+s) z^{s-1}}{\Gamma(s) \Gamma(r) [\mu_0(r+sz)]^{s+r}}. \quad (39)$$

Распределение статистики z при наличии сигнала (гипотеза H_1):

$$\omega(z/H_1) = \int_0^\infty \frac{(r\mu_0)^r}{\Gamma(r)} y^{r-1} e^{-r\mu_0 y} \times \frac{(s\mu_0)^s}{(1+b)^s \Gamma(s)} (yz)^{s-1} e^{-\frac{s\mu_0 yz}{1+b}} y dy = \frac{s^s r^r \mu_0^{s+r} \Gamma(r+s) z^{s-1}}{(1+b)^s \Gamma(s) \Gamma(r) \left(r\mu_0 + \frac{s\mu_0 z}{1+b} \right)^{r+s}}. \quad (40)$$

Выражения для вероятности ложной тревоги и правильного обнаружения получаются интегрированием плотностей (39) и (40) в

пределах от V_p до ∞ . Эти интегралы сводятся к табличному известной формулой.

$$D(V_p) = \int_{V_p}^\infty \omega(z/H_0) dz = 1 - \frac{s^s \Gamma(r+s) V_p^s}{r^s (1+b)^s \Gamma(s) \Gamma(r)} \times \times {}_2F_1\left(r+s; s; 1+s; -\frac{sv_p}{r(1+b)}\right);$$

$$F(V_p) = D(V_p/b = 0).$$

1.9.4. Ранговые непараметрические алгоритмы

В ситуациях с непараметрической априорной неопределенностью, когда неизвестен вид плотностей распределения выборочных отсчетов, а о различии распределений сигнала и помехи известны только сведения общего характера (различия в сдвиге, масштабе, виде закона распределения вероятностей), для получения устойчивых алгоритмов обнаружения сигналов пользуются ранговыми критериями проверки статистических гипотез.

Приложениям теории ранговых критериев к задачам обнаружения сигналов посвящены работы, в которых рассмотрен ряд вопросов, связанных с синтезом, анализом эффективности и технической реализацией ранговых обнаружителей. Однако в указанных работах не затрагиваются вопросы применения ранговых критериев проверки статистических гипотез в задачах с составным распределением выборочных отсчетов типа (1) п.1.8. Такие ситуации имеют место, когда прием сигналов проводится на фоне импульсных помех.

Пусть имеется выборка, в которой (предполагаем) находится сигнал, и пусть она содержит n независимых отсчетов. Сравним каждый сигнальный отсчет со своей независимой шумовой выборкой из m отсчетов и найдем ранги сигнальных отсчетов:

$$R_j = \sum_{i=1}^m u(x_j - y_{ji}), \quad j = 1, \dots, n,$$

где $u(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$ — функция единичного скачка; x_j — сигнальный отсчет; y_{ji} — шумовой отсчет.

Гипотезой назовем случай отсутствия сигнала. Определим плотность распределения шумовой выборки $\omega(y) = \omega(x, 0)$. Здесь $\omega(x, 0)$ — плотность распределения сигнальной выборки при отсутствии сигнала. Альтернативой будет наличие некоторого сигнала с определенным параметром b : $\omega(y) \neq \omega(x, b)$.

Для гипотезы вероятности появления рангов равны, ранги независимы:

$$v_{im} \{R_j\} = 1/(m + 1), \quad i = 1, \dots, m. \quad (41)$$

В случае альтернативы ранги также независимы (из-за наличия у каждого сигнального отсчета своей обучающей выборки) и распределены с некоторой функцией распределения $v_{im} (R_j, b)$, зависящей также от параметра b . Следовательно, для проверки гипотезы о наличии сигнала можно записать следующее отношение правдоподобия:

$$L(b) = \frac{\prod_{j=1}^n \prod_{i=0}^m v_{im} (R_j, b)^{e_i}}{\prod_{j=1}^n \prod_{i=0}^m v_{im} (R_j, 0)^{e_i}}, \quad (42)$$

где $v_{im} (R_j, b)$ — вероятность того, что $R_j = i$; e_i — индикатор состояния R_j ,

$$e_i = \begin{cases} 1, & R_j = i \\ 0, & R_j \neq i \end{cases}$$

Прологарифмируем (42) и, учитывая (41), получим:

$$l(b) = \ln [L(b)] = \sum_{j=1}^n \sum_{i=0}^m e_i \ln v_{im} (R_j, b) - n \ln \left\{ \frac{1}{(m+1)} \right\}.$$

Вычислим вероятности появления рангов $v_{im} \{R_j, b\}$. Их можно определить, воспользовавшись формулой полной вероятности:

$$v_{im} \{R_j, b\} = \binom{m}{i} \int_{-\infty}^{\infty} \omega(x, b) F^i(x, 0) \times [1 - F(x, 0)]^{m-i} dx, \quad (43)$$

где $\omega(x, b)$ — плотность вероятности смеси сигнала и шума в окрестности точки x ; $F^i(x, 0)$ — вероятность того, что i шумовых отсчетов меньше; $[1 - F(x, 0)]^{m-i}$ — вероятность того, что $m - i$ шумовых отсчетов больше x ; $F(x, 0)$ — функция распределения шумовых отсчетов.

Таким образом, синтезированный критерий

$$S = \sum_{j=1}^n \sum_{i=0}^m e_i a_i \geq k, \quad (44)$$

где

$$a_i = \ln \binom{m}{i} \int_{-\infty}^{\infty} \omega(x, b) F^i(x, 0) [1 - F(x, 0)]^{m-i} dx, \quad (45)$$

где $\binom{m}{i}$ — число сочетаний из m по i ; k определяется по заданной вероятности ошибок первого рода.

Найдем плотность вероятности статистики критерия S . Внутренняя сумма выражения (44) принимает дискретные значения a_1, a_2, \dots, a_m с вероятностями состояний, определяемыми выражением (43), причем события a_i независимы. Тогда характеристическая функция принимает вид:

$$\Theta(t) = \sum_{k=0}^m v_k e^{i t a_k},$$

где t — произвольный вещественный параметр; v_k — вероятность k -го состояния; i — мнимая единица.

Характеристическая функция суммы n независимых случайных величин

$$\Theta_{\Sigma}(t) = \prod_{i=1}^n \Theta_i(t) = \Theta^n(t) = \left(\sum_{k=0}^m v_k e^{i t a_k} \right)^n.$$

После возведения в n -ю степень получим:

$$\Theta_{\Sigma}(t) = \sum_{k_1=0}^n \sum_{k_2=0}^{k_1} \dots \sum_{k_{m-1}=0}^{k_{m-2}} c \times v_{0n}^{i-k} \times v_{1m}^{k_1-k_2} \times \dots \times v_{m-1}^{k_{m-1}-k_m} \times v_m^{k_m} \times \exp \{i t [(n - k_1) a_0 + (k_1 - k_2) a_1 + \dots + (k_{m-1} - k_m) a_{m-1} + k_m a_m]\},$$

$$\text{где } c = \frac{n!}{(n - k_1)! (k_1 - k_2)! \dots (k_{m-1} - k_m)! k_m!}.$$

После обратного преобразования получим плотность вероятности проверочной статистики:

$$\omega_S(S) = \sum_{k_1=0}^n \sum_{k_2=0}^{k_1} \sum_{k_3=0}^{k_2} \dots \sum_{k_m=0}^{k_{m-1}} c \times v_0^{i-k_1} v_1^{k_1-k_2} \dots \dots v_{m-1}^{k_{m-1}-k_m} v_m^{k_m} \times \delta \{S - [(n - k_1) a_0 + (k_1 - k_2) a_1 + \dots + (k_{m-1} - k_m) a_{m-1} + k_m a_m]\},$$

где $\delta(\cdot)$ — дельта-функция Дирака.

Обозначив $k_0 = n$ и приняв $k_{m+1} = 0$, можно записать выражение для плотности в более компактном виде:

$$\omega_S(S) = \sum_{k_1=0}^n \sum_{k_2=0}^{k_1} \sum_{k_3=0}^{k_2} \dots \sum_{k_{m-1}=0}^{k_{m-2}} c \times \prod_{i=0}^m v_{im}^{k_i - k_{i+1}} \times \delta \left\{ S - \sum_{i=0}^m (k_i - k_{i+1}) a_i \right\}. \quad (46)$$

Полученное выражение позволяет найти плотность вероятности статистики любого критерия вида

$$\sum_{i=1}^n \sum_{i=0}^m \epsilon_i a_i \geq k,$$

где a_i — произвольные коэффициенты.

Рассмотрим особенности синтеза ранговых критериев для задач обнаружения сигналов, когда плотности распределения выборочных отсчетов при гипотезе и альтернативе описываются составным распределением вида (1) п.1.8. Пусть распределение

$$\omega(x) = (1 - p)\omega_1(x) + p\omega_2(x) \quad (47)$$

состоит из двух непересекающихся классов, т. е. распределение (47) можно записать в виде

$$\omega(x) = (1 - p)\omega_1(x)I_1(x) + p\omega_2(x)I_2(x), \quad (48)$$

где $I_1(x), I_2(x)$ — индикаторные функции,

$$I_1(x) = \begin{cases} 1, & x \in \Omega_1, \\ 0, & x \notin \Omega_1; \end{cases} \quad I_2(x) = \begin{cases} 1, & x \in \Omega_2, \\ 0, & x \notin \Omega_2, \end{cases}$$

где Ω_1 и Ω_2 — области определения функций $\omega_1(x)$ и $\omega_2(x)$ соответственно, $\Omega_1 \cap \Omega_2 = \emptyset$.

Для реальных распределений, описывающих смеси сигналов и интенсивных импульсных помех, запись вида (48) дает небольшие погрешности и вполне может быть использована для синтеза субоптимальных алгоритмов. При подстановке (48) в (43) и после тождественных преобразований получим выражение для вычисления вероятностей рангов

$$v_{im}\{R_j, b\} = \binom{m}{l} \sum_{k=0}^{m-l} (-1)^k \binom{m-i}{k} \left\{ (1-p)^{i+k+1} \int_{\Omega_1} \omega_1(x)^{i+k} \times F_1 x dx + p^{m-l+1} \sum_{i=0}^l (-1)^i \binom{l}{i} p^i \int_{\Omega_2} \omega(x) F_2^{k+i} x dx \right\}. \quad (49)$$

Выражение (49) существенно упрощает вычисление вероятностей появления рангов, так как сводится к вычислению интегралов вида

$$I_j(l) = \int_{\Omega_j} \omega_j(x, b) F_j^l(x) dx, \quad j = 1, 2 \quad (50)$$

для каждого из распределений $\omega_j, j = 1, 2$, образующих составное распределение (47).

Рассмотрим некоторые частные случаи.

1. Для огибающей аддитивной совокупности гауссовского сигнала и импульсной помехи, действующей с вероятностью p , плотности распределения вероятностей $\omega_1(x, b)$ и $\omega_2(x, b)$ имеют вид

$$\omega_1(x, b) = \frac{x}{(1+b)\Psi} \exp \left\{ -\frac{x^2}{2(1+b)\Psi} \right\}, \quad (51)$$

$$\omega_2(x, b) = \frac{x}{(1+b)\Psi} e^{-\frac{x^2+u^2}{2\Psi(1+b)}} I_0 \left(\frac{xu}{\Psi(1+b)} \right), \quad (52)$$

где $b = \Psi_c/\Psi_w$ — отношение сигнал/шум по мощности; u — амплитуда импульсной помехи.

Условие $u \gg \sqrt{\Psi(1+b)}$ позволяет считать области определения функций $\omega_1(\cdot)$ и $\omega_2(\cdot)$ непересекающимися и применить описанный подход.

Вычисление первого из интегралов (50) дает

$$I_1(l) = \sum_{k=0}^l \binom{l}{k} (-1)^k \frac{1}{1+k(1+b)}. \quad (53)$$

Второй интеграл непосредственно не вычисляется ввиду сложности выражения для интегральной функции распределения Раиса, соответствующей плотности (52).

Для приближенной оценки I_2 воспользуемся тем обстоятельством, что при выполнении условия $u/\sqrt{\Psi(1+b)} \gg 1$ обобщенная функция распределения Релея переходит в нормальную с параметрами u и $\sqrt{\Psi(1+b)}$. Таким образом,

$$I_2(l) \approx \int_0^\infty \frac{x}{(1+b)} \exp \left\{ -\frac{x^2+u^2}{2(1+b)} \right\} \times I_0 \left(\frac{xu}{1+b} \right) \Phi \left(\frac{x-u}{\sqrt{2}} \right) dx = \int_0^\infty \frac{x}{(1+b)} e^{-\frac{x^2+u^2}{2(1+b)}} \frac{1}{2^l} \left[1 + \Phi \left(\frac{x-u}{\sqrt{2}} \right) \right]^l \times I_0 \left(\frac{xu}{1+b} \right) dx, \quad (54)$$

где $\Phi(x)$ — функция Крампа; $\Phi_1(x)$ — интеграл вероятности в форме Лапласа.

Пользуясь разложением функции Крампа, модифицированной функции Бесселя нулевого порядка в степенной ряд, биномом Ньютона и ограничиваясь ν членами разложения, получаем

$$I_2(t) = \frac{e^{-\frac{u^2}{2(1+b)}}}{2^t} \sum_{l=0}^t \binom{t}{l} \left(\frac{2}{\sqrt{\pi}}\right)^l \sum_{\lambda=0}^{lv} C_\lambda \sum_{h=0}^{2\lambda+l} (-1)^h \times \\ \times \binom{2\lambda+l}{h} u^{+h} [2(1+b)]^{\frac{2\lambda+l-h}{2}} \times \\ \times \sum_{q=0}^{\infty} \frac{u^{2q}}{(q!)^2 [2(1+b)]^q} \Gamma\left(\frac{2\lambda+l-h+2}{2} + q\right), \quad (55)$$

где

$$C_\lambda(t) = \frac{1}{\lambda a_0} \sum_{k=1}^{\lambda} (kt - \lambda + k) a_k C_{\lambda-k}; \quad C_0 = (1/\sqrt{2})^t; \\ a_k = \frac{(-1)^k}{k! (\sqrt{2})^{2k+2} (2k+1)}, \quad \text{при } k \leq v; \quad a_0 = 1/\sqrt{2}; \\ a_k = 0, \quad \text{при } k > v.$$

При вычислении выражений (53) и (54) интегрирование проводилось в пределах $0 \dots \infty$, что не соответствует предпосылке о непересекающихся областях определения функций $\omega_1(\cdot)$ и $\omega_2(\cdot)$. Однако большое различие значений $\omega_1(x)$ и $\omega_2(x)$ в каждой из точек $x \in (0 \dots \infty)$ делает возникающие при вычислениях погрешности малыми и в практических расчетах ими можно пренебречь.

2. Для гауссовской модели сигнала и помехи в случае последующего обнаружения вероятности появления рангов могут быть получены из формулы (49) при $p = 0$ с учетом вида плотности огибающей (51). После ряда простых преобразований получим

$$v_{im} = \binom{m}{i} \frac{i! (1+b)^i}{\prod_{k=0}^i [1+(m-k)(1+b)]}. \quad (56)$$

Коэффициенты a_i для выражения (44) вычислим по формуле $a_i = \ln v_{im}$.

3. Для гармонической модели сигнала и гауссовской помехи плотность вероятности помехи также описывается выражением (51) при $b = 0$, а плотность вероятности смеси сигнала и помехи можно представить в виде

$$\omega(x, b) = \frac{x}{\Psi} e^{-b} \exp\left(-\frac{x^2}{2\Psi}\right) I_0\left(\frac{xu}{\Psi}\right), \quad (57)$$

где $b = u^2/\Psi$ — отношение сигнал/шум; u — амплитуда гармонического сигнала.

Подставив (51) при $b = 0$ и (57) в (43), получим:

$$v_{im} = \binom{m}{i} \sum_{j=0}^i \frac{(-1)^j}{1+m+j-1} \binom{i}{j} \exp\left\{-\frac{m+j-i}{1+m+j-i} b\right\}, \\ a_i = \ln v_{im}.$$

По выражению (46) для гауссовской модели сигнала и помехи были построены характеристики обнаружения для синтезированного критерия (44) и алгоритма, в котором весовые коэффициенты $a_i = R_j$.

4. Для огибающей аддитивной совокупности гауссовского сигнала внутриприемного шума и ХИП с плотностью распределения (14), прошедшей через нелинейный преобразователь (16), одномерная плотность распределения описывается выражением

$$\omega(y) = \frac{(1-p)}{2\Psi(1+b)} e^{-\frac{y}{2\Psi(1+b)}} + \\ + \frac{p}{\sqrt{2\pi\Psi(1+b)}y} e^{-\frac{y}{2\Psi(1+b)}}, \quad y \geq b. \quad (58)$$

Соответствующая интегральная функция распределения

$$W(y) = (1-p) \left[1 - \exp\left(-\frac{y}{2\Psi(1+b)}\right) \right] + \\ + p\Phi\left(\sqrt{\frac{y}{2\Psi(1+b)}}\right). \quad (59)$$

Здесь

$$\Phi(V) = \frac{2}{\sqrt{\pi}} \int_0^V e^{-x^2} dx$$

— функция Крампа; b — соотношение сигнал/шум по мощности.

Подставив эти выражения в формулу (49), получим формулу для вероятностей появления рангов

$$v_{im}(R_i, b) = \binom{m}{i} \sum_{j=0}^{m-i} (-1)^j \binom{m-i}{j} \times \\ \times \sum_{l=0}^{i+j} \binom{i+j}{l} (1-p)^l p^{i+j-l} \sum_{t=0}^l (-1)^t \binom{l}{t} \times \\ \times \left[\frac{2}{\sqrt{\pi}} \right]^{i+j-l} \frac{1}{2^{i+j-l}} \sum_{k=0}^{i+j-l} C_k (l_1^k + l_2^k), \quad (60)$$

где

$$I_1^k = \int_0^{\infty} x^{k + \frac{i+j-l}{2}} e^{-ix} \frac{(1-p)}{2(1+b)} e^{-\frac{x}{2(1+b)}} dx =$$

$$= \frac{(1-p)}{2(1+b)} \left[\frac{2(1+b)}{1+l(1+b)} \right]^{k + \frac{i+j-p}{2}} \times$$

$$\times \Gamma \left(k + 1 + \frac{i+j-l}{2} \right);$$

$$I_2^k = p \int_0^{\infty} x^{k + \frac{i+j-l-1}{2}} \sqrt{\frac{2(1+b)}{\pi}} e^{-\frac{ix}{2}} e^{-\frac{x}{2(1+b)}} dx =$$

$$= g \sqrt{\frac{2(1+b)}{\pi}} \left[\frac{2(1+b)}{1+l(1+b)} \right]^{k - \frac{i+j-l-1}{2} + 1} \times$$

$$\times \Gamma \left(k + \frac{i+j-l+1}{2} \right);$$

C_k — коэффициенты при степенях x , которые получаются при возведении степенного ряда (разложения $\Phi(\cdot)$) в степень. Если число членов ряда конечно и равно v и ряд возводится в степень n , то

$$C_0 = (g_0)^n; \quad C_k = \frac{1}{k g_0} \sum_{l=2}^k (ln - k + l) g_l C_{k-l},$$

где g_k — коэффициент при x^k в разложении функции Крампа,

$$g_k = \begin{cases} \frac{(-1)^k}{k! 2^k (2k+1)} & \text{при } k \leq v; \\ 0 & \text{при } k > v. \end{cases}$$

Выражение (60) получено с применением записи бинорма Ньютона в виде конечного ряда и разложением функции Крампа в формуле (59) в степенной ряд с числом членов v . Как и в предыдущих случаях, весовые коэффициенты оптимального рангового обнаружителя определяются логарифмом соответствующих вероятностей

$$a_i = \ln(v_i(R_i, b)).$$

Результаты расчетов порогового соотношения b , при котором достигается значение вероятности правильного обнаружения сигнала $D = 0,9$, показаны на рис. 18.

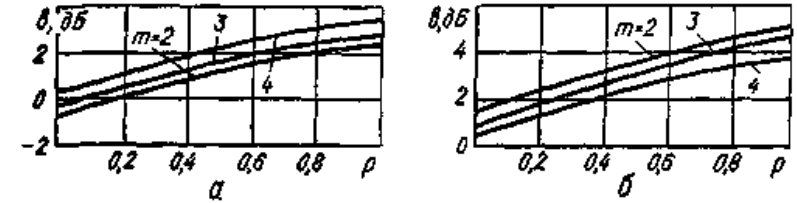


Рис. 18. Зависимости порогового соотношения сигнал/помеха рангового обнаружителя от параметра импульсной помехи p при $n = 128$ для $F = 10^{-4}$ (а), для $F = 10^{-6}$ (б)

Расчеты проведены для рангового обнаружителя, синтезированного в предположении, что при наличии сигнала соотношение сигнал/шум по мощности $b = 1$ дБ.

Таким образом, предложенная методика позволяет построить оптимальный ранговый обнаружитель для любого наперед заданного соотношения сигнал/шум. Малые объемы выборки в радиолокационном обнаружении используются при достаточно большом соотношении сигнал/шум. В этом случае гипотеза и альтернатива существенно отличаются друг от друга по сигнальному параметру b . Использование в такой ситуации оптимального рангового критерия дает существенный выигрыш по сравнению с асимптотически оптимальным ранговым алгоритмом. Выражение (46) позволяет строить точные характеристики обнаружения для любых объемов выборки.

1.9.5. Квантильные непараметрические алгоритмы

Одновыборочный алгоритм проверки гипотезы о квантиле распределения предложен в 1937 г. Кокрэном. Алгоритм проверяет гипотезу

$$H_0: X_{\Theta} = x_0, \tag{61}$$

где X_{Θ} — квантиль уровня в распределения F_x ; x_n — некоторое заданное значение этого квантиля.

Выборка из n наблюдений обрабатывается по правилу

$$S = \sum_{i=1}^n u(x_i - x_0) = \sum_{i=1}^n u(x_i - x_0), \tag{62}$$

где

$$u(z) = \begin{cases} 1, & z > 0; \\ 0, & z < 0; \end{cases}$$

— бинарный квантователь.

Статистика S считает число положительных значений среди разностей $(x_i - x_0)$. В технике этот алгоритм получил название бинарного накопителя, так как S является суммой случайных величин, принимающих два значения (0 и 1).

Для односторонней гипотезы $H_1: X_\Theta > x_0$ решающее правило $S > V_p$ является равномерно наиболее мощным.

Рассмотрим двухвыборочный квантильный алгоритм для проверки гипотезы H_0 о равенстве квантилей уровня в распределений случайных величин X и $Y - F_x$ и F_y

$$H_0: X_\Theta = Y_\Theta \quad (63)$$

против альтернативы $H_1: X_\Theta > Y_\Theta$, где X_Θ, Y_Θ — квантили уровня в распределений F_x и F_y соответственно.

Пусть имеется две выборки — рабочая \vec{x}_n и обучающая \vec{y}_{nr} объемом n и nr соответственно. Каждому выборочному значению x_i соответствует r выборочных значений y_{i1}, \dots, y_{in} .

Для решения задачи проверки гипотез H_0 и H_1 можно, согласно эмпирическому байесовскому подходу, воспользоваться статистикой (61), подставив вместо x_0 оценку квантиля уровня Θ , вычисленную по обучающей выборке y_{nr} . Двухвыборочный алгоритм

$$S = \sum_{i=1}^n u(X_i - \hat{Y}_{\Theta i}) > V_p, \quad (64)$$

где $\hat{Y}_{\Theta i}$ — оценка квантиля уровня Θ по выборке y_{i1}, \dots, y_{in} ; V_p — порог решения.

Для получения \hat{Y}_{Θ} , который в технике называется порогом квантования, могут использоваться различные алгоритмы оценивания, обеспечивающие получение непараметрических оценок квантиля.

В данном параграфе рассматривается бинарный накопитель, где для оценки порога квантования используется процедура стохастической аппроксимации. Такой подход приводит к замкнутым системам с переменным коэффициентом усиления в контуре обратной связи

$$\hat{Y}_{j+1} = \hat{Y}_j + \frac{k}{j+1} (\xi_{j+1} - \Theta), \quad (65)$$

где k — некоторая постоянная величина;

$$\xi_{j+1} = \begin{cases} 1, & y_{ij} > \hat{Y}_j; \\ 0, & y_{ij} \leq \hat{Y}_j; \end{cases}$$

где $y_{ij}, j = \overline{1, n}$ — выборочные шумовые отсчеты, полученные в j -й момент времени; Θ — заданная вероятность превышения порога шумами; \hat{Y}_j — оценка квантиля уровня в на j -м шаге (в j -й момент времени). Пусть V — истинное значение квантиля уровня $1 - \Theta$. Алгоритм (65) строит последовательность оценок $\hat{Y}_j, j = 0, 1, \dots$, которая сходится по вероятности к V , в результате

$$\lim_{j \rightarrow \infty} E (\hat{Y}_j - V)^2 = 0, \quad (66)$$

где E — символ математического ожидания.

Наряду с таким достоинством, как независимость выполнения условия (66) от параметров и вида закона распределения шума, процедуре (65) свойственны существенные недостатки.

В системе (65) с убывающей во времени обратной связью инерционность возрастает, поэтому в условиях нестационарности входного воздействия она не будет работоспособной. В этих условиях более приемлемой оказывается процедура с квазипостоянной обратной связью. Структурная схема адаптивного оценщика, использующего такую процедуру, изображена на рис. 19.

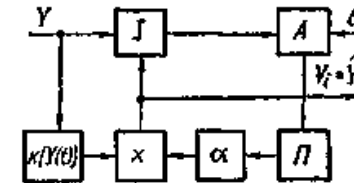


Рис. 19. Структурная схема рекуррентного оценщика квантиля случайного процесса

Здесь $k[Y(t)]$ — функционал входного процесса. Он может представлять собой оценку масштаба на интервале квазистационарности. Кольцо обратной связи содержит дискретный автомат A , преобразователь код—напряжение Π , усилитель с постоянным коэффициентом усиления α и устройство перемножения. Граф работы дискретного автомата показан на рис 20.

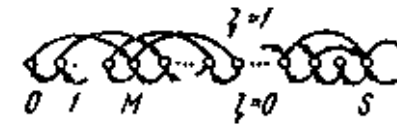


Рис. 20. Граф переходов цифрового автомата

В случае превышения входным процессом уровня квантования состояние дискретного автомата увеличивается на M , иначе — уменьшается на 1.

Условие (66) для такого алгоритма вычисления порога квантования не выполняется, т. е.

$$\lim_{l \rightarrow \infty} (\hat{Y}_l - V)^2 = c,$$

где c — действительное, не равное нулю число. Такой предел может и не существовать. Следовательно, точность, с которой система (рис. 19) отслеживает квантиль V , ограничена, что приводит к некоторой потере эффективности бинарного обнаружителя. В связи с этим рассмотрим вопросы анализа точности таких систем автоматической подстройки порога квантования (АРПК) и анализа эффективности адаптивных бинарных накопителей.

Работа системы АРПК порождает марковскую цепь переходов автомата с матрицей переходных вероятностей

$$\begin{matrix}
 & 0 & 1 & \dots & M & M+1 & M+2 & \dots & S-M-1 & S \\
 \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ S-M \\ \vdots \\ S \end{matrix} & \begin{pmatrix} 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 \\ q_1 & 0 & \dots & 0 & p_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 & 0 & p_2 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & q_{S-M} & p_{S-M} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & q_s & \dots & p_s \end{pmatrix}
 \end{matrix}$$

(67)

Число M определяет значение

$$\Theta = 1/(1 + M),$$

(68)

Приращение порога квантования при переходе автомата из состояния l в состояние $(l + 1)$

$$\Delta V = -\alpha k \{Y(t)\}.$$

(69)

В дальнейшем вместо $k \{Y(t)\}$ будем писать просто k

Вероятность переходов из i -го состояния в состояние с номером $(i+M)$ находится усреднением по плотности распределения коэффициента передачи цепи обратной связи k :

$$p_i = \int_0^\infty \{1 - F(i | \Delta V |)\} \omega(k) dk,$$

(70)

где $F(\cdot)$ — интегральная функция распределения входного процесса; $\omega(k)$ — плотность распределения коэффициента k .

Число состояний автомата s выбирается из условий практической реализуемости обнаружителя и требуемой точности оценки порога квантования.

Очевидно, распределение вероятностей состояний автомата в j -й такт подстройки определяет распределение порога квантования, а, следовательно, и оценки квантиля так как порог функционально связан с номером состояния

$$\hat{V}_i = \hat{Y}_i - \alpha k i, \quad i = \overline{1, s}.$$

(71)

Вектор вероятностей состояний автомата в j -й такт определяется согласно теории цепей Маркова

$$\vec{W}_j = \vec{W}_0 \times \Pi^j,$$

(72)

где $\vec{W}_0 = (1, 0, \dots, 0)$ — начальный вектор вероятностей состояний.

По известным элементам вектора $\vec{W}_j = (v_0^{(j)}, v_1^{(j)}, \dots, v_s^{(j)})$ распределение порога может быть записано следующим образом:

$$\varphi_j(V) = \sum_{i=1}^s v_i^{(j)} \frac{1}{\alpha i} \omega\left(\frac{V}{\alpha i}\right) + v_0^{(j)} \delta(V).$$

(73)

Соответственно вероятность превышения порога входным процессом $Y(t)$ в j -й момент вычисляется по формуле

$$\begin{aligned}
 p_j \{Y > V\} &= \int_0^\infty \times \\
 &\times \{1 - F(V)\} \varphi_j(V) dV.
 \end{aligned}$$

(74)

Рассмотрим методику на примере расчета эффективности адаптивного обнаружителя, осуществляющего накопление бинарно-квантованных сигналов (рис. 21), где k_i является функцией экстремальной статистики

$$k_i = \max \{y_{i1}, \dots, y_{im}\},$$

(75)

где m — объем шумовой выборки.

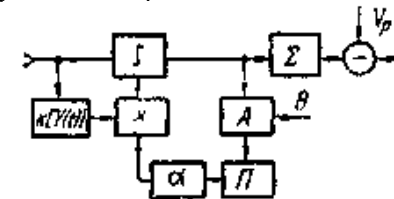


Рис. 21. Структурная схема адаптивного бинарного обнаружителя

Будем считать, что система обучается в течение r тактов. На каждом такте располагаем $m + 1$ отсчетами. Из них m отсчетов используются для оценки k , а один отсчет подвергается квантованию. На $r + 1$ такте на вход квантователя поступает отсчет смеси сигнала и помехи.

Рассмотрим обнаружение гауссова сигнала на фоне гауссового шума. Известно, что в этом случае одномерная плотность процесса после линейного детектирования описывается плотностью Релея (51), а оптимальный порог квантования выбирается из условия $1 - F(V) = 0,203$. Поэтому в формуле (68) $M = 4$.

Особенность решения задачи состоит в определении элементов матрицы (67). С учетом вида плотности распределения экстремальной статистики (75) и (51) находим

$$p_i = \int_0^{\infty} e^{-\frac{(k-\alpha i)^2}{2\Psi}} m \frac{k}{\Psi} e^{-\frac{k}{2\Psi}} \left(1 - e^{-\frac{k}{2\Psi}}\right)^{m-1} dk = m \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} \frac{1}{(\alpha i)^2 + j + 1}. \quad (76)$$

Формула (76) справедлива для установившегося режима в разомкнутом контуре, т. е. когда выборка $y_{i1}, y_{i2} \dots y_{im}$ — однородна. Если имеет место скачок дисперсии входного процесса, т. е. в момент времени $t_j, j \leq r$ отсчетов принадлежат распределению с параметром Ψ_j , плотность распределения вероятностей экстремальной статистики

$$\omega(k) = (m-h) \omega_1(k) F_1^{m-h-1}(k) \cdot F_2^h(k) + h \omega_2(k) F_2^{h-1}(k) F_1^{m-h}(k), \quad (77)$$

где $\omega_i(k)$ и $F_i(k), (i = 1, 2)$ — одномерные дифференциальные и интегральные функции распределения Релея с параметром Ψ_i . С учетом формул (76) и (77) получаем

$$p_{ij} = I_{1,i} + I_{2,i}, \quad (78)$$

где

$$I_{1,i} = \frac{m-h}{\Psi_1} \sum_{j=0}^{m-h-1} (-1)^j \binom{m-h-1}{j} \times \sum_{l=0}^h (-1)^l \binom{h}{l} \frac{\Psi_1 \Psi_2}{(1+j) \Psi_2 + [l + (\alpha i)^2] \Psi_1},$$

$$I_{2,i} = \frac{h}{\Psi_2} \sum_{j=0}^{m-h} (-1)^j \binom{m-h}{j} \times \sum_{l=0}^{h-1} (-1)^l \binom{h-1}{l} \frac{\Psi_1 \Psi_2}{j \Psi_2 + [l+1 + (\alpha i)^2] \Psi_1}.$$

При анализе ситуации, когда в контуре оценки параметра масштаба имеется переходной процесс, следует вычислять элементы матрицы (67) для каждого j -го момента времени. В этом случае марковская цепь переходов автомата A будет неоднородной и формула (72) примет вид:

$$\bar{W}_j = \bar{W}_0 \times \Pi_1 \times \Pi_2 \times \dots \times \Pi_j. \quad (79)$$

Плотность распределения порога и вероятность превышения его сигнальным отсчетом в j -й момент определяется по формулам (73) и (74) с учетом

$$F(V) = 1 - \exp\left[-\frac{V^2}{2\Psi(1+b)}\right],$$

где W — дисперсия шума; b — соотношение сигнал/шум по мощности.

Графики зависимости вероятности превышения порога при $b = 0$ от времени для различных значений α и m в случае переходного процесса (скачок Ψ с 0 до 1) показаны на рис. 22.

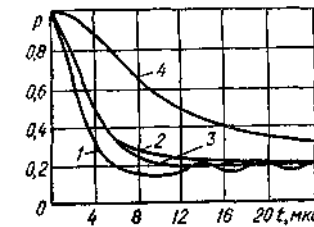


Рис. 22. Графики зависимости вероятности превышения порога при $b = 0$ от t :

$1 - m = 10; \alpha = 0,5, 2 - m = 5, \alpha = 0,3, 3 - m = 10, \alpha = 0,3; 4 - m = 10, \alpha = 0,1$

Зависимость $P = P(t)$ характеризует скорость сходимости последовательности распределений оценок $\varphi_j(V), j = 1, 2$ к распределению, при котором выполняется условие

$$E(\xi - \theta) = 0, \quad (81)$$

где E — символ математического ожидания;

$$\xi = \begin{cases} 1, & y_j > V_j, \\ 0, & y_j \leq V_j. \end{cases}$$

Для построения характеристик обнаружителя (рис. 21), использующего накопление бинарно-квантованных сигналов, воспользуемся нормальной аппроксимацией закона распределения суммы. На рис. 23 показаны графики зависимости порогового сигнала b для вероятности обнаружения $D=0,9$ от объема выборки n ,

просчитанные согласно изложенной методике, для значения вероятности ложной тревоги $F=10^{-4}$.

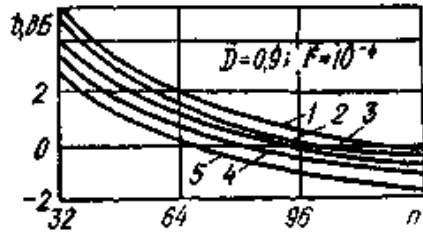


Рис. 23. Графики зависимости порогового сигнала b от n .

1 — $m = 10, \alpha = 0,3$; 2 — $m = 5, \alpha = 0,3$; 3 — $m = 10, \alpha = 0,3$; 4 — $m = 10, \alpha = 0,1$

Здесь же для сравнения приведена кривая 5, отражающая зависимость порогового сигнала от объема выборки для случая, когда значение порога квантования определено точно из условия $p = 0,2$. Очевидно, при проектировании бинарных обнаружителей следует исходить из разумного компромисса между требованиями инерционности системы АРПК, диктуемыми условиями эксплуатации и неизбежными потерями в пороговом сигнале, возникающими вследствие адаптации в условиях априорной неопределенности.

1.9.6. Алгоритмы обнаружения, использующие расслоение выборки

Процедура расслоения выборки используется в двух аспектах: для получения оценок неизвестных параметров по обучающей или рабочей выборке; для расслоения рабочей выборки на две подвыборки с целью раздельной обработки частей по разным алгоритмам в соответствии с формулами (7а, 7б). Рассмотрим второй аспект.

Используем представление функции правдоподобия выборки x_1, \dots, x_n в виде

$$L(x_1 \dots x_n | b) = \prod_{i=1}^n \omega_{\xi_i}(x_i, \Theta_{\xi_i}, b), \quad (82)$$

где b — соотношение сигнал/шум; ξ_i принимает значение 1, если x_i принадлежит основному распределению с плотностью $\omega_1(x, \Theta_1, b)$ и значение 2, если x_i принадлежит распределению с плотностью $\omega_2(x, \Theta_2, b)$. Алгоритм проверки гипотезы H_0 : (сигнал отсутствует) против альтернативы H_1 : $b = b_0$ (сигнал есть) при известном векторе

$\Xi = (\xi_1 \dots \xi_n)$ определяется сравнением отношения правдоподобия с порогом V_p

$$I = \frac{\prod_{i=1}^n \omega_{\xi_i}(x_i, \Theta_{\xi_i}, b_0)}{\prod_{i=1}^n \omega_{\xi_i}(x_i, \Theta_{\xi_i}, b = 0)} > V_p. \quad (83)$$

Если вектор Ξ неизвестен и неизвестно его распределение, то можно пользоваться методом максимального правдоподобия и вычислять обобщенное отношение правдоподобия. Алгоритм имеет вид

$$I = \frac{\max_{\Xi} \prod_{i=1}^n \omega_{\xi_i}(x_i, \Theta_{\xi_i}, b = b_0)}{\max_{\Xi} \prod_{i=1}^n \omega_{\xi_i}(x_i, \Theta_{\xi_i}, b = 0)} > V_p. \quad (84)$$

Оценивание вектора Ξ является расслоением выборки на две подвыборки. Процедура расслоения позволяет записать функцию правдоподобия в виде (97) п.1.8, что приводит к следующему представлению алгоритма (84)

$$I(x_1, \dots, x_n) = \frac{\prod_{i=1}^q \omega_1(x_{k_i}, \Theta_1, b = b_0) \prod_{i=q+1}^n \omega_2(x_{k_i}, \Theta_2, b = b_0)}{\prod_{i=1}^q \omega_1(x_{k_i}, \Theta_1, b = 0) \prod_{i=q+1}^n \omega_2(x_{k_i}, \Theta_2, b = 0)} = I_1(x_{k_1}, \dots, x_{k_q}) \cdot I_2(x_{k_{q+1}}, \dots, x_{k_n}), \quad (85)$$

где $(k_1 \dots k_n)$ — некоторая перестановка чисел $(1, \dots, n)$, или после логарифмирования в эквивалентной форме

$$l'(x_1, \dots, x_n) = l'(x_{k_1}, \dots, x_{k_q}) + l'(x_{k_{q+1}}, \dots, x_{k_n}). \quad (86)$$

Могут применяться различные методы оценивания вектора Ξ . Наиболее очевидным, но не простым, является метод прямого перебора на множестве перестановок чисел $1, \dots, n$. Если $\omega_2(x)$ существенно больше $\omega_1(x)$, как в задаче с «выбросами», возможно применение процедур классификации, описанных в разд. 1.8.6. Обобщенная структурная схема алгоритма обнаружения, использующего процедуру расслоения выборки, изображена на рис. 24.

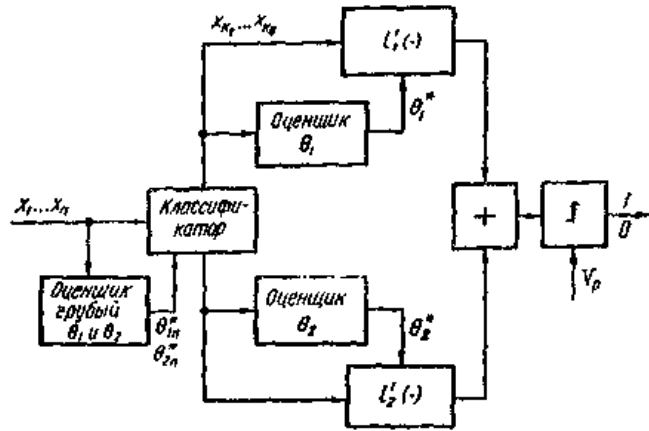


Рис. 24. Структура алгоритма обнаружения с расслоением выборки

Алгоритм осуществляет предварительное оценивание параметров Θ_1 и Θ_2 по нерасслоенной выборке; расслоение выборки на основе одной из процедур классификации; уточнение оценок параметров Θ_1 и Θ_2 ; обработку подвыборок по алгоритмам $l'_1(\cdot)$ и $l'_2(\cdot)$; вычисление статистики $l'(\cdot)$ и принятие решения. Так как при классификации возможны ошибки, алгоритмы $l'_1(\cdot)$ и $l'_2(\cdot)$ должны удовлетворять требованиям устойчивости. Их синтез должен проводиться с учетом распределений оценок $\hat{\Theta}_1$ и $\hat{\Theta}_2$.

Рассмотрим пример построения последетекторного алгоритма обнаружения гауссовского сигнала на фоне внутриприемного шума и ХИП с неизвестной вероятностью появления на основе процедуры классификации.

Плотность распределения отсчетов на выходе детектора для рассматриваемого случая

$$\omega(x, \Psi, u, b) = (1 - p)\omega_1(x, \Psi, b) + p\omega_2(x, \Psi, u, b), \quad (87)$$

где

$$\omega_1(x, \Psi, b) = \frac{x}{\Psi(1+b)} e^{-\frac{x^2}{2\Psi(1+b)}},$$

$$\omega_2(x, \Psi, u, b) = \frac{x}{\Psi(1+b)} e^{-\frac{x^2+u^2}{2\Psi(1+b)}} I_0\left(\frac{ux}{\Psi(1+b)}\right),$$

$I_0(\cdot)$ — модифицированная функция Бесселя нулевого порядка.

Задачу будем решать при следующих допущениях: амплитуда импульсной помехи u существенно больше дисперсии внутриприемного шума ($u \gg \sqrt{\Psi}$), дисперсию шума Ψ будем считать известной; вероятность присутствия импульсной помехи p мала ($p \ll 1$).

Сформулированные допущения позволяют воспользоваться для предварительной оценки дисперсии смеси шума и сигнала $\Psi_1 = \Psi(1+b)$ алгоритмом

$$\hat{\Psi}_1 = \frac{1}{2r} \sum_{i=1}^r x_{(i)}^2, \quad r > n. \quad (88)$$

Необходимая глубина цензурирования определяется по оценке вероятности принадлежности r -й порядковой статистики одной из генеральных совокупностей Γ_1 или Γ_2 с плотностями распределения ω_1 и ω_2 соответственно. Это можно сделать, воспользовавшись формулой

$$p\{x_{(r)} \in \Gamma_i\} = \frac{1-p}{B(r, n-r+1)} \int_{-\infty}^{\infty} \omega_i(x) [(1-p)F_1(x) + pF_2(x)]^{r-1} \times \\ \times [1 - (1-p)F_1(x) - pF_2(x)]^{n-r} dx, \\ i = 1, r; B(\cdot) — бета функция.$$

Величину r выбирают так, чтобы вероятность $p\{x_{(r)} \in \Gamma_2\}$ была достаточно малой.

Применим оценку (88) в алгоритме классификации (99) п.1.8

$$\xi_k = \begin{cases} 1, & x_{(k)} < L(2r\hat{\Psi}_1)^{\frac{1}{2}}; \\ 2, & x_{(k)} > L(2r\hat{\Psi}_1)^{\frac{1}{2}}, \end{cases} \quad (89)$$

где L — некоторый коэффициент.

Численное значение L рассчитывается с учетом задания величины вероятности «ложного» обнаружения выброса, т. е. так, чтобы вероятность превышения порога $L(2\hat{\Psi}_1 r)^{\frac{1}{2}}$ порядковой статистикой $x_{(s)}$ ($s \geq r$) при условии, что импульсных помех нет (выборка однородна), была не выше заданной величины. Запишем без вывода основные соотношения, необходимые для вычисления L .

Пусть s -я порядковая статистика $x_{(s)}$ сравнивается с порогом классификации $L(2\hat{\Psi}_1 r)^{\frac{1}{2}}$, где $\hat{\Psi}_1$ вычислена согласно формуле (88) по первым r порядковым статистикам выборки объемом n , принадлежащей распределению с плотностью $\omega_1(x)$. Плотность распре-

деления вектора порядковых статистик $x_{(1)}, x_{(2)}, \dots, x_{(r)}, x_{(s)}$ имеет вид

$$\omega(x_{(1)}, x_{(2)}, \dots, x_{(r)}, x_{(s)}) = n! \left[\prod_{i=1}^{r+1} \omega_i(x) \right] \frac{[F_1(x_{(s)}) - F_1(x_{(r)})]^{s-r-1}}{(s-r-1)!} \quad (90)$$

Воспользуемся преобразованием

$$\left. \begin{aligned} x_{(1)} &= R \cos \varphi_{r-1} \cos \varphi_{r-2} \dots \cos \varphi_1; \\ x_{(2)} &= R \cos \varphi_{r-1} \dots \sin \varphi_1; \\ x_{(3)} &= R \cos \varphi_{r-1} \dots \sin \varphi_2; \\ &\vdots \\ x_{(r)} &= R \sin \varphi_{r-1}; \\ x_{(s)} &= R \lambda. \end{aligned} \right\} \quad (91)$$

Якобиан преобразования

$$|J| = R^r \prod_{i=1}^{r-1} \cos^{i-1} \varphi_i.$$

Вводя замену переменных $t_i = \sin \varphi_i, i = \overline{1, r-1}$, получим выражение для плотности новых случайных величин $\omega(R, \lambda, t_{r-1}, \dots, t_1)$, которое позволит вычислить вероятность того, что порядковая статистика $x_{(s)}$ меньше порога $L(2\Psi r) = LR$:

$$\begin{aligned} p\{\lambda < L\} &= \int_0^L \int_{\frac{1}{\sqrt{r}}}^L \int_{\frac{1}{\sqrt{r}}}^{\min\{t, \lambda\}} \int_{\frac{1}{\sqrt{r-1}}}^{\min\left\{1, \frac{t_{r-1}}{1-t_{r-1}^2}\right\}} \dots \\ &\dots \int_{\frac{1}{\sqrt{2}}}^{\min\left\{1, \frac{t}{1-t^2}\right\}} \omega(R, \lambda, t_{r-1}, \dots, t_1) \times \\ &\times dR d\lambda dt_{r-1} \dots dt_1. \end{aligned} \quad (92)$$

Приведем явные выражения (92) для некоторых частных случаев:

$$1) n = r = s = 2, \\ p\{\lambda < L\} = 1 - 2(1 - L^2), \quad L \in \left(\frac{1}{\sqrt{2}}, 1\right);$$

$$2) n = r = s = 3,$$

$$p\{\lambda < L\} = \begin{cases} (1 - 3L^2)^2, & L \in \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{2}}\right), \\ 1 - 3(1 - L^2)^2, & L \in \left(\frac{1}{\sqrt{2}}, 1\right); \end{cases}$$

$$3) n = r = s = 4,$$

$$p\{\lambda < L\} = \begin{cases} (4L^2 - 1)^2, & L \in \left(\frac{1}{2}, \frac{1}{\sqrt{3}}\right); \\ 3 - 24L^2 + 53L^4 - 44L^6, & L \in \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{2}}\right), \\ 1 - 4(1 - L^2)^2, & L \in \left(\frac{1}{\sqrt{2}}, 1\right). \end{cases}$$

Пусть в результате классификации в первой подвыборке оказалось q отсчетов, а во второй соответственно $n - q$.

Для синтеза подалгоритмов $I_1(\cdot)$ и $I_2(\cdot)$ воспользуемся ЭБМ. Вычислим логарифм отношения правдоподобия для простых гипотез

$$\ln \frac{\prod_{i=1}^q \omega_i(x_{(i)}, b)}{\prod_{i=1}^q \omega_i(x_{(i)}, b=0)} = -q \ln(1+b) + \frac{\sum_{i=1}^q x_{(i)}^2}{2\Psi} \left(\frac{b}{1+b}\right).$$

При известной дисперсии шума Ψ алгоритм обработки вычисляет статистику

$$s_1 = \frac{1}{\Psi} \sum_{i=1}^n x_{(i)}^2, \quad (93)$$

Так как в результате ошибок при классификации возможно появление «выбросов» в выборке, алгоритму (93) надо обеспечить свойство устойчивости. Это можно сделать, модифицировав (93) следующим образом

$$I_1(\bar{x}, \Psi) = \sum_{i=2}^q \varphi_1(x_{(i)}, \Psi), \quad (94)$$

где

$$\varphi_1(x_{(i)}, \Psi) = \begin{cases} x_{(i)}^2, & x_{(i)} < C_1 \sqrt{\Psi}; \\ x_{(i)}^2, & x_{(i)} > C_1 \sqrt{\Psi}; \end{cases}$$

$$\bar{x}_1 = \frac{1}{q} \sum_{i=1}^q x_{(i)}, \quad C_1 \text{ — некоторый коэффициент.}$$

Запишем логарифм отношения правдоподобия для отсчетов второй подвыборки с плотностью распределения $\omega_2(x, \Psi, u, b)$. Воспользуемся разложением $\ln l_0(x)$ в степенной ряд и ограничимся линейными членами:

$$\ln \frac{\prod_{i=q+1}^n \omega_2(x_{(i)}, b)}{\prod_{i=q+1}^n \omega_2(x_{(i)}, b=0)} \approx - (n-q) \ln(1+b) + \frac{\sum_{i=q+1}^n (x_{(i)} - u)^2}{2\Psi(1+b)}.$$

При известной амплитуде импульсной помехи u и дисперсии шума Ψ алгоритм обработки вычисляет статистику

$$S_2 = \sum_{i=q+1}^n \frac{(x_{(i)} - u)^2}{\Psi}.$$

В соответствии с ЭБМ заменим в этом алгоритме неизвестный параметр u его оценкой

$$\hat{u} = \frac{1}{n-q} \sum_{i=q+1}^n x_{(i)}, \quad (95)$$

а свойство устойчивости алгоритму $l'_2(\cdot)$ обеспечим, модифицировав статистику S_2 так же, как и в первом случае. Таким образом,

$$l'_2(\bar{x}, \Psi, \hat{u}) = \sum_{i=q+1}^n \varphi_2(x_{(i)}, \Psi), \quad (96)$$

где

$$\varphi_2 = (x_{(i)}, \Psi) = \begin{cases} \frac{(x_{(i)} - \hat{u})^2}{\Psi}, & (x_{(i)} - \hat{u})^2 > C_2 \Psi; \\ \bar{x}_i^2, & (x_{(i)} - \hat{u})^2 > c_2 \Psi. \end{cases}$$

В целом алгоритм обработки, использующий процедуру классификации, определяется формулами (88), (89), (94), (96), а его структура соответствует рис. 24.

На рис. 25 показаны характеристики обнаружения описанного алгоритма, вычисленные методом статистических испытаний для разных значений вероятности появления помехи p .

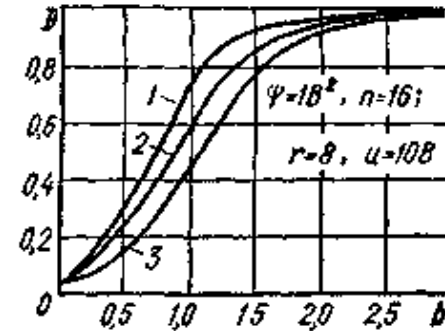


Рис. 25. Характеристики обнаружителя, использующего расслоение выборки:

- 1 — Ξ — известен $p = 0,1$;
- 2 — Ξ — неизвестен, $p = 0,1$;
- 3 — Ξ — неизвестен, $p = 0,2$

Проведем синтез двухвыборочного алгоритма обнаружения, статистика которого свободна от значений мешающих параметров u и Ψ в модели (87). Рассмотрим две независимые выборки $(x_1^c \dots x_n^c)$ и $(x_1^w \dots x_m^w)$. Первую выборку назовем рабочей или сигнальной, вторую — опорной или обучающей. Запишем функцию правдоподобия объединенной выборки

$$\begin{aligned} L(x_1^c, \dots, x_n^c, x_1^w, \dots, \\ \dots, x_m^w, \Theta, b) = \\ = \prod_{i=1}^n \omega_{\xi_i}(x_i^c, \Theta, b) \times \\ \times \prod_{i=1}^m \omega_{\xi_i}(x_i^w, \Theta, b=0), \end{aligned} \quad (97)$$

где ξ_i принимает значение 1 или 2 в зависимости от принадлежности выборочного значения x_i генеральной совокупности с распределением ω_1 или ω_2 , Θ — вектор параметров помехи. В нашем случае $\Theta = [u, \Psi]$. Относительно объединенной выборки проверяется сложная гипотеза $H_0 : b = 0$, против сложной альтернативы $H_1 : b = b_0 \neq 0, \Theta \in \Omega_\Theta$, где Ω_Θ — пространство значений вектора параметров.

Пусть $\xi = \{\xi_i, \xi_j\}$, $i = 1, n, j = 1, m$ определен в разд. 1.8.6, $\Xi = \{\xi_i, \xi_j\}$, $i = 1, n, j = 1, m$ известен.

Тогда функция правдоподобия объединенной выборки с одномерным распределением (87) имеет вид

$$\begin{aligned}
 & L(x_1^c, \dots, x_n^c, x_1^u, \dots, x_m^u, u, \Psi, b) = \\
 & = \prod_{i=1}^{r_c} \omega_1(x_{k_i}^c, u, \Psi, b) \prod_{j=1}^{r_u} \omega_1(x_{l_j}^u, u, \Psi, b=0) \times \\
 & \times \prod_{i=r_c+1}^n \omega_2(x_{k_i}^c, u, \Psi, b) \prod_{j=r_u+1}^m \omega_2(x_{l_j}^u, u, \Psi, b=0) = \\
 & = \left[\frac{1}{\Psi} \right]^{m+n} \left[\frac{1}{1+b} \right]^n \prod_{i=1}^n x_{k_i}^c \prod_{j=1}^m x_{l_j}^u \times \\
 & \times e^{-\frac{1}{2\Psi} \left\{ \sum_{j=1}^{r_u} (x_{l_j}^u)^2 + \sum_{i=r_u+1}^m (x_{l_i}^u - u)^2 \right\} +} \\
 & + \frac{1}{1+b} \left[\sum_{i=1}^{r_c} (x_{k_i}^c)^2 + \sum_{i=r_c+1}^n (x_{k_i}^c - u)^2 \right] \}. \quad (98)
 \end{aligned}$$

Здесь $\{x_{k_i}, x_{l_j}\}, i = 1, \dots, r_c, j = 1, \dots, r_u$ — множество выборочных значений рабочей и опорной выборок, из генеральной совокупности Γ_1 с плотностью распределения ω_1 ; $\{x_{k_i}, x_{l_j}\}, i = r_c + 1, \dots, n, j = r_u + 1, \dots, m$ — множество выборочных значений из генеральной совокупности Γ_2 с плотностью распределения ω_2 .

Структура решающего правила при известном векторе Ξ и неизвестном векторе параметров $\{u, \Psi\}$ определяется отношением функций правдоподобия для конкурирующих гипотез, усредненных по возможным значениям параметров,

$$\lambda = \frac{\int_{-\infty}^{\infty} \int_0^{\infty} L(x_1^c, \dots, x_n^c, x_1^u, \dots, x_m^u, u, \Psi, b) dud\Psi}{\int_{-\infty}^{\infty} \int_0^{\infty} L(x_1^c, \dots, x_n^c, x_1^u, \dots, x_m^u, u, \Psi, b=0) dud\Psi} \quad (99)$$

Подставляя выражение (98) в формулу (99) и проводя интегрирование, получим

$$\begin{aligned}
 & \lambda(x_1^c, \dots, x_n^c, x_1^u, \dots, x_m^u, b) = \\
 & = \left[\frac{n - r_c + (1+b)(m - r_u)}{(1+b)(n - r_c + m - r_u)} \right]^{n+m} \times
 \end{aligned}$$

$$\begin{aligned}
 & (n + m - r_c - r_u) \left(\sum_{i=1}^n (x_i^c)^2 + \sum_{j=1}^m (x_j^u)^2 - \right. \\
 & \times \left. \left[\frac{- \left(\sum_{i=r_c+1}^n x_{k_i}^c + \sum_{j=r_u+1}^m x_{l_j}^u \right)^2}{\frac{n - r_c + (1+b)(m - r_u)}{(1+b)} \left(\frac{1}{1+b} \sum_{i=1}^n (x_i^c)^2 \right) +} \right. \right. \\
 & \left. \left. + \sum_{j=1}^m (x_j^u)^2 - \left(\frac{1}{1+b} \sum_{i=r_c+1}^n x_{k_i}^c + \sum_{i=r_u+1}^m x_{l_i}^u \right)^2 \right] \right) \quad (100)
 \end{aligned}$$

Введем обозначения

$$\begin{aligned}
 \hat{U}_0 & = \frac{1}{n + m - r_c - r_u} \left[\sum_{i=r_c+1}^n x_{k_i}^c + \sum_{j=r_u+1}^m x_{l_j}^u \right]; \\
 \hat{U}_1 & = \frac{1+b}{n + m(1+b)} \left[\frac{1}{1+b} \sum_{i=r_c+1}^n x_{k_i}^c + \sum_{j=r_u+1}^m x_{l_j}^u \right].
 \end{aligned}$$

Тогда выражение (100) можно привести к виду

$$\begin{aligned}
 & \lambda(x_1^c, \dots, x_n^c, x_1^u, \dots, x_m^u, b) = \\
 & = \left[\frac{(n + m - r_c - r_u)(1+b)}{n - r_c + (1+b)(m - r_u)} \right]^{1/2} \times \\
 & \times \left[\frac{\sum_{j=1}^{r_u} (x_{l_j}^u)^2 + \sum_{j=r_u+1}^m (x_{l_j}^u - \hat{U}_0)^2 +}{(1+b) \left[\sum_{j=1}^{r_u} (x_{l_j}^u)^2 + \sum_{j=r_u+1}^m (x_{l_j}^u - \hat{U}_1)^2 \right] +} \right]^{n+m+\frac{1}{2}} \\
 & \times \left[\frac{\sum_{i=1}^{r_c} (x_{k_i}^c)^2 + \sum_{i=r_c+1}^n (x_{k_i}^c - \hat{U}_0)^2}{\sum_{i=1}^{r_c} (x_{k_i}^c)^2 + \sum_{i=r_c+1}^n (x_{k_i}^c - \hat{U}_1)^2} \right]
 \end{aligned} \quad (101)$$

Используя в выражении (100) оценку \hat{U}_0 вместо \hat{U}_1 и вводя статистику

$$Z(x_1^c, \dots, x_n^c, x_1^w, \dots, x_m^w) = \frac{\sum_{i=1}^{r_c} (x_{k_i}^c)^2 + \sum_{i=r_c+1}^n (x_{k_i} - \hat{U}_0)^2}{\sum_{j=1}^{r_w} (x_{l_j}^w)^2 + \sum_{j=r_w+1}^m (x_{l_j} - \hat{U}_0)^2},$$

подоптимальное решающее правило можно записать следующим образом:

$$\lambda(x_1^c, \dots, x_n^c, x_1^w, \dots, x_m^w, b) = \left[\frac{(n + m - r_c - r_w)(1 + b)}{n - r_c + (1 + b)(m - r_w)} \right]^{1/2} \times \left[\frac{1 + z}{1 + b + z} \right]^{n+m+1/2} > V_p. \quad (102)$$

Учитывая, что выражение (102) является монотонной функцией переменной z , решающее правило (102) можно заменить эквивалентным тестом

$$z(x_1^c, \dots, x_n^c, x_1^w, \dots, x_m^w) = \frac{\sum_{i=1}^{r_c} (x_{k_i}^c)^2 + \sum_{i=r_c+1}^n (x_{k_i}^c - \hat{U}_0)^2}{\sum_{j=1}^{r_w} (x_{l_j}^w)^2 + \sum_{j=r_w+1}^m (x_{l_j}^w - \hat{U}_0)^2} > V_p. \quad (103)$$

Так как на стадии синтеза принято допущение $\hat{U}_1 = \hat{U}_0$, решающее правило (103) является близким к оптимальному, но не оптимальным. При этом существенно упрощается техническая реализация алгоритма обнаружения. Потери в эффективности, возникающие вследствие указанной замены, невелики.

1.9.7. Комбинированные адаптивно- непараметрические алгоритмы в сложных помеховых ситуациях

В сложных помеховых ситуациях целесообразно комбинирование методов синтеза алгоритмов обнаружения, изложенных в предыдущих параграфах. Так, например, распределение статистики (33) в примере 19 существенно зависит от вероятности появления импульсной помехи

в выборке, что затрудняет решение задачи стабилизации вероятности ошибок первого рода. Одним из возможных выходов в этой ситуации является последующая непараметрическая обработка, т. е. комбинирование методов адаптации и инвариантности.

В алгоритмах с расслоением выборки (см. разд. 1.9.6) в зависимости от априорной информации о плотностях $\omega_1(\cdot)$ и $\omega_2(\cdot)$ подалгоритмы $l'_1(\cdot)$ и $l'_2(\cdot)$ могут быть как параметрическими, так и непараметрическими.

В настоящем параграфе решается задача синтеза оптимального рангового обнаружителя радиолокационного сигнала на фоне коррелированной пассивной помехи. Известные алгоритмы обычно предполагают априорное знание вида корреляционной функции сигнала и помехи (с точностью до параметров) при нормальном распределении вероятностей мгновенных значений. Отклонения истинного распределения от нормального приводят к нарушению условия стабильности вероятности ложной тревоги.

Предлагается следующий способ решения этой задачи. Техническими средствами выборка формируется из n независимых между собой векторов, состоящих из m коррелированных величин. Независимость обеспечивается разнесением сигналов по частоте, времени или поляризации. Из каждого вектора сформируем один скалярный отсчет таким образом, чтобы статистика, составленная из n таких отсчетов, являлась достаточной. Например, при известной гипотезе и альтернативе достаточную статистику можно получить из отношения правдоподобия

$$L = \frac{\prod_{i=1}^n \omega_{c+w}(x_{i1}, \dots, x_{im})}{\prod_{i=1}^n \omega_w(x_{i1}, \dots, x_{im})} = \prod_{i=1}^n \lambda_i(x_{i1}, \dots, x_{im}),$$

где $\omega_{c+w}(\cdot)$ — плотность распределения смеси сигнала и шума; $\omega_w(\cdot)$ — плотность распределения шума.

Вектор $\lambda = (\lambda_1, \dots, \lambda_n)$ является достаточной статистикой, величины λ_i независимы и в случае отсутствия сигнала одинаково распределены. Алгоритм, обладающий свойством подобия (сохраняющий вероятность ложной тревоги), ищется среди критериев, основанных на перестановках величин λ_i . Налагая дополнительно условия инвариантности относительно всех монотонных преобразований величин λ_i , переходим к ранговым статистикам.

Рассмотрим задачу обнаружения нормального шумоподобного сигнала на фоне нормальной коррелированной помехи. Такая задача возникает при обработке сигналов на выходе аналогового синхронного

(когерентного) детектора, где плотность вероятности процесса описывается многомерным нормальным законом. С учетом сказанного выше задачу обнаружения можно сформулировать в терминах проверки гипотез о m -мерной плотности распределения процесса.

Пусть в результате наблюдений формируется матрица выборочных значений

$$\begin{pmatrix} x_{11} & \dots & x_{1i} & \dots & x_{1n} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mi} & \dots & x_{mn} & \dots & x_{mn} \end{pmatrix},$$

где

$$x_{jt}, \quad j = 1, \dots, m; \quad i = 1, \dots, n$$

— выборочные отсчеты, полученные n_i -кратным зондированием сигнального элемента разрешения по дальности и n — n_i -кратным зондированием шумовых элементов разрешения в m соседних периодах повторения зондирующего сигнала. Гипотеза H_0 (отсутствие сигнала) состоит в равенстве m -мерных законов распределения в сигнальном и шумовых элементах разрешения

$$\begin{aligned} \omega_m(x_1, \dots, x_m) &= \omega_c(x_1, \dots, x_m) = \\ &= \omega(x_1, \dots, x_m; b = 0). \end{aligned}$$

Альтернатива H_1 (наличие сигнала) состоит в том, что m -мерная функция распределения вероятностей отсчетов в сигнальном элементе разрешения отличается от таковой в элементе разрешения, где сигнала нет, значением сигнального параметра $b \neq 0$:

$$\begin{aligned} \omega_m(x_1, \dots, x_m) &\neq \omega_c(x_1, \dots, x_m) = \\ &= \omega(x_1, \dots, x_m; b). \end{aligned}$$

Проведем синтез локально оптимального подобного рангового критерия для задачи обнаружения шумоподобного нормального сигнала в коррелированных нормальных шумах при $m = 2$. Плотность распределения смеси сигнала и шума определяется выражением

$$\begin{aligned} \omega(x_1, x_2; b) &= \frac{1}{2\pi\Psi(1+b)} \sqrt{\frac{r^2}{1 - \frac{r^2}{(1+b)^2}}} \times \\ &\times \exp\left\{-\frac{x_1^2 + x_2^2 - \frac{2r}{1+b}x_1x_2}{2\left(1 - \frac{r^2}{(1+b)^2}\right)\Psi(1+b)}\right\}, \end{aligned} \quad (104)$$

где Ψ — дисперсия шума; r — коэффициент корреляции шума; b — соотношение сигнал/шум по мощности (сигнальный параметр). Легко

показать, что локально достаточной статистикой в рассматриваемой задаче проверки гипотез будет вектор

$$\lambda = (\lambda_1, \dots, \lambda_n), \quad (105)$$

где

$$\lambda_i = \frac{\frac{\partial}{\partial b} \omega(x_{1i}, \dots, x_{mi}; b) |_{b=0}}{\omega(x_{1i}, \dots, x_{mi}; 0)}.$$

Аналогично общему случаю подобный (сохраняющий постоянство вероятности ложной тревоги) тест для этой задачи ищется среди критериев, основанных на перестановках величин λ_i . Так как практическая реализация перестановочных тестов возможна лишь при небольших n , то, налагая дополнительно условия инвариантности относительно всех монотонных преобразований величин λ_i , переходим к ранговым критериям (ранг — это функция перестановок). Для этого произведем ранжировку статистики λ (105) и получим вектор рангов

$$R(\lambda) = (R_1, \dots, R_{n1}, R_{n1+1}, \dots, R_n).$$

Для плотности вероятности (104) величина λ_i определяется выражением

$$\begin{aligned} \lambda_i &= \frac{\frac{\partial}{\partial b} \omega(x_1, x_2, b) |_{b=0}}{\omega(x_1, x_2, 0)} \frac{1}{(1-r^2)} + (1+r^2) \times \\ &\times \left(x_1^2 - \left(\frac{1-r^2}{1+r^2} + 1\right) 2rx_1x_2 + x_2^2\right) / (2\Psi(1-r^2)^2). \end{aligned}$$

При известном r в качестве λ_i также можем использовать

$$\lambda_i = x_1^2 - \left(\frac{1-r^2}{1+r^2} + 1\right) 2rx_1x_2 + x_2^2.$$

Зная совместную плотность распределения величин x_1 и x_2 , можно получить плотность распределения величины λ_i .

$$\begin{aligned} \omega(\lambda_i, b) &= \frac{1}{2\pi\Psi(1+b)} \sqrt{\frac{r^2}{1 - \frac{r^2}{(1+b)^2}}} \times \\ &\times \sum_{j=1}^2 \int_{-\infty}^{\infty} \frac{\exp\left\{-\frac{u^2 + \frac{2r}{1+b}u\varphi_j(u, \lambda_i) + \varphi_j(u, \lambda_i)}{2\left(1 - \frac{r^2}{(1+b)^2}\right)\Psi(1+b)}\right\}}{|-\alpha u + 2\varphi_j(u, \lambda_i)|} du, \end{aligned} \quad (106)$$

где

$$\varphi_i(u, \lambda_i) = \frac{\lambda_i u}{u^2 - \lambda_i} + (-1)^i \sqrt{u^2 \left(\frac{\alpha^2}{4} - 1 \right) + \lambda_i},$$

$$\alpha = 2r \left(\frac{1-r^2}{1+r^2} + 1 \right).$$

Интеграл (106) может быть вычислен только численно. Зная $\omega(\lambda_i, b)$, можем записать функцию правдоподобия для вектора рангов

$$L(R(\lambda), b) = n_1! (n - n_1)! \int_{-\infty < \lambda_1, \dots} \dots$$

$$\dots \int_{\dots \lambda_n < \infty} \prod_{i=1}^n d[F(\lambda_i, 0)^{1-e_i} F(\lambda_i, b)^{e_i}],$$

где e_i — индикаторная функция, принимающая значение 1 в случае, если ранг i есть в проверяемой выборке и 0 в противном случае; $F(\lambda_i, b)$ — функция распределения величины λ_i .

Учитывая, что проверочная статистика локально оптимального критерия определяется для ранговых критериев производной функции правдоподобия по сигнальному параметру b , получаем

$$\frac{\partial}{\partial b} L(R(\lambda), b)_{b=0} = n_1! (n - n_1)! \sum_{i=1}^n e_i a_i,$$

где

$$a_i = \frac{1}{(i-1)! (n-i)!} \int_0^1 t^{i-1} (1-t)^{n-i} dQ(t), \quad (107)$$

$$Q(t) = \frac{\partial F[F^{-1}(t, 0) b]}{\partial b}$$

при $b = 0$.

Коэффициенты a_i вычислялись численно по выражению (107). Характеристики обнаружения, полученные методом статистических испытаний, показаны на рис. 26.

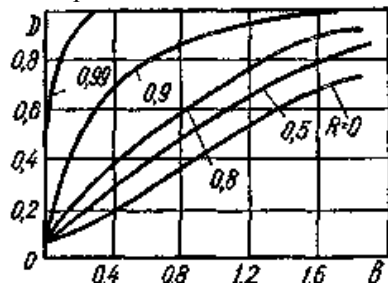


Рис. 26. Характеристики адаптивно-непараметрического обнаружителя

Предложенная процедура обнаружения может применяться и в других случаях, когда сигнал описывается векторной величиной.

1.9.8. Сравнение эффективности алгоритмов обнаружения

Рассмотрим эффективность четырех алгоритмов обнаружения, синтезированных различными методами в задаче последетекторного обнаружения шумоподобного сигнала на фоне смеси гауссовского шума с дисперсией Ψ и импульсной помехи с неизвестной вероятностью появления p и амплитудой U . Отсчеты сигнальной выборки распределены с плотностью вероятности (14). При отсутствии сигнала плотность распределения вероятностей отсчетов соответствует формуле (14) при $\Psi_c = 0$. Наблюдатель располагает двумя выборками — сигнальной и обучающей с объемами n и m соответственно.

Рассмотрим следующие алгоритмы обнаружения.

1. Алгоритм (15), использующий в соответствии с ЭБМ оценки неизвестных параметров p, U, Ψ, Ψ_c , полученные по методу моментов (17).
2. Алгоритм (38), синтезированный в соответствии с ОЭБМ, использующий L -оценки параметров масштаба $\Psi, \Psi + \Psi_c$ сигнальной и обучающей выборок. Оценки параметров p и U в этом алгоритме не используются.
3. Оптимальный ранговый алгоритм, синтезированный в соответствии с формулами (49) ... (55) для $p = 0,2$.
4. Адаптивный алгоритм обнаружения (103), синтезированный в соответствии с ОЭБМ. Для оценки параметров вектора Ξ (см. разд. 1.9.6) используется процедура расслоения выборки по критерию минимума суммы дисперсий (95) п.1.8.

Для анализа эффективности проведено статистическое моделирование отсчетов смеси сигнала и помехи (только помехи), распределенных по закону (14) с параметрами $p = 0,2, U = 10 В, \Psi = 1 В^2$ при различных значениях дисперсии сигнала Ψ_c . Сформированные выборки (сигнальная и опорная) обрабатывались в соответствии с описанными четырьмя алгоритмами. Порог решения определялся по многократным испытаниям (до 1000) таким образом, чтобы вероятность ложной тревоги была 0,01. Вероятность обнаружения при фиксированном отношении $b = \Psi_c / \Psi$ оценивалась по выборке 500.

Характеристики обнаружения сравниваемых алгоритмов при объемах сигнальной и опорной выборок $n = m = 16$ показаны на рис. 27.

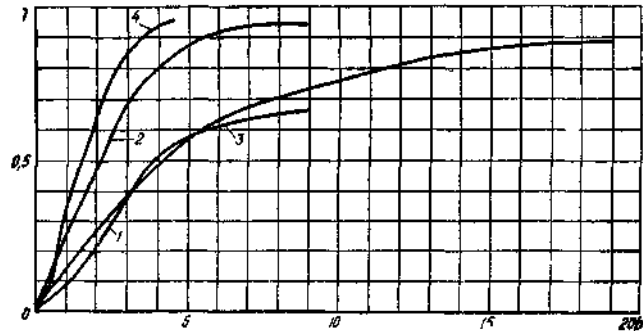


Рис. 27. Характеристики обнаружения. 4 — обнаружителя, синтезированного в соответствии с ОЭБМ; 2 — L -обнаружителя; 3 — рангового обнаружителя; 1 — обнаружителя использующего оценки по методу моментов (ЭБМ)

Данные статистического моделирования показывают, что наиболее эффективным является четвертый алгоритм обнаружения (103), синтезированный в соответствии с ОЭБМ и использующий процедуру расслоения выборки (95) п.1.8. Ему уступает в эффективности второй алгоритм, использующий L -оценки параметров Ψ , Ψ_c . Заметим, что

при заданных характеристиках входной смеси ($p = 0,2$) в этом алгоритме пришлось использовать глубокое цензурирование выборок. Отбрасывались семь наибольших отсчетов каждой из выборок. При увеличении p эффективность этого алгоритма обнаружения снижается, так как приходится увеличивать глубину цензурирования. Подобные алгоритмы эффективны при малых вероятностях появления импульсных помех.

Характеристики первого и третьего алгоритмов обнаружения идут существенно ниже. Причем на малых соотношениях сигнал/шум первый обнаружитель имеет небольшое преимущество перед третьим алгоритмом (оптимальным ранговым) обнаружения, а в области больших соотношений сигнал/шум эффективность этих алгоритмов примерно одинакова.

Результаты моделирования показали, что эффективность первого алгоритма обнаружения, построенного на основании ЭБМ, существенно зависит от качества оценок неизвестных параметров. Оценки, получаемые по методу моментов при малых объемах обучающей выборки (в нашем случае $m=16$), имеют выбросы, т. е.

распределение сильно отличается от нормального, что и объясняет низкую по сравнению со вторым и четвертым алгоритмами эффективность.

При увеличении объема обучающей выборки по эффективности первый алгоритм обнаружения асимптотически стремится к четвертому. Таким образом, использование метода моментов для оценивания неизвестных параметров совокупностей с составным распределением типа (14) целесообразно только при больших объемах обучающей выборки. Объем выборки должен быть таким, чтобы вероятность отсутствия в ней выбросов была незначительной.

Относительно оптимального рангового обнаружителя сделаем следующие замечания. Вид весовой функции рангов отсчетов существенно зависит от априорной вероятности воздействия импульсной помехи p . Если структура обнаружителя соответствует параметрам воздействующих помех, эффективность его, как следует из рис. 27, достаточно высока. Однако при несоответствии истинного и априорного значений p , принятых на стадии синтеза, эффективность оптимального рангового обнаружителя существенно снижается. Следовательно, ранговые обнаружители, применяемые в задачах с составными распределениями типа (14), должны использовать оценки вероятности воздействия импульсной помехи и в соответствии с этими оценками адаптировать свою структуру.

1.10. Устойчивые алгоритмы фильтрации сигналов

Задача фильтрации информационного параметра $\Theta_n(t)$ имеет разнообразные приложения в технике, является сложной и в целом недостаточно разработанной. Как и в задачах оценивания информационный параметр $\Theta_n(t)$ (опуская далее индекс «и» и обозначая просто $\Theta(t)$) можно интерпретировать: как сообщение $\alpha(t)$; сигнал, несущий сообщение $s\{\alpha(t), \beta(t), t\}$; как параметр плотности распределения вероятности некоторого нестационарного случайного процесса.

Различная степень априорной определенности о динамике процесса $\Theta(t)$, конкретная интерпретация $\Theta(t)$, а также вид взаимодействия с помехами (вид уравнения наблюдения $x(t) = F[\Theta(t), n(t)]$ и степень априорного знания их статистических характеристик) определяют существующее многообразие методов синтеза алгоритмов фильтрации сигналов. При

дискретном по времени формировании вектора наблюдений $\vec{x}_n = \{x_1, \dots, x_n\}$ результатом обработки является вектор оценок $\vec{\Theta}_p = \{\Theta_1, \dots, \Theta_p\}$. Параметр p отражает модельные представления о выделяемой функции $\Theta(t)$ (например, p -степень полинома, Θ_i , $i = \overline{1, p}$ — коэффициенты полинома, $p \leq n$). В задачах интерполяции — экстраполяции параметр p определяется шагом h интерполяции — экстраполяции и может быть много больше n . В задачах рекуррентной фильтрации в собственном смысле $p = n$, $\Theta_i = f_i(x_n)$, $i = \overline{1, p}$. Функции $f_i(\cdot)$ подлежат отысканию (регулярному синтезу или эвристическому поиску) с учетом выбранного критерия эффективности.

Теория фильтрации сигналов представлена условно **тремя направлениями**: классическим (метод наименьших квадратов, интерполяционные полиномы Ньютона, Чебышева, сплайны различного порядка); байесовой теорией синтеза рекуррентных алгоритмов нелинейной фильтрации марковских сигналов на фоне марковских помех Р. Л. Стратоновича, частным случаем которой является теория линейных фильтров Калмана — Бьюси — Винера; небайесовыми, в том числе эмпирическими байесовыми (адаптивными) и эвристическими методами, не представляющими в настоящее время единой теории.

Анализ устойчивости известных алгоритмов фильтрации сигналов к аномальным наблюдениям вида (3) п.1.7 свидетельствует, что как классические, так и байесовы алгоритмы являются неустойчивыми (см., например, разд. 1.10.2). Найденные в ряде работ алгоритмы фильтрации сигналов при наличии дискретной компоненты в помехах (совокупность импульсных и гауссовых помех) предполагают априорно точное знание вероятностей появления p в формуле (3) п.1.7 и амплитуды (или дисперсии) импульсной помехи и неработоспособны, если реальные характеристики выбросов существенно отличны от априорных. Поэтому основное внимание уделяется поиску и синтезу устойчивых алгоритмов фильтрации в рамках методов третьего небайесовского направления. Рассматривается адаптивный алгоритм фильтрации полиномиального сигнала в $\Theta(t)$ на фоне аддитивной совокупности гауссовой и импульсных помех (разд. 1.10.1). Систематизируются известные и описываются новые алгоритмы фильтрации случайных процессов $\Theta(t)$ на фоне гауссовых и импульсных помех (разд. 1.10.2), синтез которых выполняется регулярным методом.

1.10.1. Устойчивые адаптивные алгоритмы фильтрации квазидетерминированных сигналов

Рассмотрим случай фильтрации квазидетерминированного сигнала вида

$$\Theta_i = \Theta(t_i) = \sum_{r=1}^N a_r \varphi_r(t_i), \quad i = \overline{1, n}, \quad (1)$$

поступающего на вход спецпроцессора обработки в аддитивной совокупности гауссова шума и потока двухполярных выбросов (аномальных ошибок), имеющих неизвестные амплитуды: $u_1 > 0$, $u_2 < 0$, постоянные в «скользящем окне» из n наблюдений, вероятности появления которых обозначим p_1 и p_2 соответственно. Запишем функцию правдоподобия соотношением

$$\begin{aligned} L(\vec{x}_n / \vec{a}_r, p_1, p_2, u_1, u_2, \sigma^2) &= \\ &= \prod_{i=1}^n \omega(x_i / \vec{a}_r, p_1, p_2, u_1, u_2, \sigma^2) = \\ &= \prod_{i=1}^n \left\{ (1 - p_1 - p_2) \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_i - \Theta_i)^2}{2\sigma^2}\right] + \right. \\ &\quad \left. + p_1 \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_i - \Theta_i - u_1)^2}{2\sigma^2}\right] + \right. \\ &\quad \left. + p_2 \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_i - \Theta_i + u_2)^2}{2\sigma^2}\right] \right\}, \quad (2) \end{aligned}$$

где $\vec{a}_r = \{a_1, \dots, a_N\}$; σ^2 — неизвестная дисперсия гауссова шума;

$$\Theta_i = \sum_{r=1}^N a_r \varphi_{ri}; \quad \varphi_{ri} = \varphi_r(t_i).$$

В такой постановке решение задачи фильтрации сводится к применению метода максимального правдоподобия теории оценивания (см. раз. 1.8). Система нелинейных уравнений метода максимального правдоподобия для неизвестных параметров в общем случае сложна. Для ее упрощения необходимо использовать двухвыборочные процедуры. Эксперимент при этом организуют таким образом, чтобы выключив датчик сообщений, получить выборку только помех и оценить их неизвестные параметры $p_1, p_2, u_1, u_2, \sigma^2$, используя методы моментов или максимального правдоподобия.

Оценки $\hat{p}_1, \hat{p}_2, \hat{u}_1, \hat{u}_2, \hat{\sigma}^2$ подставляются в формулу (2), а дифференцирование логарифма функции правдоподобия (2) по параметрам $\vec{a}_r, r = \overline{1, N}$ дает укороченную систему уравнения метода максимального правдоподобия

$$f_r = \frac{\partial \ln L(\cdot)}{\partial a_r} = \sum_{i=1}^n \omega^{-1}(x_i/(\cdot)) \left\{ (1 - \hat{p}_1 - \hat{p}_2) \times \right. \\ \times \exp \left[-\frac{(x_i - \Theta_i)^2}{2\hat{\sigma}^2} \right] (x_i - \Theta_i) \varphi_{ri} + \\ + \hat{p}_1 \exp \left[-\frac{(x_i - \Theta_i - \hat{u}_1)^2}{2\hat{\sigma}^2} \right] (x_i - \Theta_i - \hat{u}_1) \varphi_{ri} + \\ \left. + \hat{p}_2 \exp \left[-\frac{(x_i - \Theta_i + \hat{u}_2)^2}{2\hat{\sigma}^2} \right] (x_i - \Theta_i + \hat{u}_2) \varphi_{ri} \right\} = 0, \quad (3) \\ r = \overline{1, N}.$$

Решение системы (3) целесообразно искать методом скорейшего спуска, используя в качестве нулевого приближения \hat{a}_r^0 решение линеаризованной системы, которая получается при замене в выражении (3) всех экспонент константами. По найденным в соответствии с заданным критерием останова итеративного процесса оценкам \hat{a}_r восстанавливается фильтруемый процесс (1)

$$\hat{\Theta}_i = \sum_{t=1}^N \hat{a}_r \varphi_r(t_i), \quad t_i = T(i-1), \quad i = \overline{1, n}, \quad (4)$$

решаются задачи интерполяции ($t_n = T(n-1) - \Delta t, 0 < \Delta t < T$) или экстраполяции ($t_n = T(n-1) + \Delta t, 0 < \Delta t < \infty$) обрабатываемого процесса.

Рассмотренный метод обобщает известные методы фильтрации полиномиальных, полигармонических (с известными частотами компонент) и других произвольных сигналов вида (1) на случай наличия в гладкой гауссовой составляющей помехи выбросов.

Рассмотрим конкретный пример.

Пример 20. Случай фильтрации полинома 2-го порядка

$$\Theta_i = \sum_{r=1}^2 a_r t_i^r \quad (5)$$

на фоне гауссова шума и однополярных импульсных помех с неизвестной амплитудой u_i и неизвестной вероятностью появления p_i . Ставится задача разработки и оценки эффективности рекуррентного алгоритма фильтрации — экстраполяции. К качеству возможных

применений такого алгоритма можно отметить дистанционные системы прогнозирующего автоконтроля аналоговых параметров промышленного оборудования, системы предупреждения столкновений в транспортных АСУ и др.

Исходная одномерная ПРВ нестационарного процесса при сделанных предположениях имеет вид

$$\omega(x_i/a_1, a_2, p_1, u_1, \sigma^2) = \omega(x_i/(\cdot)) = \\ = (1 - p_1) N \left(\Theta_i = \sum_{r=1}^2 a_r t_i^r, \sigma^2 \right) + p_1 N(\Theta_i + u_1, \sigma^2) \quad (6)$$

как частный случай i -й компоненты в записи ФП (2) при $p_2 = 0$. На основании формул (5), (6), учитывая, что $\varphi_{ri} = t_i^n, r=1, 2$, систему уравнений метода максимального правдоподобия (3) запишем в виде:

$$f_r = \sum_{i=1}^n \omega^{-1}(x_i/(\cdot)) \left\{ (1 - \hat{p}_1) \exp \left[-\frac{\left(x_i - \sum_{r=1}^2 a_r t_i^r \right)^2}{2\hat{\sigma}^2} \right] \times \right. \\ \times \left(x_i - \sum_{r=1}^2 a_r t_i^r \right) t_i^r + \hat{p}_1 \exp \left[-\frac{\left(x_i - \sum_{r=1}^2 a_r t_i^r - \hat{u}_1 \right)^2}{2\hat{\sigma}^2} \right] \times \\ \left. \times \left(x_i - \sum_{r=1}^2 a_r t_i^r - \hat{u}_1 \right) t_i^r \right\} = 0, \quad r = 1, 2, \quad (7)$$

где будем полагать σ^2 известным, а оценки \hat{p}_1 и \hat{u}_1 найдем по обучающей (помеховой) выборке объемом m методом моментов

$$\hat{p}_1 = \hat{m}_1^2 (\hat{m}_2 - \sigma^2)^{-1}; \quad \hat{u}_1 = (\hat{m}_2 - \sigma^2) \hat{m}_1^{-1};$$

$$\hat{m}_1 = \frac{1}{m} \sum_{i=1}^m x_i, \quad \hat{m}_2 = \frac{1}{m} \sum_{i=1}^m x_i^2.$$

Система уравнений (7) решается методом скорейшего спуска. Так, для $n = m = 100; \sigma^2 = 19; p_1 = 0,5; u_1 = 76$ при истинных значениях коэффициентов полинома $a_1 = -3; a_2 = 0,007$, шаге дискретизации $T = 1$, интервале экстраполяции $\Delta t = T h = h$ результаты оценки качества алгоритма (рис. 28) свидетельствуют о его высокой эффективности.

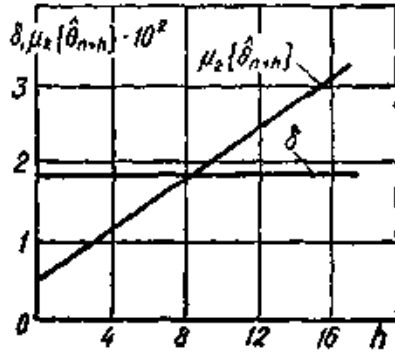


Рис. 28. Показатели эффективности алгоритма (7) фильтрации — экстраполяции полинома (5): δ — относительная смешанность оценки $\hat{\theta}_{n+h}$ при h шагах экстраполяции: $\delta = (m_1\{\hat{\theta}_{n+h}\} - \theta_{n+h}) \theta_{n+h}^{-1} \times 100$; $\mu_2\{\hat{\theta}_{n+h}\}$ дисперсия оценки θ , при h шагах экстраполяции

При этих же условиях, если пренебрегать возможностью воздействия аномальных ошибок и строить алгоритм по МНК (совпадающему для гауссовой помехи с максимальным правдоподобием), то при наличии выбросов дисперсия фильтрации—экстраполяции и смещенность оценок будет на один-два порядка большими.

1.10.2. Устойчивые алгоритмы рекурсивной фильтрации сигналов

Синтез устойчивых алгоритмов фильтрации. В основе синтеза алгоритмов рекурсивной фильтрации лежит представление фильтруемого процесса в виде решения неоднородного дифференциального либо конечно-разностного уравнения со случайной правой частью. Иными словами, фильтруемый процесс представляется как последовательность состояний некоторой динамической системы, находящейся под воздействием некоррелированного либо коррелированного гауссовского (марковского) случайного процесса.

Ограничимся случаем дискретного времени и запишем уравнение состояний

$$\Theta(i+1) = \Phi(i+1/i)\Theta(i) + W(i), \quad (4.8)$$

где $\Theta(i)$ — значение вектора оцениваемых параметров системы размерности m , характеризующих ее состояние в i -й момент времени,

$\Phi(i+1/i)$ — переходная матрица размерности $m \times m$; $W(i)$ — m -мерный вектор шума, например, с нормально распределенными компонентами и ковариационной матрицей $Q(i)$.

Результаты наблюдения состояний системы определяются уравнением наблюдения

$$X(i) = H(i)\Theta(i) + n(i), \quad (9)$$

где $H(i)$ — матрица размерности $l \times m$; $n(i)$ — погрешность наблюдений.

Матрица H может зависеть от момента наблюдений i , а может и не зависеть. Погрешности наблюдений $n(i)$ могут быть распределены в общем случае по негауссовскому закону. Заметим, что, располагая последовательностью состояний системы в $k+1$ моментов времени $\Theta(0), \Theta(1), \dots, \Theta(k)$, можно однозначно вычислить последовательность возмущающих воздействий

$$\begin{aligned} W(i) &= \Phi^{-1}(i+1/i)\Theta(i+1) - \Theta(i) = \\ &= \Theta(i+1) - \Phi(i+1/i)\Theta(i), \quad i = 0, \dots, k-1, \end{aligned} \quad (10)$$

а знание $k+1$ последовательных наблюдений $X(i), i = 0, \dots, k$ позволяет вычислить последовательность погрешностей или помех

$$n(i) = X(i) - H(i)\Theta(i), \quad i = 0, \dots, k. \quad (11)$$

Априорное знание законов распределения вероятностей возмущающих воздействий и помех дает возможность записать функцию правдоподобия

$$\begin{aligned} L = \left[\prod_{i=1}^{k-1} \omega(\Theta(i+1) - \Phi(i+1/i)\Theta(i), X(i) - \right. \\ \left. - H(i)\Theta(i)) \right] \omega_2(X(k) - H(k)\Theta(k)), \end{aligned} \quad (12)$$

где $\omega(W(i), n(i))$ — совместная двумерная плотность распределения вероятностей возмущений и помех; $\omega_2(n(i))$ — плотность распределения вероятностей помех.

Функция правдоподобия (12) зависит от наблюдаемых значений $X(i)$ и параметров $\Theta(i)$. Задача фильтрации, таким образом, сводится к задаче поиска последовательности оценок $\hat{\theta}(0), \hat{\theta}(1), \dots, \hat{\theta}(n)$, максимизирующей выражение (12), т. е. к оцениванию по методу максимального правдоподобия. Заметим, что функцию (12) можно рассматривать как функцию неизвестных параметров $\Phi(i+1/i)$ и при известных возмущениях решить задачу идентификации динамических систем.

Пусть шумы возмущений и помехи являются независимыми случайными величинами с одномерными законами распределения

вероятностей $\omega_1(W)$ и $\omega_2(n)$ соответственно. Тогда в соответствии с формулой (12)

$$L = \left[\prod_{i=0}^{k-1} \omega_1(\Theta(i+1) - \Phi(i+1/i)\Theta(i)) \times \right. \\ \left. \times \omega_2(X(i) - H(i)\Theta(i)) \right] \omega_2(X(k) - H(k)\Theta(k)). \quad (13)$$

В дальнейшем будем оперировать с логарифмом функции правдоподобия

$$l = \ln L \left[\sum_{i=0}^{k-1} (\rho_1(\Theta(i+1), \Theta(i))) + \rho_2(X(i), \Theta(i)) \right] + \\ + \rho_2(X(k), \Theta(k)),$$

где $\rho_i(\cdot) = \ln \omega_i(\cdot)$.

Оценки неизвестных параметров ищутся как решение системы уравнений

$$\frac{\partial l}{\partial \Theta(i)} = \Psi_1(\Theta(i-1), \Theta(i), \Theta(i+1)) + \\ + \Psi_2(X(i), \Theta(i)), \quad i = 1, \dots, k, \quad (14)$$

где

$$\Psi_1(\cdot) = \frac{\partial \rho_1(\Theta(i), \Theta(i-1))}{\partial \Theta(i)} + \frac{\partial \rho_1(\Theta(i+1), \Theta(i))}{\partial \Theta(i)}; \\ \Psi_2(\cdot) = \frac{\partial \rho_2(X(i), \Theta(i))}{\partial \Theta(i)}.$$

Как и в теории устойчивого оценивания (раз. 8), вид функций $\rho_i(\cdot)$ в условиях априорной неопределенности, когда распределения ω_1 и ω_2 заданы не точно (принадлежат p -окрестности некоторого распределения), может выбираться из минимаксных соображений.

При параметрической неопределенности возможно использование оценок неизвестных параметров (масштаб, сдвиг, амплитуда импульсной помехи и др.)

Для построения рекурсивных фильтров решение системы уравнений (14) ищут в рекуррентном виде. В i -й момент времени система (14) состоит из i уравнений. Возмущающее воздействие $W(i)$ в последнем уравнении не учитывается и формально на i -м шаге оценивания его можно положить равным нулю. При этом $\Theta(i+1) = \Phi(i+1/i)\Theta(i)$ и $\Psi_1(\cdot)$ зависит от $\Theta(i)$ и $\Theta(i-1)$, а уравнения системы (14) принимают вид

$$\Psi_1(\Theta(j-1), \Theta(j), \Theta(j+1)) + \Psi_2(X(j), \Theta(j)) = 0, \\ j = 1, \dots, i-1; \\ \Psi_1(\Theta(i-1), \Theta(i)) + \Psi_2(X(i), \Theta(i)) = 0. \quad (15)$$

Система (15) допускает рекуррентное решение, если известны начальные условия, т. е. значение вектора параметров в нулевой момент времени $\Theta(0)$. Рассмотрим решение системы (15) для случая, когда W — некоррелированный гауссовский шум с ковариационной матрицей $Q(i)$, а $n(i)$ — помеха с негауссовским распределением. Тогда

$$\rho_1(\Theta(i-1), \Theta(i)) = [\Theta(i) - \Phi(i/i-1)\Theta(i-1)]^T \times \\ \times \Theta^{-1}(i) [\Theta(i) - \Phi(i/i-1)\Theta(i-1)].$$

Функции

$$\Psi_1(\Theta(i-1), \Theta(i), \Theta(i+1))$$

являются линейными.

Система уравнений на i -м шаге

$$\frac{\partial}{\partial \Theta(j)} \rho_2(X(j) - H(j)\Theta(j)) + [\Theta(j) - \Phi(j/j-1)\Theta(j-1)]^T \times \\ \times Q^{-1}(j) - [\Theta(j+1) - \Phi(j+1/j)\Theta(j)]^T \times \\ \times \Phi^T(j+1/j)Q^{-1}(j) = 0, \\ \frac{\partial}{\partial \Theta(i)} \rho_2(X(i) - H(i)\Theta(i)) + [\Theta(i) - \Phi(i/i-1)\Theta(i-1)] \times \\ \times \Theta^{-1}(i)]^T Q^{-1}(i) = 0, \quad j = 1, \dots, i-1.$$

Векторную функцию векторного аргумента

$$\Psi_2(X(i) - H(i)\Theta(i)) = \frac{\partial}{\partial \Theta(i)} \rho_2(X(i) - H(i)\Theta(i)) \times \\ \times \Theta(i) \text{ разложим в ряде Тейлора в окрестности экстраполированного значения вектора оцениваемых параметров } \Theta(i) = \Theta^0(i) = \Phi(i/i-1)\Theta(i-1) \text{ и ограничившись двумя членами получим линеаризованную систему} \\ \Psi_2(X(j) - H(j)\Theta^0(j)) + \Psi_2'(X(j) - H(j)\Theta^0(j))|_{\Theta(j)=\Theta^0(j)} \times \\ \times (\Theta(j) - \Theta^0(j)) + [\Theta(j) - \Theta^0(j)]^T Q^{-1}(j) - \\ - [\Theta(j+1) - \Theta^0(j+1)]^T \Phi^T(j+1/j)Q^{-1}(j) = 0, \\ j = 1, \dots, i-1; \quad (16)$$

$$\Psi_2(X(i) - H(i)\Theta^0(i)) + \Psi_2'(X(i) - H(i)\Theta(i))|_{\Theta(i)=\Theta^0(i)} \times (\Theta(i) - \Theta^0(i)) + [\Theta(i) - \Theta^0(i)]^T Q^{-1}(i) = 0,$$

где

$$\Psi_2'(X(i) - H(i)\Theta(i)) = \begin{bmatrix} \frac{\partial \Psi_{21}}{\partial \Theta_1(i)} & \dots & \frac{\partial \Psi_{21}}{\partial \Theta_m(i)} \\ \frac{\partial \Psi_{2m}}{\partial \Theta_1(i)} & \dots & \frac{\partial \Psi_{2m}}{\partial \Theta_m(i)} \end{bmatrix} \text{ — матрица Якоби.}$$

Решение системы уравнений (16) записывается в виде

$$\hat{\Theta}(i) = \Theta(i)^0 + P(i) H^T(i) \Psi_2 \left(\frac{X(i) - H(i)\Theta^0(i)}{s(i)} \right) s(i), \tag{17}$$

где $s(i)$ — параметр масштаба распределения помехи; $P(i)$ — приближенная матрица ковариаций ошибок оценивания.

Матрица $P(i)$ вычисляется рекуррентно по формуле

$$P^{-1}(i) = P^{-1}(i/i - 1) + H^T(i) H(i) \Psi_2'((X(i) - H(i)\Theta^0(i))/s(i))/S^2(i), \tag{18}$$

где

$$P(i/i - 1) = \Phi(i/i - 1) P(i - 1) \Phi^T(i/i - 1) + Q(i). \tag{19}$$

Совокупность выражений (17) — (19) и начальных условий, состоящих в задании $\Theta(0)$, либо ковариационной матрицы начального вектора $P(0)$ определяет устойчивый рекурсивный фильтр. Определенный выше фильтр при $\Psi_2(X) = X$ переходит в линейный фильтр Калмана. Априорное знание распределений вероятностей помех наблюдения $\omega_2(X, \Theta)$ дает возможность оптимизировать структуру фильтра путем введения нелинейных преобразователей с характеристикой

$$\Psi_2(X) = \frac{\partial \ln \omega_2(X, \Theta)}{\partial \Theta}.$$

При выборе $\Psi_2(X)$ применимы рассматривавшиеся выше принципы минимакса, инвариантности, различные эвристические соображения. Так, например, $\Psi_2(X)$ может быть функцией порядковых статистик. Вид некоторых из этих функций рассматривался ранее. Если вид $\Psi_2(X)$ определен с точностью до параметров, например, неизвестна амплитуда импульсной помехи, рекуррентный фильтр должен

содержать канал оценки неизвестных параметров помех. Структурная схема устойчивого рекурсивного алгоритма фильтрации показана на рис. 29.

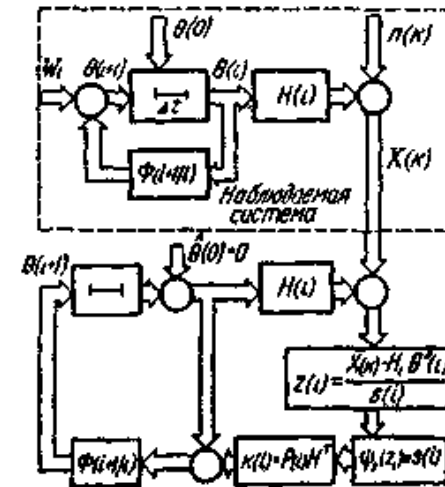


Рис. 29. Структурная схема устойчивого рекурсивного фильтра

Рассмотрим пример устойчивой фильтрации данных траекторных измерений.

Пример 21. Рекурсивный фильтр параметров траектории самолета при негауссовских погрешностях измерения координат.

Пусть движение самолета под воздействием случайного ускорения задается разностным уравнением

$$\Theta(i + 1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Theta(i) + W(i),$$

где $\Theta^T(k) = [\Theta_1(k), \Theta_2(k)]$; $\Theta_1(k)$ — координата; $\Theta_2(k)$ — скорость самолета в k -й момент времени; $W(i)$ — вектор нормальных случайных величин с ковариационной матрицей

$$Q(k) = \begin{bmatrix} 1/8 & 1/2 \\ 1/2 & 1 \end{bmatrix}.$$

Будем считать, что ошибка наблюдения координаты $n(k)$ представляет собой смесь нормированного гауссовского некоррелированного шума и скачкообразной помехи с амплитудой u , действующей с вероятностью p .

Плотность распределения вероятностей ошибки наблюдения

$$\omega_2(n) = (1-p) \frac{e^{-\frac{n^2}{2s^2}}}{\sqrt{2\pi s^2}} + p \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(n-u)^2}{2s^2}}. \quad (20)$$

Уравнение наблюдения

$$X(k) = [1 \ 0] \Theta(k) + n(k). \quad (21)$$

В начальный момент времени положение самолета известно точно ($Q_1(0) = 0$).

Найдем вид нелинейного преобразователя

$$\Psi_2 = \frac{\partial \ln \omega_2(X(i), \Theta(i))}{\partial \Theta_1(i)} = \frac{X_1(i) - \Theta_1(i)}{s} - \frac{puN(\Theta_1(i) + u, s)}{(1-p)N(\Theta_1(i), s) + pN(\Theta_1(i) + u, s)},$$

где $N(m, s)$ — плотность нормального закона распределения вероятностей с математическим ожиданием m и дисперсией s^2 .

На рис. 30 показано графическое изображение характеристики нелинейного преобразователя $\Psi_2(z)$.

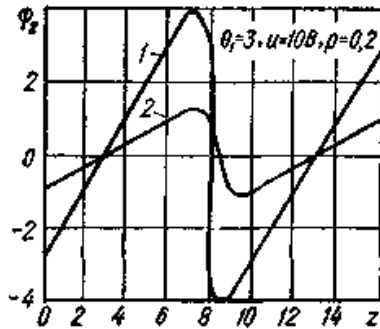


Рис.30. Характеристики нелинейного преобразователя (формула (22)) для $S^2=1$ (1); для $S^2=3$ (2)

На рис. 31 показаны результаты моделирования описанного фильтра при следующих параметрах плотности помехи (20): $s = 1, u = 10$.

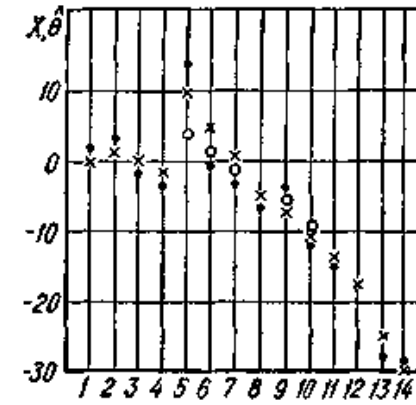


Рис.31. Реализации случайных процессов: наблюдений (\cdot); на выходе линейного рекурсивного фильтра Калмана (\times); на выходе устойчивого фильтра (O)

Здесь же приведены результаты моделирования фильтра Калмана, синтезированного в предположении отсутствия импульсной помехи. Устойчивый фильтр обеспечивает большую точность оценивания.

Другой пример нелинейного преобразователя можно получить, положив плотность распределения помех

$$\omega_2(z) = \begin{cases} (1-p)\omega(z_0)e^{[1+c_1\alpha(z-z_0)-e^{\alpha(z-z_0)}]}, & z < -k; \\ (1-p)\omega(z), & |z| < k; \\ (1-p)\omega(z_1)e^{[1-c_1\alpha(z-z_1)-e^{-\alpha(z-z_1)}]}, & z > k, \end{cases} \quad (23)$$

где $\omega(z)$ — плотность распределения нормированной нормальной случайной величины; z_0 и z_1 ($z_0 < z_1$) — концы интервала, на котором выполняется неравенство $(\omega'/\omega) \leq k$; величина k связана с p соотношением (47) п.1.8: $z = \frac{x - m_x}{\sigma_x}$ — нормированная случайная величина; m_x, σ_x^2 — математическое ожидание и дисперсия наблюдаемой случайной величины x .

Значение α вычисляется из условия нормировки

$$(1 - \rho) \omega(z_0) e^{\int_{-\infty}^{z_0} e^{c_1 \alpha(z-z_0) - \alpha(z-z_0)} dz} = \frac{1}{2} - \frac{1-\rho}{2} \int_{z_0}^{z_1} \omega(z) dz. \quad (24)$$

Интегрирование левой части выражения (24) приводит к следующему условию:

$$(1 - \rho) \omega(z_0) \frac{e}{\alpha} \Upsilon(c_1, 1) = \frac{1}{2} - \frac{(1-\rho)}{2} \int_{z_0}^{z_1} \omega(z) dz, \quad (25)$$

где $\Upsilon(\cdot)$ — неполная гамма-функция.

Распределение (23) имеет центральную часть, расположенную в пределах $z_0 < z < z_1$, которая описывается гауссовской функцией, и «хвосты», описываемые распределением типа распределения выбросов случайных процессов — первым предельным распределением.

Распределению (23) соответствует нелинейный преобразователь с функцией

$$\Psi(z) = \begin{cases} -c_1 \alpha + \alpha e^{\alpha(z-z_0)}, & z < -k, \\ z, & |z| < k; \\ c_1 \alpha - \alpha e^{-\alpha(z-z_1)}, & z > k. \end{cases} \quad (26)$$

Функция (26) имеет разрывы в точках $|z| = k$. Эти разрывы можно устранить подбором коэффициента c_1 , что приводит к уравнению

$$\alpha(c_1 - 1) = k. \quad (27)$$

Характеристика нелинейного преобразователя (26) показана на рис. 32

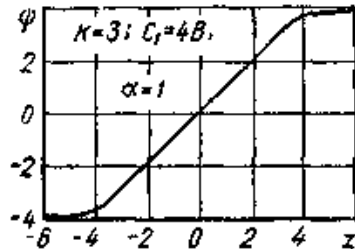


Рис 32. Характеристика нелинейного преобразования (формула (26))

Результаты моделирования рекурсивного фильтра с нелинейным преобразователем (26) для 21-го примера при воздействии импульсной

помехи с неизвестной амплитудой свидетельствуют о его высокой устойчивости (рис. 33).

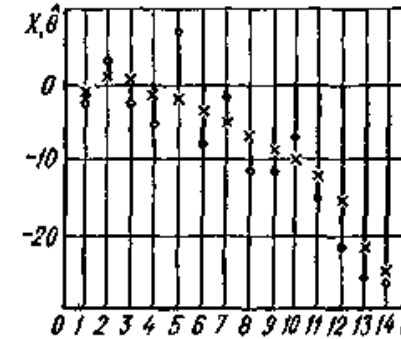


Рис. 33. Результаты моделирования рекурсивного фильтра:

O — наблюдения; x — робастные оценки

Нелинейный преобразователь (26) обеспечивает более высокую точность оценивания параметров траектории самолета при воздействии импульсных помех с неизвестной амплитудой, чем фильтр.

На практике в ряде случаев априорные сведения о фильтруемых процессах и помехах задаются в частотной области, а соответствующие оптимальные линейные фильтры называют «винеровскими». Техническая реализация таких фильтров сложна, сами фильтры не обладают устойчивостью к выбросам. В ряде работ показано, что для задачи фильтрации стационарного нормального случайного процесса с известным спектром $S_{\Theta}(\omega)$ на фоне нормального шума со спектром $S_{\Psi}(\omega)$ стационарной и нормальной коррелированной помехи с конечным спектром $S_{\eta}(\omega)$ существует эквивалентное представление во временной области, приводящее к стационарному фильтру Калмана. Т. е. задача может быть сформулирована как задача наблюдения m -мерного марковского процесса, представляющего собой значение фильтруемого процесса $\Theta_1(t_i)$ и его производных $\Theta_2(t_i) = \dot{\Theta}_1(t_i)$, $\Theta_3(t_i) = \ddot{\Theta}_1(t_i) \dots$, \dots , $\Theta_m(t_i) = \Theta_{m-1}(t_i)$ в момент времени t_i . Размерность измеряемого процесса $\Theta(t)$, размерность наблюдаемого процесса $x(t)$ и вид матриц Φ и H определяются спектром фильтруемого сигнала $S_{\Theta}(\omega)$.

В уравнении состояний (8) $W(i)$ — m -мерный вектор с нормальным распределением координат и ковариационной матрицей

$$Q_w = \begin{pmatrix} 0 & \dots & 0 \\ & & \\ 0 & \dots & \sigma_m^2 \end{pmatrix}.$$

В уравнении (9) шум наблюдения — l -мерный вектор с негауссовским распределением, описываемым составной плотностью вида

$$\omega_{\Sigma}(X, \Theta) = c_1(1 - I_n(X, t_i))\omega_1(X/\Theta) + c_2 I_n(X, t_i)\omega_2(X), \quad (28)$$

где $\omega_1(X/\Theta)$ — l -мерная условная нормальная плотность с математическим ожиданием Θ и ковариационной матрицей шумов наблюдений R , соответствующей спектру $s_{\Sigma}(\omega)$; $\omega_2(X)$ — нормальная плотность распределения с нулевым математическим ожиданием и ковариационной матрицей, определяемой спектром помехи $s_n(\omega)$; $I_n(X, t)$ — индикаторная функция помехи; c_1, c_2 — нормирующие множители.

По условию задачи спектры сигнала и помехи разнесены, т. е. использование индикаторной функции в записи плотности (28) оправдано.

Найдем вид нелинейного преобразования результатов наблюдений

$$\Psi(X, t) = \frac{\partial \ln \omega_{\Sigma}(X, \Theta)}{\partial \Theta} = \frac{c_2(1 - I_n(X, t)) \frac{\partial \omega_1(X/\Theta)}{\partial \Theta}}{c_1(1 - I_n(X, t))\omega_1(X/\Theta) + c_2 I_n(X, t)\omega_2(X)}.$$

С учетом свойств индикаторной функции

$$\Psi(X, t) = \begin{cases} \frac{\partial}{\partial \Theta} \omega_1(X/\Theta) \\ \omega_1(X/\Theta) \end{cases}, \quad I_n(X, t) = 0; \quad (29)$$

$$0, \quad I_n(X, t) = 1.$$

Устойчивый алгоритм фильтрации узкополосного сигнала на фоне высокочастотной помехи определяется выражениями (17) — (19), где $\Psi(X, t)$ имеет вид (29). Алгоритм предполагает вычисление индикаторной функции $I_n(X, t)$, т. е. содержит канал обнаружения высокочастотной помехи. Канал обнаружения помехи можно реализовать различными способами. В общем случае здесь должна производиться оценка ковариационной матрицы, на основании которой и принимается решение о принадлежности наблюдаемого вектора X области помехи.

Возможны другие, упрощенные способы, например, сравнение максимальной дисперсии фазовой координаты процесса с некоторым

порогом. Еще один способ заключается в сравнении следа ковариационной матрицы наблюдений с порогом решения. След ковариационной матрицы узкополосного случайного процесса является функцией, монотонно возрастающей при увеличении средней частоты спектра.

Эвристические процедуры устойчивой фильтрации. Рассмотрим некоторые широко применяемые алгоритмы и устройства. В практике траекторных измерений используют отбраковку аномальных измерений, вызванных машинными сбоями при предварительной математической обработке в аппаратуре первичной обработки информации (АПОИ), сбоями при передаче данных по каналам связи в распределенных вычислительных сетях АС УВД, сбоями работы самого измерительного устройства. Для исключения из совокупности траекторных измерений аномальных результатов используют закономерности изменения измеряемых параметров летательного аппарата.

Проверку результатов измерений на аномальность в процессе их получения проводят по условию нахождения измерений в допустимых пределах изменения:

$$\Theta(i)_{\min} < \Theta(i) < \Theta(i)_{\max},$$

где $\Theta(i)_{\min}, \Theta(i)_{\max}$ — допустимые пределы изменения измеряемого параметра $\Theta(i)$ летательного аппарата, $\Theta(i)$ — векторный параметр, включающий измеряемую координату и скорость. Такой метод исключения аномальных данных эквивалентен применению в фильтре функции

$$\Psi(z) = \begin{cases} z, & z \in G; \\ 0, & z \notin G, \end{cases}$$

где G — допустимая область изменений параметра объекта.

В трехмерном пространстве это соответствует выбору строга в виде параллелепипеда. Размеры строга рассчитываются по заданным вероятностям попадания летательного аппарата в строб и вероятностям пропуска аномальных наблюдений в канал оценки параметров траектории и зависят как от диапазона изменения параметров, так и от видов воздействующих помех. Применяются режимы адаптации к параметрам помех.

Алгоритм типа (29) защиты от помех реализован в аппаратном приемнике. Здесь используются различия в спектре полезного сигнала и импульсной помехи. Импульсная помеха содержит высокочастотные составляющие. Одна из этих составляющих (обычно 180 кГц) используется для опознавания импульсной помехи, действующей в приемном тракте (но сути дела, классификация в частотной области).

Сигнал опознавания служит для бланкирования тракта сигнала на время действия помехи. На выходе тракта сигнала сохраняется напряжение, предшествовавшее импульсной помехе, которое запоминается емкостным интегратором.

Существуют специализированные интегральные микросхемы для систем автоматического подавления импульсных помех, в которых объединены все основные элементы каналов опознавания, и подавления импульсной помехи и обработки пораженных помехой участков реализаций сигнала.

1.11. Моделирование алгоритмов обработки сигналов радиолокационных информационно-измерительных системах (ИИС)

1.11.1. Постановка задачи и критерии эффективности ИИС

При проектировании подсистем обработки информации в ИИС широко применяется метод имитационного моделирования, реализуемый на ЭВМ средней и высокой производительности. Отработанные на этапе моделирования программы, реализующие выбранные алгоритмы обработки сигналов, могут быть непосредственно использованы в математическом обеспечении микро- и мини-ЭВМ ИИС. Особенности применения и возможности метода имитационного моделирования рассмотрим на конкретном достаточно сложном примере оценки эффективности специального алгоритмического обеспечения (САО) радиолокационной ИИС автоматизированной системы управления воздушным движением (АС УВД), удовлетворяющего требованиям устойчивости.

Информационно-измерительная система любой АС УВД включает в свой состав сложное оборудование (первичные и вторичные РЛС), аппаратуру первичной обработки информации (АПОИ), средства вторичной обработки информации и функционирует в условиях воздействия ряда случайных внешних факторов (флуктуации сигналов, наличие помех). Анализ эффективности подобных сложных систем можно выполнить на основе построения машинной модели по цепочке: **среда — система — математическая модель — моделирующий вычислительный алгоритм** с реализацией на достаточно производительной ЦВМ.

Разрабатываемая модель должна отвечать следующим требованиям: учитывать разнообразие принимаемых сигналов и воздействующих помех, обусловленное большим количеством типов воздушных судов (ВС), различными дальностями до ВС и различными их скоростями, метеоусловиями, наличием помех от гидрометеобразований, переотражений от местных предметов и др.;

учитывать тактико-технические характеристики моделируемой ИИС (форму диаграммы направленности антенны РЛС, параметры сканирования антенной системы, типовые параметры радиочастотного тракта приемного устройства);

учитывать особенности устойчивой обработки сигналов в системах селекции движущихся целей (СДЦ), в АПОИ, особенности алгоритмов фильтрации параметров траекторий ВС п ЦИМ вторичной обработки;

точность аналитических выражений, используемых при описании элементов ИИС, и ошибки процедур статистических испытаний должны быть не хуже 10—15 % и регламентироваться требованиями пользователя исходя из точности задания входных воздействий и возможностью экономии машинного времени;

прикладные программы должны допускать различные модификации и удовлетворять требованиям возможных пользователей, в частности, допускать реализацию на современных микро- и мини-ЭВМ.

Критерии эффективности ИИС УВД. Выбор критериев эффективности ИИС производится исходя из обычных требований к критериям сложных систем: соответствие критерия основной функции системы; критичность к определяющим параметрам; доступность измерения (оценки) в процессе эксплуатации систем. Сформулированным требованиям удовлетворяют критерий достоверности отображения информации в зоне обзора РЛС ИИС УВД, оцениваемый количественно условной вероятностью правильного вывода на индикатор воздушной обстановки (ИВО) метки ВС с формуляром сопровождения; критерий точности обработки траекторий ВС, оцениваемый в общем случае ковариационной матрицей погрешностей измерения координат.

Использование моделей ИИС АС УВД возможно в двух аспектах: исследовательском (решение задач анализа действующих и перспективных ИИС в заданных областях вариации определяющих параметров, оптимизация, выбор вариантов на ранних стадиях проектирования, в частности, разработка технических требований к ИИС); эксплуатационном (прогнозирование эффективности действующих ИИС, оценка ИИС в экстремальных внешних условиях применительно к конкретным условиям эксплуатации, оптимизация размещения ИИС).

1.11.2. Алгоритмы моделирования сигнально-помеховой ситуации

Рассмотрим основные соотношения, используемые при описании и имитационном моделировании типовых радиолокационных сигналов и помех. Подмножество моделей включает: флюктуирующие (шумоподобные) сигналы и помехи; нефлюктуирующие (гармонические) сигналы и помехи; помехи типа хаотически-импульсных; помехи типа отражений от местных предметов. При аналитических расчетах эффективности алгоритмов обработки сигналов либо при их имитационном моделировании на ЦВМ исходными являются законы распределения исследуемых процессов.

Модель шумоподобных сигналов и помех. Плотность распределения вероятности (ПРВ) огибающей флюктуирующей (шумоподобной) помехи имеет вид:

$$\omega(x) = \frac{x}{\Psi} \exp\left\{-\frac{x^2}{2\Psi}\right\}, \quad (1)$$

где Ψ — мощность (дисперсия) помехи.

Соотношение (1) носит название закона Релея, интегральная форма которого (функция распределения) записывается следующим образом:

$$F(x) = 1 - \exp\left\{-\frac{x^2}{2\Psi}\right\}. \quad (2)$$

Распределение огибающей аддитивной смеси гауссовской помехи типа внутриприемного шума и шумоподобного гауссовского сигнала, порождаемого отражениями радиоимпульсов от целей,

$$\omega_c(x) = \frac{x}{\Psi(1+b)} \exp\left\{-\frac{x^2}{2(1+b)\Psi}\right\}, \quad (3)$$

для плотности вероятности и

$$F_c(x) = 1 - \exp\left\{-\frac{x^2}{2(1+b)\Psi}\right\} \quad (4)$$

для функции распределения (ФР), причем $b = \Psi_c/\Psi$ — отношение сигнал/помеха по мощности.

Алгоритм моделирования смеси сигналов и помех состоит в генерировании при помощи ЦВМ последовательности случайных чисел, распределенных по закону распределения вероятностей значений (2) или (4). Для этого используется стандартный датчик (подпрограмма) равномерных случайных чисел, имеющийся в математическом обеспечении ЦВМ и производится преобразование чисел, генерируемых этим датчиком, по закону, обратному требуемой ФР (формулы (2) или (4)). Так, для ФР (4) преобразование имеет следующий вид

$$x_i = F^{-1}(\xi_i) = \sqrt{-2 \ln(1 - \xi_i) \Psi(1 + b)}, \quad (5)$$

где ξ_i — значение случайного числа с равномерным распределением, сгенерированного на i -м такте моделирования. Алгоритм (5) используется в подпрограмме моделирования сигналов и помех.

Модель аддитивной смеси нефлюктуирующего (гармонического) сигнала и шумоподобной помехи. В такой модели предполагается, что эхо — сигнал на входе детектора приемного устройства РЛС представляет собой отрезки гармонических колебаний с амплитудой u , по времени равные длительности зондирующего импульса.

Закон распределения огибающей смеси такого сигнала и внутриприемного шума носит название обобщенного закона Релея и та закона Райса и имеет вид

$$\omega_c(x) = \frac{x}{\Psi} \exp\left\{-\frac{x^2 + u^2}{2\Psi}\right\} I_0\left(\frac{xu}{\Psi}\right) \quad (6)$$

для плотности вероятности и

$$F_c = \exp\left\{-\frac{x^2 + u^2}{2\Psi}\right\} \sum_{k=1}^{\infty} \left(\frac{x}{\Psi}\right)^k I_k\left(\frac{xu}{\Psi}\right) \quad (7)$$

для функции распределения, в которой $I_k(\cdot)$ — функция Бесселя k -го порядка от мнимого аргумента.

Для моделирования последовательности чисел с распределением вероятностей (7) использование преобразования, обратного к (7), затруднительно, что заставляет применять другие приемы. Известно, что по закону (6) распределен модуль вектора, две координаты которого x_1 и x_2 распределены по нормальному закону с дисперсией Ψ и математическими ожиданиями 0 и u соответственно. Поэтому для моделирования используют датчики (подпрограммы) пар стандартных нормально распределенных случайных чисел с математическим ожиданием 0 и дисперсией 1 и преобразуют их в соответствии с формулой

$$s_i = [(x_{i1} \sqrt{\Psi} + u)^2 + (x_{i2} \sqrt{\Psi})^2]^{1/2}, \quad (8)$$

где x_{i1} , x_{i2} — стандартные нормальные случайные числа, сгенерированные на i -м такте моделирования. Отношение сигнал/помеха по мощности такой смеси определяется формулой

$$b = u^2/2\Psi.$$

Модель хаотической импульсной помехи. Такая помеха (ХИП) представляет собой смесь случайных по амплитуде, длительности и моменту времени появления отрезков гармонического колебания и гауссовского процесса мощностью Ψ .

Плотность вероятности огибающей ХИП имеет вид:

$$\omega(x) = p_1 \omega_1(x) + p_2 \omega_2(x), \quad (9)$$

где p_1, p_2 — вероятности отсутствия и наличия в процессе составляющей ХИП, определяемые соотношениями

$$p_1 = 1 - 1/Q, \quad p_2 = 1/Q, \quad (10)$$

в которых Q — средняя скважность ХИП. Плотностям вероятностей ω_1 , и ω_2 соответствуют законы Релея и Райса, т. е. ω_1 совпадает с выражением (1), а ω_2 определяется правой частью формулы (6), где u — амплитуда гармонической составляющей процесса на входе детектора.

Соответствующий плотности (9) интегральный закон распределения можно представить в форме

$$F(x) = p_1 F_1(x) + p_2 F_2(x), \quad (11)$$

причем F_1 и F_2 определяются правыми частями выражений (2) и (7) соответственно.

Алгоритм моделирования ХИП соответствует формуле (8), слагаемое u вводится с вероятностью p_2 .

Модели отражения от местных предметов и метеопомехи.

Помехи, обусловленные отражением зондирующего сигнала от местных предметов (подстилающей поверхности, горных образований и от метеобразований), относятся к наиболее сложному виду помех, влияющих на прием сигналов от целей, как в ближней, так и в дальней зонах действия РЛС. Эти помехи характеризуются сложными пространственно-временными спектрами (пространственно-временными корреляционными функциями). Рассмотрим алгоритм моделирования нормального случайного процесса с заданным спектром $s(p)$. Для моделирования такого процесса необходимо провести факторизацию спектра $s(p)$, т. е. представить его в виде двух взаимно-сопряженных сомножителей

$$s(p) = K(p)K(-p), \quad (12)$$

один из которых $K(p)$ имеет полюсы в левой полуплоскости комплексного переменного, а $K(-p)$ — соответственно в правой полуплоскости.

Сомножитель $K(p)$ будет комплексным коэффициентом передачи искомого фильтра, формирующего нормальный случайный процесс с заданным спектром $s(p)$. Для построения фильтра надо представить

$$K(p) = \frac{\alpha_0 + \alpha_1 p + \dots + \alpha_m p^m}{\beta_0 + \beta_1 p + \dots + \beta_n p^n}, \quad m < n \quad (13)$$

и определить формирующий фильтр в виде модели сообщения, описываемой уравнением состояний фильтруемого процесса,

$$\dot{x}(t) = Fx(t) + GW(t) \quad (14)$$

и уравнением наблюдения

$$y(t) = Hx(t), \quad (15)$$

где

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ -\beta_0 & -\beta_1 & -\beta_2 & \dots & \beta_{n-1} \end{bmatrix};$$

$$H = [\alpha_0, \alpha_1 \dots \alpha_m, 0 \dots 0],$$

$W(t)$ — вектор, координаты которого являются белым шумом с ковариационной матрицей

$$Q_W = \begin{bmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ 0 & \dots & 1 \end{bmatrix};$$

где $x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \dots \\ x_n(t) \end{pmatrix}$ — векторный случайный процесс;

$y(t)$ — наблюдаемый одномерный случайный процесс.

Алгоритмы (14), (15) пригодны для моделирования в непрерывном времени, т. е. для аналогового моделирования. Для моделирования в дискретном времени на ЦВМ надо произвести соответствующее преобразование уравнений (14) и (15).

Модель сообщения в дискретном времени имеет вид

$$x(k+1) = \Phi(\Delta t)x(k) + GW(k), \quad (5.16) \quad y(k) = Hx(k), \quad (17)$$

где

$$\Phi(\Delta t) = I + F \cdot \Delta t = \begin{bmatrix} 1 & \Delta t & 0 & \dots & 0 \\ 0 & 1 & \Delta t & \dots & 0 \\ -\beta_0 \Delta t & -\beta_1 \Delta t & \dots & 1 & -\beta_{n-1} \Delta t \end{bmatrix};$$

$\Phi(\Delta t)$ — переходная матрица дискретной во времени модели сообщения (сигнала).

Последовательность $y(k)$, $k = 0, 1 \dots$ будет последовательностью значений нормального случайного процесса с заданным спектром $s(p)$, взятых с интервалом времени Δt .

Для формирования огибающей такого процесса требуется знать значение процесса в момент времени $t_i = y(t_i)$ и его первой производной $-\dot{y}(t)$. Значение огибающей вычисляется по формуле

$$u(t_i) = \sqrt{y^2(t_i) + (\dot{y}(t_i)/\omega_0)^2}, \quad (18)$$

где ω_0 — средняя частота спектра процесса.

Если наблюдаемый процесс $y(t)$ одномерный, то значение его производной может быть вычислено как первая конечная разность

$$y(t) = [y(t + \Delta t) - y(t)] / \Delta t, \quad (19)$$

а значение огибающей — по формуле (18) с учетом выражения (19). Другой путь состоит в формировании двух координат наблюдаемого процесса $y(t)$ и $\dot{y}(t)$. При этом так как спектр производной процесса $S_{\dot{y}(t)}(p) = p S_{y(t)}(p)$, то матрица наблюдения H принимает вид

$$H = \begin{bmatrix} \alpha_0 & \alpha_1 & \dots & \alpha_m & \dots & 0 \\ 0 & \alpha_0 & \alpha_1 & \dots & \alpha_m & \dots & 0 \end{bmatrix}. \quad (20)$$

При такой модели сообщения наблюдаем сразу процесс с заданным спектром и его производную и можем более точно вычислить значение его огибающей по формуле (18). Структурная схема алгоритма моделирования изображена на рис. 34.

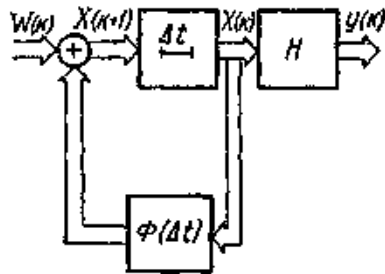


Рис. 34. Структурная схема моделирования процесса с заданным спектром и его производной

1.11.3. Имитационная модель ИИС АС УВД и ее применение

Модель должна отражать существенные для исследования функции системы, которые определяются критерием эффективности, а также чувствовать влияние параметров внешней среды (сигналы и помехи) и параметров самой системы (аппаратурное и алгоритмическое обеспечение) на характеристики качества ее работы. Обобщенная структурная схема ИИС АС УВД (рис. 35) выполняет следующие функции: сбор и первичную обработку информации о воздушной об-

становке в зоне УВД; вторичную обработку информации, т. е. обнаруживает траектории движения целей, осуществляет автоматическое сопровождение, определяет параметры движения и формирует массивы информации для устройств отображения.

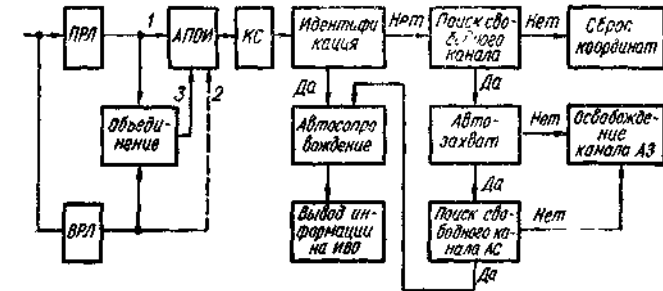


Рис. 35. Обобщенная структурная схема ИИС АС УВД

Рассмотрим коротко работу системы. Импульсные сигналы, отраженные от пассивных целей, и кодовые послышки самолетных ответчиков поступают соответственно на входы первичной (ПРЛ) и вторичной (ВРЛ) РЛС, где осуществляется аппаратурная обработка аналоговых сигналов. С выхода РПУ сигналы поступают на аппаратуру первичной обработки информации (АПОИ) либо по отдельным каналам 1 и 2, либо после логического объединения по каналу 3. АПОИ определяет координаты целей (как истинных так и ложных) и передает их по каналу связи (КС) на ЦВМ вторичной обработки. Кроме того, по КС передаются данные (о бортовом номере, высоте и т. д.), полученные от самолетного ответчика. В результате вторичной обработки для каждой из целей накапливаются данные, характеризующие ее свойства — координаты и параметры движения. Эти данные, относящиеся к одной цели, образуют в запоминающем устройстве ЦВМ вторичной обработки массив, называемый каналом автоматического сопровождения цели (АС). Сопровождение состоит в периодическом обновлении координат и параметров движения цели. Все данные, имеющиеся в канале АС, можно легко передать на индикатор воздушной обстановки (ИВО).

Остановимся на работе ЦВМ системы вторичной обработки (СВО). Каждый частный цикл работы ЦВМ СВО начинается анализом новой информации, накопленной в течение предыдущего цикла. Прежде всего делается попытка **идентификации, т. е. отождествления каждой новой отметки с какой-либо из целей, за которыми уже ведется слежение.**

Для активных отметок идентификация не представляет трудностей и производится по бортовому номеру. Если цель с данным бортовым номером появилась впервые, осуществляется автоматический ввод в сопровождение ее при наличии свободного канала АС.

Для пассивных целей идентификация производится лишь в том случае, когда отметка отождествлена лишь с одной из имеющихся целей. Если отметку можно отнести к двум или большему числу известных целей или несколько новых отметок к одной из известных целей, то возникает так называемая спорная ситуация. Наконец, если новую отметку нельзя идентифицировать ни с одной из имеющихся целей, то в действие вступает алгоритм автозахвата цели. Алгоритм автозахвата анализирует информацию о движении цели в течение нескольких циклов (периодов обзора зоны УВД) и по истечении заданного числа циклов производит проверку критерия обнаружения траектории. Если траектория обнаружена, накопленные данные передаются в свободный канал АС. Если такового не оказалось, данные о траектории сбрасываются и канал автозахвата (АЗ) освобождается.

Данные из канала АС поступают на ИВО. Одновременно в каждом из каналов АС производится проверка критерия сброса траектории с сопровождения. Если критерий сброса выполняется и каком-либо из циклов (обзоров), канал АС освобождается и вывод информации на ИВО по этому каналу прекращается.

В соответствии с приведенным описанием функционирования ИИС АС УВД разрабатываются **следующие моделирующие алгоритмы: моделирования потока ложных целей и их координат; моделирования траекторий «истинных» целей; моделирования оценок дальности и азимута в АПОИ; идентификации отметок целей; автосопровождения; автозахвата; анализа спорных ситуаций.**

Алгоритм моделирования потока ложных целей и их координат. В программе моделируется пуассоновский поток ложных целей. Интенсивность потока $\lambda = N_v F$, где N_v — число элементарных объемов разрешения в зоне обзора; F — вероятность возникновения ложной цели в одном элементарном объеме разрешения (вероятность ложной тревоги). Моделирование реализаций потока происходит следующим образом:

$$K = \min \left\{ n : \prod_{i=1}^n x_i < e^{-\lambda} \right\}, \quad (21)$$

где x — случайная величина, распределенная равномерно на интервале $0 \dots 1$.

Реализацией потока Пуассона с параметром λ будет такое число членов случайной последовательности $\{x_i\}$, $i = 1, k$, при котором будет выполнено условие (21). Этому числу присваивается идентификатор K . Координаты целей считаются распределенными равномерно по всей зоне обзора. Для каждой из K целей разыгрывается случайное число, распределенное равномерно от 1 до N_v . Целая часть этого числа определяет номер элементарного участка разрешения и цели приписываются координаты этого участка (азимут α и дальность R). Координаты ложных целей разыгрываются в каждом обзоре.

Алгоритм моделирования траекторий «истинных» целей. Для «истинных» целей координаты разыгрываются только в одном обзоре, который соответствует моменту входа в зону УВД. В последующих обзорах координаты цели меняются по заданному закону. В общем случае возможно изменение по **кривой второго порядка (параболе)**:

$$\begin{aligned} x_i &= x_{i-1} + a_1 T + a_2 (2i - 1) T^2; \\ y_i &= y_{i-1} + b_1 T + b_2 (2i - 1) T^2, \end{aligned} \quad (22)$$

где i — номер текущего обзора; T — время обзора, с; a_1, a_2, b_1, b_2 — соответственно составляющие скорости и ускорения по координатам x и y .

Перемещение цели задается в прямоугольных координатах, которые затем в программе пересчитываются в полярные по формулам:

$$R = \sqrt{x_i^2 + y_i^2}; \quad \alpha = \arctg \left\{ \frac{y_i}{x_i} \right\}. \quad (23)$$

Далее определяются координаты элементарного разрешающего объема, в котором находится цель:

$$I = [R \cdot A], \quad J = [\alpha / D],$$

где A и D — размеры элемента разрешения по дальности и азимуту соответственно; $[\cdot]$ — операция взятия целой части числа.

Координаты всех целей, ложных и «истинных», полученные в данном обзоре, хранятся в соответствующем массиве, зарезервированном в памяти ЦВМ.

Алгоритм моделирования оценок дальности и азимута в АПОИ. Моделирование оценок осуществляется следующим образом.

В программе моделируется последовательность Бернулли. Вероятность появления единицы в этой последовательности p равна вероятности обнаружения одиночного импульса РЛС.

В АПОИ АС УВД используется алгоритм обнаружения, осуществляющий бинарное квантование входного процесса. В качестве порога квантования служит оценка некоторого квантиля помех. Порог квантования определяется как максимальное из двух напряжений U_i и U_p , которые формируются по отражениям от 13 участков разрешения

по дальности, предшествующих зондируемому, и по отражениям от 13 последующих элементарных участков разрешения, соответственно. Напряжения U_L и U_p вычисляются как средние значения. Структурная схема блока квантования показана на рис. 36.

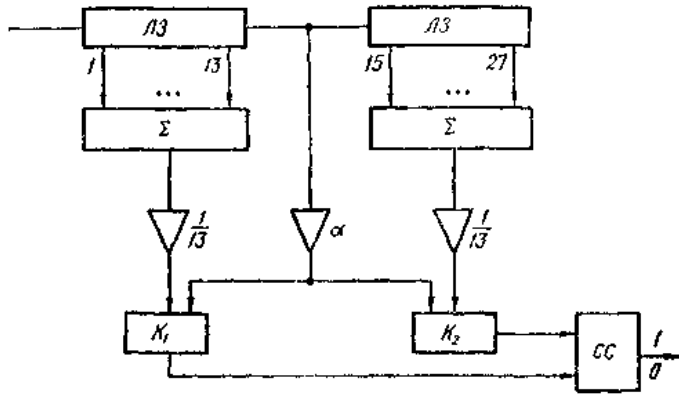


Рис. 36. Схема квантования АПОИ;

K_1 и K_2 — компараторы; CC — схема сравнения, $ЛЗ$ — линия задержки; Σ — сумматор; ∇ — усилитель

Построим аналитическую модель блока квантования АПОИ АС УВД. Вероятность превышения сигнальным отсчетом x_c порогового напряжения U определяется по формуле

$$p\{x_c > U\} = \int_0^{\infty} \omega_c(x) F_{U_L}(x) F_U(x) dx, \quad (24)$$

где $\omega_c(x)$ — плотность распределения отсчетов сигнала; F_{U_L} , F_{U_p} — интегральные функции распределения пороговых напряжений U_L и U_p . Вычислим вероятность (24) в предположении, что на входе радиоприемного устройства присутствует шумоподобный сигнал, а детектор огибающей имеет линейную характеристику. В этом случае плотность распределения сигнальных и шумовых отсчетов подчиняется распределению Релея:

$$\omega_c(x) = \frac{x}{\Psi\alpha(1+b)} \exp\left\{-\frac{x^2}{2\Psi\alpha(1+b)}\right\}; \quad (25)$$

$$\omega_{ш}(x) = \frac{x}{\Psi} \exp\left\{-\frac{x^2}{2\Psi}\right\},$$

где Ψ — дисперсия шума на входе детектора; b — соотношение сигнал/шум по мощности; α — коэффициент, учитывающий различие усиления в сигнальном и шумовом каналах.

Плотность распределения пороговых напряжений U_L и U_p может быть достаточно хорошо приближена нормальным законом с параметрами $m = \sqrt{\pi/2\Psi}$, $\mu = \frac{4-\pi}{2} \Psi$;

$$F_{U_L}(x) = F_{U_p}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\mu}} \exp\left\{-\frac{(t-m)^2}{2\mu}\right\} dt.$$

Воспользовавшись разложением $F(\cdot)$ в ряд Тейлора в окрестности точки $x = m$ и ограничившись двумя его членами, имеем

$$p\{x_c > V\} = \frac{1}{4} \int_A^B \omega_c(x) \left[1 + \frac{4}{\sqrt{2\pi\mu}}(x-m) + \frac{2}{\pi\mu}(x-m)^2\right]^2 dx + \int_B^{\infty} \omega_c(x) dx, \quad (26)$$

где A и B определяются из условия нормировки

$$A = m - \sqrt{\frac{\pi\mu}{2}}; \quad B = m + \sqrt{\frac{\pi\mu}{2}}.$$

Подставляя уравнение (25) в выражение (26) и группируя члены при одинаковых степенях x , получаем

$$p\{x_c > U\} = \frac{1}{4} \sum_{i=0}^2 J_i + \exp\left\{-\frac{B^2}{2\Psi\alpha(1+b)}\right\}, \quad (27)$$

где

$$J_i = K_i [2\Psi\alpha(1+b)]^{i/2} [\gamma(1+i/2, B^2) - \gamma(1+i/2, A^2)], \quad i = 0, 1, 2;$$

$$A^2 = A^2/(2\Psi\alpha(1+b)); \quad B^2 = B^2/(2\Psi\alpha(1+b));$$

$\gamma(\alpha, x)$ — неполная гамма-функция,

$$K_0 = 1 - 4m/\sqrt{2\pi\mu} + 2m^2/(\pi\mu);$$

$$K_1 = \frac{4}{\sqrt{2\pi\mu}} - \frac{4m}{\pi\mu};$$

$$K_2 = 2/(\pi\mu)$$

Выражение (27) является аналитической моделью блока квантования АПОИ АС УВД и используется в аналитико-имитационной модели. Число членов последовательности Бернулли берется несколько большим, чем объем пачки зондирующих импульсов, отраженных от цели за один обзор, так, чтобы началу пачки предшествовало несколько членов последовательности с параметром p ($b = 0$), затем следовало некоторое число, равное принятому объему пачки, членов последовательности с параметром

p ($b \neq 0$) и далее опять несколько членов последовательности с p ($b = 0$).

Описанный алгоритм квантования радиолокационных сигналов является параметрическим адаптивным, так как использует адаптацию порога квантования к изменениям мощности помех, т. е. обладает свойством устойчивости вероятности ложной тревоги. Однако это свойство соблюдается лишь в рамках заданной параметрической модели. При изменениях формы закона распределения вероятностей помех вероятность ложной тревоги также будет меняться.

Алгоритм не является устойчивым при воздействии импульсной помехи с распределением вероятности (9), так как оценки порогов U_L и U_p формируются здесь как выборочные средние.

Повысить устойчивость алгоритма к импульсной помехе можно, используя устойчивые оценки средних значений — выборочные медианы, т. е. порог квантования в модифицированном алгоритме определяется как максимальное из двух напряжений \hat{U}_L и \hat{U}_p , которые являются выборочными медианами, формируемыми по отражениям от 13 участков разрешения по дальности, предшествующих зондируемому и по отражениям от 13 последующих участков дальности соответственно. Структурная схема модифицированного блока квантования показана на рис. 37.

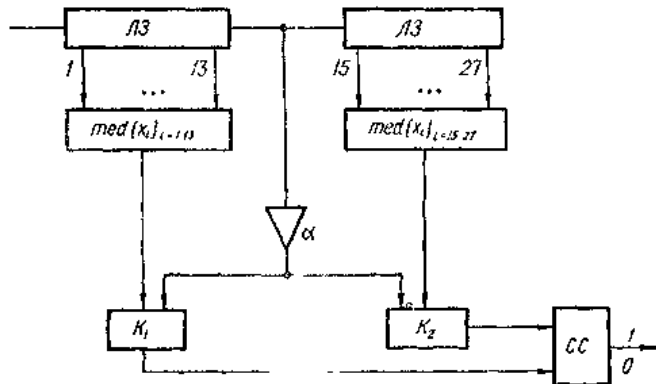


Рис. 37. Схема модифицированного блока квантования с медианным оценщиком среднего значения:

K_1, K_2 — компараторы, $СС$ — схема сравнения; $ЛЗ$ — линия задержки; $med \{x_i\}$ — схема выбора медианы; ∇ — усилитель

Модифицированный блок квантования будет использован в имитационной модели радиолокационной ИИС с целью оценки его

эффективности в условиях воздействия импульсных помех. Вероятность появления одиночного импульса РЛС на выходе блока квантования определяется по характеристике обнаружения заданной таблично в моделирующей программе.

Для сравнения устойчивости двух описанных алгоритмов квантования радиолокационных сигналов к импульсным помехам на рис. 38 показаны их мощностные характеристики.

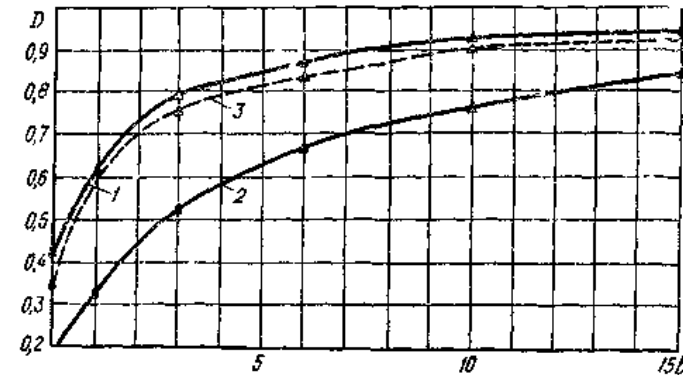


Рис. 38. Характеристики обнаружения блоков квантования ($\alpha \leq 1$): 1, 2 — блок квантования АПОИ для $p = 0,0$ и $p = 0,1$ соответственно; 3 — модифицированный блок квантования для $p = 0,0$ и $p = 0,1$

Характеристики получены методом статистических испытаний. Объем испытаний при оценке вероятности превышения порога квантования составил 1000. Кривые 1 и 2 соответствуют блоку квантования АПОИ (рис. 36). Характеристика 1 снята при отсутствии импульсной помехи, характеристика 2 соответствует воздействию импульсной помехи в каждом элементе разрешения с вероятностью $p=0,1$, характеристика 3 — модифицированному блоку квантования (рис. 37). Для модифицированного блока также вычислялись вероятности обнаружения при отсутствии импульсной помехи ($p = 0$) и в ситуации, когда помеха воздействует с вероятностью $p=0,1$. Практически расхождение между оценками вероятностей в первом и во втором случае находятся в пределах статистического разброса, определяемого объемом статистических испытаний. Таким образом, импульсная помеха со средней скважностью $Q = 1/p = 10$ практически не оказывает влияния на работу модифицированного блока квантования и в то же время существенно ухудшает характеристики немодифицированного блока квантования АПОИ.

Алгоритм определения азимута состоит в обнаружении начала пачки Q_1 , конца пачки Q_2 и вычисления среднего значения

где

$$\hat{\varphi} = \frac{Q_1 + Q_2}{2} - 6,$$

$$Q_1 = Ak \cdot \sum_{j=k}^{k+12} u_{ij} > V$$

при условии

$$\sum_{i=k-1}^{k+11} u_{ij} < V,$$

$$Q_2 = Al : \sum_{i=l}^{l+12} u_{ij} < V$$

при условии

$$\sum_{i=l-1}^{l+11} u_{ij} > V,$$

u_{ij} — результат квантования сигнала, отраженного от i -го участка разрешения по дальности в j -м зондировании; A — угол между направлениями антенны в соседних зондированиях.

Ошибка оценки дальности фомируется как случайная величина, распределенная равномерно на интервале

$$\left[-\frac{D}{2} \dots \frac{D}{2} \right].$$

Применение устойчивых алгоритмов квантования сигналов (рис. 37) обеспечивает устойчивость оценок координат ВС к импульсным помехам.

Алгоритм идентификации отметок целей. Задачей алгоритма является отнесение отметки цели с координатами x_i, y_i , предъявляемой аппаратурой первичной обработки (алгоритмом обнаружения и оценки координат), к какой-либо из сопровождаемых траекторий.

Алгоритм идентификации вычисляет значение функций

$$\Psi_{x_j} = \begin{cases} x_j - x_{i_j}^2, & R < R_{c_j}; \\ 0, & R > R_{c_j}. \end{cases} \quad (28)$$

$$\Psi_{y_j} = \begin{cases} y_j - y_{i_j}^2, & R < R_{c_j}; \\ 0, & R > R_{c_j}. \end{cases}$$

где $R = \sqrt{(x_i - x_{i_j}^2)^2 + (y_i - y_{i_j}^2)^2}$ — эвклидово расстояние цели с координатами x_i, y_i до экстраполированных координат $x_{i_j}^2, y_{i_j}^2$ j -й сопровождаемой цели в i -й периоде зондирования. Экстраполированные координаты $x_{i_j}^2, y_{i_j}^2$ образуют центр строба сопровождения, а R_{c_j} радиус этого строба.

Если координаты цели попадают в j -й строб ($\Psi_{x_j} \neq 0, \Psi_{y_j} \neq 0$), отметка считается принадлежащей j -й траектории. Алгоритм идентификации обеспечивает устойчивость оценок координат и скорости к аномальным наблюдениям.

Алгоритм автосопровождения, по сути, является алгоритмом фильтрации параметров цели — координат x, y и составляющих вектора скорости v_x, v_y . Устойчивый алгоритм фильтрации имеет вид:

$$\hat{x}_i = x_{i_j}^2 + K_{x_{ij}} \Psi_{x_j}(x_i, y_i, x_{i_j}^2, y_{i_j}^2), \quad (29)$$

$$\hat{y}_i = y_{i_j}^2 + K_{y_{ij}} \Psi_{y_j}(x_i, y_i, x_{i_j}^2, y_{i_j}^2),$$

где $x_{i_j}^2, y_{i_j}^2$ — экстраполированные на i -й момент времени значения

координат j -го сопровождаемого ВС; $K_{x_{ij}}, K_{y_{ij}}$ — коэффициенты передачи j -го фильтра по координатам x и y на i -м такте; $\Psi_{x_{ij}}(\cdot), \Psi_{y_{ij}}(\cdot)$ — нелинейные функции результатов наблюдений, полученные из соображений устойчивости оценок (см. формулу (28)).

Алгоритм автосопровождения формирует массив прямоугольных координат j -й цели $x_j(10), y_j(10)$, полученных за последние десять обзоров. Эти данные используются для вычисления сглаженных оценок координат j -й цели в текущем обзоре $x_{i_j}^2, y_{i_j}^2$, а также для экстраполяции значений координат на следующий период обзора для построения стробов АС.

Вычисляются сглаженные оценки составляющих вектора скорости по координатам x и y по методу наименьших квадратов:

$$\left. \begin{aligned} \hat{v}_{x_j} &= \frac{\sum_{i=2}^n (i-1)(x_{i_j} - x_{i_j})}{T \sum_{i=2}^n (i-1)^2}; \\ \hat{v}_{y_j} &= \frac{\sum_{i=2}^n (i-1)(y_{i_j} - y_{i_j})}{T \sum_{i=2}^n (i-1)^2}, \end{aligned} \right\} \quad (30)$$

где $x_{i_j}, y_{i_j}, i = \overline{1, n}$ — значения координат цели, хранящихся в массивах $x_j(10), y_j(10)$; T — период обзора, с.

Экстраполированные на период обзора координаты определяются по формулам:

$$x_{i+1,j}^2 = \hat{x}_{i_j} + \hat{v}_{x_j} T; \quad y_{i+1,j}^2 = \hat{y}_{i_j} + \hat{v}_{y_j} T, \quad (31)$$

где \hat{x}_{i_j} , \hat{y}_{i_j} — сглаженные значения координат в текущем обзоре, которые вычисляются по алгоритму скользящего сглаживания:

$$\left. \begin{aligned} \hat{x}_{i_j} &= x_{i_j}^? + M(x_{i_j} - x_{i_j}^?); \\ \hat{y}_{i_j} &= y_{i_j}^? + M(y_{i_j} - y_{i_j}^?). \end{aligned} \right\} \quad (32)$$

где M — коэффициент сглаживания, который вычисляется по формуле

$$M = \begin{cases} \frac{2(2i+1)}{(i+1)(i+2)}, & \text{при } \frac{2(2i+1)}{(i+1)(i+2)} \geq 0,6; \\ 0,6, & \text{при } \frac{2(2i+1)}{(i+1)(i+2)} < 0,6, \end{cases} \quad (33)$$

где i — номер точки траектории; число 0,6 выбрано согласно рекомендациям, приводимых в ряде работ.

Алгоритм автозахвата (АЗ) или, иными словами, алгоритм обнаружения траекторий, **включает следующие подалгоритмы**: идентификацию отметки цели с какой-либо из траекторий, находящейся на обнаружении; вычисление экстраполированных координат цели и построение стробов АЗ; проверку критерия обнаружения траектории, передачу траектории на АС и очистку канала АЗ; поиск свободного канала АЗ и построение первичных стробов АЗ.

Идентификация цели с траекторией, находящейся на обнаружении, происходит по признаку попадания ее в строб АЗ (см. формулы (28)). Вычисление экстраполированных координат цели происходит по алгоритму, который дается формулами (30) ... (33). Отличие алгоритма АЗ от алгоритма АС состоит в том, что здесь используются только три текущих значения координат. Траектория считается обнаруженной при наличии трех последовательных отметок от цели (**критерий три из трех**). В общем случае возможна модификация критерия обнаружения (**к из n**), которая осуществляется изменением численных параметров — размера «скользящего окна» и порога обнаружения.

При выполнении критерия обнаружения производится передача массива данных текущих и экстраполированных координат на АС. Если отметка от цели не идентифицирована, происходит поиск свободного канала АЗ и при наличии такового — построение первичного строба автозахвата. В рассматриваемой версии модели радиус первичного строба составляет 2500 м. Центр строба совпадает с координатами отметки цели. Если свободного канала нет, координаты цели сбрасываются.

Алгоритм анализа спорных ситуаций. Спорные ситуации возникают при одновременном попадании в строб j -го канала АС двух или более радиолокационных отметок, а также при попадании одной

отметки от цели в два или более стробов. В первом случае истинной считается отметка, находящаяся на минимальном расстоянии от центра строба. Остальные сбрасываются как ложные. Во втором случае отметка считается принадлежащей тому каналу АС, где расстояние будет наименьшим.

Таблица 3

Оценка точностных характеристик АС УВД с учетом первичной и вторичной обработки радиолокационной информации

δ	P на выходе блока квантования АС «Грасса»	P на выходе блока квантования модифицированного	Среднеквадратическое отклонение ошибки оценки координат, м	
1	2	3	4	5
12	0,8	0,92	818	455
13	0,82	0,933	818	378
14	0,84	0,937	731	351
15	0,86	0,940	716	322

Структура моделирующей программы является модульной и позволяет проводить всесторонние исследования по анализу эффективности радиолокационных ИИС (РИИС), влиянию параметров сигналов и помех, структуры АПОИ, алгоритмического обеспечения вторичной обработки на следующие два показателя качества системы: достоверность отображения информации в зоне обзора РЛС, оцениваемая количественно условной вероятностью правильного вывода на индикатор метки цели с формуляром; точность обработки траектории, оцениваемая в общем случае ковариационной матрицей ошибок измерения координат. Использование модели возможно в двух аспектах: исследовательском (решение задачи анализа действующих и перспективных РИИС в заданных областях вариации определяющих параметров, оптимизации, разработка тактико-технических требований к РИИС); эксплуатационном (прогнозирование эффективности действующих РИИС, оценка эффективности РИИС в экстремальных условиях эксплуатации, оптимизация размещения РИИС).

В табл. 3 содержатся результаты моделирования комплекса устойчивых алгоритмов, учитывающих специфику обработки информации в АС УВД. В колонке 4 приведены данные, характеризующие точность оценивания координат самолета при воздействии импульсных помех с вероятностью $p=0,1$ комплексом

алгоритмов с немодифицированным блоком квантования; в колонке 5 — данные, характеризующие точность комплекса алгоритмов с модифицированным блоком квантования, использующим устойчивые оценки среднего значения помехи, основанные на выборочной медиане.

Как показывают результаты счета, эффективность радиолокационной ИИС в условиях воздействия импульсных помех можно существенно увеличить, применяя устойчивые алгоритмы обработки сигнала.

2. Введение в имитационное моделирование

2.1. Введение в имитацию и моделирование

2.1.1. Общие замечания

Проблемы, с которыми приходится сталкиваться в различных сферах нашей жизни (в промышленном производстве, сельском хозяйстве, торговле, управлении и т. д.), постоянно, усложняются. Это определяет необходимость совершенствования имеющихся и разработки новых методов и процедур их решения. Далее будет показано, что имитационное моделирование является эффективным средством решения сложных проблем.

Имитационное моделирование (ситуационное моделирование) — метод, позволяющий строить модели, описывающие процессы так, как они проходили бы в действительности. Такую модель можно «проиграть» во времени как для одного испытания, так и заданного их множества. При этом результаты будут определяться случайным характером процессов. По этим данным можно получить достаточно устойчивую статистику.

Имитационное моделирование — это метод исследования, при котором изучаемая система заменяется моделью, с достаточной точностью описывающей реальную систему, с которой проводятся

эксперименты с целью получения информации об этой системе. Экспериментирование с моделью называют имитацией (имитация — это постижение сути явления, не прибегая к экспериментам на реальном объекте).

Имитационное моделирование — это частный случай математического моделирования. Существует класс объектов, для которых по различным причинам не разработаны аналитические модели, либо не разработаны методы решения полученной модели. В этом случае аналитическая модель заменяется имитатором или имитационной моделью.

Имитационным моделированием иногда называют получение частных численных решений сформулированной задачи на основе аналитических решений или с помощью численных методов.

Имитационная модель — логико-математическое описание объекта, которое может быть использовано для экспериментирования на компьютере в целях проектирования, анализа и оценки функционирования объекта.

Применение имитационного моделирования

Имитационные модели могут применяться для:

- исследования границ и структур систем с целью решения конкретных проблем,
- определения и анализа критических элементов, компонентов и точек в исследуемых системах и процессах,
- синтеза и оценки предполагаемых решений;
- прогнозирования и планирования будущего развития исследуемых систем

Процесс имитационного моделирования начинается с определения подлежащих решению проблем, что в свою очередь определяет состав и границы исследуемой системы. Построение имитационной модели исследуемой системы, хотя и зависит от специфики решаемой проблемы, требует определенной методологической схемы. Имитационный язык обеспечивает исследователя такой схемой, а также осуществляет трансляцию модели в доступную вычислительной системе форму. Компьютер, на котором исследуется разработанная

имитационная модель, выдает информацию о поведении модели, которая затем может анализироваться в процессе решения проблемы.

К имитационному моделированию прибегают, когда :

- дорого или невозможно экспериментировать на реальном объекте;
- невозможно построить аналитическую модель: в системе есть время, причинные связи, последствие, нелинейности, стохастические (случайные) переменные;
- необходимо симитировать поведение системы во времени.

Цель имитационного моделирования состоит в воспроизведении поведения исследуемой системы на основе результатов анализа наиболее существенных взаимосвязей между ее элементами или другими словами — разработке симулятора (англ. *simulation modeling*) исследуемой предметной области для проведения различных экспериментов.

Имитационное моделирование позволяет имитировать поведение системы во времени. Причём плюсом является то, что временем в модели можно управлять: замедлять в случае с быстропротекающими процессами и ускорять для моделирования систем с медленной изменчивостью. Можно имитировать поведение тех объектов, реальные эксперименты с которыми дороги, невозможны или опасны. С наступлением эпохи персональных компьютеров производство сложных и уникальных изделий, как правило, сопровождается компьютерным трёхмерным имитационным моделированием. Эта точная и относительно быстрая технология позволяет накопить все необходимые знания, оборудование и полуфабрикаты для будущего изделия до начала производства. Компьютерное 3D моделирование теперь не редкость даже для небольших компаний.

Имитация, как метод решения нетривиальных задач, получила начальное развитие в связи с созданием ЭВМ в 1950-х — 1960-х годах.

2.1.2. Модели и их построение

Модель — это описание системы. В физических науках модели обычно разрабатываются на основе теоретических законов и

принципов. Моделями могут быть масштабированные физические объекты (иконические модели), математические уравнения и взаимосвязи (абстрактные модели) или графические модели (визуальные модели). Модели эффективно используются, в частности, при описании, проектировании и анализе систем

Разработка модели — сложный процесс, который во многом является искусством, но, однако, упрощается, если:

- 1) известны физические законы, описывающие функционирование системы;
- 2) может быть разработано графическое представление системы;
- 3) можно управлять входами, элементами и выходами системы.

Моделирование сложных крупномасштабных систем чаще всего представляет собой более трудную задачу, чем моделирование физических систем. Это объясняется следующими причинами:

- 1) в распоряжении исследователя имеется мало фундаментальных законов, относящихся к рассматриваемой системе;
- 2) многие взаимосвязи между элементами в системе с трудом поддаются количественному описанию и формализации;
- 3) трудно количественно описать поведение входных элементов;
- 4) важную роль играют стохастические процессы;
- 5) неотъемлемой частью таких систем является процесс принятия решений человеком

Построение моделей

Модель представляет собой абстрактное описание системы, уровень детализации которого определяет сам исследователь. Человек принимает решение о том, является ли данный элемент системы существенным, а следовательно, будет ли он включен в описание системы. Это решение принимается с учетом цели, лежащей в основе разработки модели. От того, насколько хорошо исследователь умеет выделять существенные элементы и взаимосвязи между ними, зависит успех моделирования.

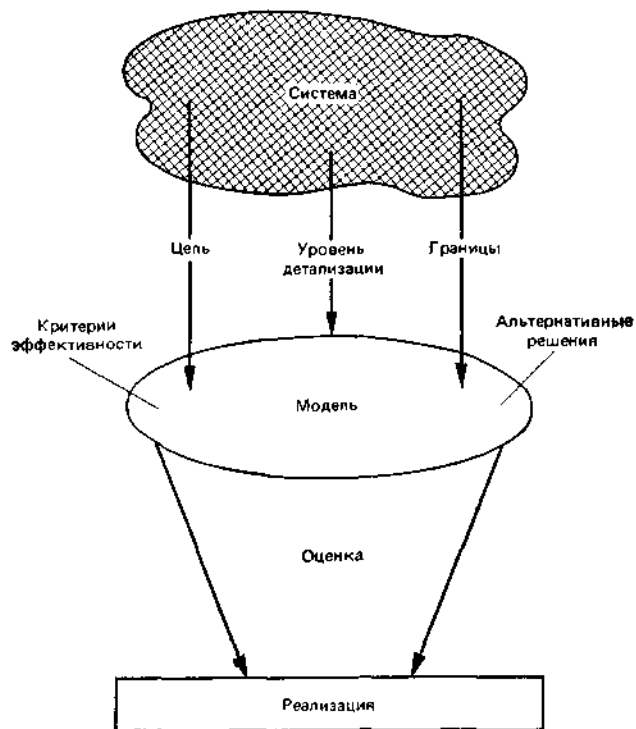


Рис. 1. Процесс построения модели.

На рис. 1 приведена схема предлагаемого подхода к построению моделей. Система рассматривается как состоящая из множества взаимосвязанных элементов, объединенных для выполнения определенной функции. Определение системы во многом субъективно, т.е. оно зависит не только от цели разработки модели, но и от того, кто именно определяет систему.

Итак, процесс моделирования начинается с определения цели разработки модели, на основе которой затем устанавливаются границы системы и необходимый уровень детализации моделируемых процессов. Выбранный уровень детализации должен позволять абстрагироваться от неточно определенных из-за недостатка информации аспектов функционирования реальной системы. В описание системы, кроме того, должны быть включены критерии эффективности функционирования системы и оцениваемые альтернативные решения, которые могут рассматриваться как часть модели или как ее входы. Оценки же альтернативных решений по

заданным критериям эффективности рассматриваются как выходы модели. Обычно оценка альтернатив требует внесения изменений в описание системы и, следовательно, перестройки модели. Поэтому на практике процесс построения модели является итеративным. После того как на основе полученных оценок альтернатив могут быть выработаны рекомендации, можно приступить к внедрению результатов моделирования. При этом в рекомендациях должны быть четко сформулированы как основные решения, так и условия их реализации.

Описанный выше подход полностью применим к построению имитационных моделей. С помощью имитационного моделирования могут строиться как агрегированные, так и детализированные модели. Имитационному моделированию также свойственна концепция итеративного построения модели, в ходе которого модель изменяется путем добавления новых или исключения некоторых ее элементов и (или) взаимосвязей между ними. Эти аспекты имитационного моделирования описываются ниже.

Как мы уже отмечали, особым видом математических моделей являются имитационные модели. Имитационная модель — это компьютерная программа, которая описывает структуру и воспроизводит поведение реальной системы во времени. Имитационная модель позволяет получать подробную статистику о различных аспектах функционирования системы в зависимости от входных данных.

Далее мы будем рассматривать только те имитационные модели, которые разрабатываются для проведения исследований с помощью цифровых ЭВМ. Таким образом, **имитационной моделью** будем называть **логико-математическое описание системы, которое может быть исследовано в ходе проведения экспериментов на цифровой ЭВМ** и, следовательно, может считаться лабораторной версией системы. После окончания разработки имитационной модели с ней проводятся машинные эксперименты, которые позволяют сделать выводы о поведении системы:

- без ее построения, если это проектируемая система;
- без вмешательства в ее функционирование, если это действующая система, экспериментирование с которой или слишком дорого, или небезопасно;

- без ее разрушения, если цель эксперимента состоит в определении пределов воздействия на систему.

Таким образом, имитационные модели могут использоваться для проектирования, анализа и оценки функционирования систем. Имитационное моделирование используется для исследования разнообразных систем, в частности городских, экономических, коммерческих, производственных, биологических, социальных, транспортных систем, систем здравоохранения и др. Наряду со статистическими имитационными методами являются наиболее распространенными средствами теории управления и исследования операций в управлении промышленными предприятиями и организациями.

В имитационном моделировании предполагается, что систему можно описать в терминах, понятных вычислительной системе. Ключевым моментом при этом является выделение и описание состояний системы. Система характеризуется набором переменных, каждая комбинация значений которых описывает ее конкретное состояние. Следовательно, путем изменения значений переменных можно имитировать переход системы из одного состояния в другое. Таким образом, **имитационное моделирование** — это представление динамического поведения системы посредством продвижения ее от одного состояния к другому в соответствии с хорошо определенными операционными правилами.

Изменения состояния системы могут происходить либо непрерывно, либо в дискретные моменты времени. Хотя процедуры описания динамического поведения дискретно и непрерывно изменяющихся моделей различны, основная концепция имитации системы — отображение изменений ее состояния с течением времени — остается той же. В следующем разделе мы проиллюстрируем это положение на примере простой дискретно изменяющейся системы.

2.1.3. Примеры имитационных моделей

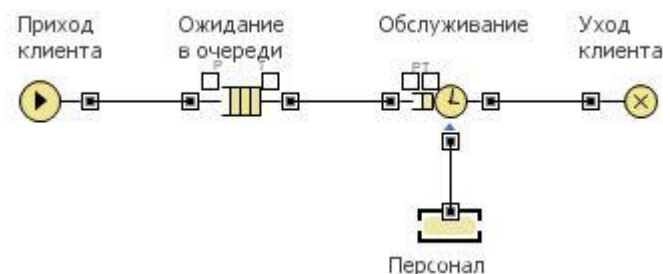
Пример 1

Применение метода имитационного моделирования можно продемонстрировать на примере работы отделения банка по обслуживанию физических лиц. Допустим, что необходимо определить минимальное количество обслуживающего персонала, которое обеспечивает требуемое качество сервиса.

Критерий качества сервиса зададим правилом: средний размер очереди клиентов не должен превышать N человек. Очевидно, что для решения поставленной задачи необходимо иметь достаточные знания о системе: какие клиенты посещают банк, какое количество клиентов приходит в течение рабочего дня, а также сколько времени занимает обслуживание одного клиента.

Хотя данная задача и может показаться специализированной, схожие проблемы возникают во многих областях, где задействованы людские и технические ресурсы. Оплата времени работы квалифицированного работника и времени использования сложной техники составляет немалую долю расходов компаний. Определение оптимального графика использования ресурсов, позволяющего системе эффективно выполнять поставленные задачи, позволяет снизить расходы, а значит увеличить прибыльность.

На первом этапе решения задачи создается модель, которая соответствует структуре и бизнес-процессам отделения банка. В ходе разработки модели учитываются только те детали, которые оказывают существенное влияние на изучаемые аспекты работы системы. Например, наличие отделения обслуживания юридических лиц или кредитного отдела не влияет на обслуживание физических лиц, поскольку они физически и функционально отделены от последнего. Схематично такую модель можно представить в виде последовательности следующих действий.



На втором этапе на вход модели подаются исходные данные: интенсивность прихода клиентов, среднее время обслуживания клиентов, количество доступного персонала. На основании этих данных модель имитирует, или воспроизводит, работу банка в течение заданного промежутка времени, например, рабочего дня.

Время	Событие
19:54	клиент № 167 пришел и встал в очередь
19:56	клиент № 168 пришел и встал в очередь
19:57	клиент № 164 закончил обслуживаться и ушел
19:57	клиент № 167 начал обслуживаться

Следующий этап заключается в анализе статистики, собранной и представленной моделью. Если средний размер очереди клиентов превышает выбранный предел в N человек, то количество доступного персонала следует увеличить и выполнить новый эксперимент.



В результате проведения серии экспериментов над моделью пользователь может определить оптимальное количество персонала. Процесс подбора параметров может быть осуществлен также и с помощью встроенного оптимизатора, который в автоматическом режиме проверяет различные сочетания и находит лучшее решение.

Пример 2.

В качестве примера имитационного моделирования рассмотрим обслуживание кассиром клиентов в банке. Клиенты прибывают в банк, ожидают обслуживания, если кассир занят, обслуживаются и затем покидают банк. Клиенты, прибывающие в банк в тот момент, когда кассир занят, ожидают в единственной к нему очереди. Для простоты мы предполагаем, что момент прибытия клиента и время обслуживания его кассиром известны для каждого клиента (табл. 1).

Таблица 1.
Моменты прибытия времена обслуживания клиентов

Номер клиента	Момент прибытия, мин	Время обслуживания, мин
1	3,2	3,8
2	10,9	3,5
3	13,2	4,2
4	14,8	3,1
5	17,7	2,4
6	19,8	4,3
7	21,5	2,7
8	26,3	2,1
9	32,1	2,5
10	36,6	3,4

Наша цель состоит в ручной имитации функционирования описанной выше системы для того, чтобы определить, какой процент времени кассир свободен и каково среднее время пребывания клиента в банке.

Так как имитационное моделирование — это динамическое отображение изменений состояния системы с течением времени, то в первую очередь должны быть определены состояния системы. Для данного примера состояние моделируемой системы может определяться состоянием кассира (свободен или занят) и числом клиентов в банке. Состояние системы изменяется в результате: 1) прибытия клиента в банк; 2) завершения обслуживания кассиром и последующего ухода клиента. Для иллюстрации имитации мы будем определять состояние системы путем обработки упорядоченных во времени событий, соответствующих прибытию и уходу клиента.

Итоги ручной имитации в соответствии с исходными данными, представленными в табл. 1, приводятся в табл. 2. При этом предполагалось, что в начальный момент времени в системе нет клиентов, кассир свободен и первый клиент прибывает в момент времени, равный 3,2 мин. В табл. 2 первый и второй столбцы взяты из табл. 1.

Время начала обслуживания, приведенное в третьем столбце, зависит от того, покинул ли предыдущий клиент банк. Оно принимается равным наибольшему значению из времен прибытия данного клиента и ухода предыдущего. Время ухода, приведенное в четвертом столбце, вычисляется как сумма соответствующего элемента третьего столбца и времени обслуживания данного клиента, которое

определяется по табл. 1. Значение времени нахождения каждого клиента в очереди и в банке вычисляется, как показано в табл. 2.

Таблица 2.

Ручное моделирование банка					
Номер клиента (1)	Момент прибытия, мин (2)	Момент начала обслуживания, мин (3)	Момент ухода, мин (4)	Время ожидания в очереди, мин (5) = (3) - (2)	Время пребывания в банке, мин (6) = (4) - (2)
1	3,2	3,2	7,0	0,0	3,8
2	10,9	10,9	14,4	0,0	3,5
3	13,2	14,4	18,6	1,2	5,4
4	14,8	18,6	21,7	3,8	6,9
5	17,7	21,7	24,1	4,0	6,4
6	19,8	24,1	28,4	4,3	8,6
7	21,5	28,4	31,1	6,9	9,6
8	26,3	31,1	33,2	4,8	6,9
9	32,1	33,2	35,7	1,1	3,6
10	36,6	36,6	40,0	0,0	3,4

Средние значения этих переменных соответственно равны 2,61 и 5,81 мин.

Таблица 2 содержит итоговую информацию, относящуюся к клиенту, но не содержит сведений о кассире и размере очереди к нему. Для получения такой информации необходимо исследовать события, связанные с этой ситуацией.

Логика обработки событий прибытия и ухода зависит от состояния системы в момент наступления этих событий. При наступлении события «прибытие клиента» в банк дальнейшая ситуация определяется состоянием кассира. Если кассир свободен, он переходит в состояние «занят» и приступает к обслуживанию клиента. При этом планируется событие «уход данного клиента» в момент времени, равный текущему времени плюс продолжительность его обслуживания. Если же кассир занят, обслуживание клиента не может начаться, и, следовательно, он встает в очередь (длина очереди увеличивается на 1). Логика обработки события «уход клиента из банка» зависит от длины очереди. Если в очереди есть хотя бы один клиент, кассир остается в состоянии «занят», длина очереди уменьшается на 1 и для первого клиента из очереди планируется событие ухода. Если же очередь пуста, кассир переводится в состояние «свободен».

В табл. 3 приводится событийно-ориентированное описание состояния кассира и числа клиентов в банке (события расположены в хронологическом порядке).

Таблица 3.

Событийно-ориентированное описание имитации работы банковского кассира

Время события, мин	Номер клиента	Тип события	Длина очереди	Число клиентов	Состояние кассира	Время простоя кассира, мин
0,0	—	Начало	0	0	Свободен	—
3,2	1	Приход	0	1	Занят	3,2
7,0	1	Уход	0	0	Свободен	
10,9	2	Приход	0	1	Занят	3,9
13,2	3	»	1	2	»	
14,4	2	Уход	0	1	»	
14,8	4	Приход	1	2	»	
17,7	5	»	2	3	»	
18,6	3	Уход	1	2	»	
19,8	6	Приход	2	3	»	
21,5	7	»	3	4	»	
21,7	4	Уход	2	3	»	
24,1	5	»	1	2	»	
26,3	8	Приход	2	3	»	
28,4	6	Уход	1	2	»	
31,1	7	»	0	1	»	
32,1	9	Приход	1	2	»	
33,2	8	Уход	0	1	»	
35,7	9	»	0	0	Свободен	
36,6	10	Приход	0	1	Занят	0,9
40,0	10	Уход	0	0	Свободен	

На рис. 1 приведены графики изменения значений этих переменных состояний во времени.

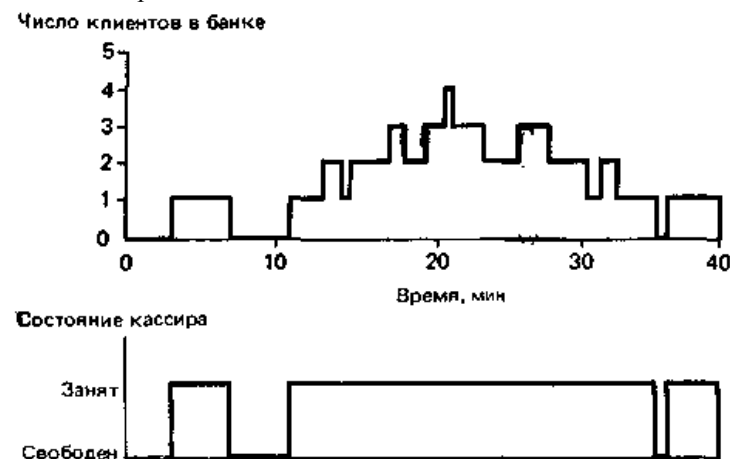


Рис.1. Графическое представление результатов моделирования работы банка.

Результаты имитации показывают, что в течение первых 40 мин работы в банке в среднем одновременно находилось 1,4525 клиента, а кассир был свободен 20% времени.

Для расположения событий прихода и ухода в хронологическом порядке необходимо вести запись (календарь) событий, подлежащих последующей обработке (будущих событий). Это осуществляется путем фиксирования моментов наступления следующего события прихода и следующего события ухода. Сравнение этих моментов определяет затем выбор одного из событий для обработки. Такой упорядоченный список событий обычно называется файлом или календарем событий.

Приведенный пример иллюстрирует целый ряд важных понятий. Мы убедились в том, что в любой момент имитационного времени модель находится в некотором состоянии. При наступлении событий состояние системы меняется в соответствии с логико-математическими отношениями, связанными с этими событиями. Таким образом, события определяют динамическую структуру системы. При заданных начальном состоянии, логической схеме обработки каждого события и методе определения случайных величин наша проблема в большей степени становится вычислительной проблемой. Существенным элементом схемы вычислений является календарь событий, который обеспечивает механизм для записи и упорядочения будущих событий. Еще одним важным моментом является то, что мы можем рассматривать изменения состояния системы в двух аспектах: 1) как процесс, с которым сталкивается клиент при обслуживании (точка зрения клиента), и 2) как последовательность событий, которые вызывают изменение состояния кассира (точка зрения кассира).

2.1.4. Процесс имитационного моделирования

Как уже говорилось выше, процесс последовательной разработки имитационной модели начинается с создания простой модели, которая затем постепенно усложняется в соответствии с требованиями, предъявляемыми решаемой проблемой. В процессе имитационного моделирования можно выделить следующие основные этапы:

1. *Формулирование проблемы*: описание исследуемой проблемы и определение целей исследования.
2. *Разработка модели*: логико-математическое описание моделируемой системы в соответствии с формулировкой проблемы.

3. *Подготовка данных*, идентификация, спецификация и сбор данных
4. *Трансляция модели*, перевод модели на язык, приемлемый для используемой ЭВМ
5. *Верификация* - установление правильности машинных программ
6. *Валидация*: оценка требуемой точности и соответствия имитационной модели реальной системе.
7. *Стратегическое и тактическое планирование*: определение условий проведения машинного эксперимента с имитационной моделью
8. *Экспериментирование*: прогон имитационной модели на ЭВМ для получения требуемой информации
9. *Анализ результатов*: изучение результатов имитационного эксперимента для подготовки выводов и рекомендаций по решению проблемы.
10. *Реализация и документирование*: реализация рекомендаций, полученных на основе имитации, и составление документации по модели и ее использованию.

Хотя некоторые из этих этапов уже обсуждались выше в связи с построением модели, мы тем не менее снова кратко рассмотрим основные этапы процесса имитационного моделирования.

Первой задачей имитационного исследования является точное определение проблемы и детальная формулировка целей исследования. Как правило, определение проблемы является непрерывным процессом, который обычно осуществляется в течение всего исследования. Оно пересматривается по мере более глубокого понимания исследуемой проблемы и возникновения новых ее аспектов.

Как только сформулировано начальное определение проблемы, начинается этап построения модели исследуемой системы. Модель включает статистическое и динамическое описание системы. В статистическом описании определяются элементы системы и их характеристики, а в динамическом — взаимодействия элементов системы, в результате которых происходят изменения ее состояния во времени.

Процесс формулирования модели во многом является искусством. Разработчик модели должен понять структуру системы, выявить правила ее функционирования и суметь выделить в них самое существенное, исключив ненужные детали. Модель должна быть простой для понимания и в то же время достаточно сложной, чтобы реалистично отображать характерные черты реальной системы. Наиболее важными являются принимаемые разработчиком решения относительно того, верны ли принятые упрощения и допущения, какие элементы и взаимодействия между ними должны быть включены в

модель. Уровень детализации модели зависит от цели ее создания. Необходимо рассматривать только те элементы, которые имеют существенное значение для решения исследуемой проблемы. Как на этапе формулирования проблемы, так и на этапе моделирования необходимо тесное взаимодействие между разработчиками модели и ее пользователями. «Первый эскиз» модели должен быть построен, проанализирован и обсужден. Во многих случаях это требует от разработчиков большой ответственности и готовности продемонстрировать свое возможное незнание исследуемой системы. Однако эволюционный процесс моделирования позволяет быстрее обнаруживать допущенные разработчиками неточности и более эффективно их конкретизировать. Кроме того, тесное взаимодействие на этапах формулирования проблемы и разработки модели создает у пользователя уверенность в правильности модели и поэтому помогает обеспечить успешную реализацию результатов имитационного исследования.

На этапе разработки модели определяются требования к входным данным. Некоторые из этих данных могут уже быть в распоряжении разработчика модели, в то время как для сбора других потребуются время и усилия. Обычно значения таких входных данных задаются на основе некоторых гипотез или предварительного анализа. В некоторых случаях точные значения одного (и более) входных параметров оказывают небольшое влияние на результаты прогонов модели. Чувствительность получаемых результатов к изменению входных данных может быть оценена путем проведения серии имитационных прогонов для различных значений входных параметров. Имитационная модель, следовательно, может использоваться для уменьшения затрат времени и средств на уточнение входных данных.

После того как разработана модель и собраны начальные входные данные, следующей задачей является перевод модели в форму, доступную для ЭВМ. Хотя для программирования имитационной модели может использоваться универсальный язык, применение специализированного имитационного языка имеет существенные преимущества. Помимо сокращения времени программирования использование имитационного языка упрощает разработку модели, так как язык содержит набор понятий для формализованного описания системы.

На этапах верификации и валидации осуществляется оценка функционирования имитационной модели. На этапе верификации определяется, соответствует ли запрограммированная для ЭВМ модель замыслу разработчика. Это обычно осуществляется путем ручной

проверки вычислений, а также может быть использован и ряд статистических методов.

Установление адекватности имитационной модели исследуемой системе осуществляется на этапе валидации. Валидация модели обычно выполняется на различных уровнях. Мы рекомендуем выполнять валидацию на уровне входных данных, элементов модели, подсистем и их взаимосвязей. И хотя валидация имитационных моделей достаточно сложна, она является существенно более легкой задачей, чем валидация моделей других типов, например моделей линейного программирования. В имитационных моделях существует соответствие между элементами модели и элементами реальной системы, поэтому проверка адекватности разработанной модели включает сравнение ее структуры со структурой системы, а также сравнение того, как реализованы элементарные функции и решения в модели и системе.

Специальные методы валидации включают установление адекватности путем использования постоянных значений всех параметров имитационной модели или путем оценивания чувствительности выходов к изменению значений входных данных. В процессе валидации сравнение должно осуществляться на основе анализа как реальных, так и экспериментальных данных о функционировании системы. Следует помнить, что имеющиеся в распоряжении исследователя реальные данные о функционировании системы являются всего лишь выборкой из того, что могло произойти в прошлом.

Условия проведения машинных прогонов модели определяются на этапах стратегического и тактического планирования. Задача стратегического планирования заключается в разработке эффективного плана эксперимента, в результате которого либо выясняется взаимосвязь между управляемыми переменными, либо находится комбинация значений управляемых переменных, минимизирующая или максимизирующая отклик имитационной модели. В тактическом планировании в отличие от стратегического решается вопрос о том, как в рамках плана эксперимента провести каждый имитационный прогон, чтобы получать наибольшее количество информации из выходных данных. Важное место в тактическом планировании занимают определение начальных условий имитационных прогонов и методы снижения дисперсии среднего значения отклика модели.

Следующие этапы в процессе имитационного исследования — проведение машинного эксперимента и анализ результатов — включают прогон имитационной модели на компьютере и интерпретацию полученных выходных данных. При использовании

результатов имитационных экспериментов для подготовки выводов или проверки гипотез о функционировании реальной системы применяются статистические методы.

Последним этапом в процессе имитационного исследования являются реализация полученных решений и документирование имитационной модели и ее использования. Ни один из имитационных проектов не должен считаться законченным до тех пор, пока их результаты не были использованы в процессе принятия решений. Успех реализации во многом зависит от того, насколько правильно разработчик модели выполнил все предыдущие этапы процессов имитационного исследования. Если разработчик и пользователь работали в тесном контакте и достигли взаимопонимания при разработке модели и ее исследовании, то результаты проекта скорее всего будут успешно внедряться. Если же между ними не было тесной взаимосвязи, то, несмотря на элегантность и адекватность имитационной модели, сложно будет разработать эффективные рекомендации.

Названные выше этапы имитационного исследования редко выполняются в строго заданной последовательности, начиная с определения проблемы и кончая документированием. В ходе имитационного исследования могут быть сбои в прогонах модели, ошибочные допущения, от которых в дальнейшем приходится отказываться, переформулировки целей исследования, повторные оценки и перестройки модели. Такой итеративный процесс позволяет разработать имитационную модель, которая дает верную оценку альтернатив и облегчает процесс принятия решения.

В заключение отметим, что имитационное моделирование — это средство, которое широко используется для решения проблем, а имитационная модель — абстрактное описание системы. Логика построения модели должна быть простой и понятной всем участникам имитационного исследования; в случае необходимости имитационная модель должна легко модифицироваться. Реализация полученных с помощью имитации рекомендаций является неотъемлемой частью имитационной методологии.

Преимущества имитационного моделирования

Применение имитационных моделей дает множество преимуществ по сравнению с выполнением экспериментов над реальной системой и использованием других методов.

Стоимость. Допустим, компания уволила часть сотрудников, что в дальнейшем привело к снижению качества обслуживания и потери части клиентов. Принять обоснованное решение помогла бы имитационная модель, затраты на применение которой состоят лишь из цены программного обеспечения и стоимости консалтинговых услуг.

Время. В реальности оценить эффективность, например, новой сети распространения продукции или измененной структуры склада можно лишь через месяцы или даже годы. Имитационная модель позволяет определить оптимальность таких изменений за считанные минуты, необходимые для проведения эксперимента.

Повторяемость. Современная жизнь требует от организаций быстрой реакции на изменение ситуации на рынке. Например, прогноз объемов спроса продукции должен быть составлен в срок, и его изменения критичны. С помощью имитационной модели можно провести неограниченное количество экспериментов с разными параметрами, чтобы определить наилучший вариант.

Точность. Традиционные расчетные математические методы требуют применения высокой степени абстракции и не учитывают важные детали. Имитационное моделирование позволяет описать структуру системы и её процессы в естественном виде, не прибегая к использованию формул и строгих математических зависимостей.

Наглядность. Имитационная модель обладает возможностями визуализации процесса работы системы во времени, схематичного задания её структуры и выдачи результатов в графическом виде. Это позволяет наглядно представить полученное решение и донести заложенные в него идеи до клиента и коллег.

Универсальность. Имитационное моделирование позволяет решать задачи из любых областей: производства, логистики, финансов, здравоохранения и многих других. В каждом случае модель имитирует, воспроизводит, реальную жизнь и позволяет проводить широкий набор экспериментов без влияния на реальные объекты.

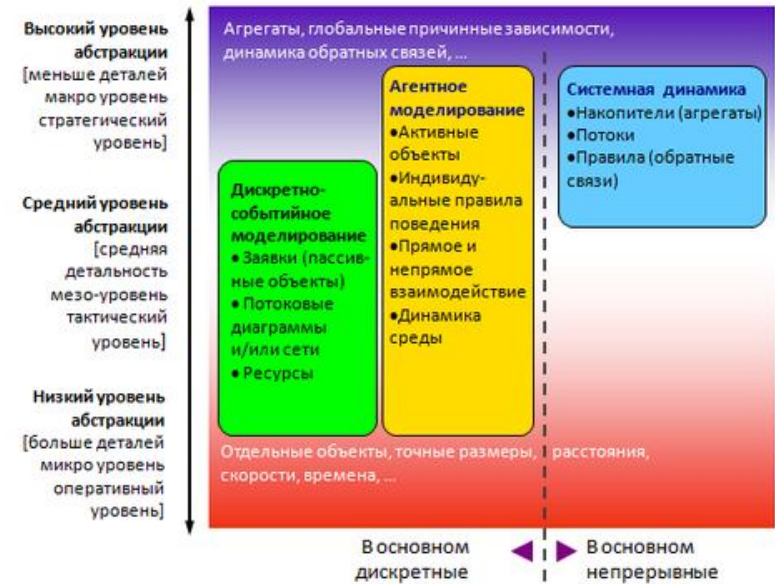
Можно выделить две разновидности имитации:

- Метод Монте-Карло (метод статистических испытаний);
- Метод имитационного моделирования (статистическое моделирование).

Виды имитационного моделирования



Три подхода имитационного моделирования



Подходы имитационного моделирования на шкале абстракции

- **Агентное моделирование** — относительно новое (1990-е-2000-е гг.) направление в имитационном моделировании, которое используется для исследования децентрализованных систем, динамика функционирования которых определяется не глобальными правилами и законами (как в других парадигмах моделирования), а наоборот, когда эти глобальные правила и законы являются результатом индивидуальной активности членов группы. Цель агентных моделей — получить представление об этих глобальных правилах, общем поведении системы, исходя из предположений об индивидуальном, частном поведении ее отдельных активных объектов и взаимодействии этих объектов в системе. Агент — некая сущность, обладающая активностью, автономным поведением, может принимать решения в соответствии с некоторым набором правил, взаимодействовать с окружением, а также самостоятельно изменяться.

- **Дискретно-событийное моделирование** — подход к моделированию, предлагающий абстрагироваться от непрерывной природы событий и рассматривать только основные события моделируемой системы, такие как: «ожидание», «обработка заказа», «движение с грузом», «разгрузка» и другие. Дискретно-событийное моделирование наиболее развито и имеет огромную сферу приложений — от логистики и систем массового обслуживания до транспортных и производственных систем. Этот вид моделирования наиболее подходит для моделирования производственных процессов. Основан Джеффри Гордоном в 1960-х годах.
- **Системная динамика** — парадигма моделирования, где для исследуемой системы строятся графические диаграммы причинных связей и глобальных влияний одних параметров на другие во времени, а затем созданная на основе этих диаграмм модель имитируется на компьютере. По сути, такой вид моделирования более всех других парадигм помогает понять суть происходящего выявления причинно-следственных связей между объектами и явлениями. С помощью системной динамики строят модели бизнес-процессов, развития города, модели производства, динамики популяции, экологии и развития эпидемии. Метод основан Джеймсом Форрестером в 1950 годах.

Области применения

- Бизнес-процессы
- Боевые действия
- Динамика населения
- Дорожное движение
- ИТ-инфраструктура
- Математическое моделирование исторических процессов
- Логистика
- Пешеходная динамика
- Производство
- Рынок и конкуренция
- Сервисные центры
- Цепочки поставок
- Уличное движение

- Управление проектами
- Экономика здравоохранения
- Экосистема
- Информационная безопасность

Свободные системы имитационного моделирования

- Scilab
- Maxima

Действующие системы имитационного моделирования

- MATLAB
- AnyLogic
- Aimsun (моделирование транспортных потоков)
- Arena (моделирование транспортных потоков)
- Business Studio (Имитационное моделирование бизнес-процессов)
- PTV Vision VISSIM (моделирование транспортных потоков и организации дорожного движения)
- eM-Plant
- Powersim
- GPSS
- NS-2
- Transyt
- Tecnomatix Plant Simulation
- simuLab
- Simplex3
- Simul8

2.2. Элементы теории вероятностей и математической статистики

2.2.1. Введение

В общем случае исследуемая система содержит ряд элементов, обладающих некоторой неопределенностью. Подобные системы обычно называют стохастическими, так как их поведение не может

быть полностью предсказано заранее. При имитации стохастических систем требуется описывать изменчивость элементов в терминах теории вероятностей. Поскольку результаты, полученные с помощью имитационной модели, также носят вероятностный характер, они требуют статистической интерпретации. Мы предполагаем, что читатель уже знаком с теорией вероятностей и статистикой, и в данном разделе приводим лишь некоторые основные вероятностные и статистические понятия, относящиеся непосредственно к имитационному моделированию.

2.2.2. Эксперимент, пространство выборки и результат

Эксперимент является хорошо отработанной процедурой или процессом, результат которого можно наблюдать, но нельзя точно предсказать. Множество всех возможных результатов называется **пространством выборки**. Если пространство выборки является **конечным** или **конечночисленным**, оно называется **дискретным**, в противном случае — **непрерывным**.

Комбинация результатов при помощи операций теории множеств, таких, например, как объединение (\cup) и пересечение (\cap), может давать новые результаты (обычно для описания комбинаций результатов используется термин «событие»). В имитационной терминологии понятие события имеет иной смысл, поэтому здесь мы его употреблять не будем). Если результат C определяется как объединение множества результатов A и множества результатов B , т. е. $C = A \cup B$, это означает, что C состоит из множества всех результатов, принадлежащих A или B . Если результат D является пересечением A и B , т. е. $D = A \cap B$, это означает, что D является множеством результатов, принадлежащих и A и B .

В качестве примера использования приведенных выше понятий рассмотрим модель банка с одним кассиром. Посетители приходят в банк, ожидают обслуживания и обслуживаются кассиром. Будем полагать, что интервалы времени между приходами посетителей и продолжительностью их обслуживания кассиром имеют разброс. Пусть наш первый эксперимент заключается в наблюдении интервалов времени между приходами посетителей в банк. Пространство выборки данного эксперимента состоит, таким образом, из всевозможных наблюдений за интервалами между приходами посетителей. **Поскольку интервал времени может быть любым неотрицательным вещественным числом, пространство выборки**

является непрерывным. Результат определяется как подмножество пространства выборки. Таким результатом может быть, например, совокупность интервалов времени продолжительностью от 8 до 9 мин.

В качестве второго примера рассмотрим эксперимент, состоящий в наблюдении числа посетителей, обслуженных в течение первого часа работы банка. **Это число может быть любым неотрицательным целым числом, следовательно, пространство выборки дискретно.** Результат же, например, может быть определен как обслуживание пяти посетителей в течение первого часа работы банка.

2.2.3. Вероятность

Вероятность является мерой возможности осуществления результата. Формально мера вероятности является функцией $P(\cdot)$, которая ставит в соответствие результатам некоторые вещественные числа и удовлетворяет следующим аксиомам:

1. $0 \leq P(E) \leq 1$ для любого результата E .
2. $P(S) = 1$, где S — пространство выборки, или «достоверный результат».
3. Если E_1, E_2, E_3, \dots — взаимно исключающие результаты, то $P(E_1 \cup E_2 \cup E_3 \cup \dots) = P(E_1) + P(E_2) + P(E_3) + \dots$.

Из этих трех аксиом и правил теории множеств могут быть выведены основные законы теории вероятностей. Однако использования этих аксиом недостаточно для вычисления вероятности результата. Обычно числовые значения вероятностей получить довольно сложно, хотя весьма полезно постулировать их существование.

В ряде простых случаев точную вероятность результата можно вычислить, используя комбинаторный анализ. Примером может служить вычисление вероятности выпадения h «орлов» при n бросаниях монеты или вычисление вероятности нахождения трех тузов в раздаче из пяти карт. Однако в большинстве случаев точная вероятность результатов не может быть вычислена. Иногда в таких случаях может быть получено приблизительное значение вероятности результата с помощью ее частотной интерпретации. Если при повторении эксперимента n раз результат E произойдет k раз, то отношение k/n будет частотой появления E . Вероятность результата E может быть выражена так:

$$P(E) = \lim_{n \rightarrow \infty} \frac{k}{n}$$

при условии существования данного предела. При выборе достаточно большого значения n частота k/n будет асимптотически стремиться к вероятности результата E . Можно показать, что полученные подобным образом приблизительные значения вероятностей удовлетворяют аксиомам вероятностей, определенным выше. На практике применение данного подхода ограничивается имеющимися возможностями или целесообразностью проведения требуемого эксперимента.

2.2.4. Случайные величины и вероятностные распределения

Функция, которая ставит в соответствие каждому результату из пространства выборки некоторое вещественное число, называется случайной величиной. **Дискретными** называются те случайные величины, которые принадлежат **конечному или счетному множеству значений**. **Непрерывные** случайные величины могут принадлежать **континууму значений**. В нашей модели банка с одним кассиром интервал времени между приходами посетителей является непрерывной случайной величиной, а число посетителей, обслуженных за первый час работы банка, — дискретной.

Вероятностное распределение представляет собой некоторое правило задания вероятности для каждого из всех возможных значений случайной переменной. Правило задания вероятности имеет две различные формы в зависимости от того, является случайная величина дискретной или непрерывной. Для дискретной случайной величины вероятность каждого ее значения задается функцией вероятности $p(x)$, определяемой как

$$p(x_i) = P(X = x_i).$$

(При определении понятий теории вероятностей и математической статистики мы будем стараться использовать прописные буквы для обозначения случайных величин, а строчные для обозначения их возможных значений. При отступлении от этого правила смысл того или иного обозначения будет обязательно поясняться.)

Для каждого возможного значения x_i функция устанавливает конкретную вероятность того, что случайная переменная X принимает значение x_i . Аксиомы вероятностей накладывают следующие ограничения на $p(x_i)$:

$$0 \leq p(x_i) \leq 1 \quad \text{для всех } i,$$

$$\sum_i p(x_i) = 1.$$

Альтернативой функции вероятности является функция распределения или кумулятивная функция распределения $F(x)$, определяемая следующим образом:

$$F(x) = P(X \leq x).$$

Здесь функция $F(x)$ определяет вероятность того, что случайная величина X примет значение не большее, чем x . Из аксиом вероятностей вытекают следующие свойства $F(x)$:

$$0 \leq F(x) \leq 1 \quad \text{для всех } x,$$

$$F(-\infty) = 0,$$

$$F(+\infty) = 1.$$

Функция распределения связана с функцией вероятности следующим образом:

$$F(x) = \sum_{x_i \leq x} p(x_i).$$

В качестве примера дискретного распределения вероятности рассмотрим эксперимент, заключающийся в трех бросаниях монеты. Пусть случайная величина X обозначает число «орлов», выпавших в результате трех бросаний. Случайная величина X может, таким образом, принимать целые значения от 0 и до 3. При этом возможны восемь результатов, из которых в одном выпадет 0 «орлов», в трех — 1 «орел», в трех — 2 «орла» и в одном — 3 «орла». На рис. 2.1 показана функция вероятности для переменной X , а на рис. 2.2 — функция распределения.

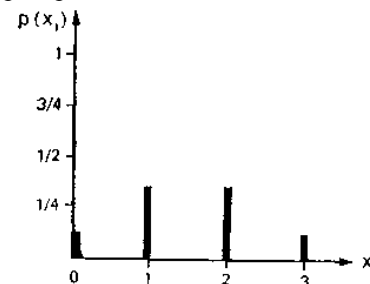


Рис. 2.1. Пример функции вероятности (дискретный случай).

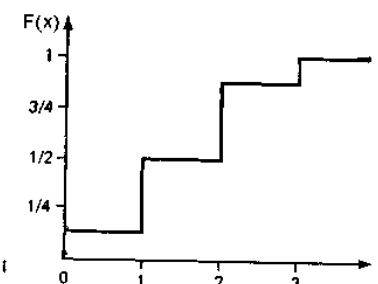


Рис. 2.2. Пример функции распределения (дискретный случай).

Для непрерывных случайных величин требуется иная форма представления вероятностного распределения. Поскольку случайная

величина может принимать любое из бесконечного несчетного множества значений, вероятность конкретного значения равна нулю. Это говорит не о том, что данное значение невозможно, а о том, что оно крайне невероятно вследствие бесконечного числа альтернативных значений. При этом, конечно, вероятность того, что переменная примет значение в интервале между точками а и b, в большинстве случаев не будет равна нулю. Следовательно, функция вероятности для дискретного случая заменяется на непрерывную функцию плотности вероятности $f(x)$, определяемую следующим выражением:

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Таким образом, функция плотности вероятности при интегрировании на интервале от а до b дает вероятность того, что случайная величина примет значение из этого интервала. В соответствии с аксиомами вероятностей функция плотности должна удовлетворять следующим условиям:

$$f(x) \geq 0, \int_{-\infty}^{\infty} f(x) dx = 1.$$

Функция распределения $F(x)$ определяется для непрерывных случайных величин следующим образом:

$$F(x) = \int_{-\infty}^x f(y) dy = P(X \leq x).$$

Функция $F(x)$ определяет вероятность того, что непрерывная случайная величина X примет значение, не большее их. В качестве примера непрерывного вероятностного распределения рассмотрим случайную величину X , которая может принимать значения на интервале от 0 до 1. На рис. 2.3 и 2.4 показаны функция плотности вероятности и функция распределения соответственно для случая, когда любое из несчетного бесконечного множества значений равновероятно.

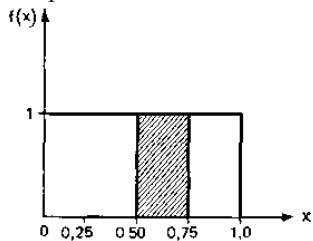


Рис. 2.3. Пример функции плотности вероятности (непрерывный случай).

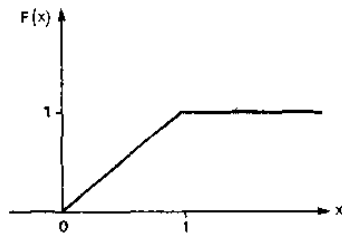


Рис. 2.4. Пример функции распределения (непрерывный случай).

Вероятность того, что случайная величина примет значение на интервале от 0,5 до 0,75, равна площади под кривой функции плотности вероятности на этом интервале. Для случайной величины на рис. 2.3 эта вероятность равна 0,25.

Случайная величина может быть одновременно и дискретной, и непрерывной. Обычно говорят, что такая случайная величина имеет «смешанное» распределение. Случайная переменная со смешанным распределением может принимать как дискретные значения с конечной вероятностью, так и континуум значений, описанных функцией плотности вероятности. На рис. 2.5 изображено смешанное распределение с дискретными значениями 1 и 2, вероятность каждого из которых равна 1/3, что показано вертикальными прямыми со стрелками. Значения между 1 и 2 описываются функцией плотности $f(x) = 1/3$.

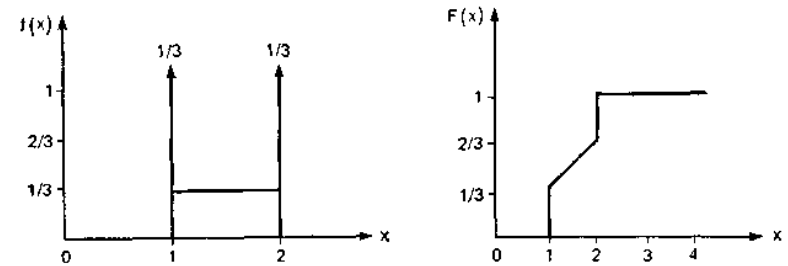


Рис. 2.5. Пример смешанного распределения.

Подобное распределение можно получить из равномерного непрерывного распределения на интервале от 0 до 3, присваивая всем значениям, большим 2, значение, равное 2, а значениям, меньшим 1, — значение, равное 1. Уравнение для функции распределения этой случайной величины имеет следующий вид:

$$F(x) = \begin{cases} 0, & x < 1, \\ \frac{1}{3} + \frac{x-1}{3}, & 1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

Из этого уравнения и рис. 2.5 видно, что $F(x)$ имеет разрывы в точках $x=1$ и $x=2$. В этих точках $P(X=x)$ равна величине скачка, который делает $F(x)$ в точке x . Например, $P(X=1) = 1/3$. В то же время при $1 < x < 2$ функция $F(x)$ непрерывна по x , а $P(X=x) = 0$.

2.2.5. Математическое ожидание и моменты

Часто необходимо охарактеризовать случайную переменную одним или несколькими значениями, которые суммируют информацию, содержащуюся в функции распределения вероятности. Математическим ожиданием случайной величины X , обозначаемым $E[X]$, является значение, определяемое следующим образом:

$$E[X] = \sum_1 x_i p(x_i), \text{ если величина } X \text{ дискретна;}$$

$$E[X] = \int_x x f(x) dx, \text{ если величина } X \text{ непрерывна.}$$

Математическим ожиданием является, таким образом, взвешенная по вероятности средняя величина всех возможных значений X , определяющая меру центральности распределения. Поэтому эта величина часто называется средним значением.

Математическое ожидание можно находить также для функций случайных величин. В частности, математическое ожидание X^n называется n -м моментом случайной переменной и определяется следующим образом:

$$E[X^n] = \sum_1 x_i^n p(x_i), \text{ если величина } X \text{ дискретна;}$$

$$E[X^n] = \int_x x^n f(x) dx, \text{ если величина } X \text{ непрерывна.}$$

Математическое ожидание является частным случаем данного выражения при $n=1$ и называется первым моментом

Вариацией n -го момента называется n -й момент среднего, который определяется выражением

$$E[(X - E[X])^n].$$

Следовательно, перед вычислением n -го момента математическое ожидание X вычитается из X .

Особое значение в теории вероятностей имеет второй момент среднего, называемый обычно дисперсией X и обозначаемый как σ^2 . Дисперсия случайной переменной является мерой разброса вероятностного распределения. Если дисперсия случайной величины мала, вся выборка лежит вблизи математического ожидания. Квадратный корень из дисперсии называется среднеквадратичным отклонением случайной величины.

Если X и Y — случайные величины, то ковариацией X и Y называется величина $Cov[X, Y]$, определяемая следующим образом:

$$Cov[X, Y] = E[(X - E[X])(Y - E[Y])].$$

Ковариация измеряет линейную связь между X и Y . Если результат X не влияет на результат Y , говорят, что X и Y независимы, а $Cov[X, Y] = 0$. В общем случае X и Y независимы тогда и только тогда, когда

$$p(y | x) = p(y) \quad (\text{в дискретном случае } p(y | x) \text{ является вероятностью того, что } Y = y \text{ при } X = x);$$

$$f(y | x) = f(y) \quad (\text{в непрерывном случае } f(y | x) \text{ является условной функцией плотности } Y \text{ при } X = x).$$

Данные выражения утверждают, что вероятностное распределение Y при наличии информации об X аналогично вероятностному распределению Y при отсутствии информации об X .

Мерой зависимости, связанной с ковариацией, является коэффициент корреляции ρ , определяемый выражением

$$\rho = \frac{Cov[X, Y]}{\sqrt{Var[X] \cdot Var[Y]}}.$$

Коэффициент корреляции может лежать в интервале от -1 до $+1$, причем нулевое значение свидетельствует об отсутствии корреляции между X и Y . Положительное значение ρ показывает, что Y увеличивается с увеличением X , а отрицательное значение — что Y уменьшается с увеличением X . Величина ρ отражает степень линейности зависимости Y от X . Если Y линейно зависит от X , то $\rho = \pm 1$. Если X и Y независимы, то график зависимости Y от X представляет собой набор случайных точек, а $\rho = 0$. На рис. 2.6 приведены типичные графики зависимости Y от X при различных значениях ρ .

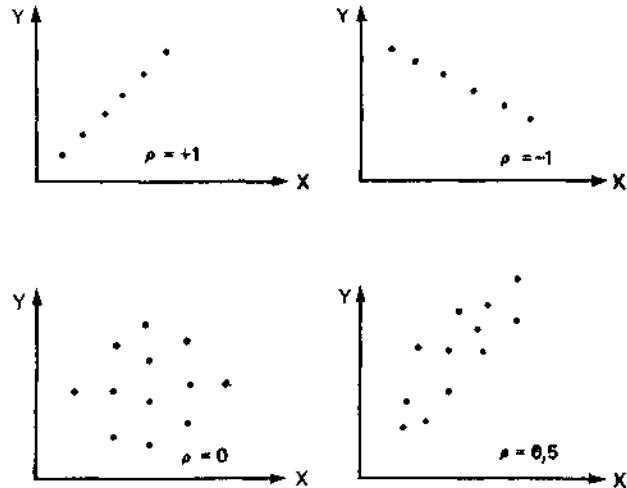


Рис. 2.6. Зависимость Y от X для различных значений ρ

2.2.6. Функции случайных величин

Функция случайной величины также является случайной величиной. В данном разделе показан ряд важных свойств функций случайных величин.

Если X и Y — случайные величины, а k — некоторая константа, то для математического ожидания могут быть выведены следующие свойства:

$$E[X + Y] = E[X] + E[Y],$$

$$E[kX] = kE[X],$$

$$E[X + k] = E[X] + k.$$

Для дисперсии же аналогичные свойства являются менее очевидными:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y],$$

$$\text{Var}[kX] = k^2 \text{Var}[X],$$

$$\text{Var}[X + k] = \text{Var}[X],$$

$$\text{Var}[kX + nY] = k^2 \text{Var}[X] + n^2 \text{Var}[Y] + 2kn \text{Cov}[X, Y].$$

Отметим, что в случае независимости случайных величин X и Y ковариация $\text{Cov}[X, Y]$ равна нулю, и, следовательно,

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y].$$

В математической статистике определенную роль играет случайная величина, называемая выборочным средним (средним по выборке) \bar{X}_I , где I — размер выборки из вероятностного распределения. Выборочное среднее определяется отношением суммы всех значений выборки к ее размеру, т. е.

$$\bar{X}_I = \frac{1}{I} \sum_{i=1}^I X_i.$$

Предположив, что все X_i независимы и одинаково распределены (НОР), можно получить следующие свойства для математического ожидания и дисперсии \bar{X}_I :

$$E[\bar{X}_I] = E[X],$$

$$\text{Var}[\bar{X}_I] = \frac{\text{Var}[X]}{I}.$$

Дисперсия среднего по выборке размером I в I раз меньше, чем дисперсия случайной величины, по которой взята выборка. Следовательно, выбрав I достаточно большим, можно уменьшить дисперсию среднего до любой малой величины.

Отметим, что зависимости, приведенные выше для дисперсии \bar{X}_I , справедливы только для случая независимых наблюдений X_i . Если наблюдения не являются независимыми, вычисление $\text{Var}[\bar{X}_I]$ требует принятия во внимание ковариации между ними. Например, в модели работы кассира в банке времена ожидания приходящих друг за другом посетителей будут коррелированы из-за наличия вероятности того, что $(i+1)$ -й посетитель будет ждать дольше, если i -й посетитель по прибытии в банк помещается в очередь, чем если i -й посетитель сразу же обслуживается. Следовательно, дисперсия среднего времени ожидания не может быть оценена простым делением дисперсии времени ожидания на число наблюдений. Подобная последовательность коррелированных наблюдений называется автокоррелированной выборкой. Далее мы уделим внимание проблеме оценки дисперсии выборочного среднего для автокоррелированных выборок.

2.2.7. Генераторы

Функции случайных величин часто называют функциями-генераторами. Существует целый ряд генераторов, однако мы рассмотрим только производящие функции вероятностей и производящие функции моментов. Производящая функция вероятностей для дискретной случайной величины имеет следующий вид:

$$A(s) = \sum_i p(x_i) s^i.$$

Если вид $A(s)$ известен, то значения $p(x_i)$ могут быть получены путем дифференцирования i раз по s при $s=0$. Математическое ожидание X можно получить из $A(s)$ взятием первой производной по s при $s=1$. Моменты более высокого порядка вычисляются аналогичным способом, но требуют комбинирования производных. Примером генератора случайной функции является Z -преобразование.

Производящая функция моментов (ПФМ) случайной величины определяется следующим образом:

$$M(s) = E[e^{sX}].$$

n -й момент вычисляется путем дифференцирования данного выражения по s :

$$\frac{d^n M(s)}{ds^n} = E[X^n e^{sX}].$$

При $s=0$ имеем $E[X^n]$. Производящая функция моментов называется характеристической функцией.

Кроме вычисления моментов случайной величины по производящим функциям ПФМ полезны для вычисления моментов I сумм независимых случайных величин. Например, если $W=X+Y$, причем X и Y независимы, то

$$E[e^{sW}] = E[e^{s(X+Y)}] = E[e^{sX}]E[e^{sY}].$$

Таким образом, ПФМ W является произведением ПФМ X и ПФМ Y . Моменты W могут быть получены затем с помощью соответствующей ПФМ. В ряде публикаций дается перечень производящих функций вероятностей и моментов.

2.2.8. Закон больших чисел и центральная предельная теорема

Поведение \bar{X}_I при бесконечно большом увеличении выборки определяется двумя весьма важными теоремами. Первой теоремой формулируется сильный закон больших чисел, действие которого интуитивно понятно и заключается в том, что по мере увеличения размера выборки I величина \bar{X}_I стремится к $E[X]$ с вероятностью, равной 1. С данной теоремой связано и действие слабого закона больших чисел, заключающегося в том, что

$$\lim_{I \rightarrow \infty} P\{|\bar{X}_I - E[X]| > \varepsilon\} = 0 \text{ для любого положительного } \varepsilon.$$

Другими словами, для любого положительного, сколь угодно малого ε вероятность того, что модуль разности \bar{X}_I и $E[X]$ превысит ε , стремится к нулю при $I \rightarrow \infty$.

Второй важной теоремой, определяющей поведение \bar{X}_I , является центральная предельная теорема. Она утверждает, что при определенных благоприятных условиях распределение суммы I независимых наблюдений X стремится к нормальному (см. разд. 2.2.9), когда $I \rightarrow \infty$, независимо от характера распределения X . Отсюда следует также, что выборочные средние асимптотически нормально распределены при достаточно больших I . При этом, однако, трудно сказать, какой размер выборки является достаточным для того, чтобы считать величину \bar{X}_I нормально распределенной. Очень часто достаточным бывает относительно небольшой размер выборки (примерно 10—15 наблюдений). Существует целый ряд модификаций центральной предельной теоремы. В частности, при определенных условиях она применима для последовательностей зависимых случайных величин. Описание этих условий приводится в приложении.

2.2.9. Распределения

В предыдущих разделах мы описывали свойства случайных величин и их распределений в общих чертах. Ниже описаны несколько конкретных распределений, весьма важных при моделировании случайных процессов. Описаны также характеристики этих распределений для того, чтобы сориентировать разработчика модели в выборе конкретного типа случайной величины при моделировании некоторого случайного процесса.

Ниже будут использованы следующие обозначения:

X —случайная величина;

$f(x)$ —функция плотности вероятности X ;

$p(x)$ —функция вероятности X ;

a — минимальное значение;

b — максимальное значение;

m — мода;

μ — математическое ожидание $E[X]$;

σ^2 — дисперсия $E[(X-\mu)^2]$;

σ — среднеквадратичное отклонение;

α — параметр функции плотности вероятности;

β — параметр функции плотности вероятности.

Для тех функций плотности вероятности, которые нельзя записать с помощью μ и σ , приводятся выражения для μ и σ .

2.2.9.1. Равномерное распределение

Функция плотности вероятности равномерного распределения задает одинаковую вероятность для всех значений, лежащих между минимальным и максимальным значениями переменной. Другими словами, вероятность того, что значение попадает в указанный интервал, пропорциональна длине этого интервала. Применение равномерного распределения часто вызвано полным отсутствием информации о случайной величине, кроме ее предельных значений. Равномерное распределение называют также прямоугольным. На рис. 2.7 приводятся характеристики равномерного распределения и график его функции плотности вероятности.

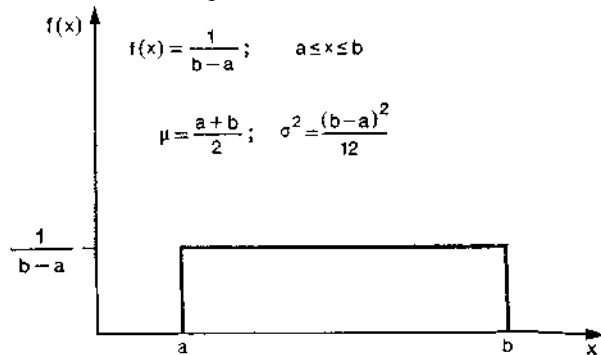


Рис. 2.7. Функция плотности вероятности и характеристики равномерного распределения.

2.2.9.2. Треугольное распределение

Треугольное распределение является более информативным, чем равномерное. Для этого распределения определяются три величины — минимум, максимум и мода. График функции плотности состоит из двух отрезков прямых, одна из которых возрастает при изменении X от минимального значения до моды, а другая убывает при изменении X от значения моды до максимума. Значение математического ожидания треугольного распределения равно одной трети суммы минимума, моды и максимума. Треугольное распределение используется тогда, когда известно наиболее вероятное значение на некотором интервале и предполагается кусочно-линейный характер функции плотности. На рис. 2.8 приведены характеристики треугольного распределения и график его функции плотности вероятности. Треугольное распределение легко применять и интерпретировать, однако для его выбора необходимы веские основания.

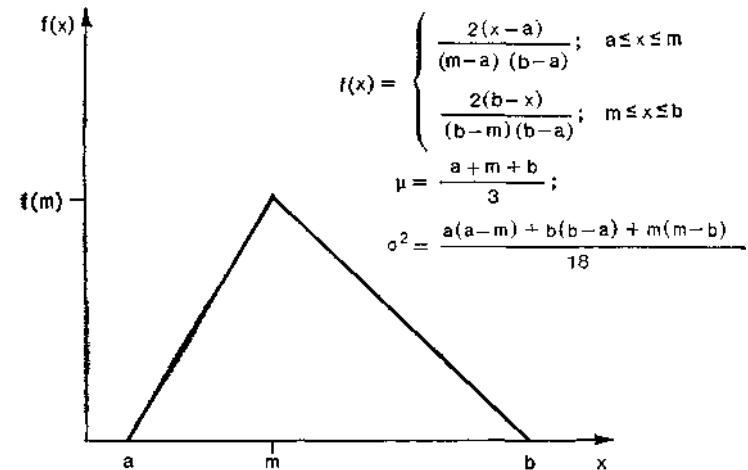


Рис. 2.8. Функция плотности вероятности и характеристики треугольного распределения.

2.2.9.3. Экспоненциальное распределение

Если вероятность того, что один и только один результат наступит на интервале Δt , пропорциональна Δt и если наступление результата не зависит от наступления других результатов, величины интервалов между результатами распределены экспоненциально. Другими словами, работа, продолжительность которой экспоненциально распределена, имеет одинаковую вероятность завершения в течение любого последующего периода времени Δt . Таким образом, работа, выполняемая за t единиц времени, имеет ту же вероятность окончания в последующий период Δt , что и только что начата работа. Подобное отсутствие временной обусловленности называется **марковским свойством** или **свойством отсутствия последствий**. Существует прямая связь между предположением об экспоненциальности распределения продолжительности работы и марковским свойством. Экспоненциальное распределение предполагает значительную вариабельность переменной. Если математическое ожидание продолжительности работы равно μ , то дисперсия равна μ^2 . По сравнению с большинством остальных распределений экспоненциальное обладает большей дисперсией. С экспоненциальным распределением легко осуществлять математические преобразования, благодаря чему оно применяется в целом ряде исследований. На рис.

2.9 представлены характеристики экспоненциального распределения и его функция плотности вероятности.

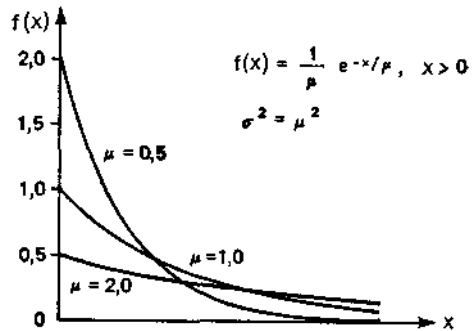


Рис. 2.9. Функция плотности вероятности и характеристики экспоненциального распределения.

2.2.9.4. Распределение Пуассона

Распределение Пуассона является дискретным и обычно связано с числом результатов за определенный период времени. Если продолжительность интервалов времени между результатами распределена экспоненциально и в каждый момент времени может произойти только один результат, то можно доказать, что число результатов на фиксированном интервале времени распределено по закону Пуассона. Другими словами, если интервалы между прибытиями распределены экспоненциально, то распределение числа прибытий будет пуассоновским. Пуассоновское распределение используется часто как аппроксимация биномиального распределения в том случае, когда оно моделирует последовательности независимых испытаний Бернулли (результаты таких испытаний могут быть типа «да-нет», «стоять-идти», «успех-неудача» и т. п.). При больших значениях математического ожидания пуассоновское распределение аппроксимируется нормальным. На рис. 2.10 приведена функция плотности вероятности распределения Пуассона при различных значениях μ .

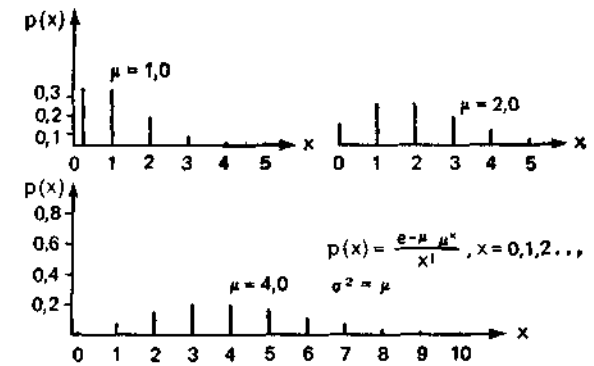


Рис. 2.10. Функция плотности вероятности и характеристики распределения Пуассона.

2.2.9.5. Нормальное распределение

Нормальное, или гауссово, распределение является наиболее важным в теории вероятностей и математической статистике. Эту роль нормальное распределение приобрело в связи с центральной предельной теоремой, которая, как это говорилось выше, утверждает, что при весьма нестрогих условиях распределение средней величины или суммы I независимых наблюдений из любого распределения стремится к нормальному по мере увеличения I . Таким образом, сумму случайных величин часто можно считать нормально распределенной.

Именно благодаря центральной предельной теореме нормальное распределение так часто применяется в исследованиях по теории вероятностей и математической статистике. Существует и другая причина частого применения нормального распределения. Его преимуществом является легкость математического трактования, в связи с чем многие методы доказательств в таких областях, как, например, регрессионный или вариационный анализ, основаны на предположении о нормальном характере функции плотности.

Как уже говорилось выше, при больших значениях среднего нормальное распределение является хорошей аппроксимацией для распределения Пуассона, которое в свою очередь является аппроксимацией биномиального распределения. На рис. 2.11 приведена функция плотности нормального распределения при различных значениях математического ожидания и среднеквадратичного отклонения.

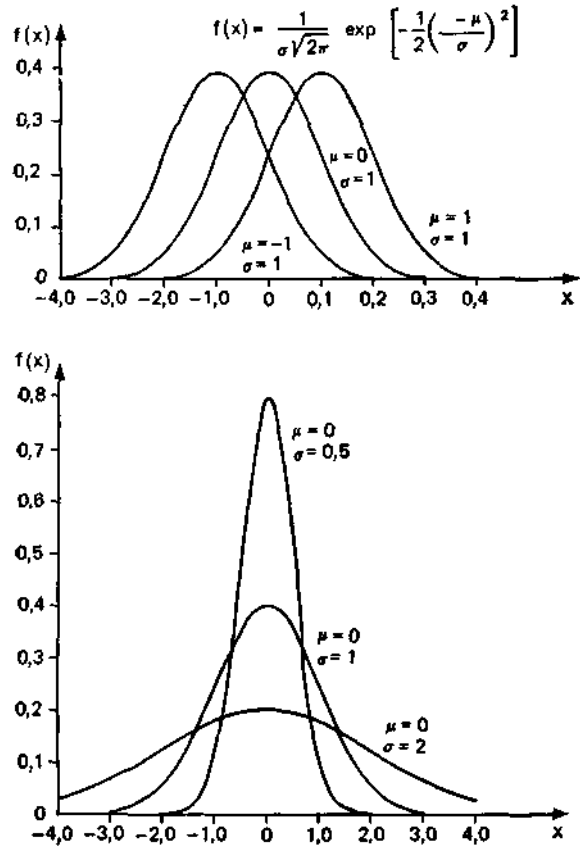


Рис 2. 11. Функция плотности вероятности и характеристики нормального распределения.

2.9.6. Логарифмическое нормальное (логнормальное) распределение

Логнормальное распределение является таким распределением случайной величины, натуральный логарифм которой нормально распределен. Это распределение пригодно для моделирования мультипликативных процессов так же, как нормальное — для аддитивных. С помощью центральной предельной теоремы можно показать, что распределение произведения независимых положительных случайных величин стремится к логнормальному.

Если после логарифмирования каждого элемента некоторого набора данных этот трансформированный набор данных нормально распределен, можно сказать, что исходные данные распределены по логнормальному закону. Логнормальное распределение широко используется для моделирования биологических и экономических систем. Оно хорошо моделирует процессы, в которых значение наблюдаемой переменной является случайной долей от значения предыдущего наблюдения. Примерами подобных процессов являются распределение личных доходов, наследства или банковских вкладов, распределение длины слов и т. п. На рис. 2.12 приведены характеристики логнормального распределения, а также график функции его плотности для различных значений среднего и дисперсии.

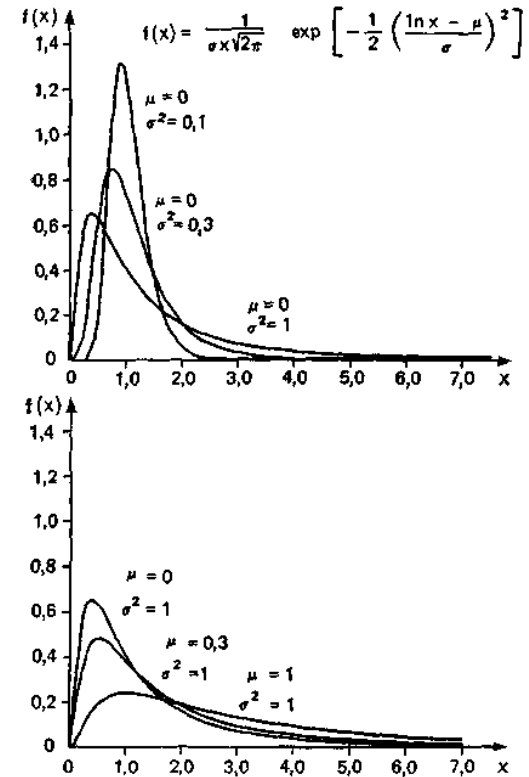


Рис. 2.12. Функция плотности вероятности и характеристики логнормального распределения. μ и σ приведены для соответствующего нормального распределения.

2.2.9.7. Распределение Эрланга

Распределение Эрланга является результатом суммирования независимых и одинаково распределенных экспоненциальных случайных величин. Оно представляет собой частный случай гамма-распределения, поэтому все, что касается функции плотности, интерпретации и замечаний относительно гамма-распределения, справедливо также и для распределения Эрланга. Это распределение широко используется в теории массового обслуживания, когда исследуется выполнение работ в течение экспоненциально распределенных промежутков времени.

2.2.9.8. Гамма-распределение

Гамма-распределение является обобщением распределения Эрланга для случая, когда число суммируемых экспоненциальных величин не является целым. Гамма-распределенная величина может принимать значения от 0 до бесконечности. Функция плотности гамма-распределения принимает различные формы при различных значениях параметров, что позволяет моделировать различные физические процессы.

Гамма-распределение можно интерпретировать также как сумму квадратов нормально распределенных случайных переменных, т. е. как χ^2 -распределение. Таким образом, χ^2 -распределение, распределение Эрланга и, следовательно, экспоненциальное распределение являются частными случаями гамма-распределения.

На рис. 2.13 приведены характеристики гамма-распределения, а также график его функции плотности для различных значений этих характеристик.

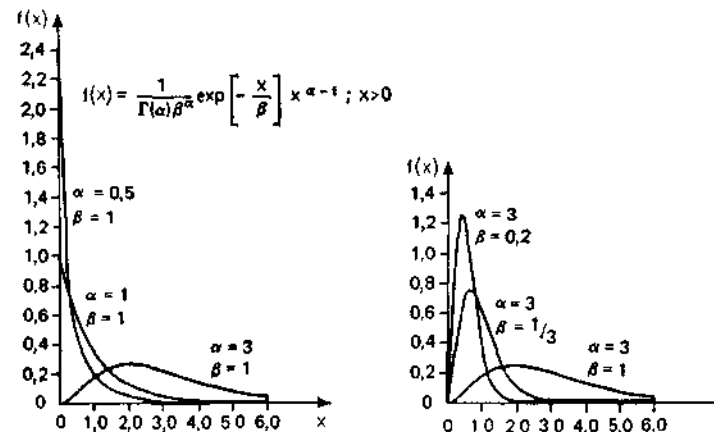


Рис. 2.13. Функция плотности вероятности и характеристики гамма-распределения.

2.2.9.9. Бета-распределение

Бета-распределение определено на конечном интервале и при различных значениях параметров описывается разными кривыми. Эти кривые могут иметь либо форму «колокола» (симметричного или асимметричного), либо U-образную форму. Для U-образных бета-функций значение функции плотности стремится к бесконечности на концах интервала. Один из простейших случаев бета-распределения называется распределением Парето и используется обычно для моделирования распределения доходов. Благодаря тому что бета-функция описывает множество кривых различного вида, она используется для моделирования множества различных данных.

(Здесь необходимо напомнить о предупреждении Феллера относительно закона логистического роста: «единственной теоретической проблемой является то, что не только логистическое распределение, но и нормальное распределение, распределение Коши и другие могут описывать один и тот же статистический материал, давая ничуть не худшие значения критерия согласия распределения... Весьма противоречивые теоретические модели могут опираться на одни и те же наблюдения».)

Поскольку бета-распределение определено на конечном интервале, это обуславливает выбор объекта моделирования. Примерами могут служить функции плотности оценок вероятностей или долей чего-либо. Часто бета-распределение применяется в качестве априорного распределения параметра биномиального процесса в байесовских

исследованиях. Кроме того, бета-распределение часто применяют в качестве описательной функции плотности продолжительности работы в сетевом моделировании. Субъективные оценки продолжительности работы, основанные на оптимистическом (а), пессимистическом (b) и наиболее вероятном (m) значениях, определяют значение оценки среднего в виде $(a+4m+b)/6$ и оценки дисперсии в виде

$$(b-a)^2/36.$$

На рис. 2.14 приведены функции плотности бета-распределения для различных значений его параметров.

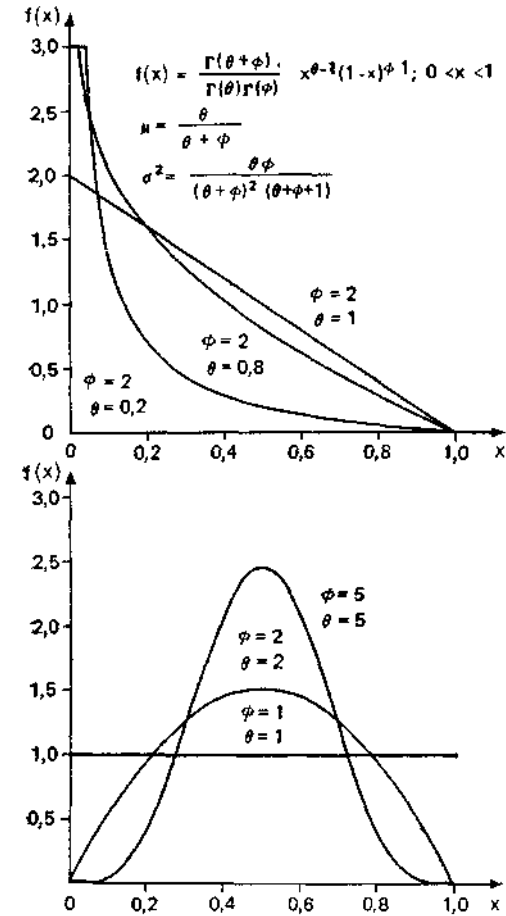
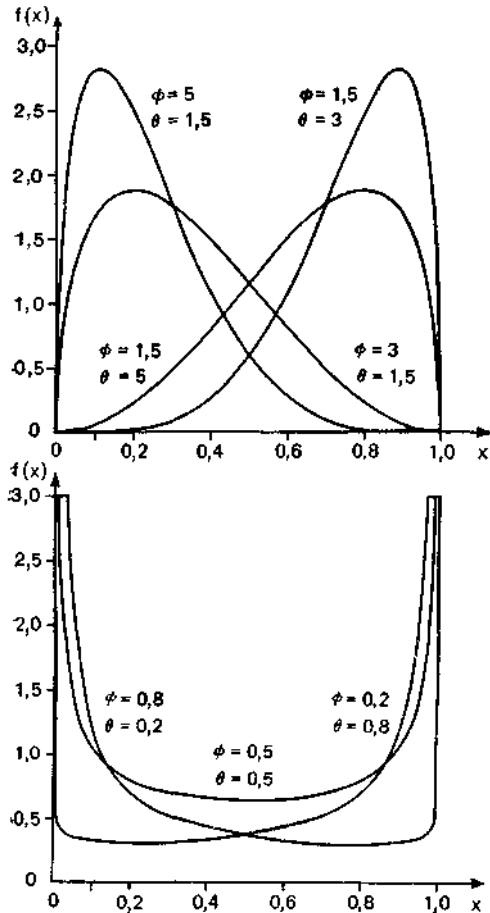


Рис. 2.14. Функция плотности вероятности и характеристики бета-распределения.

2.2.10. Генерация псевдослучайных чисел

В имитационных моделях иногда необходимо получать случайные выборки из одного или нескольких распределений, описанных в предыдущем разделе. Наиболее часто применимым на практике методом получения выборок случайных чисел из заданного

распределения на цифровом компьютере является генерация одного или нескольких случайных чисел, равномерно распределенных на интервале между 0 и 1, и последующее преобразование этого числа или чисел в новое случайное число, распределенное по желаемому закону. Независимые случайные числа, равномерно распределенные на интервале от 0 до 1, являются, таким образом, основой для генерации выборок всевозможных распределений. Мы остановимся ниже на проблеме получения случайных чисел, а затем обсудим процедуры их преобразования в случайные величины, распределенные по различным законам.

В цифровой имитации существует по крайней мере три способа получения случайных чисел. Первым методом является хранение в компьютере таблицы случайных чисел и получение затем из нее данных для имитационного моделирования. Недостаток этого метода заключается в относительно медленной скорости считывания компьютером данных с внешнего устройства ввода и в необходимости хранения большого объема табличных данных. Вторым методом является использование некоторого физического устройства, например электронного прибора, для генерации случайного шума. Недостаток этого метода заключается в невозможности повторного воспроизведения результатов имитации, а следовательно, невозможности осуществления верификации модели и направленного эксперимента с ее параметрами. Третьим методом, которому отдается предпочтение, является применение рекурсивных формул, по которым на основании i -го случайного числа вычисляется $(i+1)$ -е случайное число. Поскольку последовательность чисел вычисляется в уравнении детерминированно, они, естественно, не являются случайными, и их обычно называют псевдослучайными числами.

В дальнейшем для краткости будем называть эти числа случайными, имея в виду, что на самом деле они псевдослучайны.

Генераторы псевдослучайных чисел должны удовлетворять следующим требованиям:

1. Числа равномерно распределены на интервале $(0, 1)$ и независимы, т. е. корреляция между случайными числами последовательности отсутствует.
2. Генерируется достаточное количество неповторяющихся чисел, т. е. период (цикл) генератора довольно длинный.
3. Последовательность случайных чисел воспроизводима. Это предполагает, что различные начальные значения (корни) дают различные последовательности.
4. Генератор должен быть быстродействующим, поскольку для моделирования может потребоваться большое число чисел.

5. Желательно использование малого объема памяти. Наилучшим образом удовлетворяет данным требованиям конгруэнтный метод.

Конгруэнтный метод использует следующее рекурсивное уравнение:

$$z_{i+1} = (az_i + b) \pmod{c}, \quad i = 0, 1, 2, \dots,$$

$$r_{i+1} = z_{i+1}/c,$$

где z_0 — значение корня, а r_i является i -м псевдослучайным числом. Это уравнение определяет, что ненормализованное случайное число z_{i+1} равно остатку деленному $(az_i + b)$, на c , где z_i — предыдущее ненормализованное случайное число, z_0 — начальное значение (корень), а a , b и c — константы. Выбор значений констант a , b и c является предметом постоянных исследований. В приложении кратко излагаются правила определения констант a , b и c для конгруэнтных генераторов. Эти правила дают лишь общие рекомендации для выбора a , b и c , выбор же наилучших значений непосредственно зависит от используемого компьютера. Мы рекомендуем разработчикам моделей использовать генераторы случайных чисел, разработанные специально для компьютера, на котором затем реализуется имитационная модель. В процессе имитации часто необходимо в одной модели работать с несколькими потоками случайных чисел. Например, отдельные потоки случайных чисел могут быть использованы в системе массового обслуживания для моделирования процессов прибытия и обслуживания заявок. При этом можно генерировать одни и те же последовательности моментов прибытия заявок независимо от порядка их обслуживания и, таким образом, оценивать различные процедуры обслуживания для одной и той же последовательности заявок. Разработчику модели предоставляется возможность выбора различных значений корней генератора случайных чисел для параллельных случайных потоков.

2.2.10.1. Метод обратной функции

Случайные числа являются основой для получения величин, распределенных по заданным законам. Простейшим и наиболее фундаментальным методом, на основе которого генерируются подобные величины, является метод обратной функции. В основе этого метода лежит тот факт, что случайная величина $R = F(X)$ равномерно распределена на интервале $[0, 1]$. Другими словами, для генерации случайной величины из распределения X генерируется случайное

число r и решается уравнение $r = F(x)$ относительно значения $x = F^{-1}(r)$. Доказательство состоятельности метода очевидно и основано на следующих соображениях. Пусть $R = F(X)$ имеет функцию распределения $G(\cdot)$. Тогда для $0 \leq r \leq 1$ имеем

$$G(r) = P[F(X) \leq r] = P[X \leq F^{-1}(r)] = F[F^{-1}(r)] = r.$$

Следовательно, R равномерно распределена на интервале $[0, 1]$.

Для иллюстрации метода на примере непрерывного распределения рассмотрим генерацию экспоненциально распределенной случайной величины. Функция экспоненциального распределения имеет вид $F(x) = 1 - e^{-\lambda x}$, где $1/\lambda$ — математическое ожидание. Приравняв $F(x) = r$ и решая уравнение относительно x , получаем

$$x = -(1/\lambda) \ln(1 - r).$$

Если r равномерно распределена на интервале $[0, 1]$, из данного уравнения вытекает, что x распределена экспоненциально с ожиданием, равным $1/\lambda$.

Данный метод применим также и для дискретных распределений. Рассмотрим, например, следующую функцию вероятности:

$$p(0) = 0,25; p(1) = 0,50; p(2) = 0,25.$$

Кумулятивная функция распределения $F(x)$ показана на рис. 2.15.

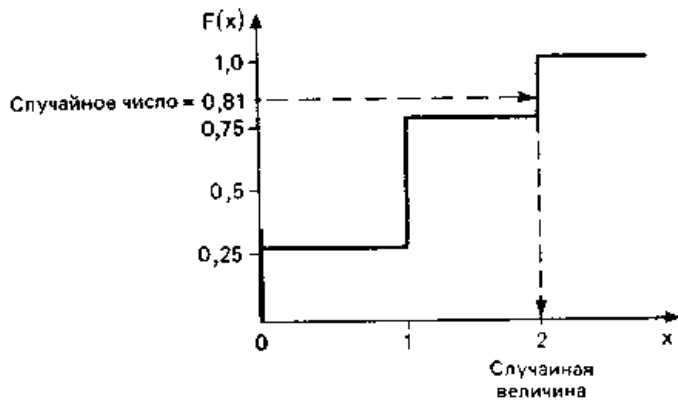


Рис. 2.15. Иллюстрация метода обратной функции для получения выборки из заданного распределения.

Для получения случайной величины из этого распределения необходимо на интервале от 0 до 1 сгенерировать случайное число и нанести его на ось ординат на графике функции распределения. Проведя горизонтальную прямую из этой точки до линии графика функции $F(x)$, а затем проведя вертикальную прямую из этой точки до оси абсцисс, получаем результирующую случайную величину. Например, случайное число 0,81 даст случайную величину, равную 2. Очевидно, что при такой процедуре 25% случайных чисел попадет на интервал $(0; 0,25)$, 50% — на интервал $(0,25; 0,75)$ и 25% — на интервал $(0,75; 1,00)$ в соответствии с заданной функцией распределения. При этом необходимо принять некоторое правило определения результирующей величины для случайных чисел, попадающих в точки разрыва функции.

Затруднения при использовании метода обратной функций происходят обычно при поиске обратного преобразования $F^{-1}(r)$. В ряде случаев метод приводит к простым преобразованиям, подобным тем, что были проделаны для экспоненциального распределения. Тем не менее для ряда непрерывных распределений представление обратной функции в явном виде отсутствует. Для всех основных распределений, не имеющих явного представления обратной функции, разработаны специальные методы генерации.

2.2.11. Тесты на случайность

Статистическая состоятельность результатов имитационного моделирования зависит от степени «случайности» применяемого генератора случайных чисел. В связи с этим разработан целый ряд статистических процедур тестирования генераторов случайных чисел. Однако, не существует некоторого набора тестов, гарантирующего пригодность конечной случайной последовательности вообще. Для данного набора тестов всегда существует удовлетворяющая ему последовательность случайных чисел, но полностью не пригодная для целого ряда частных случаев. Эта оговорка не порождает серьезных проблем, поскольку исследователь, как правило, не нуждается в дополнительных свойствах случайности, выходящих за рамки описанных выше.

Для исследования свойств случайности генераторов применяются как аналитические, так и эмпирические тесты. К ним относятся тесты: частотный, сериальный, интервальный, на сумму цифр, циклический и др. В ряде работ описаны спектральные, решеточные процедуры оценки качества конгруэнтных генераторов случайных чисел относительно их отклонения от идеальных свойств случайности.

2.2.12. Сбор и анализ данных

Важной функцией имитационного моделирования является сбор и анализ данных. Выполнение этой функции необходимо как при определении входных данных для модели, так и при получении результатов эксперимента. Ниже дается обзор ряда важных понятий математической статистики, применяемых при сборе и анализе данных.

2.2.12.1. Подготовка данных

Под подготовкой данных подразумевается процесс получения (сбора) данных об изучаемом явлении. Существует ряд методов получения исходных данных. В некоторых случаях исходные данные содержатся в существующей документации, и тогда задачей исследователя является выявление требуемых данных и организация доступа к ним. В других случаях подготовка данных может включать анкетирование, обзор проблематики или же физическое экспериментирование.

В крупномасштабных моделях, таких, как модели городских или экономических систем, требуемые данные обычно можно получить из существующей документации. Источниками данных для таких моделей служат, например, официальные отчеты, статистические сборники, а также материалы правительственных и международных организаций.

Для моделей производственных систем важным источником данных может служить финансовая и техническая документация. Хотя эта документация часто недостаточна при формировании целостной основы для оценки спроса, стоимости продукции и других важных факторов, она тем не менее служит отправной точкой при моделировании. Анкетирование и обзор предметной области также являются одними из возможных методов получения данных при моделировании производственной деятельности.

Физическое экспериментирование, как правило, является наиболее дорогостоящим и трудоемким методом получения исходных данных. Оно включает в себя измерение, запись и обработку данных. Особое внимание здесь следует уделять планированию эксперимента, которое позволяет убедиться в представительности условий эксперимента и правильности записываемых данных.

В ряде случаев исходных данных может не существовать, при этом сама природа моделируемой системы часто исключает возможность экспериментирования. Примером такой ситуации может служить

имитационное моделирование различных вариантов размещения оборудования сборочной линии. Возможным подходом к подготовке данных в таких случаях может быть предварительный синтез данных, который предполагает вычисление оценок продолжительности работ с использованием таблиц стандартных исходных данных. Таким образом, этот метод позволяет оценивать продолжительность выполнения работ еще до того, как они будут выполняться в действительности.

2.2.12.2. Описание статистических данных

Как при сборе реальных данных для определения входов модели, так и при сборе данных о функционировании системы на основе моделирования мы сталкиваемся с проблемой преобразования «сырых» данных к удобному для анализа виду. Поэтому нас интересуют способы нахождения и описания наиболее важных свойств набора данных. Эти способы обычно позволяют агрегировать данные за счет потери некоторой содержащейся в них информации.

Группировка данных. Одним из методов преобразования данных к удобному виду является группировка данных по классам (интервалам). Данные затем сводятся в таблицу, содержащую количество попаданий элементов данных в каждый класс. Такая таблица называется частотной и обычно дает хорошее общее представление о данных. Примером частотной таблицы могут служить представленные ниже данные о временах ожидания обслуживания посетителями:

Время ожидания, с	Число посетителей
0 → 20	21
20 → 40	35
40 → 60	42
60 → 80	35
80 → 100	19
100 → 120	10
> 120	10

Числа в правом столбце таблицы указывают количество попавших в каждый класс посетителей и называются частотами классов. Числа в левом столбце для каждого класса определяют интервалы значений

наблюдаемой величины и называются границами класса. Разница между верхней и нижней границами класса называется размером класса. Классы, не ограниченные сверху или снизу, называются открытыми. Классы, ограниченные с двух сторон, называются закрытыми. Довольно часто первый и (или) последний классы частотной таблицы бывают открытыми.

Существует несколько видов частотных таблиц, полезных для отображения сгруппированных данных. Одним из видов является таблица накопленных частот, получаемая последовательным сложением значений частот. Ниже приводится таблица накопленных частот для данных о времени ожидания обслуживания посетителями:

Время ожидания (с) меньше, чем	Число посетителей (с накоплением)
20	21
40	56
60	98
80	133
100	152
120	162
∞	172

Числа в правом столбце указывают общее число посетителей, время ожидания у которых оказалось меньше, чем указанная в левом столбце верхняя граница класса. Еще один вид таблиц можно получить, преобразовав частотную таблицу (или таблицу накопленных частот) в таблицу распределения частот путем деления частоты каждого класса (накопленной частоты) на общее число имеющихся в нем элементов данных. Распределения частот полезны, в частности, при сравнении нескольких законов распределений.

С целью улучшения отображения данных частоты и накопленные частоты иногда представляются графически. Наиболее общим графическим представлением является гистограмма, которая отображает в виде прямоугольников частоты классов, причем высота прямоугольников пропорциональна частотам. На рис. 2.16 приведена гистограмма времени ожидания обслуживания посетителями.

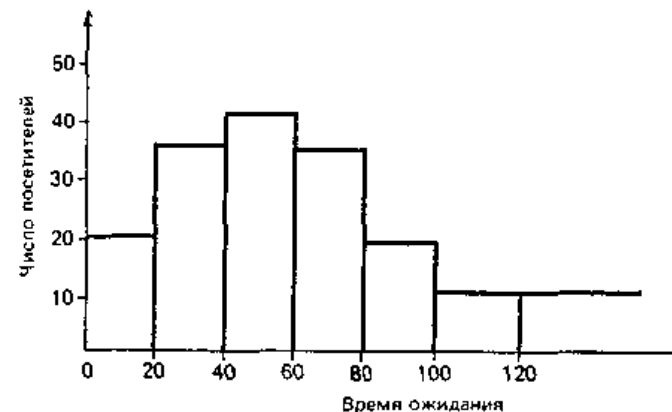


Рис. 2.16. Гистограмма распределения времени ожидания обслуживания посетителями.

Особое внимание при построении распределений частот необходимо уделять выбору числа классов и границ интервалов. При этом необходимо, учитывая природу данных и цель их использования, следовать некоторым общим рекомендациям:

1. По возможности классы должны иметь одинаковый размер, хотя первый и последний классы при этом могут быть открытыми.
2. Интервалы классов не должны перекрываться. Другими словами, каждый элемент данных должен принадлежать одному и только одному классу.
3. Не следует обычно делать меньше 5 и больше 20 классов.

Оценка параметров. Если множество элементов данных содержит множество всех возможных наблюдений, оно называется популяцией. Если же содержит только часть из них, оно называется выборкой. Одним из методов агрегации множества данных является рассмотрение данных как выборки, используемой для оценки параметров исходной популяции. Наиболее существенными параметрами популяции являются среднее, оценивающее меру центральности, и дисперсия, оценивающая меру рассеивания.

Для примера рассмотрим снова данные о времени ожидания обслуживания посетителями. Эти данные можно рассматривать как выборку из популяции, состоящей из всех возможных значений времени ожидания. Мы можем далее по данным, этой выборки оценить среднее время ожидания посетителя и дисперсию времени ожидания для всей популяции.

Для отличия параметров популяции от оценок этих параметров на основе выборки они обозначаются по-разному. Для обозначения среднего и дисперсии популяции используются обычно греческие символы μ и σ^2 соответственно. Оценки же этих параметров на основе выборки x_1, x_2, \dots, x_I обозначаются символами \bar{x}_I и s^2_x соответственно. Для разграничения в дальнейшем этих понятий характеристики популяции будем называть параметрами, а характеристики выборки — статистиками.

Прежде чем продолжить обсуждение статистических характеристик, сделаем ряд уточнений относительно обозначений, принятых для случайных величин, экспериментальных оценок случайной величины и случайной последовательности. Случайная величина до осуществления наблюдения обозначается через X_i , а после его осуществления — символом x_i . Выборочным средним \bar{X}_I будем называть случайную величину, являющуюся суммой I случайных величин до осуществления наблюдений, деленную на I . После осуществления наблюдений x_i среднее будем обозначать \bar{x}_I . Аналогично случайную величину, являющуюся оценкой дисперсии выборки до осуществления наблюдения, будем обозначать символом S^2_x , а после осуществления — символом s^2_x . Таким образом, как было условлено выше, случайные величины обозначаются прописными буквами, а их количественные оценки — строчными.

При построении оценок параметров популяции по данным выборки необходимо рассматривать два различных случая. В первом случае выборка содержит только значения самих наблюдений без учета моментов времени осуществления этих наблюдений. Примером такой выборки могут служить данные о времени ожидания обслуживания посетителями. Статистики по независимой от времени выборке называются статистиками по наблюдениям или точечными статистиками.

Во втором случае значения случайных величин определены во времени. Например, число занятых кассиров в банке является случайной величиной, значение которой меняется во времени. При этом нас интересует информация о том, какие значения принимала наблюдаемая случайная величина и на каких интервалах времени. Статистики по зависимой от времени выборке называются временными или интервальными статистиками.

В табл. 2.1 приведены формулы для вычисления как точечных, так и интервальных статистик \bar{x}_I и s^2_x .

Таблица 2.1.

Формулы для вычисления значений среднего и дисперсии по выборке

Статистика	Формула	
	Точечные статистические оценки	Интервальные статистические оценки
Выборочное среднее	$\bar{x}_I = \frac{\sum_{i=1}^I x_i}{I}$	$\bar{x}_T = \frac{\int_0^T x(t) dt}{T}$
Дисперсия выборки	$s^2_x = \frac{\sum_{i=1}^I x_i^2 - \bar{x}_I^2 I}{I - 1}$	$s^2_x = \frac{\int_0^T x^2(t) dt}{T} - \bar{x}_T^2$

Для интервального случая выборочное среднее обозначается \bar{x}_T , где T равно общей продолжительности интервала времени наблюдения. Для вычисления s^2_x существует несколько формул, однако приводимая здесь формула наиболее удобна с вычислительной точки зрения. Отметим, что для вычисления точечных статистик необходимо знать значения $\sum_{i=1}^I x_i$, $\sum_{i=1}^I x_i^2$ и размер выборки I . Аналогично для вычисления интервальных оценок необходимо знать

$$\int_0^T x dt, \int_0^T x^2 dt \text{ и } T.$$

Еще одной часто используемой при обработке данных оценкой является коэффициент вариации, равный s_x/\bar{x}_I . Он определяет отношение среднеквадратичного отклонения выборки к выборочному среднему. Коэффициент вариации применяется обычно для сравнения дисперсий нескольких наборов данных.

2.2.12.3. Подбор распределения

Предыдущий раздел был посвящен проблеме оценки параметров популяции по выборке. Аналогичной, но более сложной проблемой является идентификация распределения популяции по данным выборки. Она часто возникает в моделировании, так как вероятностные элементы необходимо представлять в модели в виде конкретных распределений. Хотя понимание свойств теоретических распределений, описанных в разд. 2.2.9, помогает разработчику модели

выдвинуть гипотезу о выборе подходящего распределения, все же необходимо проверить ее с помощью одного из статистических тестов. Наиболее пригодными из известных тестов являются тест хи-квадрат и тест Колмогорова — Смирнова. Существуют удобные для использования программные обеспечения, реализующие как графические, так в статистические тесты согласия для подбора теоретического распределения по данным выборки.

2.2.13. Статистический вывод

В имитационных исследованиях часто необходимо на основе результатов имитационных экспериментов делать некоторые выводы или прогнозы относительно поведения моделируемой системы. Поскольку имитационная модель содержит вероятностные элементы, результаты имитации являются наблюдениями случайных величин. Как следствие, любая интерпретация поведения системы на основе анализа полученных результатов имитации должна учитывать их разброс. Подобный учет осуществляется с помощью вычисления доверительных интервалов или проверки гипотез.

2.2.13.1. Доверительные интервалы

В разд. 2.2.12.2 обсуждались методы оценки параметров среднего и дисперсии популяции на основе данных выборки. Оценками параметров были некоторые числа, называемые точечными оценками. В общем случае вследствие случайного разброса оценка отличается от действительного, но неизвестного значения параметра. Недостатком точечной оценки является то, что она не позволяет лицу, принимающему решение, судить о ее точности. Мерой точности является доверительный интервал, определение которого носит вероятностный характер. Доверительный интервал характеризует вероятность попадания значения оцениваемого параметра в заданный интервал.

Наибольший интерес для имитационного анализа представляет среднее значение популяции. Классическое определение доверительного интервала для среднего значения подразумевает независимость и одинаковую распределенность наблюдений.

Следовательно, в соответствии с центральной предельной теоремой выборочное среднее \bar{X}_I распределено приблизительно нормально при достаточно больших I . Как уже говорилось выше, предположение о

независимости не является необходимым условием центральной предельной теоремы.

Если предположить, что \bar{X}_I нормально распределено, то статистика

$$Z = \frac{\bar{X}_I - \mu}{\sigma_{\bar{X}}}$$

является нормально распределенной случайной величиной со средним значением, равным нулю, и среднеквадратичным отклонением, равным единице. Кроме того,

$$P[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1 - \alpha,$$

где $Z_{\alpha/2}$ — такое значение Z , при котором площадь под кривой функции плотности вероятности нормального распределения равна $\alpha/2$. Следовательно, можно с вероятностью $1 - \alpha$ утверждать, что

$$\bar{X}_I - Z_{\alpha/2} \sigma_{\bar{X}} < \mu < \bar{X}_I + Z_{\alpha/2} \sigma_{\bar{X}}. \quad (2.1)$$

В данной формуле предполагается, что величина среднеквадратичного отклонения среднего $\sigma_{\bar{X}}$ известна, хотя это верно далеко не всегда. Если в качестве оценки $\sigma_{\bar{X}}$ принимается выборочное среднеквадратичное отклонение среднего $S_{\bar{X}}$, то, поскольку статистика

$$t = \frac{\bar{X}_I - \mu}{S_{\bar{X}}}$$

является случайной величиной с t -распределением Стьюдента с $I - 1$ степенями свободы, можно вывести следующее аналогичное соотношение, позволяющее определять доверительный интервал $1 - \alpha$ для μ с оценкой $S_{\bar{X}}$:

$$\bar{X}_I - t_{\alpha/2, I-1} S_{\bar{X}} < \mu < \bar{X}_I + t_{\alpha/2, I-1} S_{\bar{X}}, \quad (2.2)$$

где $t_{\alpha/2, I-1}$ — критическое значение t -статистики с $I - 1$ степенями свободы.

Если наблюдения X_i независимы и одинаково распределены, то выражения (2.1) и (2.2) для доверительных интервалов преобразуются путем замены

$$\sigma_{\bar{X}} = \sigma_X / \sqrt{I}, \quad (2.3)$$

$$S_{\bar{X}} = S_X / \sqrt{I}, \quad (2.4)$$

соответственно. Подобная замена позволяет вычислять доверительный интервал по данным выборки. Однако это простое соотношение между дисперсией выборки и дисперсией среднего выборки справедливо только в случае независимости наблюдений.

Методы определения $S_{\bar{X}}$ для выражения (2.2) в случае автокоррелированных наблюдений будут описаны далее. Наиболее

очевидным подходом является организация эксперимента для получения независимых наблюдений, что достигается путем повтора имитационных прогонов или группировки данных.

2.2.14. Проверка гипотез

В ряде имитационных исследований необходимо принять решение о том, справедливо или ложно определенное утверждение относительно некоторого параметра. Например, надо решить, снижает ли изменение правила распределения работ в цехе среднее время запаздывания обрабатываемых заданий. В связи с экспериментальной природой имитации мы должны учитывать случайную вариацию оценок сравниваемых параметров. Это осуществляется с помощью проверки гипотез.

В общем виде процедура проверки гипотез требует определения нулевой гипотезы (обозначается H_0) и альтернативной гипотезы (обозначается H_1). Нулевая гипотеза задается обычно с целью определения того, может ли она быть отвергнута или нет. Например, если мы хотим установить, что правило распределения работ А снижает среднее время запаздывания по отношению к правилу распределения В, нам необходимо определить нулевую и альтернативную гипотезу следующим образом:

H_0 : среднее время ожидания при правиле А равно среднему времени ожидания при правиле В.

H_1 : среднее время ожидания при правиле А меньше среднего времени ожидания при правиле В.

Затем можно использовать результаты имитационного эксперимента при правилах А и В и попытаться отвергнуть гипотезу H_0 в пользу гипотезы H_1 .

Проверка нулевой гипотезы относительно альтернативной подразумевает выбор правила решения, основанного на данных выборки и приводящего к принятию или отказу от нулевой гипотезы. Принятие нулевой гипотезы означает не то, что она справедлива, а то, что на основании данных выборки нельзя сделать уверенного заключения об отказе от нее.

При использовании данного правила решения можно сделать ошибки двух типов. Ошибка первого типа заключается в отказе от нулевой гипотезы, в то время как она верна. Ошибка второго типа состоит в принятии нулевой гипотезы, в то время как она неверна. Принятие решения определяется значением вероятностей, связанных с ошибками первого и второго типа. Эти вероятности обозначаются

обычно как α - и β -вероятности соответственно. Вероятность ошибки типа α называют уровнем значимости теста.

Критерий принятия решения формируется с помощью построения тестовой статистики, имеющей известное распределение. Тестовая статистика вычисляется по данным выборки и проверяется по правилу исключения. Если значение тестовой статистики попадает в тестовую область, нулевая гипотеза отвергается.

Тестовая статистика и правило исключения при проверке гипотез «работают» с характеристиками, приведенными в табл. 2.2. Тесты 1 и 2 проверяют, равно ли значение среднего данному значению μ_0 . Тесты 3 и 4 сравнивают значения двух средних. Поскольку нельзя сформулировать предположение о независимости наблюдений, уравнения для тестовых статистик даны в терминах σ_X и S_X .

Таблица 2.2. Тесты гипотез для средних значений

Нулевая гипотеза	Условие	Статистика теста ¹⁾	Распределение статистики теста	Число степеней свободы	Альтернативная гипотеза	Правило отказа от нулевой гипотезы
1. $\mu = \mu_0$	$\sigma_{\bar{X}}$ — известно	$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$	Стандартное нормальное	—	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$Z > Z_{\alpha}$ $Z < -Z_{\alpha}$ $ Z > Z_{\alpha/2}$
2. $\mu = \mu_0$	$\sigma_{\bar{X}}$ — неизвестно	$t = \frac{\bar{X} - \mu}{S_{\bar{X}}}$	t-стюдента	I-1	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	$t > t_{\alpha}$ $t < -t_{\alpha}$ $ t > t_{\alpha/2}$
3. $\mu_X = \mu_Y$	$\sigma_{\bar{X}}$ и $\sigma_{\bar{Y}}$ известны	$Z = \frac{X - Y}{\sqrt{\sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2}}$	Стандартное нормальное	—	$\mu_X > \mu_Y$ $\mu_X < \mu_Y$ $\mu_X \neq \mu_Y$	$Z > Z_{\alpha}$ $Z < -Z_{\alpha}$ $ Z > Z_{\alpha/2}$
4. $\mu_X = \mu_Y$	$\sigma_{\bar{X}}$ и $\sigma_{\bar{Y}}$ неизвестны	$t = \frac{\bar{X} - \bar{Y}}{\sqrt{S_{\bar{X}}^2 + S_{\bar{Y}}^2}}$	t-стюдента	Ближайшее целое к $\frac{(S_{\bar{X}}^2 + S_{\bar{Y}}^2)^2}{S_{\bar{X}}^4 \frac{I_X}{I_X + 1} + S_{\bar{Y}}^4 \frac{I_Y}{I_Y + 1}}$	$\mu_X > \mu_Y$ $\mu_X < \mu_Y$ $\mu_X \neq \mu_Y$	$t > t_{\alpha}$ $t < -t_{\alpha}$ $ t > t_{\alpha/2}$

¹⁾ Индекс I в X и Y опущен для упрощения. I — размер выборки; I_X — размер выборки для X; I_Y — размер выборки для Y; μ_0 — гипотетическое среднее; α — уровень значимости

2.2.15. Статистические проблемы имитационного моделирования

Анализ решений, принимаемых на основе анализа результатов имитационного моделирования, требует обычно получения оценок усредненного отклика имитационной модели и его дисперсии. Обе эти оценки зависят от условий эксперимента. К условиям эксперимента, которые должен определить разработчик модели, относятся начальное, или исходное, состояние имитируемой системы, момент начала сбора статистических данных, продолжительность прогона модели и число повторных прогонов. Ниже кратко описаны проблемы, связанные с заданием этих условий. Они подробно обсуждаются в последующих разделах.

2.2.15.1. Начальное состояние имитируемой системы

Любая имитационная модель в неявном виде подразумевает наличие начальных условий или исходного состояния имитации. Простейшим и наиболее общепринятым начальным состоянием является состояние «пуст и свободен», при котором имитация начинается в условиях отсутствия в модели компонентов и свободного состояния всех обслуживающих устройств. Приемлемость подобного начального условия зависит от природы моделируемой системы, а также от того, переходный или установившийся режим функционирования системы нас интересует. (Установившийся режим означает не отсутствие разброса отклика имитационной модели, а то, что вероятностный механизм, описывающий этот разброс, остается неизменным и не подвержен воздействию начальных условий.)

Если целью исследования является анализ установившегося режима работы системы, обычно всегда есть возможность улучшить качество статистических оценок с помощью выбора начального состояния, отличного от состояния «пуст и свободен». Начальные условия могут быть заданы на основе определения наиболее характерного для установившегося режима состояния системы, полученного в результате пробного имитационного прогона. Если же целью является анализ переходного режима, начальные условия должны отражать исходное моделируемое состояние системы.

2.2.15.2. Момент начала сбора статистических данных

Уменьшение смещения оценок характеристик в установившемся режиме, вызванного воздействием начальных условий, часто осуществляется с помощью метода, при котором начало сбора статистических данных задерживается до момента завершения периода «разогрева». Обычно для этой цели определяется момент отсечения, указывающий что собранные до него данные не учитываются при вычислении статистических оценок. Желаемое снижение влияния начальных условий достигается, следовательно, путем уменьшения числа собранных наблюдений в течение переходного периода имитации. Однако отсечение части данных может привести к увеличению значения оценки дисперсии среднего. Таким образом, улучшение качества оценки среднего достигается ценой увеличения разброса результатов имитационного моделирования.

Наиболее часто точка отсечения определяется по графику отклика, полученному в результате пробного прогона имитационной модели. Момент времени отсечения выбирается так, что значение отклика по графику можно считать установившимся. Существует ряд методов формализации данной процедуры в виде правила, которое включается в программу имитационной модели и автоматически определяет точку отсечения в ходе имитации. Эти правила будут описаны в другом разделе.

2.2.15.3. Продолжительность имитационного прогона и число повторных прогонов

Важным моментом планирования имитационного эксперимента является определение соотношения между продолжительностью прогона модели и числом повторных прогонов. Использование нескольких продолжительных прогонов предпочтительнее, чем использование множества коротких, так как в общем случае это дает лучшую оценку для среднего в условиях установившегося режима, поскольку меньшее число раз вносятся искажения переходного режима и отсекается меньше данных. Однако уменьшение числа наблюдений в связи с уменьшением числа повторных прогонов, с одной стороны, может увеличить оценку дисперсии среднего. Большое число коротких прогонов, с другой стороны, может внести искажения, вызываемые начальными условиями. Чем больше переходный период, тем важнее использовать более продолжительные прогоны с целью снижения влияния начальных условий.

Существует ряд методов задания продолжительности имитационного прогона. Наиболее часто, по-видимому, задается момент времени завершения моделирования. Недостатком этого метода является то, что число наблюдений, будучи случайным, может быть различным в каждом из повторных прогонов. Метод, который позволяет управлять размером выборки, заключается в задании определенного числа компонентов, поступающих на вход модели. В этом случае имитация продолжается до тех пор, пока заданное число компонентов не будет полностью обработано в модели. Таким образом, после окончания имитационного прогона система будет находиться в состоянии «пуст и свободен». Другим подходом, аналогичным данному, является задание числа компонентов, обрабатываемых в системе. При этом имитационный прогон может завершиться, когда система находится в любом, отличном от пустого состоянии. Используя такой подход, необходимо всегда обеспечивать, чтобы компоненты, оставшиеся необработанными, были типичными представителями выборки. Например, этот метод непригоден, когда в модели системы используется правило распределения работ по минимуму времени их выполнения, и, следовательно, к концу имитации в очереди могут накопиться работы, для выполнения каждой из которых требуется много времени.

Еще одним подходом к управлению продолжительностью имитационного прогона является применение правил автоматической остановки, которые позволяют автоматически отслеживать результаты моделирования через заданные интервалы времени в процессе имитации. Имитация прекращается, когда оценка дисперсии среднего становится меньше заданной величины. Более подробно правила автоматической остановки рассматриваются в другом разделе.

Если мы оцениваем дисперсию выходной переменной X с помощью повторных прогонов и если предполагаем, что X нормально распределена (если X является выборочным средним, это вполне справедливо), то число независимых повторных прогонов, которое необходимо осуществлять для достижения заданного доверительного интервала для X , будет равно

$$I = \left(\frac{t_{\alpha/2, I-1} S_x}{g} \right)^2,$$

где $t_{\alpha/2, I-1}$ — величина, взятая из таблицы критических значений t -статистики с $I-1$ степенями свободы; g — половина длины заданного доверительного интервала.

Применение этой формулы требует информации о t -статистике с $I-1$ степенями свободы и S_x . Обычно устанавливают значение I ,

проводят I повторных прогонов имитационной модели, на основе проведенных прогонов вычисляют значения t и s_x , а затем применяют приведенную выше формулу для проверки достаточности начальных предположений или для определения необходимого числа дополнительных прогонов.

Нами изложены основы теории вероятностей и математической статистики, знание которых необходимо при проведении имитационного анализа. При этом дается достаточно широкий обзор вероятностных и статистических положений, имеющих отношение к имитационному моделированию, без подробного изложения каждого из них. Представленный материал вполне достаточен для понимания цели имитационного моделирования и экспериментальной природы имитационного анализа. Он также дает возможность читателю перейти далее к изучению более тонких аспектов статистического анализа в имитационном моделировании, изложенных далее.

Приложение

Конгруэнтные генераторы

Смешанные конгруэнтные генераторы

Полный период цикла, равный 2^B , будет получен на В-битовом компьютере для генератора

$$z_{i+1} = (az_i + b) \pmod{c}$$

в том случае, если $c=2^B$, b — простое число относительно c (т. е. наибольший общий делитель b и c равен 1) и $a \equiv 1 \pmod{4}$ или $a \equiv 1 + 4k$, где k целое.

Мультипликативные конгруэнтные генераторы

Максимальный период цикла, равный 2^{B-2} , будет получен на В-битовом компьютере для генератора

$$z_{i+1} = az_i \pmod{c}$$

в том случае, если $c=2^B$, $a = \pm 3 + 8k$ или $a \equiv 1 + 4k$ для целых k и z_0 нечетно.

Такие генераторы называют мультипликативными с максимальным периодом.

Для мультипликативных конгруэнтных генераторов период, равный $c-1$, может быть получен, если $c=2^B-1$, а a — простой корень c (а является простым корнем c , если $a^{c-1} \equiv 1 + ck$, где k — целое, и для любого целого $q < c-1$ выражение $(a^q - 1)/c$ не является целым). Эти генераторы называются мультипликативными конгруэнтными с простым модулем.

2.3. Методологические подходы в имитационном моделировании

2.3.1. Введение

При разработке имитационной модели аналитику необходимо выбрать концептуальную схему для описания моделируемой системы. Эта схема базируется на определенном методологическом подходе, в рамках которого воспринимаются и описываются функциональные взаимосвязи системы. Если разработчик модели применяет имитационный язык, методологический подход обычно неявно задается этим языком. Однако, если разработчик использует универсальный язык, такой, как Фортран, ПЛ/1, Бейсик и др., он сам должен разработать такой подход. В любом случае методологический подход, применяемый разработчиком, позволяет ему четко сформулировать описание системы. В этом разделе обобщаются альтернативные подходы к имитационному моделированию.

2.3.2. Системы и модели

Система — это совокупность элементов, которые принадлежат ограниченной части реального мира, являющейся объектом исследования. Поэтому система — понятие относительное. В одном случае некоторая совокупность элементов может рассматриваться только как небольшая часть большой системы, т. е. в качестве подсистемы, а в другом та же совокупность может быть в центре интересов исследователя, т. е. рассматриваться как система. **Сфера действия любой системы и любой модели системы однозначно определяется целью**, для достижения которой она выделяется и идентифицируется. Сфера действия любой имитационной модели определяется также особенностями той проблемы, для решения которой разрабатывается эта модель.

Для установления сферы действия системы исследователь должен выявить ее **границы и состав**. При установлении границ системы выявляются не только физические, но и причинно-следственные взаимосвязи между ее элементами. На систему, которой дано предварительное определение, могут воздействовать некоторые внешние факторы. Если они существенно влияют на поведение системы, экспериментировать с такой системой не имеет смысла, и ее

следует, переопределить. Если внешние факторы частично воздействуют на систему, существуют следующие возможности:

- расширить определение системы, включив в него эти факторы;
- пренебречь этими факторами;
- трактовать их как входы в систему.

Если внешние факторы трактуются как входы в систему, предполагается, что они функционально задаются с помощью предписанных значений, таблиц или уравнений. Например, когда разрабатывается модель производственной системы фирмы и сбыт производимого этой фирмой изделия рассматривается как вход в производственную систему, в модель не включаются причинно-следственные взаимосвязи, относящиеся к процессу сбыта. Такая модель будет содержать только статистическое описание предшествующих и предполагаемых продаж, используемое в качестве входа, т. е. организация сбыта находится за границами моделируемой системы. В системной терминологии объекты, которые находятся за границами системы, но могут влиять на ее поведение, формируют окружающую среду этой системы. Таким образом, системы представляют собой совокупность взаимодействующих элементов, которые подвергаются воздействию со стороны внешних факторов. На рис. 3.1 представлена модель такой системы.

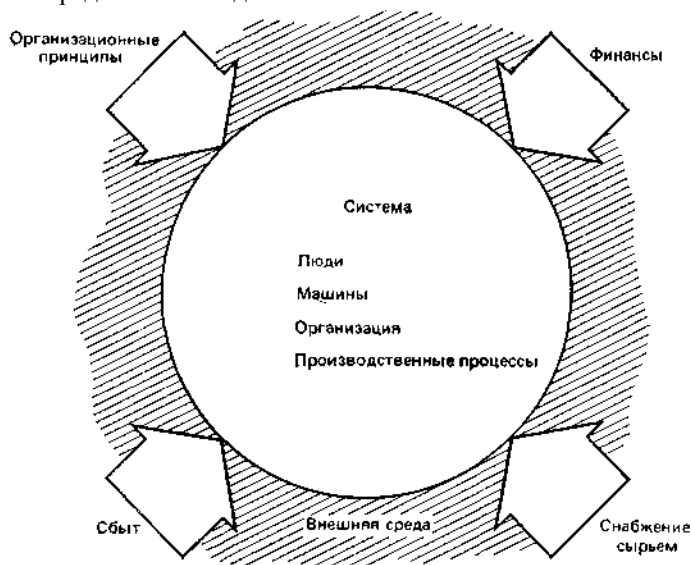


Рис 3.1. Модель производственной системы в условиях внешних воздействий.

Модели систем классифицируются на дискретно и непрерывно изменяющиеся. Отметим, что эти термины относятся к модели, а не к реальной системе. Практически одну и ту же систему можно представить в виде дискретно изменяющейся модели (далее называемой просто дискретной) либо непрерывно изменяющейся (непрерывной). Как правило, в имитационном моделировании время является основной независимой переменной. Другие переменные, включенные в имитационную модель, являются функциями времени, т. е. зависимыми переменными. Определения «дискретная» и «непрерывная» относятся к поведению зависимых переменных.

При *дискретной имитации* зависимые переменные изменяются дискретно в определенные моменты имитационного времени, называемые моментами свершения событий. Переменная времени в имитационной модели может быть либо непрерывной, либо дискретной в зависимости от того, могут ли дискретные изменения зависимых переменных происходить в любые моменты времени или только в определенные моменты.

Имитация банковской системы, обсуждаемая в раз. 2.1, является примером дискретной имитации. Зависимыми переменными в этом примере являются состояние кассира и число ожидающих в очереди клиентов. Моменты свершения событий соответствуют моментам времени, когда клиент прибывает в систему и покидает ее после окончания обслуживания кассиром. Как правило, в дискретных моделях значения зависимых переменных не изменяются в промежутках между моментами свершения событий. Пример изменения зависимых переменных в дискретной имитационной модели показан на рис. 3.2.



Рис 3.2. Графическое представление отклика дискретно-событийного имитатора.

При *непрерывной имитации* зависимые переменные модели изменяются непрерывно в течение имитационного времени. Непрерывная модель может быть либо непрерывной (рис. 3.3), либо дискретной по времени (рис. 3.4) в зависимости от того, будут ли значения зависимых переменных доступны в любой точке или только в определенные моменты имитационного времени.

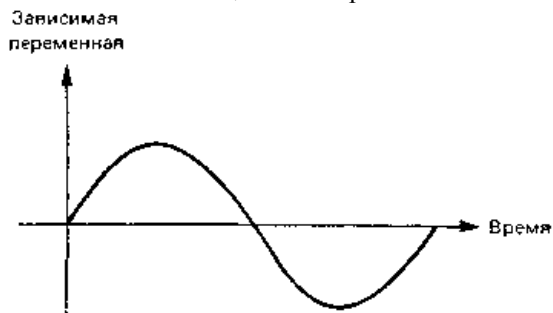


Рис 3. 3. Графическое представление отклика непрерывного имитатора.

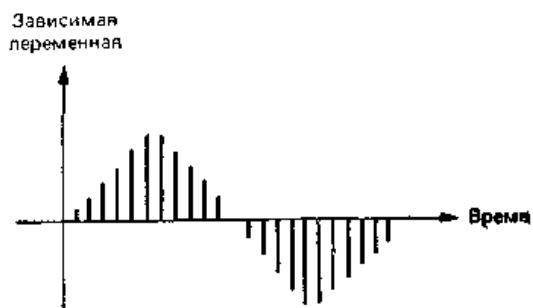


Рис. 3. 4. Графическое представление отклика непрерывного имитатора с дискретным временем.

Моделирование концентрации реагента в химическом процессе или положения и скорости движения космического корабля является примером ситуаций, когда целесообразно использовать непрерывное представление. Кроме того, в некоторых случаях полезно моделировать дискретную систему с помощью непрерывного представления, рассматривая элементы данной системы не как отдельные, а как агрегированные. Например, предпочтительнее, вероятно, популяцию определенных видов рыб в озере

моделировать с помощью непрерывного представления, хотя в реальности изменение популяции происходит дискретно.

При *комбинированной имитации* зависимые переменные модели могут изменяться дискретно, непрерывно или непрерывно с наложенными дискретными скачками. Время изменяется либо дискретно, либо непрерывно. Наиболее важный аспект комбинированной имитации заключается в возможности взаимодействия между дискретно и непрерывно изменяющимися переменными. Например, когда уровень концентрации реагента в химическом процессе достигнет предписанного уровня, процесс имитации может быть прерван. Язык комбинированной имитации должен содержать средства для определения условий возникновения таких ситуаций и моделирования их последствий. Пример изменения зависимой переменной в комбинированной имитационной модели приведен на рис. 3.5.

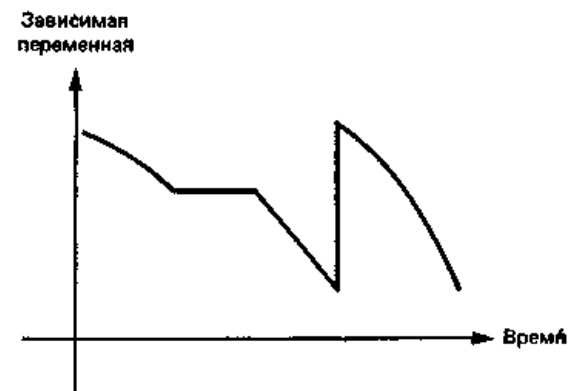


Рис. 3.5. Графическое представление отклика комбинированного имитатора.

2.3.3. Дискретное имитационное моделирование

Элементы дискретных систем, такие, как люди, оборудование, заказы, потоки материалов и т. п., включенные в имитационную модель, будем называть ее компонентами. Существует много типов компонентов, каждый из которых описывается различными характеристиками (атрибутами). Компоненты, участвующие в действиях различного типа, могут иметь одну или несколько общих характеристик, что позволяет объединять их в группы. Группы компонентов называются файлами (термин «файл» используется нами

для обозначения множества компонентов или событий.) Включение компонента в файл означает, что он логически связан с другими компонентами этого файла.

Целью дискретного имитационного моделирования является воспроизведение взаимодействий, в которых участвуют компоненты, и изучение поведения и функциональных возможностей исследуемой системы. Для этого выделяются состояния системы и описываются действия, которые переводят ее из одного состояния в другое. Говорят, что система находится в определенном состоянии, когда все ее компоненты находятся в состояниях, совместимых с областью значений, описывающих это состояние характеристик. Таким образом, **имитация** — это динамический «портрет» состояний системы во времени, т. е. воспроизведение поведения системы во времени.

При дискретной имитации состояние системы может меняться только в моменты свершения событий. Так как состояние системы не изменяется между этими моментами, полный динамический портрет состояний системы может быть получен путем продвижения имитационного времени от одного события к другому. В большинстве языков дискретной имитации используется механизм продвижения времени, основанный на поиске следующего ближайшего события.

Функционирование дискретной имитационной модели можно задать следующим образом: определяя изменения состояния системы, происходящие в момент свершения событий; описывая действия, в которых принимают участие элементы системы, или процесс, через который проходят элементы. Взаимосвязь между понятиями *событие*, *действие* и *процесс* представлена на рис. 3.6.

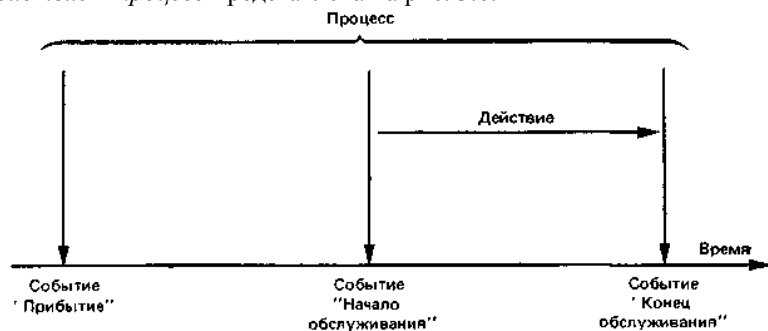


Рис. 3.6. Взаимосвязь между событиями, действиями и процессами.

Событие происходит в тот момент, когда принимается решение о начале или окончании действия. Процесс — это ориентированная во

времени последовательность событий, которая может состоять из нескольких действий. Эти представления лежат в основе **трех альтернативных методологических подходов** к построению дискретных имитационных моделей, называемых обычно *событийным подходом*, *подходом сканирования активностей* (термин «сканирование активностей» является устоявшимся в литературе по имитационному моделированию для обозначения подхода, ориентированного на действия) и *процессно-ориентированным подходом*. Указанные подходы описаны в последующих разделах.

2.3.3.1. Событийный подход

При событийном подходе система моделируется путем идентификации изменений, происходящих в ней в моменты свершения событий. Задача исследователя заключается в описании событий, которые могут изменить состояние системы, и определении логических взаимосвязей между ними. Имитация функционирования системы осуществляется путем выполнения упорядоченной во времени последовательности логически взаимосвязанных событий.

Для иллюстрации событийного подхода вновь рассмотрим гоример банка с одним кассиром, приведенный в раз. 2.1. Клиенты заходят в банк, после возможного ожидания обслуживаются кассиром и затем уходят. Состояние системы в этом примере определяется состоянием кассира и числом ожидающих обслуживания клиентов. Оно остается неизменным, за исключением моментов, когда клиент прибывает в систему или покидает ее. Поэтому событийная модель в данном случае состоит из описания действий, происходящих в момент прибытия и окончания обслуживания очередного клиента. Так как изменения состояния системы могут происходить только в эти моменты времени, использование событий «прибытие» и «конец обслуживания» полностью обеспечивает воспроизведение динамики функционирования системы.

Обсудим сначала логику события «прибытие». Операторная схема этого события имеет следующий вид:

```

ПЛАНИРОВАНИЕ СЛЕДУЮЩЕГО ПРИБЫТИЯ.
ЕСЛИ КАССИР ЗАНЯТ: ЧИСЛО ОЖИДАЮЩИХ=ЧИСЛО
ОЖИДАЮЩИХ+1; ВОЗВРАТ.
ЕСЛИ КАССИР СВОБОДЕН: ПЕРЕВОД КАССИРА В СОСТОЯНИЕ
«ЗАНЯТ»; ПЛАНИРОВАНИЕ СОБЫТИЯ «ОКОНЧАНИЕ»
ОБСЛУЖИВАНИЯ В МОМЕНТ ВРЕМЕНИ=ТЕКУЩЕЕ
ВРЕМЯ+ВРЕМЯ ОБСЛУЖИВАНИЯ; ВОЗВРАТ.
КОНЕЦ
    
```

На первом шаге планируется прибытие следующего клиента, что в ходе имитации при неоднократном обращении к этой процедуре позволяет организовать непрерывный поток прибытий. Поведение прибывшего клиента зависит от состояния системы в текущий момент времени. Если кассир занят, прибывший клиент становится в очередь, и изменение состояния системы осуществляется путем увеличения числа ожидающих клиентов на единицу. Если кассир свободен, прибывший клиент сразу же обслуживается, и изменение состояния системы осуществляется путем перехода кассира в состояние «занят». Кроме того, должно быть запланировано событие «конец обслуживания» для данного клиента в момент времени, равный текущему времени плюс время, затрачиваемое кассиром на обслуживание.

Рассмотрим теперь логику обработки события «конец обслуживания». Операторная схема этого события имеет следующий вид:

ЕСЛИ ЧИСЛО ОЖИДАЮЩИХ БОЛЬШЕ НУЛЯ: ЧИСЛО ОЖИДАЮЩИХ=ЧИСЛО ОЖИДАЮЩИХ—1; ПЛАНИРОВАНИЕ ОКОНЧАНИЯ ОБСЛУЖИВАНИЯ В МОМЕНТ ВРЕМЕНИ, РАВНЫЙ ТЕКУЩЕМУ ВРЕМЕНИ + ВРЕМЯ ОБСЛУЖИВАНИЯ; ВОЗВРАТ.

ЕСЛИ ЧИСЛО ОЖИДАЮЩИХ РАВНО НУЛЮ: ПЕРЕВОД КАССИРА В СОСТОЯНИЕ «СВОБОДЕН»; ВОЗВРАТ.
КОНЕЦ

Когда кассир заканчивает обслуживание очередного клиента, сначала проверяем, есть ли клиенты, ожидающие обслуживания. Если такие клиенты есть, их число уменьшается на единицу и планируется событие «конец обслуживания» первого из ожидающих клиентов. В противном случае кассир переходит в состояние «свободен».

При имитации функционирования банка с одним кассиром на основе использования событийного подхода мы должны воспроизвести хронологию (календарь) событий и причины, вызывающие их появление в соответствующие моменты имитационного времени. Календарь событий первоначально содержит отметку только о первом событии «прибытие». В ходе имитации возникновение других событий «прибытие» и «конец обслуживания» должно быть запланировано в календаре в соответствии с логикой функционирования системы. События выполняются в упорядоченной по времени последовательности, при этом имитационное время продвигается от одного события к другому.

Если для построения событийно дискретной модели разработчик использует какой-либо универсальный язык, например Фортран, то программирование календаря событий и механизма продвижения

имитационного времени, обеспечивающих обработку событий в хронологической последовательности, требуют значительных затрат. Так как функция продвижения времени необходима для реализации любой дискретно-событийной модели, разработан ряд имитационных языков, обеспечивающих разработчика средствами для выполнения этой и других обычно встречающихся функций.

2.3.3.2. Подход сканирования активностей

При использовании подхода сканирования активностей разработчик описывает действия, в которых принимают участие элементы системы, и задает условия, определяющие начало и окончание этих действий. События, которые начинают или завершают действие, не планируются разработчиком модели, а инициируются по условиям, определенным для данного действия. Условия начала или окончания действия проверяются после очередного продвижения имитационного времени. Если заданные условия удовлетворяются, происходит соответствующее действие. Для того чтобы было выполнено каждое действие в модели, сканирование условий производится для всего множества действий при каждом продвижении имитационного времени.

Подход сканирования активностей обеспечивает простую схему моделирования для решения целого ряда проблем. Он наиболее эффективен для ситуаций, в которых продолжительность действия определяется в зависимости от того, насколько состояние системы удовлетворяет заданным условиям. Тем не менее, так как необходимо сканировать условия для каждого действия, подход сканирования активностей менее эффективен по сравнению с событийным подходом и поэтому имеет ограниченное применение в дискретной имитации. Однако ряд имитационных языков имеет специфические черты, основанные на концепции сканирования.

2.3.3.3. Процессно-ориентированный подход

Многие имитационные модели содержат последовательности компонентов, которые возникают в них по определенной схеме, например, очередь, в которой клиенты ожидают обслуживания. Логика возникновения компонентов по требуемой схеме может быть обобщена и задана в одном операторе. Имитационный язык затем транслирует такие операторы в соответствующую последовательность событий, происходящих с компонентами модели. Имитационные языки, включающие операторы для моделирования процесса прохождения элементов через систему, обычно называются процессно-

ориентированными. Эти операторы определяют последовательность событий, которые автоматически выполняются имитационным языком, по мере того как элементы продвигаются через систему. Например, следующий набор операторов может быть использован для описания процесса в модели банка:

СОЗДАВАТЬ ПРИБЫВАЮЩИХ КЛИЕНТОВ ЧЕРЕЗ
КАЖДЫЕ T ЕДИНИЦ ВРЕМЕНИ;
ОЖИДАТЬ КАССИРА;
ПРОДВИНУТЬ ВРЕМЯ НА ПРОДОЛЖИТЕЛЬНОСТЬ
ОБСЛУЖИВАНИЯ;
ОСВОБОДИТЬ КАССИРА;
УДАЛИТЬ КЛИЕНТА;

Первый оператор генерирует прибывающих в систему клиентов через каждые T единиц имитационного времени. Величина T может быть константой или принимать случайные значения. Оператор «ожидать» определяет, что клиент ожидает до тех пор, пока кассир освободится. Этот тип оператора аналогичен понятию действия по условию, применяемого в подходе сканирования активностей. Оператор «продвинуть время» моделирует тот период времени, в течение которого клиент обслуживался кассиром. Этот тип оператора аналогичен оператору планирования в событийном подходе. Он помещает в календарь событий метку о том, что обслуживание клиента будет закончено в момент, равный текущему имитационному времени плюс время обслуживания. После завершения обслуживания клиент покидает систему и кассир освобождается. Освобождение кассира позволяет сразу же приступить к обработке какого-либо ожидающего клиента из оператора «ожидать».

Из приведенного примера видно, что процессно-ориентированный подход сочетает в себе черты событийного подхода и подхода сканирования активностей. Он обеспечивает описание прохождения компонентов через процесс, содержащий ресурсы. Простота этого подхода состоит в том, что определяемая операторами логика событий заложена в самом имитационном языке. Однако, так как мы обычно вынуждены ограничивать набор стандартных операторов языка, этот подход является менее гибким, чем событийный. Кроме того, требуется постоянный анализ состояния ресурсов после их использования.

Наиболее распространенными процессно-ориентированными языками являются GPSS, SIMULA и Q-GERT. Эти языки имеют различия в типах и синтаксисе операторов и методах их объединения. Язык GPSS основан на методе блочных диаграмм, в котором блок соответствует

определенному типу операторов. Блоки соединяются друг с другом, образуя блок-схемы. В отличие от GPSS множество операторов языка SIMULA является подмножеством общецелевого языка Алгол. В языке Q-GERT применяется графовое представление моделируемой системы.

2.3.4. Непрерывное имитационное моделирование

В непрерывной имитационной модели состояние системы представляется с помощью непрерывно изменяющихся зависимых переменных. Для того чтобы отличать непрерывно изменяющиеся переменные от дискретно изменяющихся, будем первые называть переменными состояния. Непрерывная имитационная модель создается путем задания уравнений для совокупности переменных состояния, динамическое поведение которых имитирует реальную систему.

Модели непрерывных систем часто определяются в терминах производных переменных состояния. Это объясняется тем, что иногда легче задать выражение для определения скорости изменения переменной состояния, чем сделать это непосредственно для самой переменной. Уравнения такого вида, включающие производные переменных состояния, называются дифференциальными уравнениями. Пусть, например, в процессе разработки модели мы составили следующее дифференциальное уравнение для переменной состояния s по времени t:

$$\frac{ds(t)}{dt} = s^2(t) + t^2, \quad s(0) = k.$$

Первое уравнение определяет скорость изменения s как функцию от s и t, второе уравнение — начальное условие для переменной состояния.

Цель имитационного эксперимента — определить реакцию переменной состояния в зависимости от имитационного времени.

В некоторых случаях возможно определение аналитического выражения для переменной состояния s, заданного уравнением для ds/dt. Однако на практике в большинстве случаев аналитическое выражение для s не известно. В результате мы должны получить реакцию путем интегрирования ds/dt по времени, используя уравнение следующего вида:

$$s(t_2) = s(t_1) + \int_{t_1}^{t_2} \left(\frac{ds}{dt} \right) dt.$$

Каким образом выполняется интегрирование, зависит от того, использует ли разработчик аналоговый или цифровой компьютер.

В 50-х и 60-х годах аналоговые компьютеры были основным средством реализации непрерывных моделей. Аналоговые компьютеры представляют переменные состояния в модели с помощью электрических цепей. Динамическая структура системы моделируется с помощью таких элементов, как резисторы, конденсаторы и усилители. Основным недостатком аналоговых компьютеров состоит в том, что от характеристик этих элементов зависит точность результатов. Кроме того, в аналоговом компьютере мало логических контрольных функций и отсутствуют те возможности хранения данных, которые имеются в цифровом компьютере.

Ряд непрерывных имитационных языков был разработан для цифровых компьютеров. Несмотря на то что цифровой компьютер является дискретным устройством, практически любая переменная, значение которой ограничивается только размером слова компьютера, может рассматриваться как непрерывная.

Цифровой компьютер с большой скоростью и точностью выполняет основные математические операции, такие, как сложение, умножение и логическое тестирование. Выполнение же интегрирования требует применения числовых методов интегрирования. При использовании этих методов независимая переменная (обычно время) разделяется на части, называемые *шагами*. Значения переменных состояния, требующие интегрирования, получаются путем аппроксимации производных этих переменных по времени. Точность получаемых значений зависит от порядка аппроксимационного метода и размера шага: более высокую точность дают аппроксимации высокого порядка и наименьшие размеры шагов. Так как аппроксимации высокого порядка и небольшие размеры шага требуют больше вычислений, то существует зависимость между точностью вычислений переменной состояния и затрачиваемым при этом машинным временем. Описание различных алгоритмов числового интегрирования можно найти во многих введениях в числовой анализ.

Иногда для моделирования непрерывных систем используются разностные уравнения. В этом случае временная ось разбивается на временные периоды длиной Δt . Динамика переменной состояния описывается уравнением, которое вычисляет значение переменной в период $k+1$ на основе ее значения в период k . Например, динамику переменной состояния s можно описать следующим разностным уравнением:

$$s_{k+1} = s_k + a * \Delta t.$$

Непрерывные имитационные языки для цифровых компьютеров обычно имеют либо блочную, либо операторную ориентацию. Языки с блочной ориентацией используют набор блоков, которые

функционально соответствуют элементам цепей в аналоговых компьютерах. Таким образом, знакомство разработчика с аналоговыми блочными диаграммами облегчает изучение этих языков. В большинстве непрерывных имитационных языков используется операторная ориентация, когда дифференциальные или разностные уравнения кодируются в явном виде. Преимущество ориентации на уравнения состоит в увеличении гибкости, которую обеспечивают алгебраические и логические черты этих языков. Разработан набор стандартов для непрерывных имитационных языков.

2.3.5. Комбинированные дискретно-непрерывные модели

В комбинированных дискретно-непрерывных моделях независимые переменные могут изменяться как дискретно, так и непрерывно. В рамках методологии комбинированного моделирования исследуемая система описывается с помощью элементов, их атрибутов и переменных состояния. Поведение системы имитируется путем вычисления значений переменных состояния через небольшие отрезки времени и значений атрибутов элементов в моменты свершения событий.

В комбинированном моделировании применяются два типа событий: *временные события* (события, свершение которых планируется в определенные моменты времени) и *события состояния* (эти события не планируются, а происходят тогда, когда система достигает определенного состояния). Например, как показано на рис. 3.7, событие состояния произойдет, когда переменная состояния $SS(1)$ пересечет переменную состояния $SS(2)$ в положительном направлении.

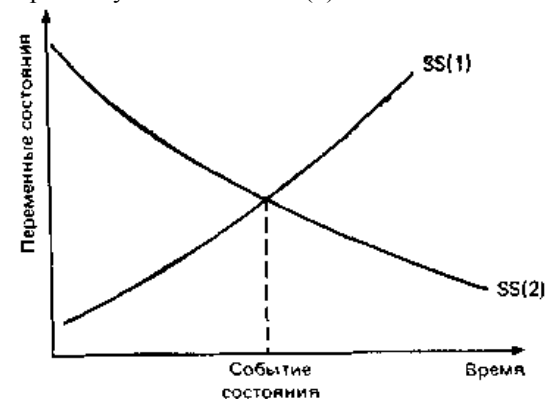


Рис. 3.7. Пример наступления события состояния.

Отметим, что понятие «событие состояния» аналогично понятию «сканирование активностей», в котором событие также не планируется, а инициируется определенным состоянием системы. Возможность возникновения события состояния должна проверяться при каждом продвижении имитационного времени.

Первым полностью документированным комбинированным языком был GASP IV. Имитационный язык GASP IV реализован на универсальном языке Фортран и базируется на методологии, в которой событийный подход для моделирования дискретных систем комбинируется с подходом, ориентированным на уравнения переменных состояния для моделирования непрерывных систем. Кроме того, существует версия GASP IV на языке ПЛ/1. В язык SIMSCRIPT также были добавлены возможности комбинированного моделирования, основанные на чертах GASP IV. Возможности сетевого моделирования включены в комбинированные языки SMOOTH и SAINT. На основе GASP IV создан также язык CROPS, ориентированный на решение сельскохозяйственных проблем. Применение комбинированных имитационных моделей для анализа систем продолжает оставаться перспективной областью для исследований, разработок и практических приложений.

2.3.6. СЛАМ: объединенная схема моделирования

В предыдущих разделах мы описали альтернативные подходы к имитационному моделированию. Каждый подход предлагает некоторую схему для описания системы, и каждый из них обладает определенными достоинствами и недостатками. Например, в дискретном имитационном моделировании процессно-ориентированный подход краток и прост в изучении схемы моделирования, но не обладает достаточной гибкостью; событийный подход, наоборот, более сложен в изучении, но обеспечивает гибкую схему моделирования.

В языке СЛАМ альтернативные подходы комбинируются и обеспечивают объединенную схему моделирования. Дискретно-изменяющаяся система моделируется на основе событийного или процессно-ориентированного подхода или их обоих. Непрерывно изменяющаяся система моделируется на основе дифференциальных или разностных уравнений. Комбинированные дискретно-непрерывные системы моделируются путем сочетания событийного и (или) процессно-ориентированного подходов с непрерывным

подходом. Язык СЛАМ содержит также ряд возможностей, которые соответствуют подходу сканирования активностей

Процессно-ориентированный подход в СЛАМ использует сетевую структуру, которая состоит из специальных символов, называемых узлами и ветвями. Эти символы отображают элементы процесса: очереди, места обслуживания и точки решений. **Задача моделирования состоит в том, чтобы построить из этих символов сетевую модель, которая графически представляет интересующую нас систему.** Другими словами, сеть — это графическое представление процесса. Элементы системы (такие, как люди и отдельные предметы) «протекают» через «сетевую модель». Разработчик переводит графическое представление системы в эквивалентную операторную модель, предназначенную для ввода в процессор СЛАМ.

При событийном подходе разработчик определяет события и те потенциальные изменения, которые происходят в системе в результате свершения этих событий. Логико-математические взаимосвязи, которые описывают изменения, связанные с каждым типом событий, представляются разработчиком в виде подпрограммы на языке Фортран. В языке СЛАМ имеется также набор стандартных подпрограмм, которые реализуют такие общие функции дискретно-событийной имитации, как планирование событий, управление файлами, сбор статистических данных и генерация случайных чисел. Управляющая программа языка СЛАМ путем продвижения имитационного времени управляет ходом имитационного процесса и инициирует вызов соответствующих подпрограмм событий в нужные моменты времени. Следовательно, пользователь полностью освобожден от необходимости установления хронологической последовательности возникновения событий.

При разработке непрерывной модели на языке СЛАМ динамическое поведение переменных состояния задается дифференциальными и разностными уравнениями, которые программируются разработчиком на языке Фортран и используют специально определенные в СЛАМ массивы данных. Значение I -й переменной состояния хранится как переменная $SS(I)$, и производная этой переменной, если это требуется, хранится как переменная $DD(I)$. Ближайшие предшествующие значения I -й переменной состояния и ее производной хранятся как переменные $SSL(I)$ и $DDL(I)$. Если в непрерывную модель включены дифференциальные уравнения, они автоматически интегрируются в СЛАМ для вычисления значений переменных состояния с определенной пользователем точностью

Важным достоинством языка СЛАМ является возможность интеграции альтернативных подходов при разработке одной и той же

имитационной модели. Существует шесть специфичных взаимодействий, которые могут иметь место между сетевым, дискретно-событийным и непрерывным подходами в СЛАМ:

1. Компоненты сетевой модели могут инициировать возникновение дискретного события.
2. События могут изменить характеристики потока компонентов сетевой модели.
3. Компоненты сетевой модели могут вызвать мгновенные изменения значений переменных состояния.
4. Переменные состояния, достигая заданных пороговых значений, могут инициировать компоненты в сетевых моделях.
5. События могут вызвать мгновенные изменения значений-переменных состояния.
6. Переменные состояния, достигая заданных пороговых значений, могут инициировать события.

Возможность создавать комбинированные сетевые событийно-непрерывные модели с взаимодействиями между подходами значительно расширяет область применения языка СЛАМ

2.4. Формализация описаний процесса функционирования сложной технической системы

2.4.1. О понятии имитационного моделирования

Как мы уже отмечали, имитация предназначена для построения некоторого идеализированного процесса функционирования системы, называемого *имитационным процессом*. Процесс функционирования реальной системы распадается на ряд процессов функционирования отдельных объектов. Эти процессы протекают одновременно, или параллельно. Задача программной имитации состоит в отображении параллельно протекающих процессов на один вычислительный процесс. Это отображение может быть выполнено различными способами в зависимости от задач и показателей эффективности.

Множество моментов времени изменения состояния конечно и может быть описано выражением

$$T = \langle t_1, t_2, \dots, t_i, \dots, t_n \rangle,$$

где t_i — i -й момент времени изменения состояния системы; t_n — общее время наблюдения за системой.

Каждому моменту времени t_i соответствует оператор H_i , вычисляющий состояние s_i где $s_i \in S$. Оператор H_i , осуществляет отображение времени t_i на один из элементов множества пространства S .

Предположим, что выполнение оператора H_i происходит мгновенно в момент времени t_i .

В общем случае оператор H можно представить в следующем виде:

$$s = H(A, t, \omega),$$

где $s \in S$ — состояние процесса Z ; A — пространство аргументов; t — текущее значение времени; ω — случайная величина, распределенная равномерно на $[0, 1]$, принимающая новое значение при каждом выполнении оператора H . Использование величины ω позволяет описать случайность в операторе.

Таким образом, для момента времени t_i

$$s_i = H_i(A_i, t_i, \omega).$$

Если для двух произвольных моментов времени t_i и t_j оказывается $s_i \in A_j$, то будем называть операторы H_i и H_j сцепленными и обозначать $H_i \infty H_j$. Практически это означает, что для вычисления состояния s_j в момент времени t_j необходимо знать состояние s_i в момент времени t_i .

Описание оператора, вычисляющего новое значение состояния системы в какой-то момент времени, является достаточно сложной задачей, тем более, что пространство состояний может иметь высокую размерность.

Предположим, что для текущего момента времени t можно задать последовательность таких подоператоров $\langle h^l / l = 1, \bar{m}_t \rangle$, что

$$s^l = h^l(A^l, t, \omega); h^k \infty h^l \text{ для всех } l > k.$$

После выполнения последовательности подоператоров $\langle h^l \rangle$ система примет состояние s , как после выполнения оператора H . Таким образом, оператор H можно рассматривать как последовательную композицию подоператоров $\langle h^l \rangle$.

Задание оператора H через последовательность сцепленных подоператоров h^l назовем *расщепленным описанием*. Расщепление оператора H не единственное и определяется структурой моделируемой системы, степенью подробности описания процесса функционирования, задачами исследования, опытом разработчика модели. Практически расщепление получают путем выделения базовой совокупности подоператоров. Построение различных композиций из этой совокупности покрывает все многообразие операторов H .

2.4.2. Алгоритмическая модель процесса

Под *функционированием системы* понимается процесс изменения ее состояния во времени. Этот процесс назовем *процессом функционирования*

$$Z = \langle T, S, F, \alpha \rangle,$$

где S — пространство состояний; F — траектория процесса ($F: T \rightarrow S$); α — отношение линейного порядка на множестве T .

В зависимости от типа процесса множество моментов времени T может быть как непрерывным, так и дискретным. Если T задано в виде упорядоченных чисел, то величину α из данного выражения можно исключить.

Пространство состояний S определяется параметрами P системы. Пусть W — множество параметров P системы. Под *параметром системы* будем понимать некоторую скалярную переменную, обладающую именем и множеством значений:

$$P = \langle W_p, \sigma(W_p) \rangle,$$

где W_p — имя параметра; $\sigma(W_p)$ — множество значений параметра. (Элементы множества $\sigma(W_p)$ скалярны.)

При этих условиях пространство состояний

$$S = \Pi \sigma(\omega) \text{ для всех } \omega \in W,$$

где Π — символ декартова произведения множеств.

Таким образом, **размерность пространства S равна числу параметров системы.**

Над процессами функционирования осуществляют операции проекции и суммирования.

Проекцию процесса Z на пространство S обозначим $\Pi_{p_s} Z$ и определим как проекцию траектории F на подпространство S' . Если $S' = \Pi \sigma(W')$, где $W' \subset W$, то эта операция описывает процесс в подсистеме W' . Проекция процесса Z на отрезок времени $\{t_1, t_2\} \subset T$ описывает подпроцесс

$$Z_1 = \Pi_{p_{\{t_1, t_2\}}} Z.$$

Операция, обратная операции проекции, называется суммированием:

$$Z = Z_1 \uparrow Z_2.$$

Так как задание процесса функционирования Z сложной системы или ее подсистемы в виде некоторого аналитического соотношения в общем случае невозможно, то используют более сложные модели. Но как показал анализ, такие модели не удовлетворяют требованиям универсальности и не позволяют описывать функционирование вычислительных систем и сетей с единых методологических позиций.

Выполним такое описание с помощью алгоритмической модели процесса (АМП).

Виды операторов. Рассмотрим дискретный процесс функционирования Z , множества моментов времени T которого счетны. Каждому моменту времени $t_i \in T$ поставим в соответствие элементарный оператор h_i , вычисляющий значение S_i в пространстве состояний S ($S_i \in S$) в соответствии с траекторией процесса F . Таким образом, для упорядочений последовательности времени $\langle t_i \rangle$ будет сформулирована во взаимно-однозначном соответствии последовательность элементарных операторов $\langle h_i \rangle$. Очевидно, что $\langle h_i \rangle$ представляет собой линейный граф, соответствующий траектории процесса F . Для полного описания процесса необходимо иметь средство, позволяющее строго регламентировать моменты времени t_i выполнения операторов h_i . В качестве такого средства в АМП введем специальный объект, называемый *инициатором*.

Постулируем следующее **утверждение**: если происходит сцепление инициатора с элементарным оператором h_i , то последний выполняется мгновенно и вычисляет соответствующее состояние системы.

Алгоритмическая модель процесса (АМП) Z имеет вид

$$\langle \{h_i\}, \beta, I \rangle,$$

где $\{h_i\}$ — совокупность элементарных операторов; β — линейный порядок на $\{h_i\}$; I — инициатор.

Каждому процессу соответствует один и только один инициатор. После сцепления с очередным оператором инициатор либо остается в этом операторе, либо находится вне АМП. В соответствующий момент времени инициатор должен сцепиться со следующим оператором. Задание момента времени сцепления отнесем к функциям текущего оператора. Тогда структура элементарного оператора будет иметь вид

$$h_i = \langle h_i^c, h_i^u \rangle,$$

где h_i^c, h_i^u — i -е элементарные операторы состояния и условия.

Оператор h_i^c выполняет функции, описанные ранее для оператора h_i , а оператор h_i^u формулирует условия, при выполнении которых происходит сцепление инициатора со следующим оператором h_{i+1} .

Будем считать, что $h_i^u \in \{h_i^c, h_i^c\}$, где оператор h_i^c задает момент времени сцепления инициатора с оператором h_{i+1} , оператор h_i^u — логическое условие. Таким образом, АМП полностью определяет процесс, поскольку содержит в явном или опосредственном виде все компоненты. Совокупность $\langle \{h_i\}, \beta \rangle$ представляет собой трек инициатора (рис. 1).

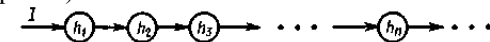


Рис. 1. Схема трека инициатора I

В результате анализа операторов h_i , выделим некоторый базовый набор операторов $\{h^{\delta}\}$, полагая, что для любого h_i можно найти соответствующий оператор h^{δ} , способный вычислить то же состояние. Тогда трек $\langle \{h_i\}, \beta \rangle$ можно свернуть путем объединения вершин в некоторый граф, соответствующий одному и тому же базовому оператору (рис. 2).

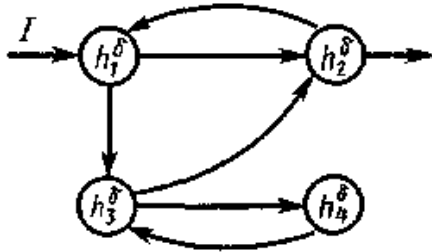


Рис. 2. Структурная схема операторов относительно инициатора I

Назовем полученный граф *структурой операторов относительно инициатора*. Однако построение трека при заданной структуре может быть выполнено неоднозначно. Для решения этой задачи введем особый тип оператора — *навигационный*, который выполняется мгновенно при сцеплении с инициатором, результатом является номер связи 1, ..., n , в соответствии с которым инициатор продолжает путь (рис. 3).

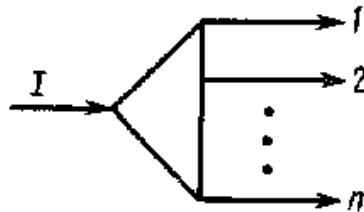


Рис. 3. Схема навигационного оператора

Навигационный оператор вычисляет лишь состояния, условная часть h^y в нем отсутствует.

Можно выделить следующие типовые структуры операторов: линейную, общего вида, полнодоступную.

Линейная структура совпадает с треком (см. рис. 1) и не требует введения навигационных операторов. **Структура общего вида** (рис. 4, а) представляет собой граф с обязательным включением навигационных операторов. Сильно связанный граф удобно представлять в виде **полнодоступной структуры**, получаемой в результате свертки всех навигационных операторов в один общий (рис. 4, б).

гационных операторов. Сильно связанный граф удобно представлять в виде **полнодоступной структуры**, получаемой в результате свертки всех навигационных операторов в один общий (рис. 4, б).

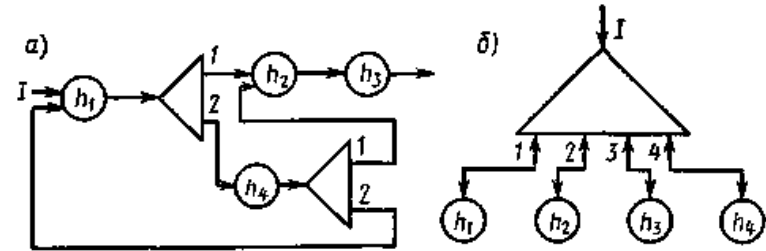


Рис. 4. Линейная структура общего вида (а) и полнодоступная (б)

Структура операторов описывает процесс либо подпроцесс и может быть представлена одним оператором общего вида H . Таким образом, возникает возможность многоуровневого описания процесса, т. е. к представлению процесса в виде суммы подпроцессов. Каждый из подпроцессов одного уровня, в свою очередь, можно представить суммой подпроцессов другого уровня. При этом необходимо помнить, что подпроцесс возникает в результате сцепления инициатора с оператором общего вида. Дробление подпроцессов прекращается, когда достигается уровень описания элементарными операторами.

Варианты сцеплений. Рассмотрим отношение доступа между основными элементами описания процесса: параметрами, операторами и инициаторами. Параметры P_1 и P_2 могут быть сцеплены между собой, что обеспечивает доступ от параметра P_1 к параметру P_2 (рис. 5, а).

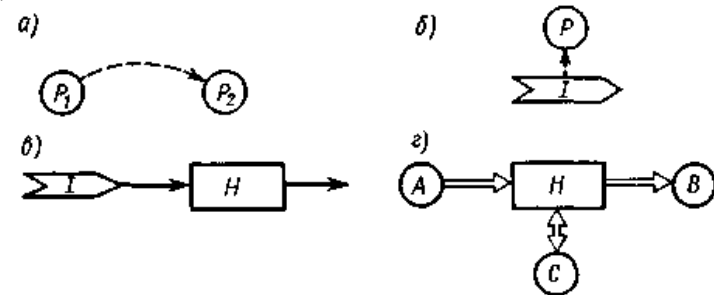


Рис. 5. Варианты сцепления:

а — параметров P_1 и P_2 , б — параметра P с инициатором I , в — инициатора I с параметром H , з — оператора H с параметрами A, B, C

Аналогично возможно сцепление параметра с инициатором (рис. 5, б). В этом случае доступ к инициатору обеспечивает и доступ к параметру. Сцепление инициатора с параметром (рис. 5, в) вызывает выполнение оператора и обеспечивает доступ оператора к инициатору (сцепление показано одинарной стрелкой). Сцепление оператора с параметрами *A, B, C* показано на рис. 5, г двойной стрелкой. Здесь параметр *A* является входным, *B* — выходным, *C* — рабочим.

Схема описания процесса, включающая сцепления между перечисленными выше элементами, называется *операторно-параметрической* (рис. 6).

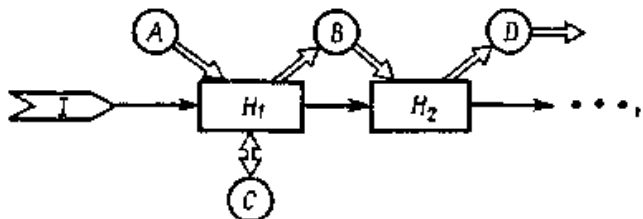


Рис. 6. Фрагмент операторно-параметрической схемы

В случае отсутствия тех или иных элементов в описании процесса возможны варианты схемы — операторная и параметрическая.

Блоки. Совокупность оператора со сцепленными параметрами представляет собой блок, который может включать как элементарный, так и укрепленный операторы. При наличии укрепленного оператора блок описывает подпроцесс или процесс. Блоки между собой взаимодействуют через общие параметры и инициаторы. Если задано разбиение операторов на блоки, то из операторно-параметрической схемы получим блочную схему описания функционирования. На каждом уровне описания может быть построена своя блочная схема.

Будем различать следующие типы блоков: А, П, К, Ф.

А - б л о к (агрегат) описывает замкнутый процесс, инициатор которого не покидает блок. С внешней средой он взаимодействует только через параметры. С помощью А-блоков описываются устройства, машины, агрегаты, функционирующие в соответствии с заданной внутренней логикой и «перерабатывающие» параметры.

П - б л о к (процессор) описывает подпроцесс для инициатора, оказавшегося с ним в сцеплении. После окончания сцепления инициатор покидает П-блок, продолжая путь в соответствии с блочной схемой. Если П-блок допускает наличие лишь одного инициатора, то

он называется *монопроцессором*. Если несколько П-блоков описывают идентичные подпроцессы и содержат однородные операторы, то можно построить один П-блок, включающий объединенные операторы. Пример построения объединенного оператора представлен на рис. 7.

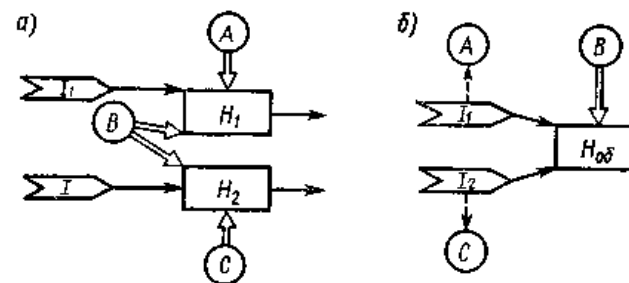


Рис. 7 Параметр построения обобщенного оператора
а — исходное описание, б — обобщенное описание

Параметры *A* и *C* доступны этому оператору лишь через соответствующие инициаторы. Каждый из них входит в локальную среду процесса, соответствующего своему инициатору. Таким образом, появление локальных сред является следствием введения объединенных операторов. П-блок, составленный из обобщенных операторов, называется *полипроцессором* и допускает описание одновременно развивающихся подпроцессов из состава различных процессов. П-блок широко используется при построении имитационных моделей, позволяя достаточно экономно описывать совокупность связанных параллельных процессов.

К - б л о к (контроллер) является разновидностью А-блока и предназначен для обработки параметров, которые либо были, либо станут инициаторами. Таким образом, К-блоки управляют протеканием процессов либо осуществляют их взаимодействие. Наиболее часто К-блоки используют для управления доступом процессов к ресурсам. Блоки типа «генератор обобщений» и «уничтожитель сообщений» являются разновидностью К-блоков.

Ф-б л о к и служат лишь для выполнения вычисления состояний. Инициатор сцепляется с Ф-блоками мгновенно. Эти блоки могут рассматриваться как процедуры, выполняющие типовые преобразования параметров

Конфликтные ситуации. Практический интерес представляет разработка способов разрешения конфликтных ситуаций,

возникающих при занятии разными процессами общего ресурса. Ресурс определяется совокупностью параметров.

Первый способ. Если каждый из процессов задан с помощью П-блока, то регулирование захвата ресурса может быть выполнено с помощью общей логической переменной для всех конфликтующих процессов. Если эта переменная принимает значение «1», то ресурс может быть захвачен каким-либо процессом, если же — «0», то ресурс не свободен и следует ожидать его освобождения. В каждом П-блоке участок, описывающий изменение состояния ресурса, начинается с проверки логической переменной. Если ее значение «1», то процесс захватывает ресурс, вычисляя значение логической переменной, равное «0». После окончания подпроцесса в ресурсе П-блок вычислит значение переменной, равное «1», и продолжит развитие процесса вне ресурса в рамках собственных локальных параметров.

Второй способ. Способ состоит в выделении ресурса в самостоятельный объект, что сокращает запись П-блоков. Процесс в ресурсе описывается с помощью П-блока. Для регулирования доступа инициаторов в ресурс создается К-блок. Все инициаторы из П-блоков при попытке захвата ресурса поступают в К-блок. Последний по заданному алгоритму определяет инициатор, поступающий в П-блок ресурса. По окончании процесса К-блок возвращает инициатор в соответствующий П-блок и выбирает новый инициатор для захвата ресурса.

• **Примечание.** Конструктивность второго варианта заключается в получении схемы, анализ которой выполняется методами теории массового обслуживания с целью получения временных характеристик. В качестве сообщений выступают инициаторы процессов.

2.4.3. Формы описания процессов

С целью повышения эффективности работы модели следует реализовать моделирующую программу на одном из алгоритмических языков, например ПЛ/1, в виде одной процедуры. При этом секции событий будут являться частями процедуры, а не отдельными процедурами, что позволит значительно сократить время передачи управления при переходе одного события к другому.

Упорядочение событий достигается с помощью списка активных времен, где указывается имя события (номер события), время события и номер транзакта. Список активных времен подвергается сортировке в порядке возрастания значений времени.

Выбор очередного события производится из начала списка. При этом управление передается секции событий, номер которой указан в элементе списка. Данные средства моделирования позволяют выполнять операции планирования и причинного вызова событий. Программа пользователя оформляется в виде процедуры на одном из алгоритмических языков.

Макрокоманды. Раздел MACRO содержит набор макроопераций, представляющих собой препроцессорные процедуры языка ПЛ/1. Обращение к ним осуществляется с помощью макрокоманд.

Макрокоманды для организации подпрограмм:

MCALL (имя подпрограммы) — обращение к подпрограмме;

MRETN — возврат из подпрограммы;

MPROC — сохранение точки возврата.

Здесь под программой подразумевается часть процедуры с определенным именем, начинающаяся командой MPROC и завершающаяся командой MRETN, обращение к которой осуществляется по команде MCALL.

Макрокоманды моделирования:

SCHEDL (соб., время, транз.) — планирование событий,

где «**соб.**» — номер подпрограммы обработки событий; «**время**»

— время совершения события; «**транз.**» — номер транзакта.

По команде SCHEDL в список активных времен помещается новый элемент.

CAUSE — выход очередного события.

По этой команде из списка активных времен удаляется элемент, находящийся в его вершине, а также устанавливаются значения параметров: TIME — время совершения события; EV — номер события; REQ — номер транзакта.

ENLIST (реж. вкл-я, спис, атр, эл-т) — включить в список, где «**реж. вкл-я**» — режимы включения в список (В — по возрастанию; У — по убыванию); «**спис.**» — имя начала списка; «**атр.**» — атрибуты, в соответствии с которыми упорядочивается список; «**эл-т**» — номер включенного элемента.

По данной команде первый элемент из списка свободных элементов включается в список элементов с указанным именем. Указатель свободных элементов передвигается на следующий элемент. Возвращаемым является номер включенного элемента, используя который, можно заменить поля включенного элемента.

ENQUE (спис, эл-т) — удалить из очереди (списка),

где «**спис.**» — имя начала списка; «**эл-т**» — номер удаленного элемента.

По этой команде удаляется элемент, находящийся в начале списка. Указатель начала списка сдвигается на следующий элемент. Удаленный элемент вставляется в начало списка свободных элементов. Возвращаемым является номер удаленного элемента, по которому можно определить значение полей удаленного элемента.

Каждая препроцессорная процедура обращается к одной из подпрограмм, находящихся в разделе SIMULATE:

SCHEDL — к # SCHEDL; CAUSE — к # CAUSE;
ENLIST — к # ENLIST; DELIST — к # DELIST

Макрокоманды операций над очередями.

ENQUE (имя очереди, транз., атрибут) — поставить транзакт в очередь соответственно с атрибутом;

HEAD (имя очереди) — удалить транзакт, находящийся в вершине очереди.

Возвращается параметр HEAD — номер транзакта. В случае отсутствия элементов в очереди HEAD = 0.

QLEN (имя очереди) — вернуть параметр QLEN — число элементов, находящихся в очереди;

QBEG (имя очереди) — вернуть параметр QBEG — номер первого транзакта в очереди.

Для пустой очереди QBEG = 0.

♦ **Примечание.** Очередь в программе описывается оператором DECLARE:

1 (имя очер. 1, имя очер. 2 ..) LITKE QUEUE

Структура подпрограммы обработки событий:

EVENT (номер соб.); MPROC;

MRETN тело подпрограммы

Описание используемых подпрограмм. Основой внутренней организации программы является пул # POOL, содержащий четыре элемента (# L1, # L2, # L3, # L4) длиной 32 байта. Из элементов пула конструируются таблица активных времен и структуры данных пользователя, например структуры, отражающие очереди и средства.

Слово # L4 всегда содержит либо указатель следующего элемента списка или очереди, либо 0, которым помечается конец списка или очереди. Незанятые элементы (нули) связаны друг с другом и образуют список свободных элементов; начало этого списка содержится в указателе #FREE_PTR. Список свободных элементов формируется во время инициализации:

ENLIST — включить элемент в список.

Рабочими переменными являются переменные, описанные ранее в макропроцедуре ENLIST:

REGIM, # BEQ — PTR, # ATR.

Из элементов пула конструируется таблица активных событий:

```
1 FUTURE (500) DEF POOLL
2 # EVTIM BIN FIVED (31)
2 # EVNUM BIN FIXED (31)
2 # TRANO BIN FIXED (31)
2 # NEXT — PTR BIN FIXED (31)
```

где #FUT_PTR — указатель начала таблицы активных событий.

Вставка производится с помощью подпрограммы #ENLIST с параметрами: #REGIM = 'B' — по возрастанию; #FUT_PTR — начало списка; #T — атрибут, #PTR — номер вставленного элемента.

По номеру вставленного элемента заполняются поля элемента — время, номер события, транзакт.

Макропроцедуры организации очереди. Очереди строятся в виде связанных списков, элементами которых являются элементы списка свободных элементов, описанные в виде:

```
#QUE_ELEM (500) DEF # POOL
#ELEMIBIN FIXED (31) — атрибут
#ELEMTR BIN FIXED (31) — номер транзакта
#EMPTY4 BIN FIXED (31) — пока пустой
#NEXT_EL BIN FIXED (31) — ссылка
```

Характеристики очереди описываются структурой:

```
1 QUEUE
2 # LNG BIN FIXED (15) — длина очереди
2 # QUB_PTR FIXED (15) — указание начала очереди
```

Для описания очередей в операторе необходимо использовать описатель LITKE, который указывает, что данная структура (характеристики конкретной очереди) состоит из элементов с такими же именами, что и структура QUEUE.

Используемые макропеременные в макропроцедурах:

```
Q QNAM — имя очереди
Q J — номер транзакта
Q I — атрибут
```

В заключение данного раздела отметим, что имитационные модели воспроизводят процесс функционирования моделируемой сложной системы. Следовательно, прежде чем приступить к созданию программной имитационной модели, необходимо описать процесс функционирования системы, подлежащей моделированию.

Процесс функционирования системы есть последовательность изменения ее состояний, упорядоченных во времени. Он распадается на множество сцепленных между собой процессов функционирования компонент этой системы. Для представления процесса используют алгоритмическую форму, называемую алгоритмической моделью процесса (АМП).

Развитие АМП позволяет ввести понятия элементарного и обобщенного операторов, классифицировать блоки на агрегаты и процессоры, выявить особенности агрегативных и процессных описаний функционирования сложных систем. На основе концепции АМП предлагается псевдоязык описания сцепленных процессов ПОСП, позволяющий на достаточно формализованном уровне производить описание функционирования сложной системы в виде совокупности процессов в различных блоках с указанием их взаимодействия.

2.5. Программные имитационные модели

2.5.1. Анализ одновременных событий

В модели каждый подоператор реализуется некоторым алгоритмом, оформленным в виде подпрограммы. Выполнение такого алгоритма назовем *событием*, а сам алгоритм — *подпрограммой события*. Таким образом, событие есть реализация подоператора. Линейную последовательность событий, соответствующих одному моменту времени, назовем *совокупностью одновременных событий* (СОС). На рис. 1 приведен пример такой последовательности в виде двух сцепленных процессов.

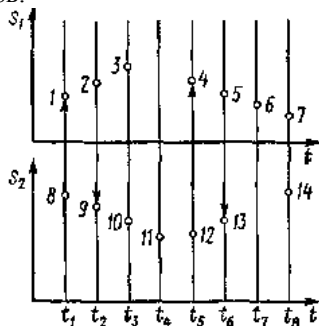


Рис. 1. Пример сцепления процессов:

1...14 – номера точек графика; S_1, S_2 – пространства состояний

Рассмотрим способ генерации СОС. Для этого определим для каждого события оператора h^l условие его реализации u^l . Будем считать, что событие выполняется в тот момент времени, когда условие u^l принимает значение «истина». В качестве аргументов, определяющих значение u^l , могут выступать любые или все параметры системы. Таким образом, выполнение какого-либо события зависит от состояния системы в целом. Так как из выполнения условия u^l однозначно следует выполнение оператора h^l , то сцепленность u^l влечет за собой отношение сцепления операторов h^l .

Утверждение 1. Совокупность одновременных событий ограничена.

Справедливость этого утверждения следует из принципа реализуемости алгоритма. Если оператор H есть композиция подоператоров, то при неограниченной последовательности подоператоров реализация H невозможна. Это означает, что либо расщепление оператора на подоператоры необходимо выполнить иначе, либо такого расщепления не существует.

Утверждение 2. Генерирование совокупности одновременных событий прекращается при выполнении условия $\bigvee u_i = 0$ для всех l .

Действительно, если это условие не выполнено, то хотя бы один его член $u_i = 1$. Но это немедленно вызовет исполнение оператора h_i и, таким образом, последовательность событий будет продолжена.

Утверждение 3. Выполнение первого события из совокупности одновременных событий может инициироваться лишь фактором времени.

Если все события в СОС, соответствующей моменту времени t_i , исчерпаны, то условие $\bigvee u_i = 0$ выполнено. Состояние системы зарегистрировано выполнением последнего подоператора из СОС. Сколько бы времени ни прошло, состояние системы не изменится. Для продолжения процесса дальше необходимо выполнение хотя бы одного подоператора. Так как все условия тождественно равны нулю, то они не могут быть причиной его выполнения. Следовательно, выполнение подоператора не может быть более определено состоянием системы, а должно диктоваться лишь внешними причинами. Такими причинами может быть изменение состояния надсистемы либо наступление какого-то критического момента времени. Время в данном случае является внешним фактором, не связанным с внутренним состоянием системы. В дальнейших рассуждениях исключим из рассмотрения наличие надсистемы, полагая, что рассматриваемая модель системы замкнута. Таким образом, первое

событие в СОС может быть инициировано лишь заданием критического момента времени. Назовем такое событие в СОС *активным*, а остальные, связанные с выполнением условий $u_i = 1$, *пассивными*.

Предположим, что в каждой СОС содержится лишь одно активное событие. Как показано выше, оно является первым. Остальные события — следствия изменения состояния системы, вызваны выполнением активного и всех предыдущих пассивных подоператоров в рамках текущей СОС.

Таким образом, введение расщепления оператора для вычисления текущего состояния в момент времени t позволило выделить структуру событий, сформулировать правила генерации последовательности событий в СОС, включая определение условия начала и конца последовательности.

Рассмотрим пример разомкнутой системы массового обслуживания (СМО) (рис. 2).



Рис. 2. Структурная схема разомкнутой СМО:
Г — генератор, Б — буфер, А — обслуживающий аппарат

Допустим, нам известны закон распределения интервалов времени между появлением сообщений из генератора и закон длительности обслуживания сообщения в аппарате. В какой-либо момент времени может измениться состояние системы в генераторе, очереди и аппарате одновременно. Для выявления более тонкой структуры такого изменения введем следующие подоператоры:

- Г — активный подоператор генерации потока сообщений;
- $\Gamma_{\text{вых}}$ — подоператор реакции на факт перехода сообщения в буфер;
- $\text{Б}_{\text{вх}}$ — подоператор реакции на поступление сообщения из генератора в очередь;
- $\text{Б}_{\text{вых}}$ — подоператор реакции на факт перехода сообщений в аппарат;
- $\text{А}_{\text{вх}}$ — подоператор реакции на поступление сообщения в аппарат;
- $\text{А}_{\text{а}}$ — активный подоператор по освобождению аппарата в связи с истечением интервала времени обслуживания;
- $\text{А}_{\text{вых}}$ — подоператор реакции на факт ухода сообщения из аппарата;
- $\text{У}_{\text{вх}}$ — подоператор уничтожения сообщений, завершивших цикл пребывания в системе.

Очевидно, что активные события могут возникать при использовании подоператоров $\Gamma_{\text{а}}$ и $\text{А}_{\text{а}}$, остальные подоператоры порождают лишь пассивные события. Таким образом, любая СОС

может начинаться лишь с выполнения подоператоров $\Gamma_{\text{а}}$ и $\text{А}_{\text{а}}$. Примеры возможных вариантов СОС, начинающиеся с $\Gamma_{\text{а}}$, для различных моментов времени приведены на рис. 3.



Рис. 3. Пример СОС для подоператоров $\Gamma_{\text{а}}$:
 а — $\text{А} = \text{'занят'}$, б — $\text{А} = \text{'свободен'}$

Очевидно, в общем случае можно получить множество различных СОС, начинающихся с выполнения одного и того же активного (однотипного) оператора. Чтобы сообщить эти варианты, необходимо перейти от трека к структуре. Для этой цели объединим однотипные СОС.

Граф, получившийся в результате объединения всех возможных однотипных СОС, называется *классом одновременных событий* (КОС). Вершинами этого графа являются имена подоператоров, дугами — возможные пути построения последовательности событий. Каждая дуга взвешивается условием, при выполнении которого к СОС присоединяется связанное с ней событие.

На рис. 4, а представлен КОС для совокупностей, начинающихся с оператора $\Gamma_{\text{а}}$.

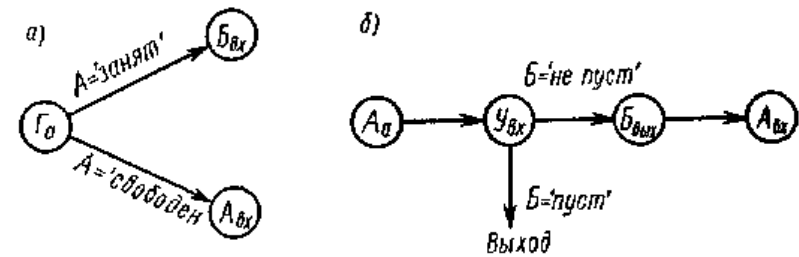


Рис. 4. Пример КОС:
 а — для блока Г; б — для блока А

Таким образом, если СОС была связана с моментом ее генерации, то КОС связан с источником активного события и генерирует совокупности для любого момента времени. На рис. 4, б представлен КОС, соответствующий активному подоператору $\text{А}_{\text{а}}$. Таким образом, для задания последовательностей событий в модели достаточно описать КОС для активных событий от блоков Г и А.

Предложенный подход к построению имитационного процесса позволяет анализировать вычислительную эффективность того или иного способа генерации СОС.

Для большинства промышленных систем имитационного моделирования — GPSS, Симула, Симскрпт, ЦСЛ, НЕДИС, SOL, СЛЭНГ характерно использование автоматического способа генерации СОС. При таком способе все подоператоры h^i имеют форму подпрограмм, которые по входу и выходу связаны только с управляющей программой. Управляющая программа выполняет функции управления, руководствуясь информацией, хранящейся в списке будущих событий и в списке условий. Структурная схема моделирующего алгоритма представлена на рис. 5.

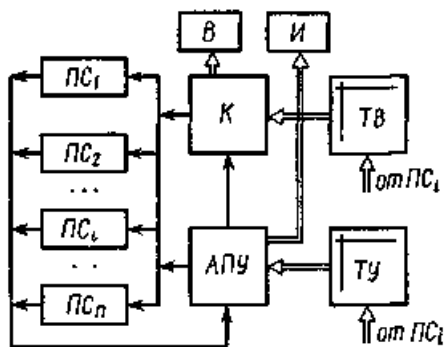


Рис. 5. Структурная схема моделирующего алгоритма сканирующего типа

При данном способе организации имитационного процесса список будущих времен содержит информацию лишь об активных событиях с указанием момента времени их активизации. Таким образом, из списка будущих времен (СБВ) извлекаются лишь первые события в СОС. Остальные, пассивные события СОС генерируются автоматически путем постоянного сканирования списка условий, где помещена информация об условиях свершения событий, указываются имена подпрограмм событий. После выполнения любой подпрограммы события управление возвращается управляющему алгоритму, который снова просматривает список условий нахождения истинного условия. Если список содержит лишь ложное условие, то это дает основание считать данную СОС исчерпанной и переходить к выполнению активного события из СБВ, начиная генерацию новой СОС. Количество машинных команд на имитацию в данном случае составит

$$N_c = n(mC + B_k + mB_y^c),$$

где n — число событий, прошедших через систему; m — среднее число событий, приходящееся на одно сообщение; C — средняя длина подпрограммы событий, выраженная в числе машинных команд; B_k — число машинных операций, выполняемых управляющей программой при обращении к СБВ; B_y^c — среднее число машинных операций, выполняемых управляющей программой при просмотре списка условий.

Другой способ генерации СОС выполняется линейным моделирующим алгоритмом (рис. 6).

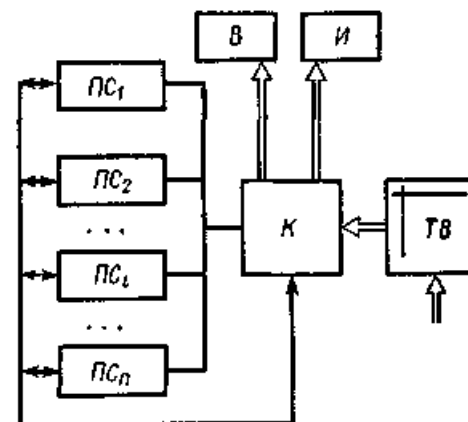


Рис. 6. Структурная схема моделирующего алгоритма линейного типа

В этом алгоритме подпрограммы событий сами определяют и вычисляют необходимые условия, организуют передачи управления друг к другу и обращаются к управляющему алгоритму лишь для записи и считывания информации из СБВ. Общее число машинных команд при линейном моделирующем алгоритме выражается формулой

$$N_n = n (mC + B_k + mB_y^d),$$

где B_y^d — число машинных операций по вычислению одного условия в составе подпрограммы события для анализа возможности дальнейшего перехода.

На практике величина $B_y^c \approx C$, в то время как $B_y^d \ll C$. Таким образом, использование линейного алгоритма по крайней мере в 2 раза

сокращает затраты машинного времени на построение имитационного процесса.

Структуры взаимосвязи подпрограмм событий в линейном алгоритме соответствуют структуре КОС. Поэтому при составлении имитационной модели ПО линейному алгоритму необходимо вначале описать все КОС, возникающие в системе, и затем отобразить их во взаимосвязях подпрограмм событий.

2.5.2. Управляющий алгоритм

Управляющий моделирующий алгоритм (УМА) выполняет следующие основные функции:

- ведение календаря событий;
- активизацию блоков;
- контроль за блоками, готовыми к передаче сообщений и за блоками, готовыми принять сообщение;
- обращение к блоку для считывания сообщения и для передачи в него сообщения;
- контроль за наступлением момента окончания моделирования;
- анализ приоритета сообщения на выходе блока и выбор наиболее приоритетного;
- прерывание обслуживания сообщения в блоке;
- анализ условий осуществления передачи по связи между блоками.

Управляющий моделирующий алгоритм построен по схеме слежения за системными списками, отражающими динамику состояния модели. Он имеет три основных и ряд вспомогательных списков

Основными списками являются:

- список будущих времен (СБВ) или календарь событий;
- список блоков, готовых выдавать сообщение (СГВ);
- список блоков, готовых принять сообщение (СГП).

Занесение очередного элемента в календарь события может быть инициировано активным блоком, определяющим интервал времени, через которое он должен быть активизирован УМА, либо оператором ПЛАНИРОВАТЬ.

В начале моделирования все активные блоки в модели заполняют первоначальное состояние СБВ.

Формирование СГВ и СГП осуществляют блоки. Элемент в список СГВ заносится блоком, у которого в выходной ячейке появилось обработанное сообщение, готовое к передаче. Занесение элемента СГП происходит по инициативе блоков, свободных от обработки сообщений. К таким блокам относятся, например, буферы неогра-

ниченного объема. В начале моделирования создается первоначальная структура СГП из блоков, которые в любой момент могут принять сообщения. Удаление информации о блоке из СГП происходит, когда блок либо начинает обработку, либо осуществляет передачу сообщения в принимающий блок.

Начальная настройка УМА состоит в первоначальном определении блоков, готовых принять сообщения, и определении ближайшего активного события. Из всего множества блоков на момент начальной настройки активизируются только блоки-генераторы. Определяется минимальное значение модельного времени, соответствующее выходу сообщения из этого генератора.

Как правило, одному моменту модельного времени соответствует несколько одновременных событий. Эта последовательность одновременных по времени событий определяется УМА, и ее нарушение может повлиять на правильность выполнения модели.

Для уменьшения времени просмотра всех имеющихся связей введен динамический список связей (ДСС), содержащий только те связи, в которых хотя бы один из блоков готов выдать или принять сообщение. Этот список обновляется, когда выбывает связь, по которой осуществлена передача или возникает блок, готовый выдать либо принять сообщение. Новая связь помещается в конец списка.

УМА выбирает очередную связь из данного списка, определяет, есть ли блоки, составляющие данный класс, готовые к выдаче или приему сообщения.

При описании оператора СВЯЗЬ необходимо рассматривать методы управления передачами сообщений между классами. Для выборки по уровням между классами задается матрица, определяющая номер выдающего и принимающего классов. Если элемент матрицы, имеющий номер строки, равный номеру уровня выдающего класса, и номер столбца, равный номеру уровня принимающего класса, есть единица, то передача разрешена.

Матрица может быть задана в виде исходных данных, либо установлена специальной программой установки значения матрицы.

Возможны два варианта задания связности без указания матрицы: 1) полностью связный и 2) взаимно однозначный.

Существует возможность закрывать передачу по всем взаимодействующим уровням с помощью оператора УСЛОВИЕ: если условие равно 1, то связь открыта, если 0, то закрыта. Кроме этого определены еще два режима работы УМА: 1) закрыть для просмотра данный уровень; 2) прекратить дальнейший просмотр по другим возможным взаимодействующим уровням.

Рассмотрим последовательность действий УМА после выбора очередной связи из списка ДСС. УМА определяет, есть ли в классах данной связи уровни групповых блоков, готовых принять и выдать сообщение. Если связь их не имеет, то происходит поиск следующей «активной связи» из списка ДСС. Если же передача по связи может произойти, УМА осуществляет проверку: был ли установлен запрос на обращение к подпрограмме пользователя оператором ПЛАНИРОВАТЬ. При наличии запроса первоначально выполняется подпрограмма, указанная в операторе ПЛАНИРОВАТЬ. Эта подпрограмма выполняется перед началом передачи по всем связям, имеющим готовность приема-передачи, затем проверяется наличие на данной связи условия. Следующая операция — проверка того, как задана дисциплина передачи по отдельным уровням классов на данной связи. Проверяется, задана ли матрица связности уровней классов либо заданы режимы безматричного взаимодействия.

В случае задания режимов ПРИОРИТЕТ или ПЕРЫВАНИЕ передача по уровням будет зависеть от значений полей сообщений, участвующих в передачах: если задана матрица, определяется элемент в данной матрице и выясняется, разрешена ли конкретная передача с уровня на уровень. Если передача была запрещена, то осуществляется поиск следующего из принимающих уровней и вновь проверяется битовый элемент матрицы. Осуществлению передачи сообщения предшествует настройка системного массива CHOISE. Следует помнить, что элементы данного массива доступны пользователю и могут быть модифицированы. Часть элементов этого массива выделяется специально для целей управления передачей сообщений по связям. Значениями элементов массива можно воспользоваться в подпрограммах, выполняемых до передачи и после передачи сообщений.

УМА позволяет изменять номера выдающего и принимающего уровней взаимодействующих классов. Этот режим, как правило, используется тогда, когда известны уровни, которые должны взаимодействовать, а применение матрицы связности невыгодно из-за холостых операций поиска соответствующего элемента.

УМА записывает в системную ячейку TRANSACT сообщение, находящееся на выходе уровня выдающего класса. Через эту ячейку пользователю дается возможность работать с полями сообщения, передаваемого по связи.

Следующим шагом УМА является обращение в подпрограмму, отмеченную на связи оператором ДО. В этой подпрограмме можно запретить передачу с уровня на уровень. Для этого можно воспользоваться восьмым элементом системного массива CHOISE.

Если CHOISE [8] = 1, то передача разрешена; если же передача была закрыта (CHOISE [8] = 0), то УМА переходит к поиску следующего уровня, готового принять сообщение В момент считывания определяется значение системной переменной NUMBER. Ее значение определяет номер уровня выдающего блока. УМА передает управление в данный блок в режиме «СЧИТАТЬ СООБЩЕНИЕ». Если в блоке предусмотрено выполнение подпрограммы пользователя в режиме «ВЫДАТЬ СООБЩЕНИЕ», то происходит вызов этой подпрограммы. Затем УМА очищает выходную ячейку блока, соответствующую данному уровню, и приступает к записи сообщения из системной переменной TRANSACT в входную ячейку принимающего блока по соответствующему уровню.

Управление передается в принимающий блок. Если предусмотрена подпрограмма в модели, выполняемая по записи в блок, то данный блок передает ей управление. К этому моменту в системной ячейке NUMBER содержится номер уровня принимающего класса.

В подпрограмме, выполненной после передачи, может быть указан режим отмены дальнейшего просмотра связи (значение CHOISE [9] = 0). Это означает, что даже если есть другие уровни, готовые выдать и принять сообщение на данной связи, то УМА перейдет к выбору следующего элемента списка ДСС, т. е. следующей связи, по которой возможна передача.

По умолчанию значение девятого элемента массива CHOISE устанавливается в 1. Эту функцию выполняет УМА при переходе к просмотру следующей связи.

Далее УМА просматривает наличие других уровней в данном классе, готовых принять сообщение. Если уровни, готовые принять, исчерпаны, осуществляется переход к просмотру следующей связи из ДСС.

Рассмотрим подробнее случай, когда передача не могла быть осуществлена по нескольким причинам: элемент матрицы равен 0; передача запрещалась в подпрограмме ДО, значение восьмого элемента массива CHOISE; в случае принудительной установки какого-либо из уровней, когда они не попали в состояние готовности.

При неизменном выдающем уровне идет поиск следующего принимающего уровня, затем цикл просмотра связи повторяется. Когда исчерпаны все принимающие уровни, то осуществляется переход к следующему уровню, готовому выдать, а просмотр блоков, готовых принять, начинается сначала. Цикл заканчивается, когда будут просмотрены все выдающие уровни. Номера уровней, по которым осуществилась передача, вычеркиваются из СГВ и СГП.

Таким образом, процесс моделирования включает три следующих вложенных цикла УМА:

- 1) внутренний цикл — просмотр уровней, по которым возможна передача;
- 2) средний цикл — выбор связи для передачи;
- 3) внешний цикл — переход к следующему модельному времени.

Выход из основного цикла возможен только тогда, когда выполнен оператор СТОП — прекращение моделирования

При анализе взаимодействующих уровней классов для данной связи возможен режим явного указания номеров уровней. Для явного указания номеров уровней служат элементы массива CHOISE [1] для выдающего уровня класса и CHOISE [2] для принимающего уровня класса.

Номера уровней должны устанавливаться в подпрограмме связи, выполняемой по метке ДО. В этом случае параметр матрицы равен ALL.

Для исключения просмотра других уровней из принимающего класса, если передача не осуществлялась, необходимо указать CHOISE [8] = 2. В противном случае будет выполнена метка ДО столько раз, сколько уровней, за минусом 1, входит в данный принимающий класс.

2.5.3. Практическая реализация моделирующих алгоритмов

Имитационные модели являются специфичным и достаточно сложным программным изделием, и их разработка должна вестись с применением прогрессивной технологии. Технология проектирования ИМ включает в себя методы и средства, обеспечивающие их создание и развитие в течение всего периода их жизни. Этот период содержит этапы проектирования, изготовления и эксплуатации ИМ. Он начинается с формирования назначения и принципов построения ИМ и завершается после полного прекращения эксплуатации модели.

Проектирование ИМ представляет собой весьма трудоемкий процесс. Большие размеры ИМ, сложность поведения ее компонент, высокая стоимость разработки требуют применения математических методов на всех этапах разработки ИМ. На всех этапах проектирования специалистам приходится рассматривать две стороны объекта проектирования: требования к ИМ со стороны внешней среды (внешнее проектирование) и организацию функционирования ИМ (внутреннее проектирование).

Формулировка технического задания. Начало работы над проектом ИМ имеет целью сформулировать задачи проектирования и организовать рабочие группы по решению поставленных задач. Для этого создается небольшая инициативная группа специалистов, которая на основе всесторонних обсуждений вырабатывает техническое задание.

Разработка технических предложений. На этом этапе предполагается выбор наилучшего решения, укомплектование группы проектировщиков системы до полного состава, составление плана всех работ над ИМ. Основными средствами исследования являются математические модели.

Работы должны осуществляться одновременно в течение срока проектирования. Этот этап характеризуется появлением большого количества новых вариантов решения задачи, новых проработок по частным вопросам отдельных направлений работы.

В результате формулируется проблема, ищутся пути ее реализации, составляется план выполнения проекта ИМ, который должен:

- содержать график распределения времени по этапам работ с указанием необходимых средств и состава исполнителей;
- предусматривать консультации специалистов по спорным вопросам;
- содержать заявки на проведение натурного и модельного экспериментов.

Эскизное проектирование. Этап начинается с разработки первого варианта ИМ.

На этапе эскизного проектирования решаются вопросы, допускающие различные ответы. Отчетная документация этапа должна содержать: подробное описание структуры ИМ; описание основных подсистем; перечень допустимых значений характеристик ИМ; хотя бы один метод физической реализации предложенного способа построения ИМ.

Испытание ИМ. Основная цель испытания состоит в получении подтверждения, что ИМ работает так, как предусматривалось при проектировании. В ходе испытания ИМ отлаживается, исключаются неизбежные ошибки программирования. Испытание проводится по плану испытаний в соответствии со специально разработанными методиками. В результате испытаний составляется отчет. В нем даются рекомендации по улучшению использования ИМ, приводятся результаты анализа различий между реальным образцом системы и ее имитационной моделью.

Техническое проектирование. В ходе разработки технического проекта уточняются и детализируются алгоритмы функционирования

компонент ИМ. На этом этапе наряду с множеством подмоделей компонент ИМ обязательны создание и испытание обобщенной имитационной модели, с помощью которой решаются вопросы внешнего и внутреннего проектирования.

Рабочее проектирование. Этап характеризуется созданием ИМ. На этом этапе должна быть разработана техническая документация, технические характеристики по эксплуатации ИМ, полное расписание всех технических решений.

В заключение отметим, что программная имитационная модель реализуется на ЭВМ в виде последовательного вычислительного процесса. Выполнение элементарного оператора называется *событием*. При реализации программы ИМ параллельные процессы функционирования моделируемой системы отображаются на один последовательный процесс вычислений, реализуемый ЭВМ. Главные задачи при осуществлении такого отображения — продвижение модельного времени и формирование для каждого момента времени класса одновременных событий. Первое событие в КОС — активное; признак окончания КОС — нулевые значения всех условий в подоператорах в составе элементарных операторов. В зависимости от подхода к построению КОС можно выделить алгоритмы моделирования сканирующего и линейного типов. Эти алгоритмы могут быть реализованы в рамках событийной, процессной или агрегативной концепций языков моделирования.

2.6. Статистические процедуры в имитационном моделировании

2.6.1. Генерирование случайных чисел

Как мы уже отмечали ранее, функционирование элементов системы, подверженных случайным воздействиям, задается *генераторами случайных чисел*, реализуемых программными методами, вырабатывающими псевдослучайные последовательности. Напомним, что *псевдослучайными последовательностями* называют вполне детерминированные числа, обладающие статистическими свойствами случайных чисел, определяемых путем их проверки специальными тестами, а также периодичностью, т. е. повторяемостью через определенные промежутки времени. При моделировании используются интервалы последовательностей псевдослучайных чисел, в которых нет одного числа, встречающегося более одного раза.

Методы генерирования случайных чисел. При генерировании случайных чисел, распределенных по любым законам, в качестве задающих используют генераторы равномерно распределенных случайных чисел в интервале 0..1. Наибольшее распространение получили следующие методы генерирования: квадратов, произведений, мультипликативный конгруэнтный и смешанный конгруэнтный методы

Метод квадратов. В квадрат возведено текущее случайное число и из результатов средних разрядов выделяется следующее случайное число.

Метод произведений. Два следующих друг за другом случайных числа умножают и из произведения средних разрядов выделяют следующее случайное число.

Мультипликативный конгруэнтный метод. В качестве текущего значения случайного числа выделяют остаток от деления произведения предыдущего случайного числа и постоянного множителя λ на постоянное число m :

$$y_i = \lambda y_{i-1} \pmod{m},$$

где λ , m — постоянные числа; y_i — i -е случайное число.

Смешанный конгруэнтный метод. Этот метод отличается от предыдущего прибавлением к остатку от деления постоянного числа μ :

$$y_i = \lambda y_{i-1} + \mu \pmod{m}.$$

Типы проверки генераторов равномерно распределенных псевдослучайных чисел. Различают три типа проверки: на периодичность, на случайность, на равномерность.

Проверка на периодичность. Проверка требует обязательного определения длины периода и отрезка аperiodичности псевдослучайных последовательностей.

Для определения *длины периода* выполняют следующие действия:
— выводят генератор псевдослучайных чисел интуитивно за пределы предлагаемого отрезка аperiodичности;
— регистрируют следующее за этим пределом случайное число;
— генерируют случайные числа и сравнивают их с зарегистрированным числом, подсчитывая количество случайных чисел, выработанных до совпадения с зарегистрированным случайным числом. Подсчитанное количество случайных чисел является длиной периода.

Для определения *отрезка аperiodичности* выполняют следующие действия:

— вырабатывают с помощью исследуемого генератора случайные числа, подсчитывая их до количества, равного длине периода. После

этого параллельно с ним генерируют случайные числа вторым генератором, аналогичным первому;

— вырабатывают случайные числа двумя генераторами, продолжая подсчитывать количество чисел, которое сгенерировал первый генератор, до сравнения между собой чисел, выработанных разными генераторами. Подсчитанное количество чисел является длиной отрезка аperiodичности.

Проверка на случайность. При проверке на случайность рекомендуется использовать совокупность тестов проверки: 1) частот; 2) пар; 3) комбинаций; 4) серий; 5) корреляции.

Первые четыре теста характеризуются общим свойством: *испытываемые псевдослучайные числа или их разряды в них классифицируются по некоторым признакам, различным для различных тестов. Полученное эмпирическое распределение сравнивается с теоретическим.* Для сравнения используются критерии согласования χ^2 и Колмогорова.

Тест проверки частот предполагает разбиение диапазона распределения на l интервалов и подсчет количества попаданий случайных чисел в выделенные интервалы. Возможно использование критериев согласия. Вероятность попаданий в заданный интервал теоретического распределения определяется по формуле

$$P_i = \int_{a_{i-1}}^{a_i} \frac{l}{a_i - a_0} dx = \frac{a_i - a_{i-1}}{a_i - a_0},$$

где a_i — верхний предел i -го интервала.

Тест проверки пар заключается в подсчете количества «1» для каждого разряда случайного числа. В этом случае используется критерий согласия χ^2 с одной степенью свободы. Теоретическая вероятность появления «1» для равномерно распределенных случайных чисел $P_i = 1/2$. Поразрядный анализ позволяет отбросить неслучайные разряды, в качестве которых часто оказываются младшие разряды с преобладанием в них «1».

Тест проверки комбинаций сводится к подсчету количества «1» в случайных числах. Можно использовать также критерии согласия. Теоретическая вероятность появления комбинации с i -м количеством «1» будет

$$P_i = \frac{k!}{i! (k-i)! 2^k},$$

где k — количество разрядов случайного числа.

Тест проверки серий заключается в подсчете количества различных длин последовательностей одинаковых значений случайных чисел.

Возможно использование критериев согласия. Теоретическая вероятность

$$P_i = R_i / N_s,$$

где R_i - количество серий i -й длины в N случайных числах; N_s — общее количество серий в N случайных числах при гипотетическом распределении.

Характерная особенность приведенных формул теста серий — независимый учет серий различной длины. Например, одна и та же серия, состоящая из четырех единиц, учитывается как одна серия из четырех, две — из трех и три — из двух единиц.

Тест проверки корреляции заключается в определении коэффициента корреляции. При этом выполняют следующие действия:

— запускают два генератора случайных чисел на отрезке аperiodичности с некоторой разницей между собой;

— подсчитывают коэффициент корреляции между этими последовательностями.

Проверка на равномерность. При проверке на равномерность можно использовать тест проверки частот, так как гистограмма частот хорошо отражает равномерность распределения случайных чисел по всему диапазону изменения. Для равномерного распределения случайных чисел $m_1 = (a_i + a_0)/2$ и $\sigma = (a_i - a_0)/(2\sqrt{3})$.

Задаемся доверительной вероятностью β того, что оценка математического ожидания m_1^* не выйдет за пределы доверительного интервала:

$$P\{|m_1 - m_1^*| < \varepsilon\} = \beta \text{ или } P\{m_1 - \varepsilon < m_1^* < m_1 + \varepsilon\} = \beta.$$

Величина β должна быть достаточно большой и составлять 0,9; 0,95; 0,99.

Методы генерирования псевдослучайных чисел. Для генерирования псевдослучайных чисел, распределенных по заданному закону, наиболее распространены методы: обратной функции; табличный; метод, основанный на функциональных особенностях генерируемых распределений.

Метод обратной функции. Метод основывается на использовании следующей теоремы:

если случайная величина имеет плотность распределения $f(z)$, распределение случайной величины

$$r = \int_a^z f(z) dz$$

является равномерным в интервале $0...1$ (a — нижняя граница диапазона изменения случайной величины z).

Чтобы получить случайное число z , плотностью $f(z)$, необходимо решить уравнение

$$r_j = \int_a^{z_j} f(z) dz.$$

Например, для экспоненциального закона

$$r_j = \int_0^{z_j} \mu e^{-\mu z} dz = -e^{-\mu z} \Big|_0^{z_j} = 1 - e^{-\mu z_j},$$

$$\ln(e^{-\mu z_j}) = \ln(1 - r_j),$$

откуда

$$Z_j = -(1/\mu) \ln(1 - r_j).$$

Достоинства метода обратной функции: точность метода; не требуется составления и хранения в памяти таблиц.

Недостатки: метод распространяется только на те функции, которые позволяют вычислить интеграл от функции плотности аналитически; использование численных методов вычисления интегралов приводит к ошибкам и большим затратам машинного времени; выражение, используемое для вычислений, содержит в себе функции вычисления логарифмов, возведения в степень, вычисления радикалов, что требует значительных затрат машинного времени.

Табличный метод. Метод применяется в трансляторе языка GPSS. В качестве аргумента используется равномерно распределенное случайное число r , в качестве функции — последовательность x_j чисел, задающих закон распределения.

С этой целью формируется таблица $\langle a_i, x_i \rangle, i = \overline{1, N}$.

Значение случайного числа Z с заданным законом распределения находят методом линейной интерполяции по формуле

$$Z = x_{i+1} + \frac{r_i - a_i - 1}{a_i - a_{i-1}} (x_i - x_{i+1}),$$

где $a_{i-1} < r_i \leq a_i, j = \overline{1, N}, i = \overline{1, l}$.

Поиск нужного интервала производится методом последовательного сравнения j -го случайного числа с границами интервалов $a_i, i = \overline{1, l}$ до выполнения условия

$$a_{i-1} < r_j < a_i.$$

Достоинства табличного метода: имеется возможность генерировать случайные последовательности с любым заданным законом; любую заданную точность можно получить при увеличении количества интервалов; требуется только одно случайное равномерное распре-

деленное число и выполнение несложных операций, занимающих мало времени.

Метод, основанный на функциональных особенностях распределений. Метод используется, как правило, в тех случаях, когда аналитически не удастся вычислить интеграл от функции плотности. Так, для генерации случайных чисел Z , имеющих специальное эрланговское распределение, можно воспользоваться k равномерно распределенными случайными числами r_i .

$$Z = -\frac{1}{k\mu} \sum_{i=1}^k \ln(r_i) = -\frac{1}{k\mu} \ln \left(\prod_{i=1}^k r_i \right).$$

Для генерации нормально распределенных случайных чисел используют центральную предельную теорему, на основании которой суммируют N равномерно распределенных случайных чисел и получают нормально распределенное случайное число. Принимают $N = 12...20$.

Если при суммировании равномерно распределенных случайных чисел в интервале $0...1$ математическое ожидание суммы $m_1 = N/2$, а среднее квадратическое отклонение $\sigma = \sqrt{N}/(2\sqrt{3})$, то для получения нормального распределения с $m_1 = 0$ и $\sigma = 1$ необходимо выполнить преобразование:

$$z = \frac{\sum_{i=1}^N r_{it} - N/2}{\sqrt{N}/(2\sqrt{3})} \sigma + m_1,$$

где σ — требуемое среднее квадратическое отклонение генерируемых случайных чисел; m_1 — требуемое математическое ожидание генерируемых случайных чисел.

Для оценки качества случайных последовательностей с заданным законом распределения наиболее часто используют тест проверки частот и метод доверительного интервала для математического ожидания.

2.6.2. Оценка точности результатов

При обработке результатов имитационного моделирования как на шаге управления, так и при проверке найденных решений уравнения баланса особое значение приобретают вопросы точности и достоверности статистических результатов.

Выборка S является конечной последовательностью значений параметра S_1, S_2, \dots, S_n , вычисленной через интервал Δt (шаг сбора статистической информации).

Выборка представляет собой временной ряд, или случайный процесс с дискретным временем.

Пусть имеется стационарный эргодический случайный процесс $f(t)$, непрерывный во времени. Однако значение процесса $f(t)$ меняется дискретно в случайные моменты времени $t_0, t_1, t_2, \dots, t_n$. Этот дискретный процесс наблюдается в течение времени T . Учитывая особенности цифрового моделирования на ЭВМ, оценим математическое ожидание процессов.

Различают три оценки математического ожидания — соответственно среднеинтегральную, средневзвешенную и среднеарифметическую:

$$m_x^a = \frac{1}{T} \int f(t) dt; \quad m_x^b = \sum_{i=1}^n c_i x_i; \quad m_x^c = \frac{1}{n} \sum_{i=1}^n x_i.$$

Известно, что средневзвешенная m_x^b и среднеарифметическая m_x^c оценки практически не различимы при достаточно больших n .

Среднеарифметическая оценка математического ожидания. Так как сбор статистики при моделировании производится через дискретные интервалы времени, то среднеарифметическая оценка математического ожидания является основной. Предположим, что временной дискретный процесс задан в дискретные моменты времени с интервалом Δ

$$f_{\Delta}(t) = (x_1, x_2, x_3, \dots, x_i, \dots, x_n),$$

где x_i — значение процесса в момент времени t_i .

Исследуем точность оценки m_x^c . Обозначим T — время моделирования.

Предположим, что автокорреляционная функция имеет экспоненциальный вид:

$$k(\tau) = \sigma^2 e^{-c|\tau|}.$$

Тогда

$$\sigma_{m_x^c}^2 = \frac{\sigma^2}{n} + \frac{2}{n} \sigma^2 \sum_{i=1}^{n-1} e^{-ci} - \frac{2}{n^2} \sigma^2 \sum_{i=1}^{n-1} i e^{-ci}.$$

Так как сбор данных производится через равные интервалы времени Δ , то $t_i = \Delta i$. Тогда получим

$$\sigma_{m_x^c}^2 = \frac{\sigma^2}{n} + \frac{2}{n} \sigma^2 \sum_{i=1}^{n-1} e^{-c\Delta i} - \frac{2}{n^2} \sigma^2 \sum_{i=1}^{n-1} i e^{-c\Delta i}.$$

Исключая промежуточные вычисления, окончательно получим

$$\sigma_{m_x^c}^2 = \sigma^2 \frac{n(1 - e^{-2c\Delta}) - 2e^{-c\Delta}(1 - e^{-c\Delta n})}{n^2(1 - e^{-c\Delta})^2}.$$

Введем следующие обозначения:

$$\sigma^2/\sigma_{m_x^c}^2 = W^a; \quad c\Delta = \theta_0; \quad c\Delta n = \theta.$$

Учитывая, что $\Delta n = T$, получим $\theta = \theta_0 n$. Тогда

$$W^a = \frac{n^2(1 - e^{-\theta_0})^2}{n(1 - e^{-2\theta_0}) - 2e^{-\theta_0}(1 - e^{-\theta})},$$

где θ_0 — число отрезков времени длиной в $K = 1/c$, помещаемое в интервале времени Δ ; θ — длительность интервала моделирования, выраженная в единицах K . Назовем величину W^a выигрышем.

Вычислив предельные значения W^a при больших и малых значениях θ_0 , окончательно получим

$$W^a = \min\{\theta/2, n\}.$$

Это выражение справедливо в области определения $\theta > 3$ либо $\theta < 0,5$.

Так как $\theta = n\theta_0$, то при $\theta_0 = 3$ и $\theta_0 = 0,5$ получим $n = 6$. Погрешность оценки в этом случае составит менее 5%. При $n > 10$ и допустимой погрешности 10% оценку можно производить, начиная с $\theta_0 < 1$. Причем, если допустимая погрешность в оценке математического ожидания составляет не более 5%, то производить сбор статистической информации чаще чем через $\theta_0 = 0,5$ не рационально вследствие излишних затрат машинного времени на выполнение процедур сбора статистики. При допустимой погрешности 10% нижний предел шага θ_0 увеличивается до $\theta_0 = 1$.

Неучет коррелирования снимаемых замеров приводит к сильно завышенным оценкам. При $\theta_0 < 1$ выигрыш в оценке математического ожидания может быть завышен в $2/\theta_0$ раз по сравнению с действительным.

2.6.3. Оценки типовых характеристик

Учитывая полученные результаты для оценок W^a , проанализируем четыре различные процедуры сбора статистической информации: увеличение числа наблюдений n при неизменном шаге θ_0 и неизменном интервале наблюдения θ ; разбиение одного наблюдения на независимые интервалы и усреднение по ним; проведение независимых повторных моделирований.

Увеличение числа наблюдений n при $\theta_0 = \text{const}$. Если $n > 3$, то W оценивается по формуле $W^a = \sigma^2/n$.

Пусть $\theta_0 > 3$, тогда $W^a \approx n$. Если число наблюдений увеличено в k раз, то $n_1 = kn$. Следовательно, $W_1^a = kn$ и выигрыш достигает $W_1^a/W^a = k$.

Пусть $\theta_0 < 0,5$. Тогда $W^a = \theta/2$, $W_1^a = k\theta/2$ и $W_1^a/W^a = k$.

Таким образом, в любом из случаев достигается увеличение точности в оценке m_x в k раз.

Увеличение числа наблюдений n при $\theta = \text{const}$. В данном случае увеличение числа наблюдений разумно лишь до значения, когда $\theta/n_1 > 0,5$. Если $\theta/n_1 = \theta/(kn) > 3$, то точность увеличивается в k раз. В противном случае — уменьшается, а при $\theta/n_1 < 0,5$ увеличения точности вообще не произойдет.

Повторные независимые наблюдения. Если в процессе моделирования выходной параметр P принимает ряд последовательных значений S_1, S_2, \dots, S_n , то выборку $\bar{S} = (S_1, S_2, \dots, S_n)$ можно считать вектором размерности n .

Если моделирование повторяется заново, но для других входных случайных последовательностей, то получим новый вектор \bar{S}_1 значений параметра P , одинаково распределенный с \bar{S} . Вообще, если моделирование повторяется k раз, то для параметра P получается последовательность k векторов $\bar{S}_1, \dots, \bar{S}_k$, имеющих одинаковую размерность и распределение.

Пусть входные случайные последовательности независимы. Тогда независимы и векторы $\bar{S}_1, \dots, \bar{S}_k$. Ниже, вплоть до особой оговорки, будем предполагать, что векторы $\bar{S}_1, \dots, \bar{S}_k$ независимы, стационарны, эргодичны и одинаково распределены. Как правило, при моделировании стационарных случаев все эти условия выполняются.

Точность оценки математического ожидания параметра P при одиночной выборке была исследована выше.

Рассмотрим случай k -повторной выборки — последовательности векторов $\bar{S}_1, \dots, \bar{S}_k$. Так как предполагается независимость выборок S_i , ($i = \overline{1, n}$), то дисперсия оценки по совокупности всех выборок составит $\sigma_k^2 = \sigma_1^2/k$, где σ_1^2 — дисперсия оценки, полученной по одной выборке \bar{S}_1 .

Сравним результаты повторных испытаний с бесповторным испытанием эквивалентной длины. Условия эквивалентны, если длина интервала при бесповторном испытании

$$T_\delta = kT_n,$$

где T_n — длина интервала моделирования для получения вектора \bar{S}_1 .

Пусть для повторной выборки $\theta_n > 3$. Тогда

$$W_n = \theta_n/2; \quad W_k = W_n k,$$

где W_k — выигрыш для k -повторной выборки.

Для бесповторной выборки $\theta_\delta = k\theta_n$. Тогда

$$W_\delta = \theta_\delta/2 = k\theta_n/2 = W_n k.$$

Таким образом, при $\theta_n > 3$ использование k -повторной выборки не дает преимуществ перед повторной выборкой эквивалентного объема. Более того, использование повторного моделирования нежелательно вследствие усложнения алгоритма моделирования из-за организации повторения и установления начальных условий.

Предположим, что $\theta_n < 0,5$. Тогда $W_n \approx 1$ и $W_k = W_n k \approx k$. Если $\theta_\delta = \theta_n k > 3$, то $W_\delta = \theta_n k/2$.

Относительный выигрыш повторной выборки перед бесповторной составит $W_k/W_\delta = 2/\theta_n$.

Учитывая, что $W_n = 1$, заключаем, что при измерениях достаточно одной точки (большее их число не дает увеличения точности оценки). Это обстоятельство позволяет неограниченно уменьшать интервал θ .

Таким образом, относительный выигрыш повторной выборки перед бесповторной не зависит от числа повторений k (при условии, что $k > 3/\theta_n$), а зависит лишь от длины интервала наблюдений за процессом при повторной выборке. Чем меньше этот интервал, тем больший выигрыш можно получить.

2.6.4. Начальный интервал моделирования

В результате имитационного моделирования замкнутых СМО, где интенсивность запросов на входе системы зависит от длины очереди к ресурсу, выявлено, что автокорреляционная функция в начале имеет тенденцию к «затягиванию» координат, т. е. $k(\tau)|_{\tau=0} = 0$, и соответственно к более медленному затуханию. Этими свойствами обладает функция Лаггера.

Аппроксимируем корреляционную функцию функцией Лаггера первого порядка:

$$R_x(\tau) = \sigma^2(c\tau + 1)e^{-c\tau}.$$

Как было показано выше, процесс $f_n(t)$ не имеет среднеинтегральной оценки, так как не имеет значений вне дискретных моментов времени. Таким образом, процесс $f_n(t)$ является усечением процесса $f(t)$ на множестве из n элементов. Процесс $f(t)$ хотя и является дискретным, но, поскольку описан в непрерывном времени и может быть замерен в произвольный момент времени, имеет среднеинтегральную оценку. Очевидно, что среднеинтегральная оценка полнее и точнее среднеарифметической, однако с увеличением объема выборки n растет точность отображения $f(t)$ на $f_n(t)$, но увеличение n приводит к увеличению длительности моделирования.

Поэтому определение объема выборки является компромиссным решением между стоимостью машинного времени, затраченного на моделирование, и точностью модели.

Среднеинтегральную m_x^m и среднеарифметическую m_x^a оценки математического ожидания процесса $f_x(t)$ будем сравнивать по значениям дисперсий этих оценок, полагая, что чем меньше дисперсия, тем точнее оценка.

Известно, что оценки m_x^m и m_x^a являются несмещенными и асимптотически нормальными. При этом среднеквадратическое отклонение оценки m_x^m процесса $f(t)$

$$\sigma_{m_x^m}^2 = \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) k_x(\tau) d\tau;$$

среднеквадратическое отклонение оценки m_x^a процесса $f_a(t)$.

$$\sigma_{m_x^a}^2 = \frac{\sigma^2}{n} + \frac{2}{n} \sum_{i=1}^{n-1} \left[1 - \frac{i}{n}\right] K_{x_a}(t_i).$$

Известно, что для получения точного значения среднеинтегральной оценки математического ожидания при моделировании должно выполняться равенство

$$\frac{1}{T} \int_0^T f(t) dt = \frac{1}{T} \sum_{i=1}^n f_a(t_i - t_{i-1}),$$

где $f(t)$ — стационарный случайный процесс, непрерывный во времени; $f_a(t)$ — временной дискретный процесс; n — число значений $f(t)$ на интервале $[0, T]$.

Опуская промежуточные вычисления, получим:

$$W^a = \theta/\theta_0;$$

$$W^a = \min\{\theta/2, \theta/\theta_0\} \text{ для } \theta_0 < 0,5 \text{ и } \theta_0 > 3;$$

$$W^a = \theta/\theta_0 \text{ для } \theta_0 > 5.$$

Таким образом, вид корреляционной функции влияет на длительность начального интервала моделирования и величину допустимой погрешности в оценке математического ожидания. Так, при шаге $\theta_0 = 0,5$ одна и та же точность достигается при интервалах моделирования для случаев:

- а) $\theta = 5$ — некоррелированность элемента выборки;
- б) $\theta = 20$ — корреляционные функции экспоненциальные;
- в) $\theta = 37$ — корреляционная функция — функция Лаггера первого порядка

Уменьшение же шага θ_0 (увеличение числа наблюдения) с 1 до 0,1 приводит к сокращению интервала моделирования лишь на 14% в случае в), а в случае а) — в 10 раз. При уменьшении θ от 5 до 0,5

интервала моделирования может быть сокращен лишь в 1,35 раза в случае в), в случае б) — в 2,5 раза, а в случае а) — в 10 раз.

Игнорирование влияния начального интервала моделирования приводит к существенному завышению ожидаемой точности оценки математического ожидания.

Методы математической статистики широко применяют в имитационном моделировании. Для генерации случайных чисел с заданным законом распределения используются псевдоустойчивые числа, равномерно распределенные в интервале $[0,1]$. Наибольшее распространение получили два метода: построение обратной функции распределения и использование специфических свойств искомого распределения. Для оценки точности результатов моделирования важно знать автокорреляционные свойства измеряемой последовательности значений. В разделе приведены аналитические выражения для вычисления статистической погрешности с учетом автокорреляционных зависимостей. Показано, что простое группирование значений не приводит к выигрышу точности. Шаг измерения контролируемой величины зависит от значения коэффициента корреляции.

Начальный интервал моделирования всегда искажает оценку стационарных значений измеряемых величин. В общем случае желательно не учитывать этот интервал в накапливаемой статистической совокупности. Однако возможны ситуации, когда отбрасывание этого интервала не улучшает точности искомых стационарных значений.

2.7. Система имитационного моделирования СТАМ-КЛАСС

2.7.1. Принципы построения

В основе имитационного процесса в языке СТАМ-КЛАСС лежит элементарный программный имитатор — блок, который имитирует процессы обработки одного сообщения в элементарной системе массового обслуживания. Как правило, блоку можно поставить в соответствие какое-либо реальное техническое устройство: каналы связи, памяти, процессоры, управляющие устройства и т. д.

Программный блок реализован в виде специальной параметрической процедуры (рис. 1).

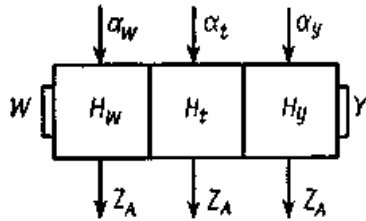


Рис. 1. Формализованная схема блока

Операторы H_W , H_y , H_t реагируют соответственно на поступление сообщения на вход W , выход Y сообщения из блока, активизацию блока по времени t . Управляющие сигналы α_i (где $i \in \{W, T, Y\}$) передают управление на соответствующие операторы. Сигналы Z_A , выходящие из блока, передают управляющую информацию в программу системной синхронизации. Они могут быть типа Z_t , Z_s , Z_r . Сигнал Z_t заносит очередной момент активизации данного блока в системный календарь; Z_s включает данный блок в очередь блоков, готовых принять очередное сообщение на вход W ; Z_r включает блок в очередь блоков, готовых передать сообщение, находящееся на выходе Y , в любой другой блок.

Кроме этого алгоритмы, реализующие операторы H , в свою очередь также могут иметь совокупность формальных параметров, позволяющих задавать режим обработки сообщения в блоке, параметры законов распределения, глобальные параметры для связи с внешним описанием либо с другими блоками.

Функционирование системы задается на базе взвешенного графа S , называемого структурой:

$$S = \langle B, L, U \rangle,$$

где B — множество вершин; L — множество упорядоченных пар (LCB^2); U — множество, состоящее из нулей и единиц, поставленных в соответствие каждой паре из L .

Физический смысл понятия структуры состоит в том, что элементам B соответствуют отдельные блоки, элементам L — линии связи передачи сообщений из блока в блок, элементам U — условия существования линий связи.

Если значение $k \in U$, соответствующее связи $l \in L$, равно 1, то связь существует и возможно прохождение сообщений из блока в блок по этой линии связи. В противном случае связь отсутствует. Очевидно, что если заданы блоки и связи между ними, а также определены

значения условий, то функционирование системы в целом задается однозначно.

С целью видоизменения структуры в процессе моделирования и создания способа ее обобщения введено понятие мультиструктуры. Мультиструктура представляет собой граф, элементами которого являются множества блоков, называемые классами (рис. 2).

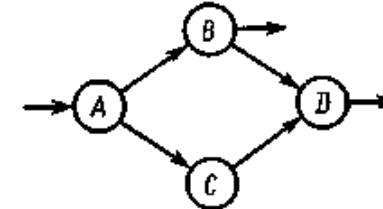


Рис. 2. Пример мультиструктуры:
A, B, C, D — имена массивов

Понятие мультисвязей и условий на них для мультиструктуры аналогично тем же понятиям в определении структуры.

Для преобразования мультиструктуры в структуру необходимо наполнить вершины мультиграфа конкретными блоками, а также раскрыть каждую мультисвязь путем задания матрицы M связности между всеми элементами двух классов (рис. 3).

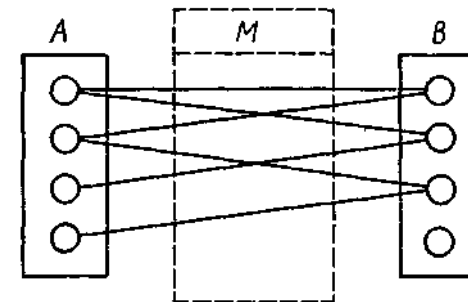


Рис. 3. Пример раскрытия мультисвязи

Таким образом, располагая средствами для видоизменения структуры системы в пределах, определяемых заданной мультиструктурой, с помощью языка СТАМ-КЛАСС строят имитационные программы обобщенных моделей, реализуя каждый раз какой-либо конкретный вариант структуры (см. приложение 2).

Символы языка. При написании программы на языке СТАМ-КЛАСС можно использовать 30 букв русского, 15 букв латинского алфавита, цифры от 0 до 9 и специальные символы.

Константы. В языке СТАМ-КЛАСС используются целая и действительная константы. Целая константа допускает две формы записи: 1) целую константу без порядка; 2) целую константу с порядком.

Примеры записи целых констант:

5; 128; 11E + 3; 500E — 1; 1E6.

Действительная константа имеет также две формы записи: 1) в виде целой и 2) дробной частей, отделяющихся друг от друга десятичной точкой.

Примеры записи действительных констант:

0,21; .21; 47.; 1.E—18; 17 62E8.

Идентификаторы. Идентификатор представляет собой последовательность не более восьми букв и цифр, начинающуюся с буквы.

Примеры записи идентификаторов:

M; Y2IP; КОНРЕАЛ; МЕТКА145.

Простые переменные, массивы, переменные с индексами. В языке СТАМ-КЛАСС используются простые переменные и переменные с индексами. Простые переменные обозначаются идентификаторами. Переменные с индексами применяются для обозначения элементов массивов. Массивы могут быть только одномерные или двумерные. Индексами могут быть целые константы или простые переменные целого типа. Индексы записываются в квадратных скобках.

Примеры записи переменных с индексами:

T[4]; МАСТ{I, J}. В{A, 7}.

Типы переменных. Переменные могут быть трех типов: целые, действительные и именные. Именные переменные используются для организации косвенной адресации и содержат не числовые значения, а адреса (ссылки) других объектов программы. Допускается формирование ссылок на простые переменные, массивы, метки и другие именные переменные.

Списки, матрицы. Для организации различных очередей, буферов, памяти при моделировании используются списки.

Каждый элемент списка содержит адрес следующего элемента $A_{с.з}$ и информации, для хранения которой отводится 8 байт. Каждый список имеет адрес первого элемента списка $A_{ис}$. Наименование списков

осуществляется с помощью простых переменных или переменных с индексами, которые содержат начала списков.

Структура моделируемой системы задается с помощью матриц смежности. Матрица смежности является битовым объектом, т. е. каждый элемент матрицы занимает 1 бит.

Номер строки матрицы соответствует номеру элемента, откуда исходит дуга, а номер столбца — номеру элемента, куда она входит. Причем, если элемент i связан с элементом j , то элемент матрицы (i, j) равен 1 и 0 — в противном случае.

Операторы, метки. Программа на языке СТАМ-КЛАСС записывается с помощью операторов. Каждый оператор начинается служебным словом, за которым следует символ «надчеркивание», и кончается символом «звездочка». В качестве названий операторов используются служебные слова:

'ОПИСАНИЕ, СИСТЕМЫ, ПЕРЕМЕННЫЕ, МАССИВЫ, СПИСКИ, МЕТКИ, КЛАССЫ, ПРОЦЕДУРЫ, ФОРМАТЫ, ВЫЧИСЛИТЬ, ОПРЕДЕЛИТЬ, СРАВНИТЬ, ЕСЛИ, ЦИКЛ, ПЕРЕЙТИ, ВЫПОЛНИТЬ, ПОДПРОГРАММА, ВЫХОД, СОЗДАТЬ, РАЗРУШИТЬ, ПЛАНИРОВАТЬ, ОТЛАДКА, СТАТИСТИКА, БЛОКИ, СБ, МАТРИЦЫ, УСЛОВИЯ, СТРУКТУРА, СВЯЗЬ, ЦЕЛЫЕ, ДЕЙСТВИТЕЛЬНЫЕ, ИМЕННЫЕ, КОНЕЦ, УПРАВЛЕНИЕ, ВВОД, СБРОС, НАПЕЧАТАТЬ, ТЕКСТ, СП, СОСЛАТЬСЯ, СТОП, КОММЕНТАРИЙ, ПАМЯТЬ, НАЧАТЬ.

• **Примечание.** Служебные слова распознаются транслятором только по первым трем буквам.

Операторы делятся на исполняемые и неисполняемые. Исполняемые операторы могут быть помечены меткой.

Меткой являются идентификаторы, которые ставятся перед оператором и отделяются от него символом «:». Примеры записи операторов:

ВВОД_А, В, С1*
МА13: СТОП_*

Арифметическое выражение. При записи арифметических выражений используются следующие символы арифметических операций: + (сложение); — (вычитание); # (умножение); / (деление); ^ (возведение в степень).

В арифметических выражениях разрешается использовать следующие функции:

SIN (синус); COS (косинус); TG (тангенс); LN (логарифм натуральный); LG (логарифм десятичный); ABS (абсолютное значение); SQRT (корень квадратный); EXP (экспонента); RAND (получение случайного числа с равномерным законом распределения в интервале 0...1).

Функция RAND может иметь только два аргумента: 0 или 1. При RAND (0) осуществляется начальная настройка на генерирование случайных чисел; при RAND (1) генерируется очередное случайное число.

Примечание. Принципы построения арифметических выражений и порядок выполнения операций аналогичны языку ФОРТРАН. Примеры записи арифметических выражений:

$(A + B) \# C^{\wedge}D + \text{SIN}(X)/M[1, K] \# CL - 1.$
 $\text{COS}(A + B^{\wedge} + C) \# \text{RAND}(1) + \text{SIN}(\text{LN}(\text{SQRT}(A1^{\wedge}2)))$

В арифметическом выражении запрещается одновременное использование объектов целого и действительного типов. Исключение составляет показатель степени, стоящий после символа «^», для действительных выражений, который может быть как целым, так и действительным.

Именуемое выражение. В качестве именуемого выражения могут использоваться метка, именная переменная или элемент именованного массива, которые содержат адреса меток.

Циклический элемент. В ряде операторов языка СТАМ-КЛАСС используется циклический элемент, определяющий работу над множеством элементов. В циклическом элементе указываются два параметра: первый параметр задает начальный номер элемента (если он равен 1, то его можно опускать), а второй — количество элементов. Параметры могут быть либо целыми числами, либо целой переменной.

Примеры записи циклических выражений: MAC [5—15] — задано 15 элементов, начиная с 5-го; КЛ [4—K] — K элементов, начиная с 4-го; МН [20] — 20 элементов, начиная с 1-го.

2.7.2. Описание элементов языка

Программа, написанная на языке СТАМ-КЛАСС, начинается строкой-комментарием, которая заканчивается символом «*». После комментария следуют основные части программы: ОПИСАНИЕ, СТРУКТУРА, УПРАВЛЕНИЕ. Заканчивается программа служебным оператором КОНЕЦ □*. Рассмотрим более подробно части ОПИСАНИЕ, СТРУКТУРА и УПРАВЛЕНИЕ.

ОПИСАНИЕ. В эту часть программы входят разделы описания системных переменных; простых переменных; массивов; меток; классов; стандартных процедур; форматов; стандартных блоков; матриц; условных переменных. Описание каждого раздела начинается с соответствующего служебного оператора. Последовательность разделов описаний фиксирована и соответствует перечисленной; кроме

того, идентификаторы объектов в различных частях описаний не должны совпадать. Описание обозначается служебным оператором ОПИСАНИЕ*.

Описание системных переменных. Системные переменные используются для задания размерностей массивов, матриц, количества экземпляров групповых блоков. Системные переменные могут использоваться и как обычные переменные. Значения системных переменных вводятся в ЭВМ перед трансляцией программы. Системные переменные могут быть только двух типов: целые и действительные.

Примеры описания системных переменных:

СИСТЕМНЫЕ*
 ЦЕ ВАР, КАБ, № 1*
 ДЕЯ ТРЕАЛ, КОЛ2*

Описание простых переменных. Простые переменные могут быть четырех типов: целые, действительные, списковые, именные.

Пример описания простых переменных:

ПЕРЕМЕННЫЕ*
 ЦЕЛ К2, L, AM*
 ДЕЯ A1, B3, X*
 СПИ СП2, 043*
 ИМЕ ИМ1, S*

Описание массивов. Массивы, так же как и переменные, могут быть четырех типов. При описании одномерных массивов в квадратных скобках указывается его размер, а при описании двумерных массивов — общая размерность, количество строк и количество столбцов. Например,

M [10] — одномерный массив, состоящий из 10 элементов;
 M2 [КОЛ КСТР, КСТЛ] — двумерный массив.

Пример описания массивов:

МАССИВЫ*
 ЦЕЛ M1 [5], MAC1 [100 20,5]*
 ДЕЯ MAC2 [КОЛ КСТР, КСТЛ]*
 СПИ ОЧ [NK], СПИ [48 NCT, NCL]*
 ИМЕ ПЕР [30]*

Описание меток. Все метки, используемые в программе, должны быть описаны.

Пример описания меток:

МЕТКИ МК1, ПОВ, КОН1: A*

Описание классов. Классы служат для описания структуры исследуемой системы в виде мультиграфа.

Пример описания классов:

КЛАССЫ⁻ ГЕН, ОЧОА, КЛОА2

Описание стандартных процедур. Система моделирования СТАМ-КЛАСС имеет библиотеку стандартных процедур. (СП). Вызов осуществляется по номерам, под которыми они хранятся в библиотеке. При описании СП устанавливается соответствие номера СП и присваиваемого ей идентификатора.

Пример описания СП с номерами 2 и 4:

ПРОЦЕДУРЫ⁻ МАТР = 2, УН = 4*

Описание форматов. Данный раздел предназначен для описания структуры динамических объектов. Динамический объект состоит из набора параметров и сообщения, которое содержит адрес начала набора параметров.

Количество байтов, выделяемое под параметры динамических объектов, кратно 8. Параметры действительного типа занимают все 8 байт. Параметры целого типа можно плотно упаковывать внутри этих байтов, указывая номера битов, которые они занимают.

Пример описания формата динамического объекта:

ФОРМАТ⁻ ФС = ТР/А/В/НАБ (1—32), ВИД (32—38),
 ФАЗА (39—52), СЧ (53—64)/С *

где ФС — название формата; / — указатель перехода на начало следующих 8 байт.

Первый параметр динамического объекта отводится под время его появления. Все параметры, перечисленные в операторе ФОРМАТ, должны быть описаны в разделе описания переменных.

Доступ к параметрам сообщения осуществляется с помощью конструкции, называемой *элементом формата*:

Ф (П, С),

где Ф — идентификатор формата; П — параметр динамического объекта; С — именная переменная, содержащая сообщение.

• **Примечание.** Если сообщение находится в стандартной переменной TRANSACT, то переменную С можно не указывать.

Элементы формата можно использовать в арифметическом выражении.

Описание стандартов блоков. В данном разделе указываются стандартные блоки (СБ), которые используются в программе. Они описываются следующим образом:

СБ⁻ □ НАЗ [КОЛ] = N (P1, P2, ..., P_n ...) *

где НАЗ — идентификатор блока; КОЛ — количество блоков данного типа (целая константа или целая системная переменная); P_i — параметры блока.

Примеры описания стандартных блоков:

БЛОКИ

СБ⁻ ГЕН [10] = 1 (M, T) *

СБ⁻ ОА {K} = 2 (M1, TOA, M2) *

Описание матриц. Матрицы описываются следующим образом:

МАТРИЦЫ⁻ НАЗ [KСТР, КСТЛ], ..., *

где ПАЗ — идентификатор матриц (целая константа или целая системная переменная).

Пример описания матриц:

МАТР⁻ МТ1 [1, 15] МОА [N1 M1] *

Изменять матрицы можно только с помощью СП.

Описание условных переменных. Условные переменные используются для управления передачей сообщений между вершинами мультиграфа. Условные переменные могут принимать значения 0 или 1, причем 0 означает запрещение передачи, а 1 — разрешение передачи сообщений. Изменяют значения условных переменных с помощью оператора СП.

Пример описания условных переменных:

УСЛОВИЯ⁻ У1, У2, УСБЛ, ПЕРС1 *

СТРУКТУРА. Эта часть программы предназначена для описания мультиграфа системы и задания последовательности передачи сообщений между вершинами мультиграфа. Она начинается со служебного оператора

СТРУКТУРА⁻ *

После этого оператора перечисляются связанные между собой вершины мультиграфа с помощью оператора

СВЯЗЬ.

Примеры использования оператора СВЯЗЬ:

* Стандартные блоки, входящие в класс К1, выдают сообщения, а СБ, входящие в класс К2, их принимают. Связь между блоками задается матрицей смежности М:

СВЯ⁻ К1—К2 МАТ⁻ М *

* Если все СБ класса К1 связаны со всеми СБ класса К2, то вместо идентификатора матрицы записывается служебный идентификатор ALL:

СВЯ⁻ К1 — К2 МАТ⁻ ALL *

* Если связь задается между блоками, имеющими одинаковые номера, то используется служебный идентификатор EQV:

СВЯ⁻ К1 — К2 МАТ⁻ EQV *

* Для описания приоритетных систем используется следующая конструкция:

СВЯ⁻ К1 — К2 МАТ⁻ М ПРИ⁻ Ф (ПРИ)*

где Ф — формат динамического объекта; ПРИ — параметр динамического объекта, имеющий значение приоритета.

При использовании данной конструкции передача сообщений между блоками осуществляется в порядке возрастания приоритета, т. е. просматриваются все блоки, входящие в класс К1, и выбирается сообщение с наименьшим значением параметра ПРИ, которое затем передается в соответствии с матрицей М блоку класса К2. Далее находится сообщение с наименьшим значением параметра ПРИ из оставшихся в классе К1, которое передается также в класс К2 и т. д.

* Для моделирования систем, имеющих дисциплину обслуживания сообщений с абсолютными приоритетами, используется следующая конструкция:

СВЯ⁻ К1 — К2 МАТ⁻ М ПРЕ⁻ Ф (ПР)*

где ПР — параметр динамического объекта, значение которого определяет необходимость прерывания.

Данная конструкция оператора обеспечивает прерывание обработки сообщения, имеющего большее значение параметра ПР, чем вновь поступающее. Прерванное сообщение помещается на выход блока.

* Управление передачей динамических объектов между классами осуществляется с помощью условных переменных:

СВЯ⁻ К1 — К2 МАТ⁻ М УСЛ⁻ У*

где У — условная переменная.

Если У = 0, то передача сообщений между классами К1 и К2 не происходит; если У = 1, то передача сообщений между классами разрешена.

* Управление передачей динамических объектов между отдельными блоками, входящими в классы:

СВЯ⁻ К1 — К2 МАТ⁻ М ДО⁻ М1 ПОС⁻ М2*

где М1 — метка подпрограммы, к которой происходит обращение при установлении связи между блоками до передачи сообщений; М2 — метка подпрограммы, к которой происходит обращение после передачи сообщений между блоками.

Для управления передачей сообщений между блоками в подпрограммах М1 и М2 используется служебный массив целого типа, который содержит следующие элементы:

SHOISE [1] — номер блока в классе К1, выдающий сообщение;

SHOISE [2] — номер блока в классе К2, принимающий сообщение; SHOISE [8] — признак передачи сообщения между блоками. Если SHOISE [8] = 1, то передача сообщения происходит; если SHOISE [8] = 0 — не происходит.

Общий вид оператора СВЯЗЬ:

СВЯ⁻ К1 — К2 МАТ⁻ М ПРИ⁻ Ф (А) ПРЕС⁻ Ф (В) ДО⁻ М1 ПОС⁻ М2 УСЛ⁻ У*

Любой из режимов ПРИ, ПРЕ, ДО, ПОС, УСЛ может отсутствовать. Порядок режимов в операторе может быть произвольным.

Сообщение, передаваемое между блоками, помещается в стандартную переменную TRANSACT.

УПРАВЛЕНИЕ. Эта часть программы предназначена для управления процессом моделирования. Она начинается служебным оператором

УПРАВЛЕНИЕ⁻ *

Операторы данной части можно разделить на две группы:

- 1) операторы, используемые для организации вычислительного процесса;
- 2) операторы, предназначенные для обеспечения процессов моделирования.

2.7.3. Вычислительные и моделирующие операторы

К вычислительным и моделирующим операторам относятся: ВЫЧИСЛИТЬ, ОПРЕДЕЛИТЬ, ПЕРЕЙТИ, СОСЛАТЬСЯ, ЕСЛИ, СРАВНИТЬ, ЦИКЛ, ВВОД, СБРОС, НАПЕЧАТАТЬ, ТЕКСТ, СП, КОММЕНТАРИЙ, СТОП, СОЗДАТЬ, РАЗРУШИТЬ, ПЛАНИРОВАТЬ, СТАТИСТИКА, НАЧАТЬ, ПАМЯТЬ.

Оператор ВЫЧИСЛИТЬ. Данный оператор содержит ряд операторов присваивания, разделенных знаком «:».

Примеры записи оператора ВЫЧИСЛИТЬ

ВЫЧ⁻ S = S + M [K, J]; A = 0*

ВЫЧ⁻ МК [I] = 5.; Ф (TP) = SITIME *

где SITIME — стандартная переменная действительного типа, содержащая текущее значение модельного времени.

♦ **Примечания:** 1. Если левая часть оператора присваивания и арифметическое выражение имеют разные типы, то результат арифметического выражения преобразуется к типу левой части.

2. При использовании именных переменных перед выполнением арифметических операций осуществляется их переименование.

Оператор ОПРЕДЕЛИТЬ. Синтаксис данного оператора совпадает с синтаксисом оператора ВЫЧИСЛИТЬ. Однако оператор ОПРЕДЕЛИТЬ предназначен для работы только с именованными переменными и не выполняет их разименование.

Примеры записи оператора ОПРЕДЕЛИТЬ:

ОПР⁻ S = TRANSACT *
ОПР ФС [СООБ] = S *

Оператор безусловного перехода ПЕРЕЙТИ. Оператор осуществляет переход на метку, заданную в явном виде, посредством простой именной переменной или элемента именованного массива.

Примеры записи оператора ПЕРЕЙТИ:

ПЕР⁻ МЕТ 6 *
ПЕР⁻ МС [K] *

Оператор СОСЛАТЬСЯ. Оператор предназначен для формирования значений именных и переменных массивов.

Пример записи оператора СОСЛАТЬСЯ:

СОС⁻ МС [5] = М1, М2, М3, М4, М5 *

Пяти первым элементам именованного массива МС будут присвоены адреса меток М1, М2, ..., М5.

Оператор условного перехода ЕСЛИ. В операторе ЕСЛИ используется логическое выражение, состоящее из двух арифметических выражений, разделенных операцией отношения: < (меньше), > (больше), = (равно), <= (меньше или равно), >= (больше или равно), \neg = (не равно).

Примеры записи логических выражений:

А1 \neq S/BC <= 10.0, И1 \neg = J + 5.

Если арифметические выражения имеют разные типы, то результаты их вычисления преобразуются к действительному типу до выполнения операции отношения.

В операторе ЕСЛИ вычисляется логическое выражение, и если оно истинно, то управление передается на метку, стоящую после служебного слова ТО, иначе идет передача управления следующему оператору.

Примеры записи оператора ЕСЛИ:

ЕСЛ⁻ A < 15. ТО⁻ М2 *
ЕСЛ⁻ NE >= 4 ТО⁻ М6 *

Оператор СРАВНИТЬ. Синтаксис оператора СРАВНИТЬ полностью совпадает с синтаксисом оператора ЕСЛИ. Отличие состоит в том, что при вычислении арифметических выражений не производится разименование именных переменных.

Пример записи оператора СРАВНИТЬ:

CPA⁻ TRANSACT = S1 TO⁻ М5 *

Оператор ЦИКЛ. Оператор ЦИКЛ предназначен для циклического повторения участка программы и имеет две формы. В первой форме задается количество повторений в явном виде. Например,

ЦИКЛ⁻ М ПОВ⁻ 20 *
М: ВЫЧ⁻ A = A + C *

Здесь выполнение оператора, помеченного меткой М, будет повторено 20 раз.

В операторе ЦИКЛ можно указывать одновременное изменение нескольких параметров цикла. Например,

ЦИКЛ⁻ М1 J = 1 (K), C = 5 (2) ПОВ⁻ A *

В этом случае осуществляется выход из цикла, если параметр цикла принимает значение, равное или большее конечного значения параметра цикла. Например,

ЦИКЛ⁻ A5 K = P (1) ДО⁻ N *

Если в операторе ЦИКЛ указано несколько параметров цикла, то условие окончания цикла задается только для последнего параметра. Например,

ЦИКЛ⁻ М I = I (1), J = K (P) ДО⁻ 100

В данном случае вывод из цикла произойдет, когда J станет равным или больше 100.

Оператор ВВОД. Он предназначен для ввода исходных данных. Например,

ВВО⁻ A, B, MAC, M1 [5], M2 [K] M3 {2—15} *

где M1 [5], M2 [K], M3 [2—15] — циклические элементы.

В результате выполнения данного оператора будут присвоены значения:

простым переменным А и В;
всему массиву MAC, размерность которого определяется описанием;
пяти первым элементам массива М1. К первым элементам массива М2 и 15 элементам массива М3, начиная со второго элемента.

Оператор СБРОС. Оператор используется для присвоения нулевых значений простым переменным, массивам, матрицам и условным переменным. Синтаксис оператора аналогичен синтаксису оператора ВВОД.

Пример записи оператора СБРОС:

СБР⁻ A, MAC, M1 [6—K], M3 [P—K] *

Оператор НАПЕЧАТАТЬ. Данный оператор используется для печати простых переменных и массивов. Синтаксис оператора аналогичен синтаксису оператора ВВОД.

Пример записи оператора НАПЕЧАТАТЬ:

НАП⁻ МАС, А, М1 [15], С, М3 [Р—К] *

При печати простых переменных каждая печатается на одной строке. При печати массива целого типа в одной строке печатается по пять чисел, а массива действительного типа — по три числа (см. приложение 3).

Оператор ТЕКСТ. Оператор используется для печати буквенно-цифровой информации, которая задается в операторе между символами «дробная черта».

Пример записи оператора ТЕКСТ:

ТЕК⁻ /КОНЕЦ МОДЕЛИ 1/ *

При выполнении данного оператора на принтере будет напечатан текст:

КОНЕЦ МОДЕЛИ 1

Оператор СП. Данный оператор используется для обращения к стандартным процедурам, описанным в части ОПИСАНИЕ.

Пример записи оператора СП:

СП МАТР (МАТ, R, J, K, ПР, МУ, МК) *
СП УСЛ (У1, 0) *

Оператор КОММЕНТАРИЙ. Оператор является вспомогательным и служит для размещения комментариев в тексте программы.

Пример записи оператора КОММЕНТАРИЙ:

М КОМ⁻ КОНЕЦ ЦИКЛ 1 *

Оператор СТОП. Данный оператор завершает выполнение программы:

СТОП⁻ *

Оператор СОЗДАТЬ. Оператор СОЗДАТЬ предназначен для выполнения классов конкретными блоками. Пример записи оператора СОЗДАТЬ:

СОЗ⁻ Г = ГЕН, ПР = ОА [3—5], ОЧ = БУФ1, БУФ2 *

По данному оператору будет сформировано три класса: класс Г, состоящий из всех блоков ГЕН, количество которых задано в части программы ОПИСАНИЕ оператором ОА; класс ПР — из пяти блоков ОА; класс — ОА—из блоков БУФ1 и БУФ2.

Оператор РАЗРУШИТЬ. Данный оператор освобождает классы от блоков, заполненных ранее оператором СОЗДАТЬ. После исполнения оператора РАЗРУШИТЬ очищенные классы можно заново наполнить другими блоками.

Пример записи оператора РАЗРУШИТЬ:

РАЗ⁻Г, ПР *

Оператор ПЛАНИРОВАТЬ. Оператор ПЛАНИРОВАТЬ задает определенные условия, при соблюдении которых будут выполняться указанные в операторе подпрограммы.

Пример записи оператора ПЛАНИРОВАТЬ:

ПЛА⁻ М1 ВРЕМЯ⁻ Т *

Подпрограмма М1 будет выполнена, когда модельное время достигнет значения, задаваемого переменной Т. Например,

ПЛА⁻ М2 СОБ⁻ МОМ *

Подпрограмма М2 будет выполняться каждый раз при переходе к новому модельному времени. Например,

ПЛА⁻ М3 СОБ⁻ СВЯ *

Подпрограмма М3 будет выполняться при каждом анализе связи между блоками:

ПЛА⁻ М4 СОБ⁻ ВЫБ *

Подпрограмма М4 будет выполняться при каждой передаче сообщения между блоками.

Оператор СТАТИСТИКА. Данный оператор предназначен для печати стандартной статистической информации, собранной в процессе моделирования.

Пример записи оператора СТАТИСТИКА:

СТА⁻ *

Оператор НАЧАТЬ. Оператор НАЧАТЬ предназначен для запуска процесса моделирования. Пример записи оператора НАЧАТЬ:

НАЧ⁻ *

Оператор ПАМЯТЬ. Данный оператор предназначен для подготовки раздела оперативной памяти ЭВМ, используемой для построения списков и хранения динамических объектов.

Пример записи оператора ПАМЯТЬ:

ПАМ⁻ *

• **Примечание.** Оператор ПАМЯТЬ должен исполняться до выполнения операторов НАЧАТЬ и СОЗДАТЬ.

Подпрограммы. В языке СТАМ-КЛАСС используются подпрограммы без параметров. Они располагаются в конце части программы «УПРАВЛЕНИЕ». Подпрограмма представляет собой набор любых операторов языка, заключенных между операторами ПОДПРОГРАММА и ВЫХОД. Причем имя подпрограммы определяется меткой, стоящей перед оператором ПОДПРОГРАММА. Например,

П1:ПОД⁻ *
ВЫЧ⁻ А = 1 *
ВЫХ⁻ *

Обращение к подпрограммам происходит по оператору ВЫПОЛНИТЬ:

ВЫП-П1 *

Задание исходных данных для моделирования. Числовые значения переменных, указанных в операторе ВВОД, задаются с помощью оператора ввода исходных данных. Правила записи числовых значений соответствуют правилам записи языка СТАМ-КЛАСС. Числа отделяются друг от друга символом «:». Двумерные массивы и матрицы вводятся по строкам. Так как матрицы являются битовыми объектами, то их значения кодируются в шестнадцатеричном виде.

Пример. Задан исходный оператор

ВВОД- А, М, МАТ *

где А — целая переменная ($A = 3$), М — двумерный массив ($M_{11} = 1,7$, $M_{12} = 2,1$; $M_{21} = 4,6$; $m_{22} = 5,3$); МАТ — матрица размерностью 5×2 (1010010101 — битовое представление матрицы; 'A 00000' — шестнадцатеричное представление).

Последовательности числовых значений имеют вид:

3; 1,7; 2,1; 4 6; 5,3, 'A5400000'

2.7.4. Пример программы имитационной модели

В качестве примера рассмотрим модель процесса обработки потоков задач на ЭВМ (рис. 4), имеющей фиксированное число разделов оперативной памяти (ОП).

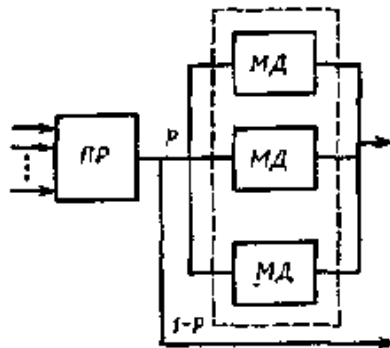


Рис. 4. Структурная схема системы обработки информации

Входные потоки задач подчиняются нормальному закону распределения, время решения задач в процессоре (ПР) — экспоненциальному закону, а время обращения к магнитным дискам (МД) — равномерному закону. Задачи, поступающие в ЭВМ, помещаются в соответствующие разделы ОП, откуда поступают в ПР в соответствии с приоритетами, причем номер приоритета соответствует номеру раздела. После обработки в ПР задачи с вероятностью $(1 - P)$ покидают систему, с вероятностью P поступают к первому свободному МД, после чего покидают систему.

На рис. 5 приведена структурная схема модели, а на рис. 6 — ее мультиграф.

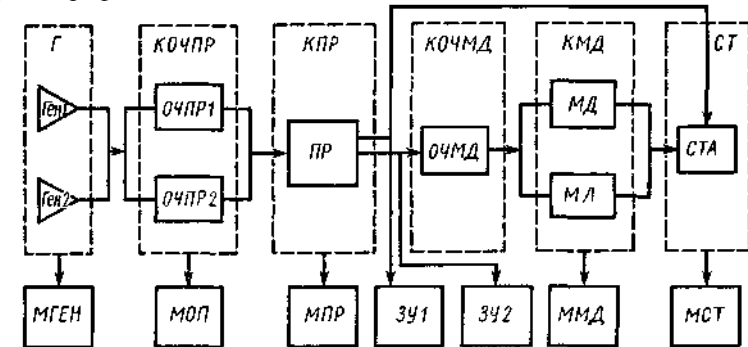


Рис. 5. Структурная схема модели

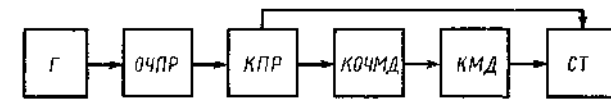


Рис. 6. Мультиграф модели

Обозначения, используемые на рисунках, соответствуют идентификаторам программы: Г — класс генераторов, имитирующий потоки задач; ГЕН — генерации динамических объектов; NF — количество генераторов в модели; МГЕН — подпрограмма вычисления интервалов времени, распределенных по нормальному закону в соответствии с массивом $ПГ_{ij}$ ($ПГ_{i1}$ — МО, $ПГ_{i2}$ — СКО, i — номер генератора); КОЧПР — класс очередей к ПР, имитирующий работу разделов ОП; ОЧПР — стандартный блок очереди с дисциплиной FIFO; МОП — подпрограмма присвоения приоритетов; КПР — класс, имитирующий работу ПР; ПР — стандартный блок, реализующий работу обслуживающего аппарата (ОА); МПР — подпрограмма вычисления времени обработки по экспоненциальному закону, а также

управления условиями передачи ($U1, U2$) в зависимости от вероятности P ; $ЗУ1, ЗУ2$ — подпрограммы вычисления $U1=0$ и $U2=0$ соответственно; *КОЧМД* — класс очередей, имитирующий очередь к МД; *ОЧМД* — стандартный блок очереди с дисциплиной FIFO; *КМД* — класс, имитирующий работу МД; *МД* — стандартный блок, реализующий работу ОА; *НМД* — количество МД; *ММД* — подпрограмма вычисления времени обработки по равномерному закону в соответствии с массивом $ПМД_{ij}$ ($ПМД_{i1}, ПМД_{i2}$ — нижняя и верхняя границы, i — номер МД); *СТ* — класс для сбора статистической информации; *СТА* — стандартный блок сбора статистической информации по времени пребывания задач в системе; *МСТ* — подпрограмма уничтожения динамических объектов и определения окончания моделирования.

Текст программы на языке СТАМ-КЛАСС приведен ниже.

МЕТКА

ПРИМЕР МОДЕЛИ НА ЯЗЫКЕ СТАМ/КЛАСС *
ОПИСАНИЕ *

```

СИС- *
ЦЕЛ- НГ, НГ2, НМА, НМА2*
ПЕР- *
ЦЕЛ- ПРИ, НОЧ*
ДЕЙ- ТР, ТОБ, КОЛ, S1, S2, P, ПАР, КС*
СПИ- СОЧ*
ЦЕЛ- I*
ДЕЙ- S*
МАС- *
ЦЕЛ- НОЧ[ 2 ]*
ДЕЙ- ПГ[ НГ2, НГ, 2 ], ПМА[ НМА2, НМА, 2 ]*
СПИ- МГЕН, МОП, МПР, ММА, МСТ, МП1, МП2 $
      ЗУ1, ЗУ2, МС1, КОН*
МЕТ- МГ1, РАБ*
КЛА- Г, КОЧПР, КПР, КОЧМА, КМА, СТ*
ПРО- ПОР=3, УН=4*
ПРО- УСА=6*
ФОР- Ф=ТР/ПРИ*
СБ- ГЕН[ НГ ]=1(МГЕН, ТР)*
СБ- ОЧПР[ 2 ]=3(МОП, СОЧ[ 1 ], НОЧ[ 1 ])*
СБ- ПР=2(МПР, ТОБ, РАБ)*
СБ- ОЧМА=3(РАБ, СОЧД, НОЧД)*
СБ- МА[ НМА ]=2(ММА, ТОБ, РАБ)*
СБ- СТА=4(КОЛ, S1, S2, МСТ, ТР)*
МАТ- ГОЧ[ НГ, 2 ]*
УСА- У1, У2*
      СТРУКТУР- *
СВЯ- Г-КОЧПР МАТ- ГОЧ*
СВЯ- КОЧПР-КПР МАТ- ALL ПРИ- Ф(ПРИ *
СВЯ- КПР-СТ МАТ- EQV ПОС- ЗУ1 УСА- У1*
СВЯ- КПР-СТ МАТ- EQV ПОС- ЗУ2 УСА- У2*
СВЯ- КОЧМА-КМА МАТ- ALL*
СВЯ- КМА-СТ МАТ- ALL*
      УПРАВЛЕН- *
ПАМ- *
ВВО- ПГ, ППР, ПМА, КС, P, ГОЧ*
ВЫЧ- S=RAND(0)*
СБР- КОЛ, S1, S2, НОЧ, СОЧ, У1, У2, НОЧД, СОЧД*
СОЗ- Г=ГЕН, КОЧПР=ОЧПР; КПР=ПР; $
      КОЧМА=ОЧМА; КМА=МА; СТ=СТА*
МГЕН.
ПОД- *
СП- ПОР(TRANSACTION, Ф)*
СБР- S*
МГ1.
ЦИК- МГ1 ПОВ- 12*
ВЫЧ- S=S+RAND(1)*
ВЫЧ- I=NUMBER*
ВЫЧ- ТР=ПГ[ I, 1 ]+(S-6.) ПГ[ I, 2 ]*
ВЫЧ- ТР=ТР+SITIME*
ВЫЧ- Ф(ТР)=ТР*
ВЫХ- *

```

```

МОП:   ПОД- *
        ВЫЧ-  $\Phi(\text{ПРН})=\text{NUMBER}^*$ 
        ВЫХ- *
МПР:   ПОД- *
        ВЫЧ-  $\text{ТОБ}=-\text{LN}(\text{RAND}(1))\#\text{ППР}^*$ 
        ЕСА-  $\text{RAND}(1)>\text{P TO} - \text{Mn1}^*$ 
        СП-   $\text{УСА}(\text{Y1},1)^*$ 
        ПЕР-  $\text{МП2}^*$ 
МП1:   СП-   $\text{УСА}(\text{Y2},1)^*$ 
МП:    ВЫХ- *
ЗУ1:   ПОД- *
        СП-   $\text{УСА}(\text{Y1},0)^*$ 
        ВЫХ- *
ЗУ2:   ПОД- *
        СП-   $\text{УСА}(\text{Y2},0)^*$ 
        ВЫХ- *
НМА:   ПОД- *
        ВЫЧ-  $\text{I}=\text{NUMBER}^*$ 
        ВЫЧ-  $\text{ТОБ}=\text{ПМА}(\text{I},1)+(\text{ПМА}(\text{I},2)-\text{ПМА}(\text{I},1))\#\text{RAND}(1)^*$ 
        ВЫХ- *
МСТ:   ПОД- *
        ВЫЧ-  $\text{ТР}=\Phi(\text{ТР})^*$ 
        СП-   $\text{УН}(\text{TRANSACT})^*$ 
        ЕСА-  $\text{КОА}<\text{КС TO}- \text{МС1}^*$ 
        ВЫП-  $\text{КОН}^*$ 
МС1:   ВЫХ- *
КОН:   ПОД- *
        СТА- *
        ТЕК- /ВРЕМЯ МОДЕЛИРОВАНИЯ/ *
        НАП- SITIME *
        НАП- /КОЛИЧЕСТВО ОБРАБОТАННЫХ ЗАЯВОК/ *
        НАП- КОА *
        ТЕК- /СРЕДНЕЕ ВРЕМЯ ОБРАБОТКИ ЗАЯВКИ/ *
        ВЫЧ-  $\text{S2}=(\text{S2}-\text{S1}^2/\text{КОА})/(\text{КОА}-1)^*$ 
        ВЫЧ-  $\text{S1}=\text{S1}/\text{КОА}^*$ 
        НАП- S1 *
        ТЕК- /ДИСПЕРСИЯ ВРЕМЕНИ ОБРАБОТКИ ЗАЯВКИ/ *
        НАП- S2 *
        ТЕК- /С.К.О.ВРЕМЕНИ ОБРАБОТКИ ЗАЯВКИ/ *
        ВЫЧ-  $\text{S2}=\text{S2}^*(0.5)^*$ 
        НАП- S2 *
        СТО- *
        ВЫХ- *
РАБ:   ПОД- *
        ВЫХ- *
        КОНЕЦ-*
```

Система имитационного моделирования СТАМ-КЛАСС ориентирована на агрегативную концепцию описания модулируемой системы. В ее основе лежат блоки-имитаторы библиотечного набора. Типичным объектом моделирования являются сети массового обслуживания. Особенность языка моделирования СТАМ-КЛАСС — возможность описания групповых агрегатов, входящих в классы агрегатов. Таким образом, возможно описание взаимодействия между множествами агрегатов. Программа включает в себя разделы: описание, управление

и структуру. Сбор статистической информации производится автоматически управляющим алгоритмом, либо пользователем системы с помощью вычислительных операторов. Система моделирования имеет препроцессор МАКРО-СТАМ, позволяющий компоновать программу из библиотечных фрагментов с последующей параметрической и структурной настройкой.

2.8. Система имитационного моделирования GPSS

2.8.1. Принципы построения имитационных программ

Объекты языка GPSS подразделяют на категории и типы.

Наименование категорий: операционная, аппаратная, динамическая, вычислительная, статистическая, запоминающая, группирующая. Наименование типов объекта: блоки, сообщения, устройства памяти, логические ключи, арифметические и булевские переменные, функции, очереди, таблицы, ячейки, матрицы ячеек, группы, списки.

Блоки. С объектами связаны определенные совокупности блоков, описывающих функционирование самой моделируемой системы либо содержащих дополнительную информацию о порядке моделирования. Моделирование заключается в продвижении сообщений (транзактов) от блока к блоку. Это продвижение создает блок GENERATE. Каждое продвижение сообщения является *событием* в модели. Комплекс программ, планирующий выполнение событий, реализующий функционирование блоков модели, регистрирующий статистическую информацию о прохождении сообщений, называется *симулятором*. Симулятор регистрирует время наступления каждого из известных на данный момент событий и выполняет их с нарастающей временной последовательностью. Задержку сообщений во времени, в соответствии с заданным законом распределения, может произвести блок ADVANCE. Реализация событий в модели может быть заблокирована невыполнением некоторых условий; эти события будут выполнены при благоприятном изменении блокирующих условий.

Симулятор обеспечивает отсчет модельного времени в принятых единицах, называемых *абсолютным условным временем*. С каждым сообщением связано *относительное условное время*, отсчет которого начинается при входе сообщения в моделируемую систему и

заканчивается при выходе сообщения из системы. Вывод сообщений из системы осуществляет блок TERMINATE.

Операционная категория. Блоки и сообщения — два основных типа объектов языка GPSS. Практически все изменения состояния модели происходят в результате входа сообщений в блоки и выполнения ими своих функций. С блоками связаны операторы, управляющие процессом моделирования.

Оператор SIMULATE указывает на необходимость проведения моделирования. При его отсутствии производится только трансляция программы.

Оператор START указывает на получение исходных данных и начало моделирования. Окончание моделирования производится при обнулении счетчика количества выводимых сообщений, задаваемого в поле А. Поле С определяет интервал выдачи промежуточной статистики.

Наборы управляющих операторов позволяют стирать накопленные данные, повторять выполнение программы, изменять параметры блоков (см. приложение 1).

Аппаратная категория. Язык GPSS оперирует тремя группами оборудования: устройствами, памятью и логическими ключами.

К группе *устройств* относятся блоки SEIZE, RELEASE, PREEMPT, RETURN. Введение в моделирующую программу описания устройства позволяет автоматически регистрировать статистическую информацию.

Группу *памятей* образуют блоки ENTER, LEAVE и оператор описания памяти STORAGE. Введение в моделирующую программу памяти позволяет автоматически регистрировать статистическую информацию.

Для управления *ключами* используется оператор LOGIG. Предусмотрено три режима изменения состояния ключа: сброс в «0»; установка в «1»; инвертирование изменения состояния ключа на противоположное.

Динамическая категория. Динамические объекты — это сообщения (транзакты). В процессе моделирования они создаются, порождают другие сообщения, собираются и уничтожаются. Каждому сообщению соответствует набор параметров, количество которых может быть установлено до 100. Если количество параметров не называется, то оно принимается равным 12. Сообщениям можно присваивать приоритет от 0 до 127; если приоритет не назван, то он принимается равным 0. С динамической категорией объектов связаны блоки, основные из которых можно подразделить на пять групп.

Группа задержки состоит из единственного блока ADVANCE, группа создания и уничтожения сообщений — из блоков GENERATE, TERMINATE, SPLIT, ASSEMBLE; группа изменения маршрутов сообщений — из блоков TRANSFER, LOOP, GATE, TEST. Блок TRANSFER имеет шесть основных режимов использования.

• **Примечание.** Блок GATE в устройствах использует следующие условия перехода: U — используется, NU — не используется, I — занято прерывающим сообщением, NI — не занято прерывающим сообщением

Блок GATE в памятьях использует следующие условия перехода SF — заполнена, SE — пуста, SNF — не заполнена, SNE — не пуста

Группа синхронизации сообщений включает в себя блоки MATCH и GATHER. Сопряженные блоки MATCH не допускают продвижения сообщения, поступившего первым, пока не поступило второе сообщение.

• **Примечание.** Блок GATHER задерживает сообщения до тех пор, пока не соберется указанное количество сообщений

Блоки ASSIGN, INDEX, MARK, PRIORITY составляют группу изменения атрибутов сообщений

Вычислительная категория. В вычислительной категории используются объекты трех видов: арифметические переменные, логические (булевские) переменные и функции. Арифметические переменные описываются блоком VARIABLE в режиме целых чисел и FVARIABLE в режиме с плавающей точкой. Название карты описывают арифметические действия над стандартными числовыми атрибутами (СЧА).

Аргументы и результаты рассматриваются как целые числа. При вычислении используется пять алгебраических операций: «+» (сложение); «—» (вычитание); «*» (умножение); «/» (деление с отбрасыванием остатка); деление на нуль не считается ошибкой и дает результат, равный нулю; «%» (деление по модулю, при котором частное отбрасывается и сохраняется остаток, который считается положительным).

Допускается использование не более пяти скобок.

Блок FVARIABLE описывает арифметические сочетания различных СЧА, вычисляемых в режиме с плавающей запятой. Он аналогичен блоку VARIABLE, за исключением того, что от аргументов, а также от промежуточных результатов не берется целая часть. Целая часть выделяется только от результата.

Блок BVARIABLE представляет собой логическое высказывание, состоящее из некоторой совокупности стандартных числовых или

логических атрибутов. При вычислении используется три типа операторов: логические, условные и булевские.

Логические операторы связаны с состоянием объектов оборудования, задаваемых стандартными логическими атрибутами СЧА, и при выполнении некоторых условий принимают значение «1».

Например, запись

BVARIABLE FN2

означает, что булевская переменная B3 принимает значение «1», если устройство 2 не обслуживает прерывания, и значение «0», если оно обслуживает прерывание.

Условные операторы производят алгебраическое сравнение СЧА и констант. Имеются следующие типы условных операторов: 'E' (равно); 'L' (меньше); 'G' (больше); 'NE' (не равно); 'LE' (меньше или равно); 'GE' (больше или равно).

Например, запись

2 BVARIABLE M1'LE'P6

означает, что B2 принимает значение «1», если время пребывания сообщения в системе меньше или равно числовому значению параметра 6. Мнемоника условия заключается в кавычки.

Булевские операторы: «+» (логическое сложение); «*» (логическое умножение).

Функции описываются с помощью блока FUNCTION. Основные типы функций — кусочно-линейная и кусочно-постоянная.

В случае если значение аргумента попадает между точками, то значение функции определяется методом линейной интерполяции:

$$Y_i = (Y_{i+1} - Y_i) \frac{r_i - X_i}{X_{i+1} - X_i}, \quad i = (r_i^*, n),$$

где r_i — j -я реализация равномерно распределенного случайного числа в интервале $[0,1]$; n — количество интервалов.

Пример:

3 FUNCTION RN1, C5
0,0/ 35,1 1/ 42,1 7/0 75,2 2/1 0,3 8/

Если значение аргумента попадает за точками, то значение функции определяется по значению функции, заданной правой границей интервала. Например,

2FUNCTION RN4, D3
3,11/ 7,24/1 0,40

Статистическая категория. В языке GPSS используется два типа статистических объектов: очереди и таблицы.

Очередь выделяется блоками QUEUE и DEPART. Частоты попадания заданного СЧА регистрируются блоком TABULATE. Характеристики таблиц вводятся блоком TABLE.

Для сохранения в модели числовой информации отводятся специальные ячейки основной памяти SAVEVALUE. Совокупности ячеек можно организовать в матрицы. Для записи информации в ячейки служит блок SAVEVALUE, в матрицы — MSAVEVALUE. Матрица описывается с помощью оператора MATRIX. Присвоение начальных значений ячейкам и матрицам осуществляется блоком INITIAL, с помощью которого можно устанавливать также логические ключи в состояние «1». Например, блок INITIAL □□ LS3 установит третий логический ключ в состояние «1».

Для обеспечения гибкости программных моделей можно использовать косвенную адресацию объектов. При записи обозначения * перед числом N индексный номер объекта определяется значением N-го параметра сообщения. Например, запись SEIZE □□ *4 означает поступление сообщения в устройство, номер которого определяется значением четвертого параметра; запись SAVEVALUE □□ *5 + 1 означает, что к переменной, номер которой определяется значением пятого параметра, прибавляется единица.

Группирующая категория. Из объектов группирующей категории рассмотрим блоки LINK и UNLINK, позволяющие переводить сообщения из списка текущих активных событий в списки пользователя неактивных событий и обратно. Использование списков пользователя позволяет моделировать различные дисциплины обслуживания сообщений.

2.8.2. Синтаксис элементов языка

Алфавит. Алфавит языка GPSS состоит из латинских букв от A до Z; цифр от 0 до 9 и специальных символов.

Русские буквы могут использоваться только в комментариях.

Числа. В языке GPSS различают два типа чисел: полные и действительные. Признаком действительного числа является десятичная точка. Числа могут занимать в памяти ЭВМ два байта (полусловные) и четыре байта (полнословные).

Идентификаторы. Они должны содержать не более пяти алфавитно-цифровых символов, причем первые три символа должны быть буквами. Идентификаторы используются для формирования имен объектов и блоков. Именованное объектов в GPSS может выполняться

двумя способами: в виде числового имени, когда для идентификации объектов используются целые числа, и в виде символического имени, т. е. идентификатора.

Стандартные числовые атрибуты. В процессе моделирования GPSS автоматически регистрирует и корректирует определенную информацию различных объектов, используемых в модели. Доступ к этой информации осуществляется с помощью СЧА (табл. 1).

Таблица 1

Объект	СЧА	Назначение
Блоки	N W	Счетчик входов Счетчик текущего содержимого
Время	CI	Значение относительного времени
Генераторы случайных чисел	RN	При использовании в качестве аргумента функции представляются дей-

Продолжение табл. 1

Объект	СЧА	Назначение
		ствительным числом в диапазоне 0,000000—0,999999, в остальных случаях — целым числом в диапазоне 000—999
Матрицы сохраняемых величин	MH (i, j) MX (i, j)	Величина элемента, стоящего в строке i и в столбце j; MH—полусловная матрица; MX—полнословная матрица
Многоканальные устройства	R S SA SC SR SM ST	Остающаяся емкость Текущее содержимое Среднее содержимое Счетчик числа входов Коэффициент использования Максимальное содержимое Среднее время задержки на единицу емкости
Очереди	Q QA QC QM QZ QT QX	Текущее содержимое Среднее содержимое Счетчик общего числа входов Максимальное содержимое Счетчик числа нулевых входов Среднее время пребывания (на основе QC) Среднее время пребывания (на основе QZ)
Переменные	BV V	Значение булевой переменной Значение арифметической переменной
Приборы	F FC FR FT	Состояние прибора (1 — занят, 0 — свободен) Счетчик числа занятий Коэффициент использования Среднее время задержки на одно занятие
Сохраняемые величины	X XH	Значение сохраняемой величины (X — полнословная, XH — полусловная)
Таблицы	TB TC TD	Среднее значение взвешенных входов Количество невзвешенных входов Стандартное отклонение невзвешенных входов
Транзакты	P PR MI	Значение параметра Значение приоритета Время пребывания в модели

Продолжение табл. 1

Объект	СЧА	Назначение
	MP	Время с момента входа в блок
Функции	FN	Значение функции
Цепи пользо- вателя	CA	Среднее содержимое
	CC	Общее число входов
	CH	Текущее содержимое
	CM	Максимальное содержимое
	CT	Среднее время пребывания в цепи на один вход

Для указания конкретного объекта, по которому необходимо получать требуемую информацию, за каждым из СЧА должно следовать числовое или символическое имя этого объекта. Если используется символическое имя, то между СЧА и именем ставится знак \square . Например,

R5 — оставшаяся емкость пятого устройства;

QQLIST — текущее содержимое очереди с именем LIST.

В GPSS можно использовать косвенную адресацию при задании конкретного объекта. В этом случае номер объекта определяется значением параметра текущего транзакта. Например,

QA*6 — среднее содержимое очереди, номер которой содержится в шестом параметре;

V*4 — значение арифметической переменной, номер которой содержится в четвертом параметре.

Правила записи программы. Программа, написанная на языке GPSS, записывается в соответствии со следующими правилами:

— поле имени блока или объекта занимает со 2-й по 6-ю колонки;

— название блока или оператора описания занимает с 8-й по 18-ю колонки;

— операнды занимают с 19-й по 71-ю колонки. Операнды разделяются запятыми.

Правила записи комментариев. Комментарии располагаются после описания операндов и отделяются от них одним или более пробелом. Если комментарии занимают всю строку, то в 1-й колонке должен стоять символ «*».

Описание блоков и операторов ввода. Блоки и операторы ввода описываются в таблицах, имеющих структуру (табл. 2).

Таблица 2

Операнд	Описание операнда	Значение операнда	Значение по умолчанию
1	2	3	4

Примечание. В колонке 3 используются следующие сокращения: К — целое число; И — символическое имя (идентификатор); СА — обозначает К; СЧА j, СЧА \square имя или СЧА * j.

В колонке 4 используются обозначения: ДНВ — действие, указанное в колонке 2, не выполняется; знак «*» — действие является обязательным.

Вычислительные средства

FUNCTION — оператор определения функции (табл. 3):

ИФ FUNCTION *Ap*, *KП*

Таблица 3

1	2	3	4
ИФ	Имя функции	И, К	—
<i>Ap</i>	Аргумент функции, используемый для получения значений функции	СЧА j, СЧА \square имя СЧА * j	—
<i>KП</i>	Количество пар координат, определяющих функцию, задается следующим образом: СК — для непрерывных функций ДК — для дискретных функций	СК ДК	— —

Оператор описания функции:

$$X_1, Y_1/X_2, Y_2/.../X_k, Y_k/$$

где X_i, Y_i — значения аргументов и функций.

Информация набирается оператором с 1-й по 71-ю колонку.

Пример задания непрерывной функции, состоящей из 24 значений, используемой для генерации случайных чисел в соответствии с экспоненциальным законом распределения со значением среднего, равным 1:

EXPON FUNCTION RNI, C24
 0 0/.1,.104.2..222/ 3,.355/.4,.509/ 5, 69/.6,.915/.7,1 2/.75,1 38
 .8,1.6/.84,1 83/.88,2.12/9.2 3/.92,2 52/.94,2.81/.95,2.81/.96,3.2
 .97,3.5/.98,3 9/.99,4 6/.995,5.3/.998,6.2/.999,7/.9998,8

Пример задания функции для генерации случайных чисел в соответствии с нормальным законом распределения, имеющим математическое ожидание 0 и стандартное отклонение 1:

NORM FUNCTION RNI, C25
 0,—5/.0003,—4/ 00135,—3/ 00621,—2.5/ 02275,—2
 06681,2—1 5/.11507,—1.2/ 15866,—1/.21186,—8/.27425,— 6
 34458,— 4/.42074,— 2/ 5.0/ 57926, 2/ 65542, 4
 .72575, 6/.78814, 8/ 84134,1/.88493,1 2/ 93319,1 5
 .97725,2/.99379,2 5/.99865,3/.99997,4/1,5

VARIABLE — оператор определения целой переменной (табл. 4)

III VARIABLE AB

Таблица 4

1	2	3	4
ИП	Имя переменной	И, К	—
AB	Арифметическое выражение, определяющее переменную		

Арифметическое выражение (AB) представляет собой комбинацию числовых данных и операций. В качестве числовых данных используются CA, в качестве операций — специальные знаки: + (плюс); — (минус); / (деление); * (умножение); @ (деление по модулю). Числовые данные, участвующие в арифметическом выражении, являются целыми.

Примеры определения переменных:

ANN VARIABLE R5 + Q2/5

TIM VARIABLE (P2—P4)/2

FRIT VARIABLE L + PR — P10/(SCFIX + 2)

FVARIABLE — оператор определения действительной переменной:

III FVARIABLE LB

Определение действительных переменных аналогично определению целых переменных VARIABLE, однако при вычислении AB используются числа с плавающей запятой. После вычисления AB его дробная часть отбрасывается.

BVARIABLE — оператор определения булевой переменной (табл. 5):

III BVARIABLE LB

Таблица 5

1	2	3	4
ИП	Имя переменной	И, К	—
ЛВ	Логическое выражение		

Логическое выражение (ЛВ) представляет собой комбинацию числовых данных, операторов отношения, логических операций и логических атрибутов. В качестве числовых данных используются CA; в качестве операторов — отношения:

'G' (больше), 'GE' (больше или равно);
 'L' (меньше), 'LE' (меньше или равно);
 'E' (равно), 'NE' (не равно);

в качестве логических операций:

+ (логическое сложение), * (логическое умножение);

в качестве логических атрибутов:

LR — логической переключатель сброшен, LS — установлен;

FU — прибор используется, FNU — не используется;

FI — находится в состоянии прерывания; FNI — не находится в состоянии прерывания;

SE — многоканальное устройство пусто, SNE — не пусто; SF — многоканальное устройство заполнено, SNF — не заполнено.

Примеры определения переменных:

BLUE BVARIABLE RCHANL 'GE'3
 THETA BVARIABLE C1'GE'500 + (SE SPA * LS @AMB)

Если булева переменная истинна, то ее значение равно 1, если ложна, то 0.

SAVEVALUE — оператор изменения сохраняемых величин (табл. 6):

SAVEVALUE ИВ [±], C_n, T

Таблица 6

1	2	3	4
ИВ	Имя изменяемой сохраняемой величины. Если после ИВ стоит знак «+», то значение ИВ увеличивается на C _n ; если указан знак «—», то ИВ уменьшается на C _n ; если знак не указан, то ИВ присваивается значение C _n .	И CA	
C _n	Параметр, используемый для модификации сохраняемой величины	CA	
T	Тип сохраняемой величины (Н — полусловная)	Н	Полнословная

MATRIX — оператор описания матриц (табл. 7):

MATRIX *ИМ* *T*, *K*_{стр}, *K*_{стб}

Таблица 7

1	2	3	4
<i>ИМ</i>	Имя матрицы	М, К	—
<i>T</i>	Тип матрицы (X — полнословная, Н — полусловная)	Х, Н	—
<i>K</i> _{стр}	Количество строк в матрице	К	—
<i>K</i> _{стб}	Количество столбцов в матрице	К	—

MSAVEVALUE — блок изменения значения элемента матриц (табл. 8):

MSAVEVALUE *ИМ*, *N*_{стр}, *N*_{стб}, *C*_в, *T*

Таблица 8

1	2	3	4
<i>ИМ</i>	Имя матрицы. Если после <i>ИМ</i> стоит знак «+», то значение элемента матрицы увеличивается на <i>C</i> _в ; если знак «-», то значение элемента матрицы уменьшается на <i>C</i> _в ; если знак опущен, то элементу матрицы присваивается значение <i>C</i> _в .	И, СА	—
<i>N</i> _{стр}	Номер строки матрицы	СА	—
<i>N</i> _{стб}	Номер столбца матрицы	СА	—
<i>C</i> _в	Величина, используемая для изменения значения элемента матрицы	СА	—
<i>T</i>	Тип матрицы	Н	Полнословная

LOGIC — блок изменения логических переключателей (табл. 9):

LOGIC *X* *ИП*_л

Таблица 9

1	2	3	4
<i>ИП</i> _л	Имя логического переключателя	И, СА	—
<i>X</i>	Оператор действия, который может принимать следующие значения: R — сбросить логический переключатель, S — установить логический переключатель, I — инвертировать логический переключатель	R S I	—

INITIAL — оператор установки начальных значений

Установка значений сохраняемых величин (табл. 10):

INITIAL *И*₁, *З*₁/*И*₂, *З*₂/ *И*_н, *З*_н

Таблица 10

1	2	3	4
<i>И</i> ₁	Имя сохраняемой величины (X — полнословная, Н — полусловная)	Х, ХН, Х □ имя ХН □ имя	—
<i>З</i> ₁	Начальное значение (может быть как положительное, так и отрицательное)	+К, -К	—

Установка значений матриц (табл. 11):

INITIAL *И*₁(*N*_{стр}¹, *N*_{стб}¹), *З*₁/ *И*_н(*N*_{стр}^н, *N*_{стб}^н), *З*_н

Таблица 11

1	2	3	4
<i>И</i> ₁	Имя матрицы (MX — полнословная, MN — полусловная)	MX, MN, MX □ имя MN □ имя	—
<i>N</i> _{стр} ¹ <i>N</i> _{стб} ¹ <i>З</i> ₁	Номер строки матрицы Номер столбца матрицы Начальное значение	К К +К, -К	— — —

Вместо *N*_{стр} и *N*_{стб} можно задавать диапазон их изменения, элементы которого имеют одинаковые значения:

INITIAL *MX* *З*(1, 1-5), *4/MN* □ *РЕТЕ* (1-5, 1-10), -1

Установка значений логических переключателей (табл. 12):

INITIAL *И*₁/*И*₂/ *И*_л/ *И*_н

Таблица 12

1	2	3	4
<i>И</i> ₁	Имя логического переключателя	LS, LS □ имя	—

Логические переключатели, перечисленные в данной операторе, устанавливаются в положение «Установлено», остальные переключатели находятся в положении «Сброшено».

Генерация и уничтожение транзактов

GENERATE — блок генерации транзактов (табл. 13):

GENERATE $t_{cp}, t_m, t_n, K_T, Pr, K_n, PP$

Таблица 13

1	2	3	4
t_{cp}	Среднее значение интервала времени между последовательными прохождениями транзактов	CA	0
t_m	Разброс интервала времени относительно t_{cp} . Если в качестве t_m используется функция, то интервал времени между последовательными порождениями транзактов $\Delta t = t_{cp}f(RN),$ где f — функция, RN — случайное число в интервале $[0, 1]$ В остальных случаях интервал времени вычисляется в соответствии с равномерным законом распределения в интервале $[t_{cp} - t_m, t_{cp} + t_m]$	CA	0
t_n	Момент времени, в который должен появиться первый транзакт	CA	1

Продолжение табл. 13

1	2	3	4
K_T	Количество транзактов, которое должно быть сгенерировано, после чего генерация транзактов прекращается	CA	∞
Pr	Значение приоритета генерируемых транзактов, которое может лежать в диапазоне 0...127, причем большее значение соответствует более высокому приоритету	CA	0
K_n	Количество параметров транзакта (≤ 100)	CA	12
PP	Размер памяти, отводимый под один параметр транзакта, равен 4 байтам	F	2 Байта

TERMINATE — блок уничтожения транзактов (табл. 14):

TERMINATE N_3

Таблица 14

1	2	3	4
N_3	При уничтожении транзактов вычисляется $Sч = Sч - N_3$, где $Sч$ — счетчик завершенный (задается в блоке START). При $Sч = 0$ моделирование прекращается	CA	0

Изменение параметров транзактов

ASSLNG — блок изменения значений параметров (табл. 15):

Таблица 15

1	2	3	4
N_4	Номер параметра транзакта, подлежащего изменению. Если задан знак «+», то к параметру транзакта прибавляется целая часть значения (PF_n), если задан «-», то вычитается; если знак не указан, то параметру транзакта присваивается значение P	CA	—
P	Величина, используемая для изменения значения параметра транзакта	CA	—
F	Имя функции, используемой для модификации величины	CA	1

PRIORITY — блок и изменения приоритета (табл. 16):

PRIORITY Pr, B

Таблица 16

1	2	3	4
Pr	Значение приоритета присваиваемое транзакту	CA	—
B	При наличии данного операнда интерпретатор переставляет транзакт в цепи текущих событий так, что он оказывается в конце нового приоритетного класса, и снова начинает просмотр цепи текущих событий	BUFFER	ДНВ

Ансамбли транзактов

SPLIT — блок расщепления транзактов (табл. 17):

SPLIT K_T, IB, N_T, K_n

Таблица 17

1	2	3	4
K_T	Число дополнительно порождаемых транзактов — «потомков»	СА	—
<i>ИБ</i>	Имя блока, куда будут направлены транзакты-«потомки»; транзакт-«родитель» поступает в следующий блок	СА	—
N_T	Номер параметра транзакта-«родителя» и транзакта-«потомков», которые принимают следующие значения: у транзакта-«родителя» значение N_T увеличивается на 1, т. е. $N_T^{род} = N_T^{род} + 1$ у 1-го «потомка»; $N_T^{1пот} = N_T^{род} + 2$ у 2-го «потомка»; $N_T^{2пот} = N_T^{род} + 3$ у третьего поколения и т. д.	СА	ДНВ
K_{II}	Число параметров, которое должен иметь каждый потомок	СА	Равно числу параметров родителей

ASSEMBLE — блок соединения транзактов (табл. 18):
ASSEMBLE K_T

Таблица 18

1	2	3	4
K_T	Количество членов ансамбля, объединяемых в один транзакт. Когда количество вошедших членов ансамбля станет равно K_T , транзакт, прибывший первым, выходит из блока ASSEMBLE	СА	—

GATHER — блок сбора транзакта (табл. 19):
GATHER K_T

Таблица 19

1	2	3	4
K_T	Количество членов ансамбля, накапливаемых в данном блоке. Когда количество вошедших членов ансамбля станет равно K_T , они выходят из блока GATHER	СА	—

MATCH — блок синхронизации транзактов (табл. 20):
MATCH *ИБ*

Таблица 20

1	2	3	4
<i>ИБ</i>	Имя блока, сопряженного с данным. Если сопряженный блок содержит транзакт, являющийся членом ансамбля, к которому принадлежит транзакт, вошедший в блок MATCH, то оба транзакта продолжают движение. В противном случае транзакт задерживается в блоке MATCH	И, СА	—

ADVANCE — блок задержки транзактов (табл. 21):
ADVANCE t_{cp}, t_m

Таблица 21

1	2	3	4
t_{cp}	Среднее значение интервала времени, на которое задерживается транзакт	СА	0
t_m	Разброс интервала времени относительно t_{cp} ; вычисляется, как и в блоке GENERATE	СА	0

Приборы

SEIZE — блок занятия прибора (табл. 22):
SEIZE *ИП*

Таблица 22

1	2	3	4
<i>ИП</i>	Имя прибора, подлежащего занятию транзакта	И, СА	

RELEASE — блок освобождения прибора (табл. 23):
RELEASE *ИП*

Таблица 23

1	2	3	4
<i>ИП</i>	Имя освобождаемого прибора	И, СА	

PREEMT — блок захвата приборов (табл. 24):
PREEMT *ИП, Пр, ИБ, N_T, AB*

Таблица 24

1	2	3	4
<i>ИП</i>	Имя захватываемого прибора	СА, И	

Продолжение табл. 24

1	2	3	4
<i>Пр</i>	Условие захвата прибора. Если указан параметр <i>PR</i> , то захват прибора происходит при условии, что вновь поступающий транзакт имеет более высокий приоритет	<i>PR</i>	Захват без учета приоритета
<i>ИБ</i>	Имя блока, в который будет послан прерванный транзакт	СА, И	ДНВ
<i>N_t</i>	Номер параметра прерванного транзакта, в который помещается значение времени, оставшегося транзакту до окончания обслуживания на приборе	СА	ДНВ
<i>АВ</i>	Если задан параметр <i>RE</i> , то прерванный транзакт теряет право на автоматическое восстановление обработки в приборе	<i>RE</i>	ДНВ

RETURN — блок возврата захваченного прибора (табл. 25):
RETURN *ИП*

Таблица 25

1	2	3	4
<i>ИП</i>	Имя освобождаемого прибора, т. е. возврат прибора ранее прерванному транзакту	И, СА	

Многоканальные устройства

STORAGE — карта описания емкости устройства (табл. 26):
ИУ STORAGE E_м

Таблица 26

1	2	3	4
<i>ИУ</i>	Имя многоканального устройства	И, К	
<i>E_м</i>	Емкость многоканального устройства	К	

ENTER — блок входа в устройство (табл. 27):
ENTER *ИУ, E_м*

Таблица 27

1	2	3	4
<i>ИУ</i>	Имя многоканального устройства	И, СА	
<i>E_м</i>	Занимаемая емкость устройства	СА	И

LEAVE — блок выхода из устройства (табл. 28):
LEAVE *ИУ, E_м*

Таблица 28

1	2	3	4
<i>ИУ</i>	Имя многоканального устройства	И, СА	
<i>E_м</i>	Освобождаемая емкость устройства	СА	И

Блоки проверки условий

TRANSFER — блок передачи транзактов
Безусловный режим (табл. 29):
TRANSFER *ИБ*

Таблица 29

1	2	3	4
<i>ИБ</i>	Имя блока, в который переходит транзакт	И, СА	—

Статический режим (табл. 30):
TRANSFER *ВП, ИБ₁, ИБ₂*

Таблица 30

1	2	3	4
<i>ВП</i>	Вероятность передачи транзакта на <i>ИБ₂</i>	СА	—
<i>ИБ₁</i>	Имя блока, куда передаются транзакты с вероятностью (1- <i>ВП</i>)	И, СА	Переход в следующий блок
<i>ИБ₂</i>	Имя блока, куда передаются транзакты с вероятностью <i>ВП</i>	СА, И	—

Условный режим (табл. 31... 34):

TRANSFER BOTH, ИБ₁, ИБ₂

Таблица 31

1	2	3	4
BOTH	Задаёт режим, при котором транзакт сначала пытается войти в блок с именем ИБ ₁ . Если вход невозможен, то транзакт пытается войти в блок с именем ИБ ₂ . Если транзакт не может войти и в этот блок, то транзакт остаётся в блоке TRANSFER	BOTH	—
ИБ ₁	Имя блока	И, СА	—
ИБ ₂	Имя блока	И, СА	—

TRANSFER P, N_T, C_M

Таблица 32

1	2	3	4
P	Задаёт режим, при котором номер блока, куда передаётся транзакт, определяется значением N _T + C _M	P	—
N _T	Номер параметра транзакта	СА	—
C _M	Смещение	СА	0

TRANSFER SBR, ИБ, N_T

Таблица 33

1	2	3	4
SBR	Задаёт режим, при котором транзакт переходит в блок с именем ИБ, а номер самого блока TRANSFER помещается в N _T	SBR	—
ИБ	Имя блока, куда передаётся транзакт	И, СА	—
N _T	Номер параметра транзакта	СА	—

TRANSFER ALL, H, B, Ш

Таблица 34

1	2	3	4
ALL	Задаёт режим, при котором транзакт последовательно пытается войти в блоки с номерами	ALL	—
H	Первый блок, подлежащий проверке	И, К	—
B	Последний блок, подлежащий проверке	И, К	—
Ш	Значение шага	К	—

Если транзакт не может перейти ни к одному из указанных блоков, он остаётся в блоке TRANSFER.

SELECT — блок выбора элементов (табл. 35)

Режим отношения

SELECT X, N_T, MIN, MAX, C_p, C_a, ИБ

Таблица 35

1	2	3	4
N _T	Номер параметра транзакта, в который записывается номер члена группы, соответствующего заданному условию	СА	—

Продолжение табл. 35

1	2	3	4
MIN MAX	Наименьший и наибольший номера из множества членов просматриваемой группы	И, СА	—
C _p	Значение, с которым должно сравниваться значение	СА	—
C _a	Имя просматриваемого множества элементов	И, СА, кроме MX, MN	—
ИБ	Имя блока, в который передаётся транзакт, если ни один элемент просматриваемого множества не отвечает заданному условию	И, СА	Транзакт переходит в следующий блок
	Оператор отношения задаёт условия выбора элемента: G — C _a > C _p , LE — C _a ≤ C _p ; GE — C _a ≥ C _p , L — C _a < C _p ; E — C _a = C _p , NE — C _a ≠ C _p	G GE E NE LE E	

Режим работы минимального или максимального элемента (табл. 36):

SELECT X N_T , MIN, MAX, C_a

Таблица 36

1	2	3	4
N_T , MIN MAX C_a	Имеют тот же смысл, что и в режиме отношения	И, CA	—
X	Если MIN, то ищется элемент с минимальным значением C_a ; если MAX, то с максимальным значением C_a		—

Логический режим (табл. 37):

SELECT X N_b , MIN, MAX, ИБ

Таблица 37

1	2	3	4
N_T , MIN MAX ИБ	Имеют те же значения, что и в режиме отношения	И, CA	—
X	Логический указатель, задающий условие, которое должно выполняться, может принимать следующие значения: LS — логический переключатель установлен LR — логический переключатель сброшен U — прибор используется NU — прибор не используется SF — многоканальное устройство заполнено SNF — многоканальное устройство не заполнено SE — многоканальное устройство пусто SNE — многоканальное устройство не пусто I — на приборе произошло прерывание NI — на приборе не произошло прерывания	LS LR U NU SF SNF SE SNE I NI	—

COUNT — блок подсчета элементов (табл. 38):
COUNT X N_b , MIN, MAX, C_p , C_a

Таблица 38

1	2	3	4
N_T	Номер параметра транзакта, в который заносится количество элементов, удовлетворяющих данному условию	CA	—

Отдельные параметры имеют то же значение, что и в блоке SELECT:

TEST — блок сравнения атрибутов (рис. 39):

TEST X A_1 , A_2 , ИБ

Таблица 39

1	2	3	4
A_1	Имя первого стандартного атрибута	CA	—
A_2	Имя второго стандартного атрибута	CA	—
ИБ	Имя блока, в который передается транзакт, если условие сравнения не выполняется	И, CA	Проверка в режиме отказа
X	Оператор основания задает операцию сравнения G — $A_1 > A_2$, NF — $A_1 \neq A_2$ GE — $A_1 \geq A_2$, LE — $A_1 \leq A_2$ E — $A_1 = A_2$, L — $A_1 < A_2$	G GE E NE LE L	

GATE — блок проверки состояния элементов

Проверка состояния логических переключений (табл. 40).

GATE X ИП, ИБ

Таблица 40

1	2	3	4
ИП	Имя логического переключателя	И, CA	—
ИБ	Имя блока, в который переходит транзакт, если проверяемое условие не выполняется	И, CA	Проверка в режиме отказа
X	Логический указатель задает условие проверки LS — логический указатель «установлен» LR — логический указатель «сброшен»	LS	—

Проверка многоканальных устройств и приборов (табл. 41):
GATE X ИУ, ИБ

Таблица 41

1	2	3	4
ИУ	Имя многоканального устройства	И, СА	
ИБ	Имя блока, в который переходит транзакт, если проверяемое условие не выполняется	И, СА	Проверка в режиме отказа
X	Логический указатель задает условие проверки U — проверка занятости прибора NU — проверка незанятости прибора SF — проверка заполненности многоканального устройства SNF — проверка незаполненности многоканального устройства SE — проверка того, что многоканальное устройство пусто SNE — проверка того, что многоканальное устройство не пусто I — проверка прибора на использование захватчиком NI — проверка прибора на неиспользование захватчиком	U NU SF SNF SE SNE I NI	

Проверка состояния парности (табл. 42)
GATE X ИБ₁, ИБ₂

Таблица 42

1	2	3	4
ИБ ₁	Имя блока, в котором ищется транзакт, находящийся в состоянии парности с вошедшим	И, СА	
ИБ ₂	Имя блока, в который переходит транзакт, если проверяемое условие не выполняется	И, СА	Проверка в режиме отказа
X	Логический указатель задает условие проверки M — находится ли транзакт в состоянии парности NM — отсутствует ли в блоке член ансамбля, находящийся в состоянии парности	M NM	

LOOP — блок организации цикла (табл. 43):
LOOP N_T, ИБ

Таблица 43

1	2	3	4
N _T	Номер параметров транзакта, значение которого используется для организации количества повторений (параметр цикла). При входе транзакта в данный блок N _T уменьшается на 1. Если N _T становится равным 0, то транзакт переходит в следующий блок, иначе транзакт переходит в блок с именем ИБ	СА	—
ИБ	Имя блока, в который переходит транзакт, если N _T ≠ 0	И, СА	—

Стандартная статистика. В процессе выполнения программы модели интерпретатор собирает стандартную статистическую информацию, которая автоматически распечатывается по окончании моделирования.

По **приборам** печатается следующая информация: имя прибора; загрузка; число обслуживаний; среднее время одного обслуживания; номер транзакта, находящегося на обслуживании в данный момент времени; номер транзакта, который захватил прибор в данный момент.

По **очередям** печатается информация: имя очереди; максимальная длина очереди за время моделирования; средняя длина очереди; общее число входов в очередь; общее число входов в очередь без последующего ожидания (нулевые входы); процент нулевых выходов; среднее время пребывания в очереди с учетом всех входов и без учета нулевых входов; имя таблицы, содержащей распределение времени пребывания; текущее значение длины очереди.

По **многоканальным устройствам** печатается информация: имя многоканального устройства; емкость; среднее содержимое; средняя загрузка; общее число занятий устройства; среднее время пребывания; текущее и максимальное содержимое.

Статистика по очередям

QUEUE — блок занятия очереди (табл. 44):
QUEUE ИО, К

Таблица 44

1	2	3	4
ИО	Имя очереди	И, СА	
К	Количество мест в очереди, занимаемое транзактом	СА	I

DEPART — блок освобождения очереди (табл. 45):

DEPART ИО, К

Таблица 45

1	2	3	4
ИО	Имя очереди	И, СА	
К	Количество мест в очереди, освобождаемое транзактом	СА	1

Построение гистограмм

TABLE — оператор описания таблицы (табл. 46):

ИТ TABLE ИП, Н_л, Ш, КИ, ВИ

Таблица 46

1	2	3	4
ИТ	Имя таблицы	И, К	—
ИП	Имя переменной, значение которой табулируется. Если указан параметр IA, то осуществляется построение гистограмм интервалов времени между моментами поступления транзактов в данную точку. Если указан параметр RT, то осуществляется построение гистограммы интенсивности прихода транзакта в данную точку, причем интенсивность определяется относительно временного интервала	СА	—

Продолжение табл. 46

1	2	3	4
Н _л	Левая граница первого интервала таблицы	К	—
Ш	Ширина интервалов таблицы	К	—
КИ	Количество интервалов таблицы, увеличенное на 2 (в первый интервал попадают значения ИП от $-\infty$ до Н _л , а в последний интервал — от Н _л + КИ·Ш-2 до $+\infty$). Если перед значением К ставится символ W, то таблица является взвешенной	К, WK	—
ВИ	Временной интервал для параметра RT	К	—

QTABLE — оператор описания таблицы времени пребывания в очереди (табл. 47):

ИТ QTABLE ИО, Н_л, Ш, КИ

Таблица 47

1	2	3	4
ИТ	Имя таблицы	И, К	—
ИО	Имя очереди, для которой строится распределение времени пребывания транзактов в очереди	И, К	—

Остальные параметры имеют те же значения, что и в операторе TABLE.

MARK — блок отметки (табл. 48):

MARK N_t

Таблица 48

1	2	3	4
N _t	Номер параметра транзакта, в который заносится момент времени входа транзакта в данный блок	СА	—

TABULATE — блок табулирования (табл. 49):

TABULATE ИТ, ВК

Таблица 49

1	2	3	4
ИТ	Имя таблицы, в которую заносится табулируемая величина, указанная в карте TABLE, в момент входа транзакта в данный блок	И, СА	—
ВК	Весовой коэффициент, задающий число раз, которое табулируемая величина должна занести в таблицу при каждом входе в данный блок. При использовании данного параметра в карте TABLE параметр К должен быть задан как WK	СА	1

Цепи пользователя

LINK — блок ввода транзакта в цепь пользователя (табл. 50):

LINK ИЦ, К_p, ИБ

Таблица 50

1	2	3	4
ИЦ	Имя цепи пользователя	И, СА	—
Кр	Критерий присоединения транзакта к цепи пользователя: K_p может принимать следующие значения: FITO — встать в конце цепи LITO — встать в начало цепи P_j — войти в цепь непосредственно перед транзактом с большим значением j -го параметра	FITO LITO P_j	
ИБ	Имя блока, куда переходит транзакт, если он не присоединяется к цепи пользователя	И, СА	Транзакт всегда присоединяется к цепи

UNLINK — блок вывода транзакта из цепи пользователя (табл. 51):

UNLINK ИЦ, ИБ, C_t , H_n , P_p , ИБ₂

Таблица 51

1	2	3	4
ИЦ	Имя цепи пользователя	И, СА	
ИБ ₁	Имя блока, в который переходит выведенные из цепи транзакты	И, СА	—
C_t	Число выводимых транзактов. Если задан параметр ALL, то выводятся все транзакты, отвечающие приведенным ниже условиям	СА ALL	
H_n , P_p	Определяют условия вывода транзактов из цепи пользователя	(см. табл. 5.52)	
ИБ ₂	Имя блока, куда переходит транзакт — инциатор вывода, если из цепи не выводится ни один транзакт	И, СА	Транзакт всегда переходит в следующий блок

Таблица определения условий вывода транзактов (табл. 52):

Таблица 52

H_n	P_p	Условия вывода
Не задан	Не задан	Транзакты выводятся из начала цепи
BACK	»	Транзакты выводятся из конца цепи
СА	»	Из цепи выводятся те транзакты, у которых значение параметра (номер которого задан H_n) равно значению того же параметра транзакта-инициатора
СА	СА	Из цепи выводятся те транзакты, у которых значение (номер которого задан H_n) равно P_p
BV_j $BV_{имя}$	Не задан	Из цепи выводятся те транзакты, для которых булевская переменная принимает значение «истина» при ее подсчете с использованием приоритета и параметров этих транзактов

Служебные операторы

SIMULATE — оператор моделирования

Оператор SIMULATE должен быть первым оператором программы модели. Если данный оператор не выполняется, то выполнение модели не производится.

END — оператор конца программы

Оператор END должен быть последним оператором программы модели.

START — оператор начала моделирования (табл. 53):

START \sqsubset Сч, ПП, N_n , ПЦ

Таблица 53

1	2	3	4
<i>Сч</i>	Начальное значение счетчика числа завершений	К	
<i>ПП</i>	Признак подавления печати. Если задан параметр <i>NP</i> , то стандартная печать в конце моделирования производиться не будет		Печать стандартной информации
<i>N_n</i>	Задаёт число завершений, через которое будет выдаваться промежуточная печать	К	ДНВ
<i>ПЦ</i>	Признак печати цепей. Печать цепей производится, если на месте <i>ПЦ</i> стоит 1	1	ДНВ

2.8.3. Типовые конструкции программ

На примерах простых имитационных моделей, приведенных ниже, рассмотрим типовые конструкции программ на языке GPSS, которые наиболее часто встречаются при построении моделей сложных технических систем.

Пример 1. Процесс прохождения заявок, поступление которых подчиняется равномерному закону с интервалом 8 ± 2 единицы времени, а обработка — равномерному закону со средним временем 5 ± 3 единицы, можно описать программой:

```

SIMULATE
GENERATE      8,2   генерация транзактов
SEIZE         SER   занять прибор
ADVANCE       5,3   задержать транзакт
RELEASE       SER   освободить прибор
TERMINATE     1     уничтожить транзакт
START         100   окончить моделирование
END           после обработки 100 транзактов
    
```

Пример 2. Пусть в условии примера 1 требуется получить статистические данные об очереди заявок, ожидающих обслуживания в приборе:

```

SIMULATE
GENERATE      8,2
QUEUE         JOB
SEIZE         SER
DEPART        JOB
ADVANCE       5,3
RELEASE       SER
TERMINATE     1
START         100
END
    
```

Пример 3. На обработку поступает два потока заявок: первый — с интервалом 10 ± 1 единица, второй 25 ± 2 единицы. Второй поток прерывает обработку заявок первого. Время обработки заявок первого потока 7 ± 3 единицы, второго 15 ± 1 единица:

```

SIMULATE
GENERATE      10,1
SEIZE         SER
ADVANCE       7,3
RELEASE       SER
TERMINATE     1
GENERATE      25,2
PREEMPT       SER
ADVANCE       15,1
RETURN        SER
TERMINATE     1
START         1
END
    
```

Пример 4. Пусть в условиях примера 1 необходимо произвести параллельную обработку заявки двумя приборами:

```

SIMULATE
GENERATE      8,2
SPLIT         1,THIS
SEIZE         SER1
ADVANCE       5,3
RELEASE       SER1
TRANSFER     ,THAT
    
```

```

THIS SEIZE      SER2
  ADVANCE      5,3
  RELEASE      SER2
THAT ASSEMBLE   2
  TERMINATE    1
  START        100
  END
    
```

Пример 5. Пусть в условиях примера 1 сообщения могут обрабатываться на одном из двух приборов; на первом — со временем 5 ± 3 единицы, на втором — 7 ± 2 единицы. Причем предпочтительнее обработка на первом приборе:

```

SIMULATE
GENERATE      8,2
TRANSFER     BOTH,THIS,THAT
THIS SEIZE    SER1
  ADVANCE    5,3
  RELEASE    SER1
  TRANSFER   ,HERE
THIS SEIZE    SER2
  ADVANCE    7,2
  RELEASE    SER2
HERE TERMINATE 1
  START      100
  END
    
```

Пример 6. Пусть в условиях примера 5 заявки поступают на обработку к блоку THIS с вероятностью 0,7 и к блоку THAT с вероятностью 0,3;

```

SIMULATE
GENERATE      8,2
TRANSFER     3,THIS,THAT
THIS SEIZE    SER1
  ADVANCE    5,3
  RELEASE    SER1
  TRANSFER   ,HERE
THIS SEIZE    SER2
  ADVANCE    7,2
  RELEASE    SER2
  TERMINATE  1
  START      100
  END
    
```

Пример 7. Пусть в условиях примера 1 заявки поступают на обработку с ограниченным числом мест в очереди, равным 2. Если очередь заполнена, то заявки покидают систему:

```

SIMULATE
MEM STORAGE   2
GENERATE      8,2
GATE SNF      1,FINAL
ENTER         MEM
SEIZE         SER
LEAVE         MEM
ADVANCE       5,3
RELEASE       SER
FINAL TERMINATE 1
  START       100
  END
    
```

Пример 8. На обработку поступает поток заявок с интервалом 10 ± 1 единица времени. Нечетные заявки обрабатываются на первом приборе со временем 15 ± 2 единицы, четные заявки — на втором приборе со временем 17 ± 3 единицы:

```

SIMULATE
GENERATE      10,1
LOGIC I       LOCK
GATE LS       LOCK,THAT
SEIZE         SER1
RELEASE       SER1
TERMINATE     1
THAT SEIZE    SER2
  ADVANCE    17,3
  RELEASE    SER2
  TERMINATE  1
  START      100
  END
    
```

Пример 9. Пусть в условиях примера 1 каждая заявка проходит 5 циклов обработки на приборе со временем 1 ± 1 единица:

```

SIMULATE
GENERATE      8,2
ASSIGN        2,5
SEIZE         SER
WAIT ADVANCE  1,1
  LOOP        2,WAIT
    
```

```
RELEASE SER
TERMINATE 1
START 100
END
```

Если некоторую последовательность блоков GPSS приходится использовать в нескольких местах программной модели, то целесообразно эту последовательность блоков оформить в виде макроопределения. В этом случае необходимо использовать управляющие операторы STARTMACRO (начало макро) и ENDMACRO (конец макро).

Следует помнить, что макроопределение не является подпрограммой, а используется только для упрощения кодирования часто повторяющихся последовательностей блоков. При этом операнды блоков, входящих в макроопределение, могут иметь фиктивные параметры, например #A, #B и т. д.

После определения макро, в те места программной модели, в которые должна быть вставлена последовательность блоков, составляющих макроопределение, необходимо поставить только один оператор MACRO с фактическими параметрами. В процессе работы интерпретатора, когда макроопределение подставляется на место оператора MACRO, фиктивные параметры операндов блоков, входящих в макро, заменяются фактическими, указанными в соответствующем макровывозе.

Макровывоз, т. е. оператор MACRO, имеет формат

ИМЯ MACRO A, B, C, ...

где ИМЯ — имя макроопределения (должно содержать не более пяти символов, причем первые три обязательно должны быть буквами латинского алфавита); MACRO — служебное слово; A, B, C, ... — значения операндов A, B и т. д.; количество операндов не должно превышать десяти.

Имя макроопределения записывается со 2-й позиции, служебное слово — с 8-й, а значения операндов — с 19-й. Операнды отделяются друг от друга запятыми.

Рассмотрим пример использования макросредств.

Пример 10. На языке GPSS требуется написать участок программной модели, имитирующей работу пяти последовательно соединенных одноканальных систем массового обслуживания (СМО). Время обслуживания требований в каждой СМО подчиняется экспоненциальному закону со средним значением 8...14 единиц времени соответственно. Для упрощения кодирования программы используется макро. В

качестве макроопределения выделяем набор блоков GPSS, описывающих работу одной СМО. Присвоим макроопределению имя SER:

```
SER STARTMACRO
    QUEUE #A
    SEIZE #A
    DEPART #A
    ADVANCE #B, FN$XPDIS
    RELEASE #A
ENDMACRO
```

где #A — параметр, определяющий номер СМО (очереди и обслуживающего аппарата), #B — параметр, определяющий значение среднего времени обслуживания требований.

Записываем участок программной модели, имитирующей работу пяти одноканальных СМО, с помощью макросредств языка GPSS.

```
SER MACRO 1,8
SER MACRO 2,9
SER MACRO 3,10
SER MACRO 4,11
SER MACRO 5,14
```

Здесь значение операнда A соответствует номеру СМО, а операнда B — среднему времени обслуживания требований в этой СМО. В результате работы интерпретатора выделенные пять MACRO в программе модели заменим на последовательность из 25 блоков GPSS. При этом фиктивные значения операндов блоков в макрорасширении будут заменены на фактические. Участок программы имеет вид

```
QUEUE 1
SEIZE 1
DEPART 1
ADVANCE 8, FN$XPDIS
RELEASE 1
QUEUE 2
SEIZE 2
DEPART 2
ADVANCE 9, FN$XPDIS
RELEASE 2
```

Система моделирования GPSS использует процессную концепцию описания функционирования сложной системы. Элементарные операторы называются в этой системе блоками и выполняют узкоспециализированные функции. Типичный объект моделирования — ресурсные

системы с дискретным временным множеством измерения состояний. В языке GPSS открыт доступ пользователя к инициаторам — транзактам. Блоки выполняют функции: вычислительные, проверки условий, навигационные, статистические, управляющие. Взаимодействие между блоками производится через транзакты, а также стандартные числовые атрибуты. Приводятся типовые программные конструкции для описания наиболее часто встречающихся ситуаций в практике моделирования.

Приложения

Приложение 1

Управление процессом моделирования. При исследовании различных технических систем методом имитационного моделирования возникает необходимость выполнения многократных прогонов модели с изменением ее параметров. В GPSS имеется возможность организовать многократное выполнение модели за один этап пакетной обработки заданий. Причем, от прогона к прогону можно выполнять следующие изменения исходной модели: переопределять операнды одного или нескольких блоков модели без изменения типов самих блоков; переопределять типы блоков; расширять модель за счет добавления одного или более независимых сегментов модели.

После модификации модели, которая задается за оператором START, необходимо подготовить модель к повторному прогону с помощью управляющего оператора CLEAR. Он осуществляет обнуление сохраняемых величин и всей статистической информации, собранной на предыдущем прогоне, очистку цепей и удаление из модели всех транзактов.

Оператор обнуления. Формат оператора CLEAR:

CLEAR C₁, C₂, ..., C_i, ...

где C_i — имена сохраняемых величин, которые не подлежат обнулению.

Если имена не заданы, то обнуляются все сохраняемые величины.

Оператор сброса статистики. Начальные условия моделирования могут заметно отличаться от тех условий, которые имеют место при достижении системой стационарного состояния, поэтому статистическая информация, собираемая в переходный период, может оказать существенное влияние на результаты, получаемые в конце моделирования. Для исключения переходного периода можно поступить таким образом:

- 1) выполнить моделирование до возникновения стационарных условий работы системы,
 - 2) сбросить собранную до этого времени статистику без изменения состояния модели;
 - 3) продолжить моделирование, собирая статистику, на которую уже не влияет переходный процесс.
- Сбросить собранную статистику можно с помощью управляющего оператора RESET. Формат оператора:

RESET A₁, A₂, ..., A_i, ...

где A_j — мнемонические имена объектов, на которые действие оператора не распространяется.

Задание таких объектов осуществляется следующими мнемоническими обозначениями: F_i — устройства, O_i — очереди, S_i — памяти, TV_i — таблицы, CH_i — списки пользователя, где i — номер соответствующего объекта.

Если имена объектов в операторе RESET не заданы, то происходит сброс статистики по всем объектам модели.

Необходимо учитывать особенности:

- 1) генераторы случайных чисел не устанавливаются в исходное состояние;
- 2) счетчики текущих значений каждого блока устанавливаются равными числу транзактов, находящихся в блоке. Счетчики числа иходов в блоки устанавливается в ноль;
- 3) таймер относительного времени устанавливается в ноль, а таймер абсолютного времени не изменяется.

Управление просмотром цепи текущих событий. Интерпретатор GPSS при определенных условиях возобновляет просмотр цепи текущих событий с первого элемента. В частности, если некоторый активный транзакт входит в блоки SEIZE, RELEASE, ENTER, LEAVE, LOGIC, PRIORITY, то интерпретатор заново начинает просмотр цепи текущих событий после того, как данный транзакт остановится.

Однако иногда бывает необходимо при достижении транзактом заданной точки модели немедленно начать просмотр цепи текущих событий, а не ждать, пока этот транзакт войдет в блок, который вызовет данное событие. В GPSS существует специальный блок, который предназначен для возобновления просмотра цепи текущих событий — BUFFER. Данный блок не имеет параметров.

Приложение 2

Макросредства системы моделирования СТАМ-КЛАСС позволяют создавать программные модели из макроопределений, написанных на

базовом языке СТАМ-КЛАСС. Макроопределения составляются и отлаживаются заранее и помещаются в личные или общественные библиотеки исходных текстов. Для вызова макроопределений из библиотеки и подстановки их в текст служат специальные операторы макровывода. При подстановке макроопределения возможна его параметрическая настройка, т. е. замена отмеченных участков текста макроопределения (формальных параметров) произвольными строками символов, задаваемыми в макровыводе (фактическими параметрами). Кроме того, в текст программы может быть подставлено не все макроопределение, а лишь определенные его фрагменты, список которых также задается в макровыводе. Таким образом реализуется условная генерация текста макрорасширения.

Следует также отметить, что в теле макроопределения могут встречаться другие макровыводы, однако при этом не должно возникать замкнутых цепочек вызовов, т. е. рекурсий.

В макросредствах системы моделирования СТАМ-КЛАСС используются следующие базовые понятия.

макроопределение — набор операторов базового языка СТАМ-КЛАСС и операторов макроязыка, имеющий имя и хранящийся в библиотечном наборе данных на магнитном диске,

макрорасширение — часть макроопределения, состоящая только из операторов базового языка СТАМ-КЛАСС и помещаемая в выходной текст макропроцессора вместо соответствующего макровывода;

макровывод — оператор макроязыка, указывающий имя макроопределения, структуру требуемого макрорасширения и значения параметров;

параметр — специальный идентификатор макроопределения (первым символом должен быть $\&$ (например $\&Mi$), заменяемый в процессе макрообработки строкой символов, заданной в макровыводе в качестве значения этого параметра;

фактор — имя списка идентификатора (элементами этого списка могут быть другие факторы);

макробиблиотека — библиотечный набор данных, состоящий из макроопределений.

Макропроцессор работает в пакетном режиме. Обращение к макропроцессору оформляется в виде шага задания.

Результатом работы макропроцессора являются программы, составленные на языке СТАМ-КЛАСС.

Макропроцессор позволяет вставлять в произвольные места исходной программы стандартные фрагменты текста, хранимые в библиотечном наборе данных на магнитном диске; заменять отдельные

идентификаторы во вставляемых фрагментах на какой-либо другой текст; производить перекомпоновку программы в зависимости от задаваемых программистом параметров.

Макропроцессор работает следующим образом. Исходный текст программы просматривается, начиная с его первого символа. До тех пор, пока в ходе просмотра не встретится какой-либо оператор языка макропроцессора, каждый символ исходного текста без изменений переносится в выходной текст. Действия, задаваемые оператором языка макропроцессора, производятся непосредственно в тот момент, когда этот оператор встречается в ходе просмотра исходного текста. После выполнения оператора языка макропроцессора продолжается просмотр исходного текста.

Если при просмотре текста встречается параметр, то он заменяется строкой символов, заданной в качестве значения этого параметра.

Если в тексте непосредственно перед параметром или вслед за ним стоят другие символы или другой параметр, то в выходном тексте макропроцессора символы, заменяющие данный параметр, сцепляются с другими символами или с теми символами, которые соответствуют другому параметру. В том случае, когда программист хочет сцепить некоторый параметр с буквой, цифрой или точкой, он должен непосредственно за параметром поставить точку. Эта точка не появляется в выходном тексте макропроцессора.

Обработка текста программы макропроцессором осуществляется в четыре этапа: уплотнение текста, генерация, суффиксация и сортировка. Каждый этап работы выполняется соответствующим сегментом макропроцессора.

Первый сегмент производит первичную обработку исходного текста программы: уплотнение за счет подавления кратных пробелов и удаление из текста служебных символов базового языка СТАМ-КЛАСС.

Второй сегмент выполняет собственно макрогенерацию текста программы: распознавание макровывода; поиск в макробиблиотеке соответствующего макроопределения; замену в тексте программы макровывода макрорасширением в соответствии со структурой, заданной в макровыводе, и подстановку значений параметров.

Третий сегмент устраняет возможную коллизию имен при неоднократном вызове одного и того же макроопределения. Для этого с каждым идентификатором макрорасширения сцепляется порядковый номер соответствующего макровывода. Нумерация макровыводов производится последовательно по мере их обнаружения в тексте программы.

Четвертый сегмент форматизует выходной текст макропроцессора в соответствии с требованиями, предъявляемыми транслятором с базового языка СТАМ-КЛАСС.

В корневого сегмент макропроцессора входят модули, необходимые на всех этапах обработки текста. Они выполняют следующие функции:

- обмен информацией с внешней памятью;
- вывод листингов всех этапов работы макропроцессора;
- вывод сообщений макропроцессора;
- лексический анализ текста;
- определение режимов работы макропроцессора.

При обращении к макропроцессору режимы его работы задаются в поле PARM оператора EXEC языка управления заданиями ОС ЭВМ.

Предусмотрены следующие режимы работы макропроцессора:

- вывод листинга любого из этапов обработки текста;
- подавление листинга любого из этапов обработки текста;
- трассировочная печать значений входных и выходных параметров любого модуля макропроцессора.

Описание макроязыка. Для написания операторов макроязыка и их параметров приняты следующие обозначения:

- специальные символы, используемые в операторах макроязыка: амперсанд (&); звездочка (*); запятая (,); точка (.); круглые скобки (); знак «равенство» (=); знак «подчеркивание» (_); знак «процент» (%).

- прописные буквы русского и латинского алфавитов для обозначения названия операторов макроязыка и их параметров;

- арабские цифры от 0 до 9;

- косая черта (/), используемая для того, чтобы показать возможность выбора одного из перечисленных значений.

Например, запись:

СУФФИКСАЦИЯ__ДА/НЕТ*

означает, что оператор должен быть записан либо как

СУФФИКСАЦИЯ__ДА*,

либо как

СУФФИКСАЦИЯ__НЕТ*;

- квадратные скобки [], показывающие, что можно либо выбрать значения, указанные в квадратных скобках, либо содержимое квадратных скобок может быть опущено. Например, запись

ФАКТОР__идентифик. 1 = идентиф. 2, идентиф. 3 ...*

означает, что справа от знака равенства может стоять либо один идентификатор, либо последовательность идентификаторов, разделен-

ных запятыми; многоточие указывает на повторение конструкции, помещенной в квадратные скобки.

Операторы макроязыка. Управление работой макропроцессора осуществляется посредством следующих операторов:

макроопределения	— МАКРО*;
конца макроопределения	— МКОНЕЦ ;
начала фрагмента текста	— ФРАГМЕНТ ;
конца фрагмента текста	— КФ ;
управления суффиксацией	—СУФФИКСАЦИЯ
условной генерации	— ФАКТОР ;
описания общих переменных	
макровызова	— ГЛОБАЛЬНЫЕ;

С помощью операндов, задаваемых в операторах макроязыка, указывается структура выходного текста макропроцессора, значения параметров и другая управляющая информация.

Оператор макроопределения МАКРО указывает начало макроопределения и задает начало текста макроопределения, его имя, структуру макрорасширения по умолчанию и значения параметров, принимаемые по умолчанию. Синтаксис оператора:

**МАКРО__имя — макрораспределения { (имя — фрагмента/
/фактор, ...) [, параметр = значение] ... ***

где «имя — макрораспределения» — имя, используемое в соответствующем макровывозе; «имя — фрагмента» — идентификатор, входящий в список операндов оператора ФРАГМЕНТ данного макроопределения; «фактор» — идентификатор, стоящий слева от знака равенства в одном из операторов ФАКТОР данного макроопределения; «параметр» — произвольный идентификатор, которому предшествует знак &; «значение» — произвольная строка символов.

Пример:

МАКРО__ТЕКСТ (А, С, Т), & М = МЕТКА*

Оператор конца макроопределения МКОНЕЦ указывает конец текста макроопределения в тексте программы. Синтаксис оператора:

МКОНЕЦ_ [комментарий] *

Оператор ФРАГМЕНТ задает начало фрагмента текста макроопределения и список идентификаторов. Упоминание хотя бы одного идентификатора из этого списка в макровывозе или в операторе МАКРО позволяет включить данный фрагмент в выходной текст макропроцессора. Синтаксис оператора:

ФРАГМЕНТ_идент. 1 / парам. 1 [, идент. 2 / парам. 2] ... *

Пример:

ФРАГМЕНТ_А, & Т, В4 *

Оператор КФ указывает конец фрагмента текста макроопределения.

Синтаксис оператора:

КФ _ [комментарий] *

Фрагменты текста могут быть вложены друг в друга, но не должны пересекаться.

Если ни один из идентификаторов оператора ФРАГМЕНТ не был указан при вызове макроопределения или в операторе МАКРО, то соответствующий фрагмент текста макроопределения не включается в выходной текст макропроцессора. Фрагменты, вложенные в пропускаемый фрагмент, также не включаются в выходной текст макропроцессора, даже если их идентификаторы указывались при вызове.

Оператор ФАКТОР присваивает имя списку идентификаторов. Это позволяет сократить список операндов при вызове макроопределения, так как указание имени списка равносильно указанию всего списка идентификаторов. Синтаксис оператора:

ФАКТОР _ *идент./парам.* = *идент.* 1/*парам.* 1 [, *идент.* 2/*парам.* 2] ... *

Пример. Пусть начало макроопределения имеет вид:

МАКРО _ **ТЕКСТ** *
ФАКТОР _ T1 = A, B, C *
ФАКТОР _ T2 = A, M *

Если в исходном тексте используется макровывод % **ТЕКСТ** (T2) *, то в выходной текст макропроцессора включаются фрагменты A, B и C.

Если в исходном тексте используется макровывод % **ТЕКСТ** (T2) *, то в выходной текст макропроцессора включаются фрагменты A и M;

Если в исходном тексте используется макровывод % **ТЕКСТ** (T1, T2) *, то в выходной текст макропроцессора включаются фрагменты A, B, C и M.

Оператор СУФФИКСАЦИЯ позволяет отключать и вновь включать механизм суффиксации. В исходном состоянии суффиксация включена. После оператора

СУФФИКСАЦИЯ _ **НЕТ** *

текст программы не суффиксируется до ближайшего оператора

СУФФИКСАЦИЯ _ **ДА** *,

который восстанавливает исходный режим.

Оператор ГЛОБАЛЬНЫЙ задает список идентификаторов, которые не должны суффиксироваться.

Если имеется необходимость отменить суффиксацию лишь для отдельных идентификаторов, то их следует перечислить в списке операндов оператора ГЛОБАЛЬНЫЕ. Синтаксис оператора:

ГЛОБАЛЬНЫЕ _ *идент.* 1/*парам.* 1 [, *идент.* 2/*парам.* 2] ... *

Пример:

ГЛОБАЛЬНЫЕ _ A, & M4, ТОВР *

Оператор макровывода указывает имя макроопределения, структуру макрорасширения и значения параметров.

При обнаружении в тексте программы макровывода макропроцессор находит в макробιβлиотеке требуемое макроопределение, обрабатывает его в соответствии со значениями параметров макровывода и помещает полученное макрорасширение в выходной текст. Синтаксис оператора:

% *имя* - *макроопределения* [(*имя*—фрагмента/фактор, ...)]
 (, *макропараметр* = *значение*) ... *

Здесь «ими — макроопределения» — имя, используемое в операторе МАКРО соответствующего макроопределения; «имя—фрагмента» — идентификатор, входящий в список операндов оператора ФРАГМЕНТ соответствующего макроопределения; «фактор» — имя списка идентификаторов в одном из операторов ФАКТОР соответствующего макроопределения; «макропараметр» — один из параметров соответствующего макроопределения, записанный без знака &; «значение» — произвольная строка символов.

При вызове макроопределения список фрагментов, включаемых в выходной текст макропроцессора и таблица значений параметров формируются на основе информации, которая имеется в операторах макровывода и МАКРО. Если информация для некоторого поля таблицы параметров или списка фрагментов задана в обоих операторах, то используется только один оператор. Оператор макровывода имеет в этом смысле более высокий приоритет по сравнению с оператором МАКРО.

Приложение 3

ОПЕРАТОР—ПЕЧАТЬ предназначен для формального вывода значений переменных, массивов и литералов. Синтаксис оператора ПЕЧАТЬ:

ПЕЧ _ (список объектов печати) (список форматов вывода) *

В списке объектов печати перечисляются имена переменных, элементов массивов, разделяемые запятыми. Список форматов должен соответствовать последовательности объектов в списке печати. В список форматов могут входить записи литералов (символьная информация) и специальные знаки — разделители, управляющие печатью

Допустимы следующие форматы вывода

2. Какие выводы можно сделать о характеристиках реальной системы как основе использования имитационной модели?

Первый вопрос связан с пониманием свойств самой модели и выяснением того, работает ли она в соответствии с задуманной схемой. Интересно также выявление чувствительности выходных данных модели к изменению ее входных параметров. Разброс выходных данных зависит прежде всего от точности подбора заложенных в модель вероятностных распределений.

Второй вопрос связан с состоятельностью модели и ее полезностью. Ответ на него включает описание характеристик системы и их статистическую обработку. Другими словами, строятся таблицы и графики, которые анализируются так, как если бы они были построены по данным обследования реальной системы, т. е. теми же методами статистического анализа. Если принятие решения основано на вероятности появления того или иного результата или на среднем значении некоторой величины, требуемые количественные оценки могут быть получены, исходя из результатов имитации. Именно с этой целью имитационный анализ используется в большинстве его современных приложений. При этом оценка разброса случайной величины выполняется так же, как и для реальной системы. Данные одного имитационного прогона представляют собой единичную выборку или временную последовательность случайного процесса, и это является ничуть не более нежелательным фактом, чем то, что запись предыстории представляет собой лишь отдельную временную последовательность.

Ответ на вопрос первого типа должен содержать подробный статистический анализ, обеспечивающий получение информации о точности и чувствительности модели. Обычно для этой цели либо повторяют имитацию несколько раз, либо имитируют более продолжительный период времени. При этом имитационная модель рассматривается как стохастическая и предполагается, что случайные элементы модели приведут к результатам, также носящим вероятностный характер. Руководители на производстве обычно не владеют анализом подобного рода, так как для этого необходимо знать специальную статистическую терминологию. Кроме того, более точные результаты могут быть получены путем изменения условий эксперимента, например путем увеличения числа прогонов.

Поскольку вопрос второго типа непосредственно связан с особенностями системы, а следовательно, и ее модели, то здесь трудно рекомендовать какие-то универсальные средства анализа, помимо стандартных статистических процедур. Методы же, используемые при

получении ответа на вопрос первого типа, широко применяются на практике, и ниже рассматриваются наиболее полезные из них.

2.9.1.1. Определения понятий

В процессе имитации необходимо фиксировать (запоминать) результаты наблюдений за интересующими исследователя переменными. Любое наблюдение представляет собой временную выборку, а, следовательно, наблюдения могут рассматриваться как случайные величины. Поэтому введем стандартный набор определений, относящийся к случайным величинам.

Пусть I — число интервалов, итераций или отдельных наблюдений. Слово «группа» и «интервал» будем применять в одном и том же смысле в данном разделе для обозначения любого из названных выше терминов.

T_i — время окончания i -го интервала ($i = 1, 2, \dots, I$), причем T_0 — момент начала первого интервала.

N_i — число наблюдений на i -м интервале ($i = 1, 2, \dots, I$).

$X_i(t)$ — значение X в момент времени t на i -м интервале;
 $t \in [T_{i-1}, T_i]$.

$X_i(n)$ — значение X для n -го наблюдения на i -м интервале
 $(n = 1, 2, \dots, N_i)$.

Примером $X_i(t)$ может быть текущий объем запаса на складе в момент t или число посетителей в системе в момент t . Выше эти переменные назывались непрерывными зависимыми от времени переменными. Примером $X_i(n)$ может служить время нахождения в очереди n -го посетителя или запас на складе при поступлении n -го заказа. Отметим, что I , T_i и N_i обычно рассматриваются как константы, однако в некоторых случаях могут быть случайными величинами.

Стохастическим процессом называется упорядоченный набор случайных величин. Таким образом, $\{X_i(t), t \in [0, \infty)\}$
 $\{X_i(n), n = 1, 2, \dots, \infty\}$ — стохастические процессы.

Реализацией стохастического процесса является набор выборочных траекторий данного стохастического процесса.

Временной ряд — это конечная реализация стохастического процесса. В терминах имитации в результате каждого прогона образуются временные ряды изучаемых стохастических процессов.

Стохастический процесс называется *стационарным*, если совместное распределение случайных величин, принадлежащих любой его реализации, является независимым во времени, т. е. исходный вероятностный механизм процесса инвариантен времени. Это свойство

называется также свойством *стационарности в узком смысле*. Особый тип стационарности называется *ковариационной стационарностью*, требующей, чтобы все средние значения μ_t и ковариации R_{st} случайных величин процесса были конечными, а ковариации, разделенные h единицами времени, были равны, т. е.

$$\mu_t = E [X_t(t)],$$

$$R_{st} = E [(X_t(s) - \mu_s) (X_t(t) - \mu_t)]$$

и $R_{st} = R_{rq}$, если $|t - s| = |q - r| = h$.

Ковариационно стационарный процесс называется также стационарным в широком смысле, или среднеквадратично стационарным, или стационарным второго порядка.

Методы проверки последовательностей на стационарность еще неполностью разработаны. Простейшим и наиболее часто применяемым методом проверки является изучение физики процессов и факторов, лежащих в основе того явления, о котором собираются данные. Если основные физические факторы, определяющие явление, инвариантны времени, то обычно предполагается стационарность результатов. Если же мы уверены в наличии некоторых тенденций или факторов сезонности, имеется возможность воспользоваться целым рядом методов, устраняющих влияние этих зависимых от времени факторов.

Эргодическим называется такой процесс, свойства случайных величин которого могут быть оценены по результатам одного временного ряда. Ковариационно стационарный процесс является эргодическим по величинам среднего и автоковариации, если выполняются два следующих условия:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=-T+1}^{T-1} R_s = 0,$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=-T+1}^{T-1} R_s^2 = 0.$$

Можно показать, что последовательность выборочных средних $\{X_s, s = 1, 2, \dots, \infty\}$ будет эргодической, если $\text{Var}[\bar{X}_s] \rightarrow 0$ при $s \rightarrow \infty$.

Особая значимость этого результата для исследователя имитационной модели состоит в том, что выборочное среднее асимптотически стремится к математическому ожиданию, если дисперсия

выборочного среднего стремится к нулю по мере увеличения величины выборки.

2.9.2. Дисперсия выборочного среднего $\text{VAR}[\bar{X}_I]$ и ее значение

Выборочное среднее является арифметическим средним I случайных величин X_i и определяется следующим образом:

$$\bar{X}_I = \frac{\sum_{i=1}^I X_i}{I}.$$

Обозначение \bar{X}_I указывает, что среднее является случайной величиной, зависящей от суммы I случайных величин. Обычно в имитационных исследованиях интерес представляет сравнение значений \bar{X}_I для различных альтернатив с помощью проверки гипотез или определение доверительных интервалов для значения \bar{X}_I отдельной альтернативы. При решении любой из этих задач необходимо уметь вычислять дисперсию выборочного среднего $\text{Var}[\bar{X}_I]$. Разработаны методы оценки $\text{Var}[\bar{X}_I]$ на основе выходных временных рядов, полученных с помощью имитации, и процедуры уменьшения оценок $\text{Var}[\bar{X}_I]$, т. е. понижения дисперсии, которые позволяют делать более точные выводы о \bar{X}_I . Эти вопросы рассмотрены в разд. 2.9.3 и 2.9.4 соответственно. В данном разделе приведены основные сведения и математические формулы, которые важны для понимания значения дисперсии выборочного среднего.

2.9.2.1. Обозначения

В нашем исследовании \bar{X}_I мы предполагаем, что X_i — это случайная величина, связанная с интервалом или группой i , т. е. средняя группы i . В этом разделе мы будем использовать понятие «группа», которое может соответствовать одиночному наблюдению либо некоторому множеству наблюдений, осуществляемых на отдельном подынтервале прогона или на всем прогоне (дублирование). То, как определяется понятие «группа» в конкретном случае, зависит от применяемой в данном имитационном исследовании процедуры определения выборочного среднего, которая в свою очередь основывается на проверке гипотез или задании доверительного интервала. Ниже приведены различные определения X_i

Замещенное наблюдение: $X_i = \begin{cases} 1 & \text{при положительном исходе} \\ & \text{в группе } i; \\ 0 & \text{при отрицательном исходе} \\ & \text{в группе } i. \end{cases}$

Среднее i -й группы по времени:

$$X_i = \frac{1}{T_i - T_{i-1}} \int_{t=T_{i-1}}^{T_i} X_i(t) dt.$$

Среднее i -й группы по наблюдениям:

$$X_i = \frac{1}{N_i} \sum_{A_i}^{A_{i+1}} X_i(n),$$

где

$$A_i = \sum_{j=1}^{i-1} N_j \text{ и } A_1 = 0.$$

Отметим, что два последних определения предполагают вычисление среднего группы. Для упрощения процесса вывода формул мы не будем придавать значения этому факту (в разд. 2.9.3 предполагаются конкретные методы вычисления для среднего группы).

2.9.2.2. Формулы для вычисления $\text{VAR}[\bar{X}_I]$

Из определения дисперсии можно вывести следующие выражения (суммирование везде осуществляется от 1 до I , а $\mu = E[X_i]$):

$$\begin{aligned} \text{Var}[\bar{X}_I] &= E[(\bar{X}_I - E[\bar{X}_I])^2] = \\ &= E\left[\left(\frac{\sum X_i}{I} - E\left[\frac{\sum X_i}{I}\right]\right)^2\right] = \\ &= \frac{1}{I^2} E\left[\left(\sum X_i - E\left[\sum X_i\right]\right)\left(\sum X_i - E\left[\sum X_i\right]\right)\right] = \\ &= \frac{1}{I^2} E\left[\sum (X_i - E[X_i]) \cdot \sum (X_j - E[X_j])\right] = \\ &= \frac{1}{I^2} E\left[\sum \sum (X_i - E[X_i])(X_j - E[X_j])\right] = \\ &= \frac{1}{I^2} \sum \sum E[(X_i - E[X_i])(X_j - E[X_j])] = \\ &= \frac{1}{I^2} \sum \sum \text{Cov}[X_i, X_j] = \\ &= \frac{1}{I^2} \left(\sum \text{Var}[X_i] + \sum_{i \neq j} \sum \text{Cov}[X_i, X_j]\right). \end{aligned} \tag{1}$$

Если X_i и X_j независимы для всех i и j , а $\text{Var}[X_i] = \sigma^2$ для всех i , то

$$\text{Var}[\bar{X}_I] = \frac{1}{I^2} \sum \sigma^2 = \frac{1}{I} \sigma^2. \tag{2}$$

Из этого уравнения видно, что при условии независимости $I * \text{Var}[\bar{X}_I]$ является константой, а величина $\text{Var}[\bar{X}_I]$ в I раз меньше, чем σ^2 . Ниже мы кратко остановимся на этом выводе.

При предположении о независимости и при условии слабой регулярности X_i центральная предельная теорема утверждает, что при больших I распределение величины $\sqrt{I}(\bar{X}_I - \mu)/\sigma$ сходится к нормальному с математическим ожиданием 0 и дисперсией 1, т. е. к $N(0, 1)$. Если величина X_i сама нормально распределена, величина \bar{X}_I также нормально распределена, а величина $(\bar{X}_I - \mu)/\sqrt{S^2_x/I}$ имеет t -распределение с $I - 1$ степенями свободы, где S^2_x — оценка σ^2 . На основе этого можно построить доверительный интервал для \bar{X}_I . Отметим, что, если X_i — среднее группы, предположение о его нормальном распределении вполне состоятельно. В ряде работ предлагается следующее выражение для дисперсии S^2_x :

$$\text{Var}[S^2_x] = \sigma^4 \left(\frac{2}{I-1} + \frac{\nu^2}{I} \right), \tag{3}$$

где γ^2 — коэффициент эксцесса (отношение четвертого центрального момента к квадрату второго центрального момента минус 3). С помощью уравнения (3) можно определить общую величину разброса, ожидаемую при оценке дисперсии основного процесса.

Если нельзя предположить, что $\text{Cov}[X_i, X_j] = 0$, но можно считать процесс ковариационно стационарным, то, обозначив $\text{Cov}[X_i, X_j] = R_{j-i} = R_h$ и используя уравнение (1), можно получить

$$\text{Var}[\bar{X}_I] = \frac{1}{I} \sum_{h=1}^{I-1} \left(1 - \frac{|h|}{I}\right) R_h. \quad (4)$$

Обозначив $\sigma^2 = R_0$ и $R_h = R_h$ и подставив их в уравнение (4), получим

$$\text{Var}[\bar{X}_I] = \frac{1}{I} \left\{ \sigma^2 + 2 \sum_{h=1}^{I-1} \left(1 - \frac{h}{I}\right) R_h \right\}. \quad (5)$$

Процедуры оценки $\text{Var}[\bar{X}_I]$ с использованием этого уравнения рассмотрены в разд. 2.9.3.5.

Если автоковариация экспоненциально затухает (обычным и вполне разумным предположением является то, что $R_h = R_0 \alpha^h$ для $0 < \alpha < 1$), можно показать, что

$$\lim_{I \rightarrow \infty} I \text{Var}[\bar{X}_I] = \sum_{h=-\infty}^{\infty} R_h = m. \quad (6)$$

Из уравнения (6) видно, что по мере увеличения числа групп дисперсия $\text{Var}[\bar{X}_I]$ уменьшается пропорционально $1/I$.

Сравнивая уравнения (6) и (2), можно отметить, что разброс основного процесса σ^2 зависит от суммы всех ковариаций. Ниже мы будем обозначать эту величину буквой m . Поскольку значение m не зависит от числа групп I , наличие оценки m для процесса позволяет оценивать $\text{Var}[\bar{X}_I]$ при любом значении I .

2.9.2.3. Интерпретация $\text{VAR}[\bar{X}_I]$

Как уже говорилось выше, $\text{Var}[\bar{X}_I] \rightarrow m/I$ при больших I и при соответствующих предположениях. На основании этого можно изобразить распределение \bar{X}_I во времени, как показано на рис. 1.

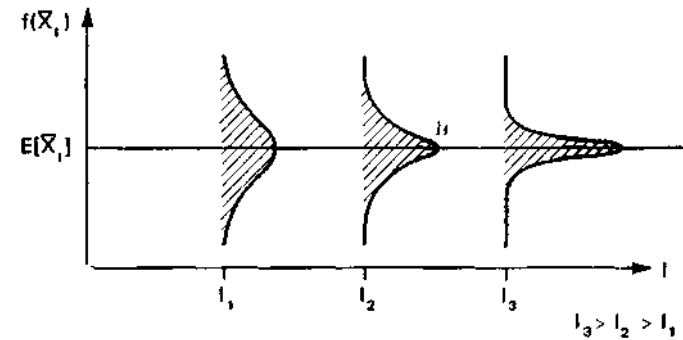


Рис. 1. Распределение \bar{X}_I .

Так как \bar{X}_I — случайная величина, то оценки являются значениями выборки. При осуществлении повторных имитационных прогонов возможны иные значения \bar{X}_I с точностью до $f(\bar{X}_I)$ и, следовательно, иные значения $\text{Var}[\bar{X}_I]$. Изменение продолжительности прогона или числа групп приводит к изменениям $f(\bar{X}_I)$, т. е. $f(\bar{X}_I)$ зависит от I и распределения X_i . Для прогона одной и той же длины могут получаться различные оценки при различном количестве групп. Более того, три различных распределения на рис. 1 имеют одно и то же значение $E[\bar{X}_I]$. Но так бывает далеко не всегда, поскольку могут использоваться и чаще всего используются смещенные оценки. Таким образом, необходимо наличие определенных критериев для сравнения оценок $\text{Var}[\bar{X}_I]$ при различных условиях эксперимента. Такими критериями сравнения во многих исследованиях являются m и среднеквадратичная ошибка среднего выборочного, обозначаемая $\text{MSE}[\bar{X}_I] = E[(\bar{X}_I - \mu_X)^2]$.

К сожалению, формула для вычисления m для немарковских процессов не выведена. Для модели с очередью $M/M/1$ в ряде работ получено $m = 6840$ при скорости обслуживания $\mu = 5$. Если же $\lambda = 2,0$ и $\mu = 2,25$, то $m = 361$. Для модели $M|M|\infty$ $m = 2\lambda/\mu^2$. В ряде работ разработана процедура получения значений m для марковских процессов с конечным числом состояний.

Для марковских процессов с конечным числом состояний можно также показать, что m является функцией λ и μ и что $m(a\lambda, a\mu) = (1/a)[m(\lambda, \mu)]$. Интуитивно ясно, что этот результат остается в силе и для марковских процессов со счетным числом состояний, как это имеет место в модели $M/M/1$. Следовательно, при вычислении m для очереди $M/M/1$ с $\lambda/\mu = 0,9$, где $\lambda = 4$, $\mu = 40/9$ и с $a = 8/9$ имеем $m = 9/8 (6840) = 7695$.

$MSE[\bar{X}_1]$ является критерием, который объединяет $Var[X_1]$ и смещение \bar{X}_1 . Это видно из следующей выкладки:

$$\begin{aligned} MSE[X_1] &= E[(X_1 - \mu_x)^2] = \\ &= E[(\bar{X}_1 - E[\bar{X}_1] + E[\bar{X}_1] - \mu_x)^2] = \\ &= E[(\bar{X}_1 - E[\bar{X}_1])^2] + E[(E[\bar{X}_1] - \mu_x)^2] = \\ &= Var[X_1] + (Смещ[\bar{X}_1])^2. \end{aligned} \quad (7)$$

Когда имеет место смещение \bar{X}_1 (смещение может существовать, если процесс не является ковариационно стационарным, или если X_i вычисляется с использованием оценки отношения (см. раз. 2.9.3.3)) вероятность попадания теоретического среднего в доверительный интервал (покрытие доверительным интервалом) отличается от предписанного значения из-за разницы, вызванной смещением. «Покрытие» определяется как вероятность того, что теоретическое среднее покрывается интервалом, центром которого является $E[\bar{X}_1]$ или \bar{X}_1 . Когда используются несмещенные оценки, $MSE(\bar{X}_1)$ уменьшается до величины $Var[\bar{X}_1]$, а «покрытие» становится равным величине, характеризующей доверительный интервал, т. е. доверительному уровню. Теперь рассмотрим процедуры оценки $Var[\bar{X}_1]$.

2.9.3. Процедура оценки $VAR[\bar{X}_1]$

Как уже говорилось в предыдущем разделе, дисперсия выборочного среднего имеет фундаментальное значение для достоверной интерпретации результатов имитационного моделирования. Оценка $Var[\bar{X}_1]$ по результатам одного имитационного прогона является сложной задачей из-за зависимости наблюдений выборки, используемых при вычислении \bar{X}_1 . Этой процедуре $Var[\bar{X}_1]$ посвящено довольно много исследований, в результате которых сформулируем следующие пять подходов к решению данной проблемы:

1. *Повторные прогоны* (дублирование). Проводится несколько прогонов, каждый из которых рассматривается как группа. По результатам i -го прогона получают значение X_i , а значение $Var[\bar{X}_1]$ оценивается с помощью уравнения (2).

2. *Подынтервалы*. Прогон делится на несколько равных групп (подынтервалов), и X_i вычисляется как среднее для i -й

группы. X_i полагают независимыми и используют уравнение (2) для оценки $Var[\bar{X}_1]$.

3. *Регенеративные циклы*. Имитационный прогон делится на независимые циклы путем задания состояний, при достижении которых цикл стартует заново. $Var[\bar{X}_1]$ оценивается на основе значений наблюдений в независимых циклах.

4. *Параметрическое моделирование*. Результаты имитационного прогона аппроксимируются некоторой функцией или уравнениями. Оценка $Var[\bar{X}_1]$ выводится из уравнений, моделирующих результаты имитации.

5. *Спектральная ковариационная оценка*. По результатам выборки оценивается автоковариация, которая используется затем в спектральном анализе или подставляется в уравнение (5) для оценки $Var[\bar{X}_1]$.

Рассмотрим каждый из этих подходов в отдельности.

2.9.3.1. Повторные прогоны

Подход предусматривает вычисление на i -м прогоне значения случайной величины X_i . Как говорилось в разд. 2.9.2.1, величина X_i может представлять собой среднее число элементов в системе на i -м прогоне, среднее время нахождения посетителя в системе, или число успешных исходов биномиальной переменной в течение прогона. В качестве оценки интересующего нас параметра используется средняя величина значений X_i по I прогонам, т. е.

$$\bar{X}_1 = \frac{\sum_{i=1}^I X_i}{I}, \quad (8)$$

Затем с помощью стандартной процедуры получается оценка $Var[X_1]$:

$$S^2_x = \frac{1}{I-1} \sum_{i=1}^I (X_i - \bar{X}_1)^2 = \frac{1}{I-1} \sum_{i=1}^I X_i^2 - \frac{1}{I(I-1)} \left(\sum_{i=1}^I X_i \right)^2. \quad (9)$$

С помощью уравнения (9) оценивается разброс случайной выборки, получаемой на каждом прогоне. Поскольку прогоны являются независимыми, оценка дисперсии выборочного среднего может быть получена следующим образом:

$$S^2_{\bar{x}} = S^2_x / I. \quad (10)$$

На основе значений $S^2_{\bar{x}}$ и \bar{X}_1 , полученных по результатам наблюдений и с помощью центральной предельной теоремы, можно

сделать заключения о вероятностном характере изучаемых параметров. На основе данных теоретических результатов можно также осуществить и проверку гипотез.

Процедура повторных прогонов обладает важным свойством независимости выборок. Другим достоинством метода является возможность его использования как для терминального анализа, так и для анализа установившегося режима (под терминальным здесь понимается анализ, проводимый на характерном конечном (заключительном) интервале времени (в ряде работ утверждается, что из практики имитационного моделирования можно сделать вывод, что значительная часть имитаций реального мира относится к типу терминальных; это обстоятельство позволяет использовать классический статистический анализ). К недостаткам метода можно отнести следующее:

- 1) в каждом повторном прогоне содержатся данные, относящиеся к начальному переходному периоду, которые не должны использоваться при анализе стационарного режима;
- 2) каждый повторный прогон дает только одну выборку X_i , а это означает, что значительная информация об изучаемой переменной остается неизвлеченной из данных. Примером такой ситуации является вычисление X_i в качестве среднего по прогону.

2.9.3.2. Подынтервалы

Подход к оценке дисперсии \bar{X}_1 с использованием подынтервалов заключается в делении отдельного имитационного прогона на группы. Если каждая группа содержит b наблюдений $X_i(n)$, то выборочное среднее группы X_i определяется следующим образом:

$$X_i = \frac{\sum_{n=1}^b X_i(n)}{b} \quad (11).$$

Если подынтервалы независимы, для получения оценок $E[\bar{X}_1]$, $\text{Var}[X_i]$ и $\text{Var}[\bar{X}_1]$ используются уравнения (8) — (10) соответственно (можно показать, что если подынтервалы независимы, то m , относящееся к X_i и обозначаемое как $m(X_i)$, равно $\frac{1}{b} + m(X_i(n))$, где $m(X_i(n))$ является значением m , относящимся к отдельному наблюдению.)

Предположение о независимости традиционно выдвигается в имитационном анализе, даже если существует автокорреляция между значениями в конце одного подынтервала и в начале следующего. Автоковариация может вызвать положительную ковариацию между групповыми средними. С увеличением размера группы ковариация между выборочными средними по группам должна уменьшаться. Существуют процедуры определения размера группы, сводящие к минимуму ковариацию между выборочными средними соседних групп. Рекомендуется использовать изложенную ниже процедуру Фишмана для определения b .

Процедура Фишмана заключается в пересчете групповых оценок при делении размера b пополам до тех пор, пока не будет принята нулевая гипотеза о том, что $X_{i,b}$ для $i = 1, 2, \dots, I_b$ независимы и одинаково распределены. Здесь индекс b указывает на наличие независимости X , от размера группы b . Фишман рекомендует использовать следующую статистику для проверки гипотез:

$$C_b = 1 - \frac{\sum_{i=1}^{I_b-1} (X_{i,b} - X_{i+1,b})^2 / 2}{\sum_{i=1}^{I_b} (X_{i,b} - \bar{X}_{I_b})^2}.$$

где I_b — число групп, если размер группы равен b .

При больших b величина C_b приблизительно равна оценке коэффициента автокорреляции между соседними группами.

Если $X_{i,b}$ независимы и нормально распределены, то C_b имеет нулевое математическое ожидание, дисперсию $(I_b - 2)/(I_b^2 - 1)$ и распределение, близкое к нормальному при $b < 8$. Таким образом, если данные условия выполняются, можно применять стандартный тест с использованием нормальных таблиц. Если $\{X_{i,b}\}$ имеет монотонную функцию автоковариации, используется односторонний тест; в противном случае следует применять двусторонний тест.

Для использования изложенного выше подхода при определении размера группы требуется уточнить несколько процедурных деталей. Значения наблюдений $x_i(n)$ должны записываться в порядке вычисления $x_{i,b}$ для различных значений b . Перед применением теста должны отсекаются все нетипичные значения $x_i(n)$, полученные в начале прогона (см. разд. 2.9.5.2). Необходимо установить начальную величину размера группы. Фишман рекомендует начинать с $b = 1$. Если между $X_i(n)$ наблюдается существенная корреляция, рекомендуется задавать большие начальные значения для b . И наконец, последнее замечание, заключающееся в том, что последующая проверка гипотез может увеличить вероятность принятия нулевой гипотезы тогда, когда она является ложной.

К достоинствам метода подынтервалов следует отнести то, что для получения оценки дисперсии выборочного среднего необходимо провести только один имитационный прогон, в котором отсекается только одна группа данных переходного периода. Недостатком процедуры является поиск такого размера группы, который делает подынтервалы независимыми. Следует отметить, что при фиксированном количестве наблюдений увеличение размера выборки уменьшает число групп и, следовательно, может привести к большим значениям оценок $\text{Var}\{\bar{X}_I\}$. Другим недостатком является необходимость задания границ группы. При вычислении средних по группам особую осторожность необходимо проявлять в случае, когда одно наблюдение «растягивается» на несколько групп; например, посетитель прибывает в систему на i -м интервале, а покидает ее на $(i + 1)$ -м.

2.9.3.3. Регенеративный метод

Регенеративный метод подобен методу подынтервалов. Имитационный прогон в этом методе также делится на отдельные участки, которые называются циклами. Цикл начинается, когда система достигает некоторого специфического состояния, начиная с которого будущее ее поведение не зависит от предыдущего. При повторном достижении системой этого состояния текущий цикл заканчивается, давая тем самым одно независимое наблюдение изучаемых величин. С помощью заданных подобным образом циклов модель генерирует независимые выборки; следовательно, успешно решается проблема ковариации, возникающая при использовании подынтервалов. Однако при этом встает другая проблема, состоящая в том, что длина цикла заранее не определена и является случайной величиной.

Наиболее часто в качестве точки регенерации в моделях массового обслуживания используют достижение системой состояния, когда все обслуживаемые устройства свободны и очередь, на обслуживание пуста. Если прибытие каждого следующего посетителя обрабатывается некоторым фиксированным способом, то точкой регенерации и, следовательно, началом регенерационного цикла будет прибытие посетителя в пустую систему. В моделях управления запасами возможной точкой регенерации может быть переход склада в состояние, когда запас достигает контрольного уровня.

Каждый цикл имитационного прогона будет независимым в соответствии с определением; следовательно, можно строить оценки

выборочного среднего по данным циклам. Введем следующие обозначения: Y_i — исследуемое на i -м цикле значение, например сумма времен ожидания посетителей на i -м цикле (исследуемая величина может быть также интегрированной по времени переменной, например интегрированное по времени число посетителей в системе в течение i -го цикла); L_i — длина i -го цикла, например число посетителей или продолжительность цикла.

Если X_{ik} является k -й выборкой на i -м цикле и имитация продолжается до тех пор, пока не завершатся I циклов, можно записать два следующих уравнения:

$$Y_i = \sum_{k=1}^{L_i} X_{ik}, \quad (12)$$

$$\sum_{i=1}^I L_i = N, \quad (13)$$

где N — общее число выборок (случайная величина)

Среднее всех выборок имитационного прогона \bar{X}_I вычисляется по формуле

$$\bar{X}_I = \frac{\sum_{i=1}^I \sum_{k=1}^{L_i} X_{ik}}{N}. \quad (14)$$

Подстановкой переменных из уравнений (12) и (13) в уравнение (14) покажем, что \bar{X}_I может рассматриваться как отношение средних величин цикла

$$\bar{X}_I = \frac{\sum_{i=1}^I \sum_{k=1}^{L_i} X_{ik}}{N} = \frac{\sum_{i=1}^I Y_i}{\sum_{i=1}^I L_i} = \frac{\sum_{i=1}^I Y_i / I}{\sum_{i=1}^I L_i / I} = \bar{Y}_I / \bar{L}_I. \quad (15)$$

Поскольку число выборок в цикле является случайной величиной, мы не можем устанавливать ни число циклов, ни общее число выборок. В подобном случае используется оценка отношения, которая является смещенной.

Оценка дисперсии \bar{X}_I вычисляется по формуле

$$S^2_{\bar{X}} = \frac{S^2}{(\bar{L}_I / I)^2}, \quad (16)$$

где

$$S^2 = S^2_Y - 2\bar{X}_1 S_{YL} + \bar{X}_1^2 S^2_L,$$

$$S^2_Y = \frac{1}{I-1} \sum_{i=1}^I (Y_i - \bar{Y}_1)^2,$$

$$S^2_L = \frac{1}{I-1} \sum_{i=1}^I (L_i - \bar{L}_1)^2,$$

$$S_{YL} = \frac{1}{I-1} \sum_{i=1}^I (Y_i - \bar{Y}_1) (L_i - \bar{L}_1).$$

Как указывалось выше, \bar{X}_1 является смещенной оценкой. Для устранения смещения оценок отношения можно использовать оценку типа «складного ножа», которая ослабляет смещение в 1/I раз. Эта оценка для выборочного среднего вычисляется по формуле

$$\bar{J} = \frac{1}{I} \sum_{i=1}^I J_i, \tag{17}$$

где J_i называется псевдозначением оценки и вычисляется следующим образом:

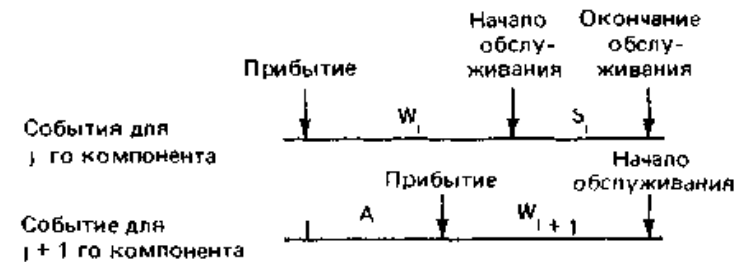
$$J_i = I\bar{X}_1 - (I-1) \sum_{j=1, j \neq i}^I Y_j / \sum_{j=1, j \neq i}^I L_j.$$

Величины J_i полагаются независимыми и одинаково распределенными, поэтому для X_1 можно построить доверительный интервал с использованием оценок \bar{J} и S_{J_i}/\sqrt{I} .

К достоинствам регенеративного метода можно отнести получение для каждого цикла независимых и одинаково распределенных величин, а, следовательно, для проверки гипотез и вычисления доверительных интервалов могут применяться стандартные статистические процедуры. Однако в процессе использования таких процедур необходимо определить точку регенерации, при которой ожидаемый интервал времени между возвратами конечен, а выборка наблюдений цикла достаточна для построения разумного доверительного интервала. Закон арксинуса, приведенный в приложении в конце данного раздела, показывает, что это далеко не всегда легко сделать. Еще одним достоинством метода является отсутствие проблемы задания начальных условий, поскольку сбор статистических данных начинается при достижении точки регенерации, которая определяет, в частности, и начальные условия. Недостатки процедуры — дополнительные вычисления и смещения оценок выборочного среднего.

2.9.3.4. Параметрическое моделирование

Параметрическое моделирование предполагает построение модели, описывающей результаты имитации. Значения оценок исследуемых величин получаются на основе вычислений в соответствии с параметрической моделью. Процедура параметрического моделирования включает этап сбора значений выборки по результатам имитации, а затем этап подбора зависимостей (в виде уравнений), описывающих совокупность полученных данных. Этот подход подобен тому, когда на основе данных о реальной системе пытаются составить ее аналитическое описание с помощью уравнений. Рассмотрим применение данного подхода на примере модели системы с одним обслуживающим устройством, в которой посетители обслуживаются по правилу «первым пришел — первым обслужен». Пусть исследуемой величиной является время ожидания обслуживания посетителем. Запишем уравнение, в котором время ожидания (j + 1)-го посетителя выражено через время ожидания j-го посетителя, интервал времени между прибытиями посетителей и продолжительность обслуживания — случайные величины. Это уравнение может быть выведено на основе анализа приведенной ниже пиктограммы:



Предполагается, что (j + 1)-ый компонент прибывает до момента окончания обслуживания j-го компонента. Приравнивая значения переменных обеих строк пиктограммы, можно получить следующее уравнение:

$$W_{j+1} = \begin{cases} W_j + S_j - A_j, & \text{если } A_j < W_j + S_j; \\ 0 & \text{в противном случае,} \end{cases}$$

где W_j — время ожидания в очереди j-го компонента; S_j — время обслуживания j-го компонента; A_j — интервал между прибытиями j-го и (j + 1)-го компонента.

Если мы проимитировали нашу систему обслуживания, получив данные по наблюдениям W_j , S_j и A_j , и смогли, как описано выше,

построить модель для W_{j+1} , эту модель можно использовать для получения информации о времени ожидания обслуживания посетителями.

Приведенная выше процедура построения модели дает общее представление о подборе зависимости по данным выборки. До сих пор исследования по параметрическому моделированию результатов имитации проводились главным образом с помощью моделей авторегрессии. Существует обширный анализ подобных моделей. Поскольку значения, полученные в результате имитации, могут рассматриваться как данные временных рядов, можно рекомендовать при идентификации и оценке модели пользоваться методологией Бокса — Дженкинса. Методология Бокса — Дженкинса заключается в построении параметрической модели данных выборки с использованием следующего уравнения:

$$Y_t = \sum_{k=1}^p \phi_k Y_{t-k} - \sum_{i=1}^q \theta_i U_{t-i} + U_t; \quad t = 1, 2, \dots, n, \quad (18)$$

где

$$Y_t = X(t) - E[X(t)]$$

и U_t — белый шум, т. е. $E[U_t] = 0$, а

$$F[U_t, U_{t-1}] = \begin{cases} \sigma^2_U, & t = 0; \\ 0 & \text{в противном случае.} \end{cases}$$

Модель, основанная на уравнении (18), называется комбинированной авторегрессионной моделью скользящего среднего. Если из уравнения (18) удалить все члены с коэффициентом θ_i , получится авторегрессионная модель порядка p . Авторегрессионная модель задает в качестве выражения для Y_t линейную комбинацию предыдущих значений временного ряда плюс компонента белого шума.

Если из уравнения (18) удалить все члены с коэффициентом ϕ_k , будет получена модель скользящего среднего, которая задает в качестве выражения для Y_t линейную комбинацию q прошлых ошибок. Она называется моделью скользящего среднего порядка q .

Методология Бокса — Дженкинса дает возможность построить процедуру для определения порядка авторегрессионной модели p и порядка модели скользящего среднего q . Кроме того, разработаны и запрограммированы процедуры получения наилучших оценок ϕ_k и θ_i . Программы, осуществляющие вычисления подобного рода, имеются в Центре данных Университета шт. Огайо и в Университете шт. Висконсин. Программа вычисляет значения θ_k и ϕ_1 , а также оценку

дисперсии белого шума σ^2_U . На основе этих оценок можно вычислить оценку дисперсии выборочного среднего:

$$s^2_{\bar{X}} = \frac{\hat{m}}{n}, \quad (19)$$

где

$$\hat{m} = \hat{\sigma}^2_U \frac{\left(1 - \sum_{i=1}^q \theta_i\right)^2}{\left(1 - \sum_{k=1}^p \phi_k\right)^2}$$

Практика показала, что параметрическое моделирование временных рядов, полученных в результате имитационного моделирования, не дает надежных оценок дисперсии среднего. Причина этого — нестационарный характер временных рядов или отсутствие нормальности распределения отдельных наблюдений. Значительно лучшие результаты получались при построении параметрических моделей с использованием временных рядов, состоящих из групповых средних. Применение оценок групповых средних позволяет «смягчить» предположения о стационарности и нормальности.

Достоинством параметрического моделирования является то, что выводится уравнение, описывающее исследуемую переменную, дальнейший анализ которого может дать новую информацию об имитируемой системе. Основным недостатком является отсутствие гарантии состоятельности параметрической модели. Построение параметрической модели для имитационной модели уводит анализ как бы на шаг дальше от реальной системы, требуя при этом от аналитика еще большей осторожности.

2.9.3.5. Оценка ковариаций и применение спектрального анализа

В разд. 2.9.2.2 показано, что

$$\text{Var}[\bar{X}_1] = \frac{1}{T} \left[R_0 + 2 \sum_{h=1}^{T-1} \left(1 - \frac{h}{T}\right) R_h \right].$$

Следовательно, если получены оценки автоковариаций R_h , можно определить

$$s^2_{\bar{X}} = \frac{1}{T} \left[\hat{R}_0 + 2 \sum_{h=1}^{T-1} \left(1 - \frac{h}{T}\right) \hat{R}_h \right].$$

В литературе для вычисления \hat{R}_h предлагаются три альтернативных уравнения:

$$A_h = \frac{1}{I-h} \sum_{i=1}^{I-h} (X_i - \bar{X}_I) (X_{i+h} - \bar{X}_I),$$

$$B_h = \frac{1}{I} \sum_{i=1}^{I-h} X_i X_{i+h} - \frac{1}{I-h} \left(\sum_{i=1}^{I-h} X_i \right) \left(\sum_{i=1}^{I-h} X_{i+h} \right),$$

$$C_h = \frac{1}{I} \sum_{i=1}^{I-h} (X_i - \bar{X}_I) (X_{i+h} - \bar{X}_I).$$

Если временной ряд короткий, а его конечные точки сильно отличаются от \bar{X}_I , в качестве оценки можно рекомендовать B_h . В имитационных исследованиях временные ряды обычно достаточно длинные, поэтому B_h в качестве оценки не применяется.

Оценка A_h интуитивно привлекательна, поскольку усредняет $I-h$ значений, тем не менее она дает большую среднеквадратичную ошибку, чем оценка C_h . Недостатком же оценки C_h является то, что она представляет собой смещенную оценку. Обычно приходят к выводу, что для временных рядов с большим I при оценке R_h должна использоваться оценка C_h .

Часто исследования по оценке автоковариации касаются главным образом выборочных величин $X_i(t)$ или $X_i(n)$. Эти переменные и оценки их автоковариаций сильно коррелированы. Таким образом, если полученная оценка R_0 велика, следует ожидать относительно большое значение оценки R_h . Это явление, наблюдаемое в имитационных экспериментах, описано в ряде работ. Использование групповых величин X_i , как это рекомендовано в предыдущем случае, должно ослабить корреляцию оценок автоковариаций.

Альтернативой описанной выше процедуре является применение спектрального анализа. Спектр определяется следующим образом:

$$g(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} R_h e^{-i\lambda h}, \quad -\pi \leq \lambda \leq \pi, \quad (20)$$

а спектральная плотность имеет вид

$$f(\lambda) = g(\lambda) / R_0.$$

Обратное преобразование осуществляется по формуле

$$R_h = \int_{-\pi}^{\pi} g(\lambda) e^{i\lambda h} d\lambda$$

и может быть использовано для получения значений R_h . При $h = 0$ имеем

$$R_0 = \int_{-\pi}^{\pi} g(\lambda) d\lambda.$$

Следовательно, можно рассматривать дисперсию процесса как совокупность неперекрывающихся составляющих на частоте λ . Большие значения $g(\lambda)$ отображают разброс процесса на частоте λ или периодичность, равную $2\pi/\lambda$.

Как отмечалось в разд 2.9.2, для целого ряда систем можно показать, что

$$\lim_{I \rightarrow \infty} I \text{Var} [\bar{X}_I] = m = \sum_{h=-\infty}^{\infty} R_h.$$

Присвоив λ нулевое значение и подставив его в уравнение (20), получим

$$g(0) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} R_h,$$

и, следовательно,

$$m = 2\pi g(0).$$

Таким образом, если можно оценить $g(0)$, то может быть получена и оценка для m .

Проведены большие исследования в области получения оценок спектра для временных рядов. Основная трудность при этом заключается в том, как определить число ковариаций, для которых необходимо выполнять вычисления, а также вид взвешивающей функции, применяемой к оцененной автоковариации, полученной по результатам конечного числа наблюдений. Взвешивающим функциям посвящен ряд работ; в частности, в ряде работ рассмотрены прямоугольные и дисперсионные функции.

Достоинством спектральных методов является то, что они хорошо разработаны. Основным недостатком является ненадежность спектральных точечных оценок, т. е. оценок при $\lambda = 0$. Этот недостаток можно частично устранить, группируя данные в отдельное наблюдение.

2.9.4. Методы понижения дисперсии

Дисперсия выборочного среднего является мерой надежности результатов, получаемых при повторном проведении имитационного

эксперимента. Уже было показано, что более длительные прогоны дают меньшие оценки $\text{Var}[X_i]$. Таким образом, значение $\text{Var}[X_i]$ в некотором смысле зависит от используемых в ходе эксперимента процедур и вычислений. Методы понижения дисперсии предназначены для уменьшения значений оценки $\text{Var}[\bar{X}_I]$ с помощью задания на основе априорной информации особых условий проведения эксперимента.

2.9.4.1. Дополняющая выборка

Уравнение (1) для $\text{Var}[X_i]$ содержит величину $\text{Cov}[X_i, X_j]$. Если $\text{Cov}[X_i, X_j]$ можно сделать отрицательной, то $\text{Var}[X_i]$ уменьшится. Поскольку X_i и X_j являются функциями псевдослучайных чисел, вполне допустимо, что при $X_i = f(\gamma_1, \gamma_2, \dots, \gamma_q)$ и $X_j = f(1-\gamma_1, 1-\gamma_2, \dots, 1-\gamma_q)$ можно получить отрицательное значение ковариаций между X_i и X_j . Возможность получения отрицательной ковариаций зависит от функции f , которая отражает преобразование с помощью имитационной модели случайных чисел в значения выборки. Очевидно, что в общем случае нельзя гарантировать положительный результат использования метода дополняющих значений. Однако на практике в экспериментах с дополняющей выборкой обычно наблюдается понижение дисперсии. Генерация дополняющего потока случайных чисел $1-\gamma_1, 1-\gamma_2, \dots, 1-\gamma_q$ легко достигается при использовании мультипликативного конгруэнтного генератора случайных чисел вида

$$z_k = az_{k-1} \bmod(c), \quad k = 1, 2, \dots,$$

$$\gamma_k = z_k/c.$$

Можно показать, что если в качестве начального значения последовательности случайных чисел используется $z'_0 = c - z_0$, то

$$z'_k = c - z_k$$

и, следовательно,

$$\gamma'_k = \frac{z'_k}{c} = 1 - \gamma_k.$$

Не рекомендуется применять дополняющую выборку внутри группы или даже внутри прогона. Манипулирование группами для получения дополняющих выборок может вызвать искажение базового процесса и привести к непредсказуемым результатам. К тому же существует целый ряд более приемлемых процедур. Например, проводят пару независимых прогонов, причем во втором используют дополняющие

потоки случайных чисел. Пусть в последовательности прибытий k -й интервал времени между прибытиями вычисляется по числу γ_k в первом парном прогоне и по числу $1-\gamma_k$ во втором. При таком способе вычисление значения дисперсии для $2l$ прогонов упрощается в связи с объединением значений парных прогонов. Если X'_i является дополняющим значением X_i , то

$$\text{Var}[\bar{X}_I] = \text{Var}\left(\frac{\sum_{i=1}^I (X_i + X'_i)}{2I}\right) = \text{Var}\left(\frac{\sum_{i=1}^I U_i}{I}\right),$$

где

$$U_i = \frac{X_i + X'_i}{2}.$$

В объединении подобного рода величина ковариации не увеличивается.

Другим способом, не противоречащим идее дополняющей выборки, является переключение в парных прогонах потоков случайных чисел, применяемых для моделирования интервалов между прибытиями и продолжительностей обслуживания. Дополняющий характер этого метода можно объяснить, например, тем, что длительные интервалы между прибытиями снижают потенциальную загрузку, в то время как длительное обслуживание ее увеличивает. В ряде работ подытожены результаты применения дополняющей выборки. Как правило, эти результаты получены на имитационных моделях малой размерности. В ней предлагается также выраженная в процентах мера оценки понижения дисперсии, т. е. если обозначить через $\text{Var}_R[\bar{X}_I]$ оценку, полученную с помощью метода понижения дисперсии, то

$$\text{Процент понижения дисперсии} = \frac{\text{Var}[\bar{X}_I] - \text{Var}_R[\bar{X}_I]}{\text{Var}[\bar{X}_I]} \cdot 100.$$

В ряде работ понижение дисперсии оценивается отношением

$$\text{Понижение дисперсии} = \frac{\text{Var}[\bar{X}_I]}{\text{Var}_R[\bar{X}_I]}.$$

В заключение отметим два важных момента. Во-первых, хотя корреляция между дополняющими случайными числами действительно равна $-1,0$, корреляция между наблюдениями, основанными на этих числах, как правило, не равна $-1,0$. Корреляция, равная -1 , имеет место для выборок из симметричного распределения. В ряде работ показано, что корреляция между дополняющими выборками из экспоненциального распределения равна $-0,645$. Подобное уменьшение абсолютной величины отрицательной корреляции можно наблюдать и для других распределений.

Во-вторых, если в модели имеют место отношения второго (или более высокого четного) порядка, введение отрицательной корреляции может привести не к уменьшению, а, наоборот, к увеличению дисперсии выборочного среднего.

2.9.4.2. Общие потоки случайных чисел

Обычной практикой в имитационном моделировании является использование ранее полученных исходных данных в качестве источника. Примером может служить имитация работы вычислительного центра по данным о поступлении работ на ЭВМ для определения момента поступления и характеристик работ. **Имитационное моделирование с использованием ранее полученных исходных данных называют иногда моделированием, управляемым по предыстории**. Поскольку записанный ранее сценарий прибытий представляет собой единственный временной ряд, очевидно, что повторное его использование уменьшит дисперсию результата имитационного моделирования. При запуске различных имитационных прогонов с одного и того же корневого случайного числа (т. е. используя один и тот же случайный поток) можно добиться понижения дисперсии. При работе с данными предыстории и общими потоками необходимо учитывать, что модель не полностью отображает разброс характеристик исследуемой системы. **Аналитик должен быть уверен, что полученный таким образом отдельный временной ряд адекватно отображает моделируемый стохастический процесс.** Применение общих потоков наиболее заманчиво при сравнении альтернатив. В этом случае нас интересует дисперсия разности между выборочными средними, т. е.

$$\text{Var} [\bar{X}_1^{(1)} - \bar{X}_1^{(2)}] = \text{Var} [\bar{X}_1^{(1)}] + \text{Var} [\bar{X}_1^{(2)}] - 2 \text{Cov} [\bar{X}_1^{(1)}, \bar{X}_1^{(2)}],$$

где $\bar{X}_1^{(k)}$ — выборочное среднее для альтернативы k. При использовании общих потоков $\text{Cov} [\bar{X}_1^{(1)}, \bar{X}_1^{(2)}]$ должна быть положительной, и в результате достигается понижение дисперсии. Использование общего потока в данном случае предполагает, что сгенерированная последовательность одинаковым образом воздействует на обе альтернативы. Однако необходимо проявлять предельную осторожность в тех случаях, когда общие потоки применяются одновременно с методом дополняющей выборки, поскольку в ряде подобных ситуаций наблюдалось увеличение дисперсии.

2.9.4.3. Использование априорной информации

Приведенная в приложении к данному разделу теорема Рао-Блэкуэлла интерпретируется следующим образом: **понижение дисперсии может быть получено при оценке выборочного среднего с помощью априорной случайной величины**. Одной из процедур, реализующих данный подход, является применение в процессе оценки аналитических зависимостей. Рассмотрим эту процедуру на двух примерах.

Хорошо известно, что для широкого класса моделей массового обслуживания выполняется формула Литтла:

$$L = \lambda W,$$

где L — ожидаемое число заявок в системе, W — ожидаемое время нахождения заявки в системе, λ — реальная интенсивность поступления заявок в систему (т. е. число поступивших за единицу времени заявок, принятых к обслуживанию, в конечном счете обслуженных).

В течение прогона среднее число заявок в очереди \bar{N} будет равно произведению наблюдаемой интенсивности поступления λ_0 на среднее время ожидания в очереди всех заявок \bar{T} , т. е. $\bar{N} = \lambda_0 \bar{T}$. Это уравнение можно получить, исходя из соображения, что интегрированное по времени число заявок в очереди равно сумме времен ожидания (предполагается при этом, что все заявки покинули очередь до конца прогона). Основываясь на приведенной информации и предполагая, что значение λ известно заранее, можно вычислить

$$\text{Var} [\bar{N}] = \lambda^2 \text{Var} [\bar{T}].$$

Таким образом, это уравнение косвенно оценивает дисперсию среднего числа заявок в системе посредством умножения оценки $\text{Var} [\bar{T}]$, полученной по результатам имитации, на величину λ^2 . Средством понижения дисперсии в данном случае является применение теоретической интенсивности поступления λ при оценивании $\text{Var} [\bar{N}]$. В реальной имитации, когда используется процедура выборки, вычисленная по выборке интенсивность применима для всего имитационного прогона. В этом случае, если рассматривать λ как независимую случайную величину, можно написать уравнение

$$\text{Var} [\bar{N}] = \text{Var} [\lambda] \text{Var} [\bar{T}] + E^2 [\lambda] \text{Var} [\bar{T}] + \text{Var} [\lambda] E^2 [\bar{T}]$$

и, следовательно, ожидать понижение дисперсии в случае, когда λ — константа, т. е. $\text{Var} [\lambda] = 0$.

Приведенные выше соображения дают основания утверждать, что должно наблюдаться понижение дисперсии. Однако остается еще ряд вопросов. Например, следует ли прямо оценивать $\text{Var}[\bar{N}]$ и использовать ее для косвенной оценки $\text{Var}[\bar{T}]$ или наоборот? В ряде работ рассматриваются **пять уравнений**, определяющих первые моменты для величин: **число заявок в системе, длина очереди, время пребывания заявки в системе, время пребывания ее в очереди и загрузка системы**. С помощью оценок, полученных в регенеративных процедурах, в работах показано аналитически (при предположении стационарного характера исследуемых величин), что в системах массового обслуживания типа M/G/1 более эффективно (т. е. с меньшей дисперсией выборочных средних) использовать косвенные оценки, основанные на дисперсии времени пребывания в системе.

Другой пример использования априорной информации, в котором проанализирована система управления запасами, допускающая просроченные заказы. В подобных ситуациях необходимо оценивать ожидаемое число просроченных заказов. Однако просроченные заказы могут возникать редко, и поэтому требуются длительные прогоны модели для получения оценок среднего количества просроченных заказов с низкой оценкой дисперсии. Получено выражение для вычисления ожидаемого количества просроченных заказов при известном состоянии склада в начале периода, предшествующего получению запроса. Показано, что для t-го периода

$$E[B_t|A_t] = E[D_t] - A_t + \sum_{d=0}^{A_t} (A_t - d) P[D_t = d],$$

где B_t — количество просроченных заказов на t-м периоде, A_t — состояние склада в начале t-го периода, D_t — объем запроса на t-м периоде.

Оценка среднего количества просроченных заказов по результатам имитации производится с помощью наблюдения значений A_t и решения приведенного выше уравнения для $E[B_t|A_t]$ при $t = 1, 2, \dots, T$. Затем получают средние величины для всех T периодов.

Описанная процедура дает понижение дисперсии в 3,89 и 8,79 раза при двух различных наборах параметров. Она является прямым приложением теоремы Рао-Блэкуэлла и использует априорную информацию, касающуюся распределения запроса D_t при оценке среднего количества просроченных заказов.

Использование априорной информации с целью понижения дисперсии весьма привлекательно, так как позволяет совместно применять аналитические и экспериментальные процедуры. Поскольку

всегда возможно прямое оценивание, легко осуществима проверка эффективности данного метода понижения дисперсии. А так как всегда обсуждается надежность результатов, понятно, что должно использоваться множество способов оценки дисперсии выборочного среднего. В разд. 2.9.1 был поставлен вопрос об ожидаемом разбросе при повторном имитационном эксперименте. Этим неявно предполагается, что при сборе и анализе статистических данных для вычисления дисперсии применяются одни и те же процедуры.

При использовании априорной информации возникает проблема вывода уравнений для вычисления выборочных данных.

2.9.4.4. Использование управляющих переменных для понижения дисперсии

Принцип метода управляющих переменных состоит в выявлении переменной Y, имеющей положительную ковариацию с исследуемой переменной X. Если такая управляющая переменная существует и мы умеем задавать для нее теоретические математические ожидания, то можно добиться понижения дисперсии исследуемой переменной. Для того чтобы убедиться в этом, рассмотрим следующее уравнение, объединяющее выборочные средние величины \bar{X}_1 и \bar{Y}_1 и формирующее новую случайную величину \bar{Z}_1 :

$$\bar{Z}_1 = \bar{X}_1 + (E[Y] - \bar{Y}_1).$$

Ясно, что $E[\bar{Z}_1] = E[\bar{X}_1]$, если оценка $E[Y]$ является несмещенной. Для дисперсий можно записать выражение

$$\text{Var}[\bar{Z}_1] = \text{Var}[\bar{X}_1] + \text{Var}[\bar{Y}_1] - 2 \text{Cov}[\bar{X}_1, \bar{Y}_1].$$

Из этого уравнения видно, что $\text{Var}[\bar{Z}_1] < \text{Var}[\bar{X}_1]$, если $\text{Var}[\bar{Y}_1] < 2 \text{Cov}[\bar{X}_1, \bar{Y}_1]$.

В области теории управляющих переменных проведены обширные исследования. Изучен обобщенный случай с многими переменными и весовыми коэффициентами:

$$\bar{Z}_1 = \bar{X}_1 + \sum_{k=1}^K w_k (E[Y_k] - \bar{Y}_{1k}).$$

Однако еще недостаточно публикаций по практической реализации метода управляющих переменных. Обычно в качестве управляющих переменных рассматривают входные переменные (предполагая, что выходные переменные положительно коррелированы с входными) и

выводят требуемые зависимости при некоторых ограничивающих предположениях об имитационной модели.

Идея метода управляющих переменных весьма проста. Тем не менее качественная оценка этого метода может быть дана только на основе результатов его дальнейшего применения на практике.

2.9.4.5. Другие методы понижения дисперсии

В данном разделе, посвященном методам понижения дисперсии, необходимо также упомянуть о процедурах стратифицированной выборки и выборки по значимости. Понижение дисперсии применяется в методе Монте-Карло и в стандартных выборочных экспериментах. В ряде работ приведен достаточно полный обзор этих процедур и опыта их применения содержится. На основании обзора литературы можно прийти к выводу, что для применения в перспективных имитационных разработках эти методы требуют дальнейшего совершенствования, поэтому ниже приводится лишь их краткое описание.

Процедура стратифицированной выборки. Процедура стратифицированной выборки включает выявление величины u , с помощью которой могут быть заданы классы стратификации G_k . Случайные наблюдения X_i стратифицируются путем проверки u_i соответствующих i -м наблюдениям. Если $u_i \in G_k$, то X_i попадает в k -ую страту. При этом предполагается, что $p_k = P(u \in G_k)$ известна, и выборочное среднее на основе стратификации вычисляется следующим образом:

$$\bar{X}_{ST} = \sum_k p_k \bar{X}_k,$$

где \bar{X}_k — выборочное среднее для k -й страты. Можно показать, что \bar{X}_{ST} является несмещенной оценкой μ_X :

$$\text{Var} [\bar{X}_{ST}] = \sum_k p_k^2 \text{Var} [\bar{X}_k] \leq \text{Var} [\bar{X}],$$

т. е. достигается понижение дисперсии с помощью стратификации. Чем больше абсолютная разность между средними страт μ_k и средним популяции μ_X , тем большее понижение дисперсии можно получить.

Процедура значимой выборки. Процедура значимой выборки состоит в переопределении исследуемой переменной с помощью новой функции плотности, задающей большие веса тем значениям X , которые дают наибольший вклад в ожидаемое значение. Например, предположим, что мы собираемся оценить ожидаемое значение $g(X)$,

где X — случайная величина с функцией плотности $f(x)$. В соответствии с законом *бессознательного статистика* имеем

$$E [g(X)] = \int_x g(x) f(x) dx.$$

Внеся под интеграл плотность $h(x)$, получим

$$E [g(X)] = \int_x \frac{g(x)f(x)}{h(x)} h(x) dx.$$

В значимой выборке этот результат используется следующим образом: в соответствии с $h(x)$ отбираются значения x и вычисляются для отобранных x значения выражения $[g(x)f(x)]/[h(x)]$; оценка выборочного среднего получается делением суммы полученных величин на число сгенерированных значений.

2.9.5. Стратегии запуска

Наличие начальных условий запуска имитационной модели обуславливает тот факт, что результаты продолжительного прогона могут отличаться от результатов, полученных вскоре после его начала. Если моделируемая система имеет естественный момент завершения работы, безусловно, присутствует переходный период в процессе ее функционирования, но полученные в течение него данные (хотя и отличаются от данных, полученных в конце) могут рассматриваться как представительные результаты обследования реальной системы. Однако, когда оценивается только стационарный режим функционирования, полученные в начале данные будут искажать оценки стационарных характеристик. Эти вопросы рассматриваются в данном разделе.

Стратегии запуска используются при задании начальных условий имитационной модели и процедуры определения точки отсечения d , в которой начинается сбор выборочных данных для вычисления оценок. В общем случае при задании начальных условий необходимо стремиться к такому запуску, при котором требуется отсечение минимального объема данных, т. е. используется малое значение d . При задании точки отсечения нужно руководствоваться двумя соображениями. С одной стороны, стирание значений, полученных в начале прогона, уменьшает смещение результирующих оценок. С другой стороны, стирание этих значений может увеличить оценку $\text{Var} [X_i]$, поскольку она будет вычислена по меньшему количеству наблюдений. Последнее положение предполагает, что стираемые значения являются наблюдениями, имеющими разброс,

подобный разбросу наблюдений установившегося режима. Однако это может быть и не так.

Из сказанного выше следует, что при определении стратегии запуска **необходим компромисс между уменьшением смещения и понижением дисперсии**. Таким образом, вполне естественно использование среднеквадратичной ошибки и доверительных интервалов в качестве оценки стратегий запуска. Процедура подобной оценки разработана и изложена в различных работах, где приводятся также обзор исследований по данной тематике и ряд деталей, относящихся к анализу упомянутого выше компромисса.

В этом разделе мы предлагаем правила задания начальных условий и ряд процедур отсечения. Вначале сделаем несколько замечаний. Прежде всего необходимо рассматривать стратегии запуска в совокупности с процедурами оценки. Если для получения оценок используются регенеративные методы, стратегия запуска весьма проста: модель стартует в точке регенерации, а поэтому первый цикл начинается немедленно и усечение не требуется. Если же процедура оценки построена на применении отдельных временных рядов, запуск осуществляется только один раз за весь прогон, и поэтому отсечением не следует пренебрегать. Кроме того, при проведении повторных прогонов стратегия запуска используется неоднократно, что обуславливает необходимость особой тщательности ее определения.

Следует сказать также несколько слов об исследованиях, ведущихся в области стратегий запуска. Теоретические результаты пригодны только для хорошо изученных моделей малой размерности, в которых разброс значений выборки в течение переходного периода слабо отличается от разброса характеристик в стационарном режиме. Поэтому теоретические исследования направлены здесь на то, чтобы выяснить, надо ли вообще производить отсечение. Практические же разработки показывают, что отсечение является разумной стратегией, которой нужно следовать в большинстве случаев. Это особенно важно для моделей работы заводского цеха или конвейерной системы, в которых необходимо воспроизвести множество операций, прежде чем система будет полностью «загружена». Обратимся опять к проблеме задания начальных условий.

2.9.5.1. Задание начальных условий

Идеальным способом задания начальных условий могла бы быть выборка из распределений установившегося режима той имитационной модели, для которой определяются начальные условия. Повторное применение этой процедуры давало бы больше уверенности в хорошем

качестве статистических оценок установившегося режима. Очевидно, однако, что это уже явный перебор, поскольку знание распределений стационарного режима исключает необходимость использования в дальнейшем имитационного моделирования. Чтобы избежать этой ситуации, предлагаются три правила задания начального состояния модели:

1. Запустить модель с состояния «пуст и свободен».
2. Запустить модель с модальных значений установившегося режима.
3. Запустить модель со средних значений установившегося режима.

Преимуществом первого правила является простота его применения. Недостаток — непредставительность начального состояния. Для моделей малой размерности, например для системы массового обслуживания типа $M|M|1$, состояние «пуст и свободен» типично, поскольку оно является модальным состоянием.

Второе правило задает в качестве начального наиболее вероятное (т. е. с максимальной вероятностью наступления) состояние. Экспериментальный анализ, проведенный для ряда моделей, показал, что это наилучшее правило. Основным недостатком его является невозможность определения модального состояния для модели большой размерности.

Третье правило рекомендует в качестве начального задавать ожидаемое или усредненное состояние модели. Преимуществом этого правила является то, что усредненное состояние может быть оценено с помощью пробного исследования или анализа похожей аналитической модели. Интуитивно понятно, что запуск модели с ожидаемого состояния обеспечит представительный разброс начальных наблюдений. Тем не менее недостаточно публикаций, демонстрирующих, что задание усредненного состояния в качестве начального дает улучшенные статистические оценки.

2.9.5.2. Процедуры отсечения

Простейшая процедура отсечения состоит в определении момента времени, когда должен инициироваться сбор данных. Практически в имитационных моделях подобное правило реализуется уничтожением в момент отсечения всех собранных значений выборки.

Вопрос, возникающий при задании точки отсечения, состоит в том, как определить время, когда необходимо осуществить отсечение. Одним из наиболее часто применяемых подходов является проведение пробного прогона и выбор момента отсечения по его результатам. Не пользуясь, как правило, формальными методами, исследователь работает с такими количественными характеристиками, как:

количество последовательных уменьшений или увеличений значений выборки; разность между последовательными групповыми средними или последовательными накопленными средними; пересечение значениями выборки средних величин. Многие авторы пытаются формализовать эти понятия и сформулировать правило, которое можно было бы использовать в имитационном моделировании для определения момента отсечения.

Краткий перечень этих правил представлен в табл. 1.

Таблица 1.

Правила отсечения	
Автор	Правило
Конвэй	Установить значение d таким образом, чтобы $x(d+1)$ не было ни максимальным, ни минимальным значением из $\{x(n): n=d+1, \dots, N\}$
Фишман	Установить значение $d=n$, где $\{sgn(x(t) - \bar{x}_n): t=0, 1, 2, \dots, n\}$ содержит k прогонов (k — параметр, который необходимо определить). Это правило соответствует установлению $d=n$, когда значения \bar{x}_n временного ряда $\{x(t): t=0, 1, 2, \dots, n\}$ пересекли \bar{x}_n по крайней мере $(k-1)$ раз
Шрайбер	Установить значение $d=n$, когда все групповые средние k последних групп размера b попадают в интервал размером ϵ
Фишман	Установить d так, чтобы число отброшенных наблюдений было «эквивалентно» одному независимому наблюдению, где отношение числа зависимых наблюдений к числу независимых равно m/R_0 . Следовательно, $d = m/R_0 - 1$
Гордон	Провести k повторных прогонов для вычисления $Var[\bar{x}_n]$, $n=1, 2, \dots$. Установить $d=n$, начиная с которого $Var[\bar{x}_n]$ снижается в пропорции $1/n$
Гафариан и др.	Установить d равным наименьшему n , для которого $x(n)$ не является ни минимумом, ни максимумом из всех предыдущих наблюдений $\{x(t): t=0, 1, 2, \dots, n\}$

В обозначениях, применяемых в данной таблице, опущен индекс номера группы и используются строчные буквы, поскольку правила основываются непосредственно на наблюдаемых значениях выборки.

Многие оценки проводились на моделях малой размерности, и было показано, что отсечение можно не проводить. Исходя из изложенных выше соображений, эти результаты нельзя распространять на модели большой размерности.

Применяя предложенные правила отсечения, необходимо помнить о следующих четырех имеющих место на практике фактах:

1. Ожидаемое значение выборочного среднего меньше ожидаемых значений переменной процесса, если система первоначально пуста, и выше, если система полностью загружена.
2. Пересечения средних не столь вероятны, как это можно предполагать.
3. Правила отсечения крайне чувствительны к заданию параметров. К тому же процедуры определения параметров для многих предлагаемых правил не разработаны.
4. При длительном переходном периоде применение правила отсечения может оказаться дорогостоящим из-за расточительного расхода машинного времени.

Эти факты приводятся в пользу отказа от процедур отсечения и прямого задания момента отсечения при инициализации статистических переменных.

2.9.6. Правила остановки

Определение продолжительности имитационного прогона, выраженной числом групп, является довольно сложной проблемой. Если мы можем допустить, что \bar{X}_I не смещено, а $Var[\bar{X}_I] = \sigma_x^2/I$, то число групп I , необходимое для получения доверительного интервала $1-\alpha$, в который должно попадать μ_x , может быть вычислено по стандартной статистической формуле. Пусть

$$P[\bar{X}_I - g \leq \mu_x \leq \bar{X}_I + g] \geq 1 - \alpha,$$

где g — заданная половина доверительного интервала. Если положить $Z = \sqrt{I}(\bar{X}_I - \mu_x)/\sigma_x$, получим

$$P\left[|Z| \leq \frac{g\sqrt{I}}{\sigma_x}\right] \geq 1 - \alpha,$$

причем это равенство выполняется, начиная с некоторого наименьшего значения I , например со значения I^* . Предположив, что I^* при этом достаточно велико, чтобы удовлетворять центральной предельной теореме, получим

$$I^* = \left(\frac{\sigma_x}{g} Z_{\alpha/2}\right)^2,$$

где

$$Z_{\alpha/2} \text{ такое, что } \frac{1}{\sqrt{2\pi}} \int_{Z_{\alpha/2}}^{\infty} e^{-y^2/2} dy = \alpha/2.$$

Уравнение для I^* требует знания σ_x . Обычно задается g , выраженное через σ_x , т. е. $g = v\sigma_x$ для $v > 0$. В этом случае I^* может

вычисляться без задания σ_x . Величина I^* для различных комбинаций v и α приводится в табл. 2.

Таблица 2

I^*	α		
	0.02	0.05	0.10
v			
0.01	54 093	38 416	27 060
0.10	541	384	271
0.20	135	96	68
0.50	22	15	11

Из этой таблицы видно, что для получения доверительного интервала $(\bar{X}_I - 0,1\sigma_x, \bar{X}_I + 0,1\sigma_x)$, в который с 95%-ной вероятностью попадает μ_x , необходимо иметь около 400 групп. Аналогичный анализ проводится и при определении размера выборки с целью получения заданного доверительного интервала для дисперсии или для значения вероятности.

В этом разделе предлагается в качестве выборочной величины X_i применять групповое среднее. В связи с этим необходимо сформулировать ряд предположений, чтобы быть уверенными о состоятельности используемой процедуры. В частности, если предположение о независимости неприемлемо, можно считать $\text{Var}[\bar{X}_I] = m/I$ и заменить σ^2_x в предыдущем уравнении на m .

Обычно вместо s_x используют σ_x (или \hat{m} вместо m), и тогда $\sqrt{I}(\bar{X}_I - \mu)/s_x$ имеет t -распределение. В имитационных исследованиях I бывает обычно достаточно большим, чтобы рассматривалась нормальная аппроксимация t -распределения. Для задания значения I^* перед запуском имитационной модели необходимо знание величины s_x . В ряде случаев осуществляют пробные исследования для получения значения s_x , на основе которого оценивают I^* . Более общим подходом является применение последовательного правила остановки.

Это правило задает условие, выполнение которого свидетельствует о достижении желаемой цели. Показано, что если X_i являются независимыми и одинаково распределенными нормальными величинами, то

$$P\{\bar{X}_{I^*} - g \leq \mu \leq \bar{X}_{I^*} + g\} \geq \begin{cases} 0,928 & \text{для } 1 - \alpha = 0,95, \\ 0,985 & \text{для } 1 - \alpha = 0,99. \end{cases} \quad (21)$$

Величина I^* задается следующим образом:

$$I^* = \min \{I : I \geq 3, \text{ нечетно}; s^2_x \leq I g^2 / t^2_{\alpha/2, I-1}\}, \quad (22)$$

где $t^2_{\alpha/2, I-1}$ соответствует квантилю $1 - \alpha/2$ t -распределения Стьюдента с $I - 1$ степенями свободы, например $t_{0,025,10} = 2,228$.

Уменьшение доверительного интервала связано с тем, что тест осуществляется последовательно. Предложено использовать уравнение (22) без учета требования нечетности I , которое вытекает из-за несходимости процедуры в случае четности I . Применение уравнения (22) в имитационном эксперименте требует наличия таблицы значений t -распределения, задания g и размера группы, а затем периодической проверки вычисляемых значений s_x до тех пор, пока они не станут меньше заданных в уравнении (22). При выполнении этого условия доверительный интервал устанавливается в соответствии с уравнением (21).

При необходимости работы с относительными величинами (т. е. когда $-v\mu \leq \bar{X}_I - \mu \leq v\mu$, где $v > 0$) можно показать, что правило остановки, записанное в виде

$$I^* = \min \{I : s^2_x \leq [(Iv\bar{X}_I/t_{\alpha/2, I-1})^2 - 1]/(I - 1)\}, \quad (23)$$

даст предельный доверительный интервал

$$\lim_{v \rightarrow 0} P\{\bar{X}_{I^*}/(1+v) \leq \mu_x \leq \bar{X}_{I^*}/(1-v)\} = 1 - \alpha.$$

Отметим, что при больших I и $g = v = \bar{x}_I$ правило остановки, определяемое формулой (23), будет аппроксимацией правила, определяемого уравнением (22).

Поскольку при использовании правил остановки, определяемых уравнениями (22) и (23), наблюдается уменьшение вероятности попадания в доверительный интервал, рекомендуется в особо ответственных случаях устанавливать значение I^* не как минимальное число групп, для которых выполняется условие относительно s^2_x , а как значение I , следующее вторым после минимального. Обычно это приводит к увеличению числа групп всего лишь на 1, но иногда это увеличение может быть и большим. Таким способом можно компенсировать как уменьшение (в соответствии с уравнением (21)) вероятности попадания в доверительный интервал, так и свойственное процедуре вычисления s^2_x занижение оценки дисперсии, связанное с предположением о независимости и одинаковом распределении групповых наблюдений.

Кроме методов определения размера выборки с целью достижения заданных параметров доверительного интервала существует еще ряд практических вопросов, касающихся остановки имитационного прогона. Например, что делать с компонентами, присутствующими в модели в момент завершения прогона? Ответ на этот вопрос тесно

связан с существом решаемой задачи. Если такие компоненты являются представительными и о них собран достаточный статистический материал, дальнейшая их обработка не требуется. Однако, если они нетипичны или же о них не собрано достаточно информации, должна быть предусмотрена их обработка. Например, в модели цеха, в котором последовательность выполнения работ определяется по правилу выбора самой непродолжительной из имеющихся работы, в конце прогона могут остаться незавершенными те работы, время выполнения которых относительно велико. Оставив такие работы невыполненными, можно получить значительное смещение оценки времени пребывания работ в системе. Необходимо проявлять осторожность во избежание подобной ситуации.

Более общая процедура задания условия остановки основана на понятии предельной отдачи. Смысл его в том, что прогон должен быть остановлен в том случае, если предельное улучшение (потенциальная польза прогона) становится ниже предельной стоимости продолжения прогона. Этот подход вполне разумен, однако оценка потенциальной пользы и вычисление предельной стоимости являются весьма сложными проблемами.

2.9.7. Планирование эксперимента

Имитационный прогон — это эксперимент, в котором осуществляется оценка функционирования системы при заданном наборе условий. В терминах теории планирования эксперимента условия называются **факторами**, а некоторое конкретное значение фактора называется **уровнем**. В данном разделе только очерчен круг проблем, касающихся планирования эксперимента. Статистические методы, связанные с планированием эксперимента, достаточно полно изложены в работах А.Е.Кононюка. Такие разработки, как анализ дисперсии (ANOVA), тест Шапиро-Вилка для проверки предположения о нормальности или тест Ньюмана-Койлса для выявления значимого отличия между парами оценок средних, могут применяться в имитационных исследованиях так же, как они применяются и в других областях.

Основной проблемой в имитационных экспериментах является определение пространства выводов о реальной системе, сделанных на основе данной имитационной модели. Выработка априорных предположений о том, насколько смело можно использовать полученные выводы и результаты, а также их разумная интерпретация — вот те аспекты планирования эксперимента, которые наиболее часто

игнорируются в имитационных исследованиях. Возможной причиной этого является наличие в эксперименте многообразных альтернативных факторов; и огромного числа средств управления экспериментом. В предыдущих разделах мы обсудили ряд средств управления, таких, как задание начальных условий и процедуры выборки, определение продолжительности прогона, задание размера группы и процедуры оценки.

Целями имитационного эксперимента в основном являются:

- 1) получение знаний о влиянии управляемых факторов на результаты эксперимента;
- 2) оценка изучаемых параметров системы;
- 3) выбор на множестве альтернатив;
- 4) определение уровней факторов, при которых достигается оптимальный отклик.

При наличии множества факторов подход к достижению первых двух целей заключается в выборе одного из возможных планов проведения эксперимента и осуществлении дисперсионного анализа выбранного плана. План эксперимента определяет комбинацию уровней и для каждой комбинации задает число повторных прогонов имитационной модели. Используя данные эксперимента, одновременно с оценкой факторов ошибки определяют параметры гипотетической модели. Затем строятся графики взаимодействия для определения совместного влияния различных факторов. Значимость каждого фактора определяется на основе производной модели, после чего могут быть вычислены оценки исследуемых параметров системы. Эта процедура напоминает подход параметрического моделирования для одной исследуемой характеристики, описанной выше в данном разделе.

При выборе альтернативы используются статистические процедуры ранжирования и отбора. Много работ содержат обзор состояния разработок в этой области и дают рекомендации по их применению в имитационном анализе. Разработано множество процедур определения необходимого для выбора альтернативы размера выборки, генеральное среднее которой с заданной вероятностью превосходит на заранее определенную величину следующее лучшее генеральное среднее. Тестовые процедуры осуществляют вычисление среднего по выборке заданного размера и отбор наибольшего наблюдаемого среднего. Данный подход разработан Беххофером и называется подходом выявления зоны безразличия.

Другим подходом является группировка альтернатив в статистически эквивалентные подмножества. Процедуры отбора

тодмножеств и сравнение их с помощью подхода выявления зон безразличия приводятся в ряде работ.

Последним моментом, имеющим отношение к планированию эксперимента, является выбор наилучшей альтернативы. Эта дроблема отличается от описанных выше тем, что в процессе выбора определяются значения управляемых переменных, которые максимизируют или минимизируют целевую функцию. Например, при анализе периодической инспекции системы запасов имитацию можно использовать для определения контрольного уровня запасов, точки перезаказа или времени между ревизиями, которые минимизировали бы среднемесячную стоимость эксплуатации склада.

Хотя принципы оптимизации на базе использования имитационных моделей в основном те же, что и на базе аналитических моделей, существует ряд отличий, которые следует иметь в виду. Поскольку имитационный отклик обычно содержит случайную составляющую, то зависимые от отклика целевая функция и ограничения также будут случайными величинами. Поэтому необходимо в вероятностной форме задавать ограничения на отклики и осуществлять статистическую интерпретацию значений целевой функции.

Существует два основных подхода к оптимизации с помощью имитации. Первый подход заключается в прямой оценке с помощью имитационной модели независимых переменных. В работах эти методы разделены на **три категории: математически тривиальные методы, такие, как эвристический поиск, полный перебор или случайный поиск; методы, пригодные для унимодальных целевых функций, такие, как покоординатный спуск или метод конфигураций; методы для мультимодальных целевых функций.**

Вторым подходом к оптимизации с помощью имитации является методология поверхности отклика. Методы данного класса осуществляют аппроксимацию поверхности результатов наблюдений с использованием факторного планирования в окрестности начальной точки поиска. Далее оптимизационный алгоритм, например градиентный метод, используется для определения оптимальных значений управляемых переменных, удовлетворяющих подобранной зависимости. Полученные оптимальные значения аппроксимирующей поверхности формируют следующую точку поиска.

В заключение отметим, что из двух особых аспектов анализа результатов имитации — точности и надежности получаемых оценок выборки — основное внимание в данном разделе уделялось надежности. Установлена важность $\text{Var} [\bar{X}_I]$ для имитационных исследований и дана рекомендация оценивать надежность результатов имитации на основе наблюдения средних по группам или циклам, а не

по отдельным значениям выборки. Представлены пять методов оценки дисперсии выборочных средних $\text{Var} [\bar{X}_I]$ по I группам.

Описаны методы понижения дисперсии, стратегии запуска, правила остановки и планирование имитационных экспериментов. В общем в данном разделе даны как детальные практические рекомендации, так и характеристики важнейших статистических проблем, с которыми сталкивается исследователь в своей работе.

Приложение

Теоремы, применяемые в имитационном моделировании

Ниже представлены теоремы и уравнения, имеющие непосредственное отношение к анализу результатов имитационного моделирования

Теорема полной вероятности

Вероятность результата В, который может возникнуть лишь при условии возникновения одного из несовместимых результатов A_i , образующих полную группу, равна сумме произведения условных вероятностей на вероятности возникновения A_i , т. е.

$$P [B] = \sum_i P [B|A_i] P [A_i].$$

Аналогичный результат получается и для математического ожидания случайной величины Y

$$E [Y] = \sum_i E \{Y|X = x_i\} P [X = x_i].$$

Совместные вероятности

Вероятность возникновения совместного результата, описываемого множеством случайных величин, равняется произведению условных вероятностей

$$P \{Y_1, Y_2, \dots, Y_n\} = P \{Y_1\} P \{Y_2|Y_1\} P \{Y_3|Y_1, Y_2\} \dots P \{Y_n|Y_1, \dots, Y_{n-1}\}.$$

Если предположить, что случайные величины обладают марковским свойством, т. е. $P \{Y_j|Y_1, Y_2, \dots, Y_{j-1}\} = P \{Y_j|Y_{j-1}\}$, получим

$$P \{Y_1, Y_2, \dots, Y_n\} = P \{Y_1\} P \{Y_2|Y_1\} P \{Y_3|Y_2\} \dots P \{Y_n|Y_{n-1}\}.$$

В случае независимости величин имеем

$$P \{Y_1, Y_2, \dots, Y_n\} = P \{Y_1\} P \{Y_2\} \dots P \{Y_n\}.$$

Случайная сумма независимых случайных величин

Если X_1, X_2, \dots, X_K — независимые и одинаково распределенные случайные величины, а K — случайная дискретная величина, не зависящая от X_i , то для суммы

$$Y = \sum_{i=1}^K X_i$$

выполняются следующие соотношения.

$$E[Y] = E[X] E[K],$$

$$Var[Y] = E[K] Var[X] + Var[K] E^2[X].$$

Формула замены переменных

Пусть дано

$$Y_j = g_j(W_1, W_2, \dots, W_n), \quad j=1, 2, \dots, n,$$

тогда f_y — совместная плотность распределения Y_j — имеет вид

$$f_y(y_1, y_2, \dots, y_n) = f_w(w_1, w_2, \dots, w_n) \frac{1}{|J|},$$

где $f_w(\cdot)$ — совместная плотность распределения W_1, W_2, \dots, W_n ; J — якобиан, представляющий собой определитель матрицы

$$\begin{pmatrix} \frac{dg_1}{dw_1} & \frac{dg_1}{dw_2} & \dots & \frac{dg_1}{dw_n} \\ \vdots & \vdots & & \vdots \\ \frac{dg_n}{dw_1} & \frac{dg_n}{dw_2} & \dots & \frac{dg_n}{dw_n} \end{pmatrix},$$

а $|J|$ — абсолютное значение J . Эта формула задает процедуру преобразования случайных величин.

Асимптотическая нормальность рекуррентных событий

Если рекуррентное событие является интервальным, а интервал между событиями имеет конечные значения среднего и дисперсии σ^2 , то T_r и N_t распределены асимптотически нормально, где T_r — интервал времени до наступления r -го события, причем $E[T_r] = r\mu$ и $Var[T_r] = r\sigma^2$; N_t — количество наступлений события за t единиц времени, причем $E[N_t] = t/\mu$ и $Var[N_t] = t\sigma^2/\mu^3$. Например, если при

имитации время между прибытиями имеет $\mu=10$ и дисперсию $\sigma^2 = 4$, то число прибытий за $t = 1000$ единиц времени распределено асимптотически нормально со средним 100 и дисперсией 4.

Приведенное выше утверждение представляет собой центральную предельную теорему для последовательности зависимых величин и может быть использовано при проверке допустимости генераторов входных данных имитационной модели.

В дополнение к приведенной центральной предельной теореме для рекуррентных событий существует центральная предельная теорема, задающая условия нормальности выборочного среднего стационарного стохастического процесса.

Первый закон арксинуса

Пусть имеется биномиальная случайная величина Y_n с $P\{Y_n=1\}=1/2$ и $P\{Y_n=-1\}=1/2$ для $n=1, 2, \dots, N$. Рассмотрим последовательность частных сумм

$$Z_n = \sum_{i=1}^n Y_i$$

для всех моментов времени, включая N . При фиксированном α ($0 < \alpha < 1$) сосредоточим внимание на результате эксперимента, в котором $Z_n > 0$ не более, чем в $N\alpha$ в моментах времени, т. е. нас интересуют эксперименты, в которых последовательность $\{Z_n; 1 \leq n \leq N\}$ расположена на положительной стороне оси более чем в α процентов из N моментов времени. При $N \rightarrow \infty$ вероятность подобного результата стремится к

$$\frac{2}{\pi} \sin^{-1}(\sqrt{\alpha}).$$

Например, вероятность того, что доля времени меньше чем $\alpha = 0,976$ равна 0,90. Иными словами, с вероятностью 0,90 процесс проведет около 97,6% времени на той стороне оси, куда он попал в самом начале.

Другой результат состоит в следующем: на отрезке времени длиной $2N$ вероятность того, что число частных сумм $\{Z_n\}$, равных нулю, будучи по крайней мере $\alpha\sqrt{2N}$, стремится к

$$\sqrt{\frac{2}{\pi}} \int_0^\alpha e^{-q^2/2} dq \quad \text{при } N \rightarrow \infty,$$

Например, если сделано 10 000 выборок Y_n , то с вероятностью 0,50 можно утверждать, что $\sum Y_n = 0$ будет наблюдаться не более чем 68 раз. Аналогичный результат, приведенный Феллером, показывает, что

число смен знака пересечений в последовательности частных сумм в течение N единиц времени увеличивается пропорционально \sqrt{N} , т. е. на интервале, равном $100N$ единиц времени, можно ожидать увеличения числа смен знака только в 10 раз по сравнению с интервалом продолжительностью в N единиц. Эти теоремы иллюстрируют нетривиальное поведение даже относительно простых стохастических процессов. Приведенные выше соотношения отображают потенциальные трудности, связанные с использованием в имитационной модели проведения статистического анализа возврата в некоторое состояние, или пересечения некоторого условия состояния.

Теорема Рао-Блэкуэлла

Пусть X и Y — случайные величины, такие, что Y имеет среднее μ и дисперсию $\sigma_Y^2 > 0$. Пусть $E\{Y|x\} = \phi(x)$. Тогда

$$E\{\phi(X)\} = \mu, \text{ а } \sigma_{\phi(X)}^2 \leq \sigma_Y^2.$$

Эта теорема утверждает, что если нас интересуют статистические свойства случайной величины Y и мы можем определить зависимую величину $\phi(X)$, которая является математическим ожиданием Y , обусловленным величиной X , то можно оценить μ на основании математического ожидания $\phi(X)$, причем дисперсия этой оценки будет по крайней мере не больше дисперсии прямой оценки. Эта теорема определяет, таким образом, стоимость использования априорной информации при оценке выборочных средних.

2.10. Применения имитационного моделирования

2.10.1. Области применения

Как уже говорилось, имитационные и статистические методы моделирования находят широкое применение в различных отраслях промышленности. В этой главе кратко описаны конкретные разработки, иллюстрирующие полезность метода имитационного моделирования и дающие представления о тех проблемных областях, в которых имитация является составной частью реального процесса принятия решений, и о тех вопросах, на которые она позволяет находить ответ. Конкретные процедуры разработки представленных имитационных моделей изложены в работах, на которые имеются соответствующие ссылки.

Основное внимание в данном разделе уделено целям разработки и результатам использования моделей. Задача проектирования и разработки моделей для промышленных приложений не является трудной. Обычно подобные модели имеют большую размерность, но концептуально они не более сложные, чем представленные ниже. Важным моментом при разработке конкретной имитационной модели является понимание того, что она может быть построена с помощью различных подходов. Следовательно (и это иллюстрируют приведенные ниже примеры), в процессе моделирования не следует придерживаться некоторого жестко заданного набора правил. Наоборот, применение новых подходов и правил можно только приветствовать.

2.10.2. Анализ потока запросов в филиале фирмы по страхованию жизни и имущества

Информационная система филиала фирмы централизованно обрабатывает поступающие страховые требования в вычислительном центре. Документы различных типов проходят через 14 функциональных подразделений (отделов) филиала, в котором работают 150 сотрудников. Система моделировалась в виде сложной системы обслуживания с целью выявления узких мест в потоке запросов и выяснения влияния различных стратегий управления. Модель на языке Q-GERT использовалась для оценки последствий конкретных действий, которые предполагает осуществить руководство. Были исследованы, в частности, следующие действия:

1. Изменение объема выполняемой филиалом работы каждого типа.
2. Изменение состава и требований к поступающей информации.
3. Изменение правил определения приоритета запросов, обрабатываемых в одном и том же функциональном подразделении.
4. Изменение потоков информации в системе.
5. Перераспределение персонала различных подразделений.
6. Изменение общего времени обработки запросов клиентов благодаря повышению квалификации персонала в конкретных областях.

2.10.3. Исследование функционирования терминала грузового аэропорта

Совершенствование управления системой авиаперевозок требует разработки способа измерения производительности грузового аэропорта. В частности, необходимо оценить воздействие флуктуации

потребностей в перевозке грузов на возможность их своевременного удовлетворения. Важным фактором является также использование ресурсов.

Грузы доставляются на терминал аэропорта на автомобилях или самолетах. Поступивший груз разгружается и сортируется по типу, месту назначения и приоритету отправки. Рассортированный груз направляется в места промежуточного хранения, где он накапливается для формирования партии для отправки. Здесь производится подсортировка грузов, взвешивание, проверка и хранение. При этом считается, что подсортированный груз находится в состоянии «готов к перевозке».

Когда готовый к перевозке груз отбирается для формирования партии, он передается к месту погрузки, где объединяется с другими грузами в партию для отправки. Партия грузится затем в самолет.

Модель данной системы на языке Q-GERT была разработана с целью получения ответов на следующие вопросы:

1. Стоит ли внедрять в аэропорту автоматическое оборудование для повышения мощностей обработки грузов?
2. Где следует размещать новое оборудование?
3. Сколько самолетов могут одновременно загружаться в аэропорту?
4. Какие дополнительные ресурсы потребуются в условиях случайного увеличения объема воздушных перевозок?

2.10.4. Оценка эффективности планируемых капитальных вложений в сталелитейной промышленности

На заводе Лакаванна корпорации «Бетлехем стил» рассматривался вопрос о разработке новых устройств для усовершенствования процесса выплавки стали. В план были включены новые операции, состоящие в расплавлении лома и десульфурации горячего металла. Анализ проводился с целью определения потребности в дополнительных вагонетках для горячего металла, обеспечивающих новые операции.

С целью определения способов эффективного использования выделенных капитальных вложений была разработана имитационная модель на языке GASP IV, включающая моделирование различных операций, связанных с доставкой горячего металла от доменных печей к основным кислородным печам. Схематическая диаграмма имитируемых моделью операций показана на рис. 1.

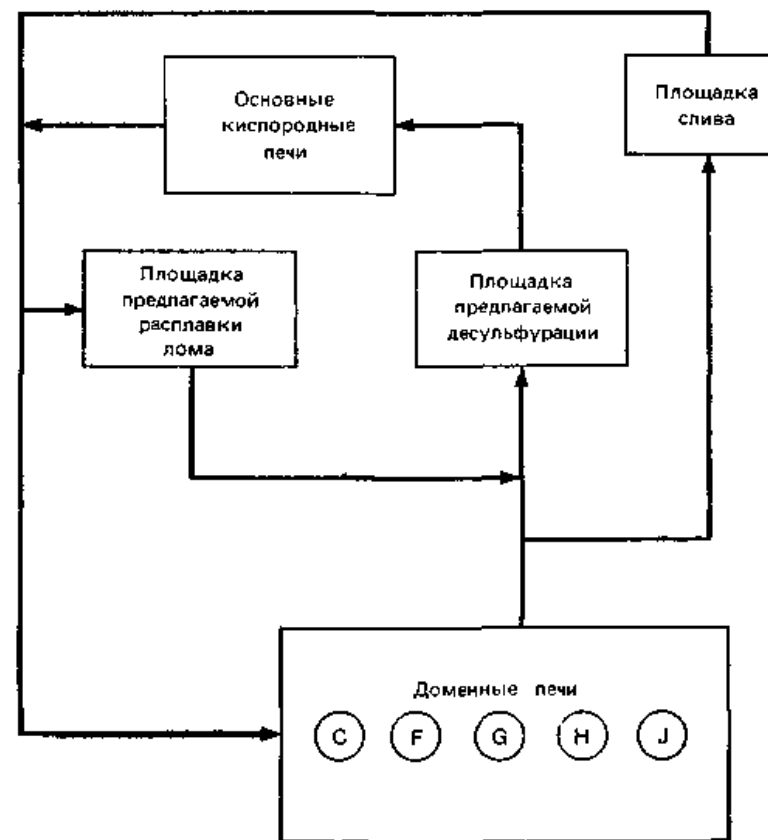


Рис. 1. Движение вагонетки от доменных печей к основным кислородным печам.

Вагонетка представляет собой средство доставки материала, которое транспортирует металл через ряд операций, возвращается назад и вновь повторяет все действия. Потребность в вагонетках для горячего металла зависит от времени плавки металла в доменных печах и в новых печах для лома. Продолжительность плавки планируется, однако реальное время зависит от характеристик горячего металла. Нехватка вагонеток для выплавленного металла является весьма опасной и дорогостоящей ситуацией, которая может привести к аварии печи.

Модель содержит правила планирования маршрута вагонеток через операции десульфурации (если они необходимы) к ковшам, используемым для разлива металла в кислородные печи. Вагонетки требуются также для перевозки горячего металла от плавильных печей для лома. В модель, кроме того, включено правило принятия решения в том случае, когда число вагонеток для разлива плавки из доменной печи оказывается недостаточным. В такой ситуации вагонетка сливает горячим металл и немедленно возвращается к доменной печи.

Имитационное моделирование данной системы в отличие от других проведенных исследований, рекомендовавших дополнительное приобретение трех вагонеток, показало, что, изменив правила распределения существующего количества вагонеток, можно удовлетворять требованиям, связанным с введением новых операций. Таким образом, результатом имитационного анализа были рекомендации, которые помогли избежать лишних, капитальных вложений объемом более 1 млн. долл. Кроме того, были предложены процедуры, позволяющие осуществить дальнейшее усовершенствование процесса выплавки стали.

2.10.5. Проектирование установки для очистки сахарного сиропа

В 1975 г. фирма «А. Е. Стэйли мэньюфэкчуринг» отвела в г. Лафайетт (шт. Индиана) место для строительства сахарного завода, на которое было отпущено 85 млн. долл. Завод был спроектирован с учетом возможности гибкой организации производства, что позволяло бы фирме реагировать на изменение спроса на сахарный сироп различных типов. На рис. 2 показана схема производственного процесса, которая была разработана и предложена в 1975 г. Как видно из схемы, процесс состоит из двух этапов выпаривания, одного этапа очистки на угольных фильтрах и одного этапа ионообмена. Блок-схема отображает также требуемые материальные потоки от этапа к этапу. В производственном процессе в больших объемах используется вода, которая насыщается сахаром и фактически также является сиропом. Кроме того, для сбалансирования операций в процессе применяются накопительные емкости. С целью определения размера емкостей-накопителей, выпаривателей и ионообменников была разработана дискретно-непрерывная модель этого процесса на языке GASP IV.

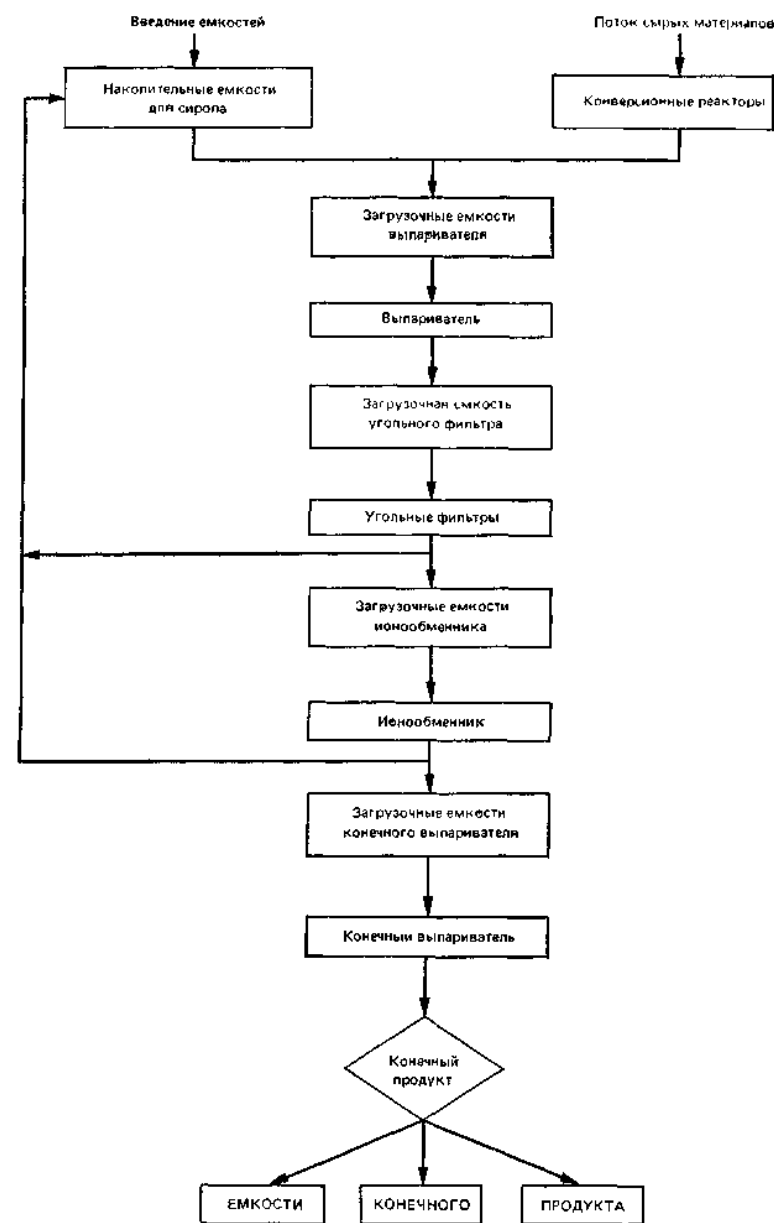


Рис. 2. Блок-схема работы проектируемой установки для очистки сахарного сиропа

На модели проводились эксперименты, которые позволили определить наиболее эффективные с точки зрения стоимости размеры различных устройств. Поскольку оборудование завода узкоспециализировано и имеет большие размеры, время обкатки обычно продолжается около 1,5 года. Поэтому единственным приемлемым способом оценки предлагаемого проекта является имитационное моделирование. Кроме размеров устройств исследовались параметры ручной и автоматизированной систем управления производством. Модель системы управления отображала установку вентилях, регулирующих потоки сиропа из различных емкостей, в каждой из которых поддерживается определенный уровень концентрации. Предполагалось, что концентрация в емкостях измеряется каждый час. **Имитационная модель была объединена с моделью линейного программирования**, определявшей положение клапанов в зависимости от уровней концентрации в емкостях с целью максимизации дохода, связанного с выпуском конечного продукта (конечный продукт является функцией распределения компонентов сахаридов, зависящей от концентрации сиропа в емкостях и от управляющих параметров процесса). Таким образом, используя имитационное моделирование на заключительной стадии проектирования, удалось определить наиболее экономически эффективную стратегию управления по обеспечению необходимого качества продукции и объема производства. Результаты моделирования были учтены при разработке процедур управления.

2.10.6. Оптимальное размещение людских ресурсов в системе

На фирме «Вестерн электрик» с целью определения наилучшей стратегии распределения людей по рабочим местам в соответствии с динамически изменяющимися целями управления была разработана обобщенная имитационная модель на языке GASP IV. Модель использовалась для определения потребностей в людских ресурсах для целого ряда производственных процессов. Одним из таких процессов была намотка скрученного одножильного изолированного проводника. Моделировалась производственная линия, состоящая из 16 скручивающих машин. Каждая машина имеет две головки. Обе головки должны выдавать проводник одинаковой длины, однако по остальным параметрам они независимы. Две приводные бобины подают на машину одножильный изолированный провод. Из машины выходит скрученный провод, который в свою очередь подается на приемные

бобины, установленные на выходе. На любой из 16 машин можно получать провод длиной 10 различных номиналов. Машины останавливаются во время перерывов на второй завтрак и обед, а также в течение коротких промежутков времени могут работать вхолостую (без провода).

Имитационная модель была разработана для исследования функционирования системы в условиях, когда оператор назначается для обслуживания определенного числа головок. В модели предусмотрены следующие события: подход оператора к головке, снятие полностью намотанной бобины, окончание установки пустой приемной бобины и перезапуск машины. Если подающая бобина пуста, работа головки автоматически прекращается и осуществляются события, связанные с установкой полной бобины. Операторы, назначенные более чем на две головки, с целью осуществления необходимых операций должны ходить от одной машины к другой.

В результате моделирования было определено число бобин стандартной длины каждого типа как функция числа головок, назначенных каждому оператору. Кроме того, выводился график загрузки в зависимости от числа головок, назначенных оператору. С помощью графиков определялось оптимальное распределение головок по операторам. Данная имитационная модель использовалась также для моделирования других ситуаций, в которых требовалось оптимальное распределение операторов по рабочим местам.

2.10.7. Анализ риска при строительстве трубопровода

Строительство трубопровода состоит из следующих основных этапов: 1) подготовка места для укладки трубы; 2) укладка трубы и 3) сварка секций трубы. Сопутствующими работами при строительстве трубопровода являются: сборка, разборка и перевозка опорных конструкций; строительство дорог и других транспортных коммуникаций; земляные работы. Строительство трубопровода, например, на Аляске, кроме того, происходит в суровых погодных условиях, которые необходимо учитывать при планировании строительных работ. Была разработана сеть на языке Q-GERT, состоящая из действий по строительству трубопровода и действий по строительству коммуникаций. Модель учитывала воздействие погодных условий на ход строительства. Анализ риска производился программой анализа Q-GERT с целью определения вероятности завершения строительства трубопровода в срок. Проводился также

стоимостный анализ с целью определения последствий срыва сроков строительства. Анализ показал, что нарушение допустимых сроков весьма вероятно. Оценивался также эффект от изменений расписания работ и темпов строительства.

2.10.8. Анализ работы производственной химической установки

Имитационная модель на языке GASP IV, моделирующая производственную химическую установку, состоит из последовательности реакторов, насосов, емкостей и систем фильтрации (рис. 3). Установка производит партии продуктов различных типов с целью удовлетворения запросов потребителей. Описываемая модель процесса является дискретно-событийной, причем события отображают начало и окончание каждого этапа исследуемого процесса. Дискретно-событийная модель является в данном случае вполне допустимой, поскольку состояние системы не меняется до тех пор, пока партия не будет полностью обработана на каждом этапе. Таким образом, событие окончания обработки партии может быть запланировано на основе информации о том, когда началась обработка партии, и характеристик ее обработки на каждом этапе. Составной частью программ событий являются сложные процедуры планирования для определения положения только что обработанной партии и партии, которая должна обрабатываться следующей. Имитационная модель была использована для решения следующих задач:

- 1 Определение воздействия различных сценариев спроса потребителей на требования к производительности установки.
- 2 Определение воздействия требований на время выполнения заявки на производственные затраты, уровень запасов и планирование выпуска продукции.
- 3 Определение воздействия различных процедур планирования производства.
- 4 Определение воздействия различных конфигураций системы (число и объемы емкостей, фильтров, реакторов и трубопроводов).

В качестве примера результата использования модели можно привести следующий. Было определено, что основным узким местом являлся процесс фильтрации. Увеличив вдвое мощность системы фильтрации, можно ожидать 80%-ного увеличения выхода продукта.

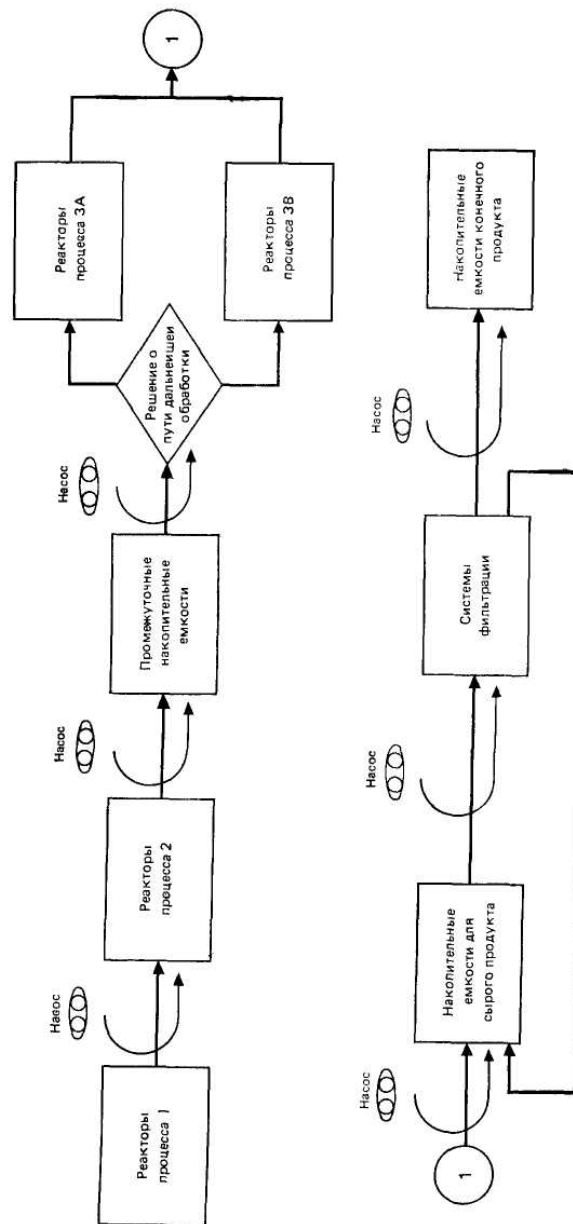


Рис. 3. Общая схема процесса производства

Однако последующее увеличение мощности фильтров указывало на возможность получения только маргинального выигрыша в выходе продукта. Дополнительный анализ показал также, что объемы некоторых емкостей могут быть уменьшены без существенного воздействия на выход продукта. Кроме того, данная имитационная модель, так же как и другие, позволила получить ряд оценок, используемых при анализе функционирования сложной системы

2.10.9. Имитационное моделирование автоматизированного оптового склада

Фирма «Филип Морис» построила новые производственные и складские помещения, оборудованные компьютерами для управления потоком продукции. Руководство фирмы интересовало мощность склада готовой продукции и его узкие места, поскольку оно внедрило иерархическую систему управления, включающую пять мини-компьютеров. Была разработана комбинированная дискретно-непрерывная модель данного объекта на языке GASP IV

С завода на склад поступают ящики с готовой продукцией, где они сортируются, а затем направляются в складское помещение с высокой плотностью размещения для хранения до поступления запроса от потребителей или оптовых магазинов

На рис 4 показана общая схема потока продукции через складскую систему. Дальнейшее описание представленной системы является краткой выдержкой из работы [15]. Поступающие с завода ящики проходят через лазерное сканирующее устройство, которое определяет, на какой из двух ярусов склада следует направить ящик. Затем ящики по двум длинным конвейерам направляются к другому сканирующему устройству, определяющему ряд, в который следует поместить ящик. Далее ящик перемещается до нужного ряда, причем его движение автоматически контролируется оптическим устройством.

После того как ряд заполнен ящиками (заполнен весь ряд плюс пять дополнительных ящиков), он переводится в состояние, позволяющее осуществить автоматический поиск и доставку ящиков. Контрольное сканирующее устройство, установленное на выходном конвейере, проверяет правильность кодов ящиков перед формированием отправки. Сформированный комплект ящиков направляется затем к контрольному устройству, которое обнаруживает случайные ошибки комплектования и на котором комплекты с обоих ярусов могут объединяться и помещаться на один и тот же погрузчик в случае поломки

одного из них. Ящики, не прошедшие проверку на выходном сканирующем устройстве, направляются в зону ручной упаковки. Прошедшие проверку ящики направляются затем в зону автоматической упаковки, где упаковываются по комплектам.

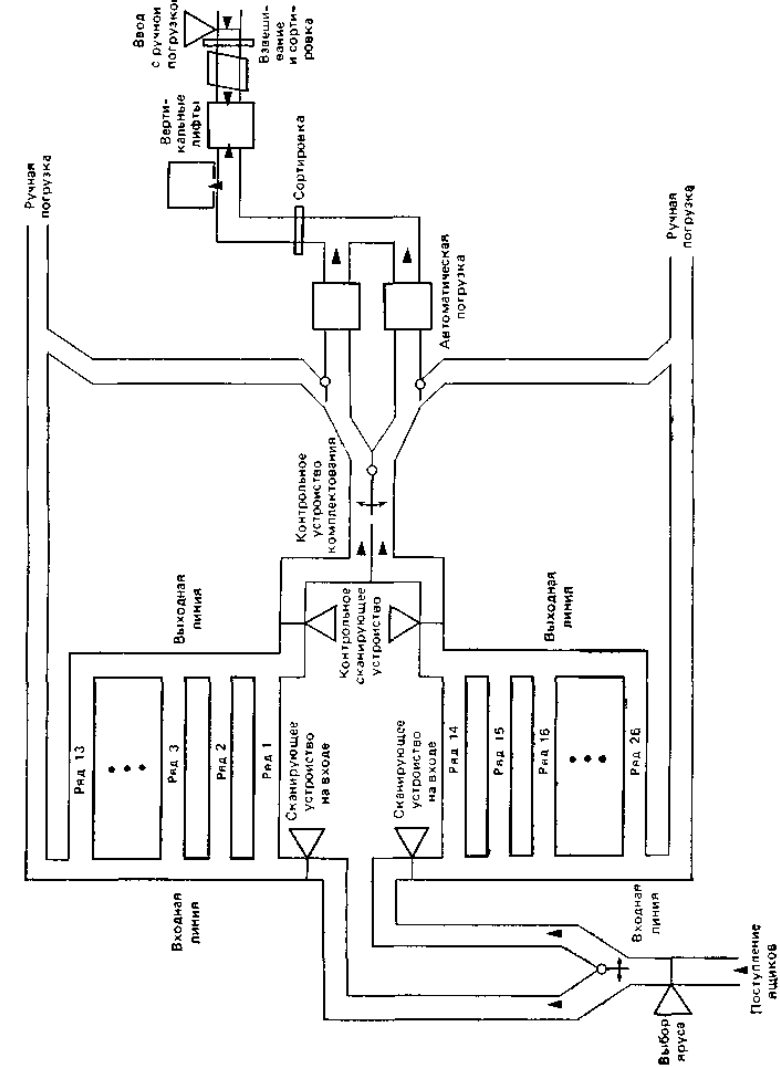


Рис. 4. Общая схема складирования готовой продукции

Специальная логика процесса и наличие ряда конвейерных линий между контрольным устройством и погрузчиком поддерживают определенное расстояние между ящиками в тех местах, где их накапливается много, а также обеспечивают требуемое объединение комплектов ящиков при функционировании в аварийном режиме

С помощью данной имитационной модели можно сравнить эффективность двух конвейеров различного типа. Роликовые конвейеры включаются и выключаются нечасто, поскольку для них не слишком важна регулярность поступления ящиков. Следовательно, погрузка и разгрузка ящиков могут моделироваться дискретными событиями. Ленточные же конвейеры, наоборот, часто включаются и выключаются из-за уменьшения потока ящиков, поэтому время перемещения ящика по конвейеру трудно определить заранее. Конвейеры этого типа моделировались с помощью переменных состояния, отображавших положение конвейеров. Для отображения ящиков на конвейере использовались указатели. Данный подход был реализован в имитационной модели склада готовой продукции на языке GASP IV.

Результаты исследования показали, что при определенной организации конвейерных операций можно повысить загрузку упаковщиков до 84%. На величину загрузки отрицательное воздействие оказывают временные характеристики некоторых элементов оборудования. Максимальная величина загрузки была достигнута путем добавления двух новых фотоэлектрических датчиков и изменения места расположения других, установленных ранее. Результаты исследования были использованы на практике. При этом наблюдалось ожидаемое улучшение качества функционирования системы в условиях сбоев в работе оборудования.

Впоследствии данная имитационная модель была развита и использовалась для оценки эффективности различных алгоритмов управления движением и размещением продукции на автоматизированном складе. Применение имитационной модели в процессе планирования позволило достигнуть впечатляющего повышения эффективности работы системы.

2.10.10. Анализ процесса литья и запаса отливок

В корпорации «Бетлехем стил» существует централизованный литейный цех для получения отливок в разовых литейных формах для всех заводов корпорации. Спрос на отливки зависит от их использования и отбраковки на различных заводах. Наличие широкого

диапазона типов и размеров отливок определяется главным образом необходимостью получения максимально возможного выхода конечной продукции из каждой отливки. Повышение эффективности использования отливок достигается обычно за счет увеличения их размеров. Ограничением при этом являются технические возможности оборудования, установленного в централизованном литейном цехе.

Для определения потребностей в увеличении производственных мощностей литейного цеха была разработана имитационная модель процесса литья в разовых литейных формах на языке GASP IV. На рис. 5 показаны операции, составляющие процесс литья.

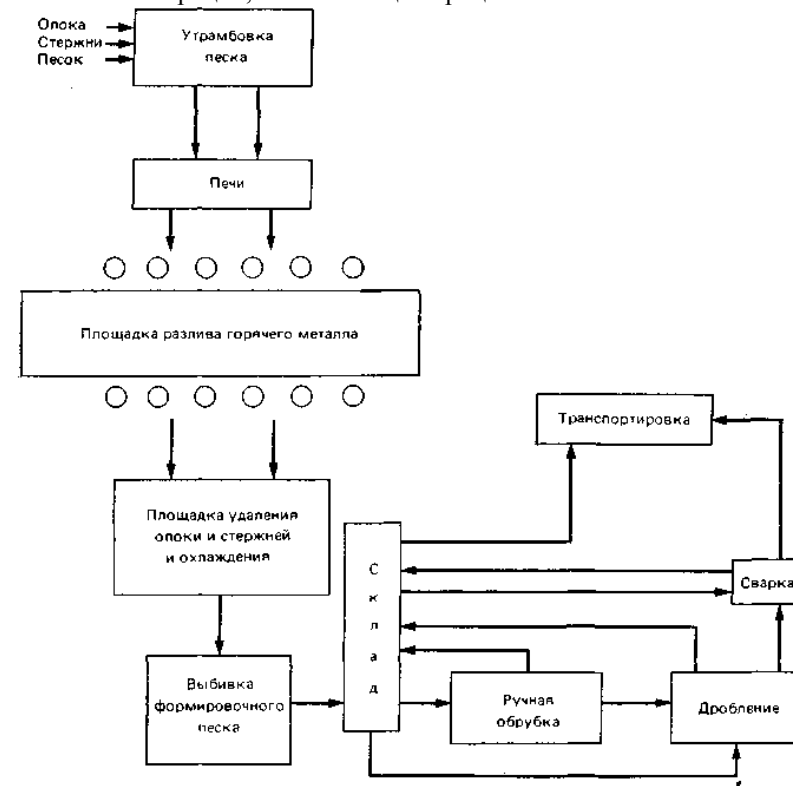


Рис. 5. Операции, выполняемые для получения отливок в разовых литейных формах.

Эта дискретно-событийная модель отображает операции загрузки и утрамбовки кварцевого песка в опоках, разогрева опок в печах, разлива горячего металла, охлаждения, извлечения отливок, удаления шлака,

обрубки, дробления и транспортировки. Сбор данных, задание правил диспетчеризации, верификация модели и анализ результатов проводились совместно с управляющим централизованного фонда и главным диспетчером, а также при участии инженеров металлургических заводов корпорации.

Анализ спроса на отливки — часть проведенного исследования. Для этого были собраны данные за предыдущий трехлетний период. Они содержали значения уровней складских запасов отливок, объемы производства стальных отливок различных типов, а также данные о доле брака. На основе этой информации была разработана процедура сглаживания колебаний спроса на отливки. С помощью имитационной модели централизованного литейного цеха было показано, что сглаживание колебаний спроса на отливки различных типов по всем предприятиям корпорации «Бетлехем стил» позволяет оставить производственные мощности централизованного литейного цеха прежними при высвобождении части мощностей, которые могут использоваться для удовлетворения спроса в пиковых ситуациях. Данная разработка позволила избежать капитальных затрат на установку новых мощностей и одновременно снизить общий уровень запасов отливок в корпорации.

В предыдущих разделах основное внимание уделялось применению имитационных моделей в конкретных областях принятия решений. Имитационные модели могут также использоваться для анализа проблем, разработки новых систем или прогнозирования будущего развития.

Литература

1. А. с. 532101, СССР, МКН³ С 06 15/36. Устройство для распознавания случайных процессов.
2. А. с. 631929, СССР, МКИ³ С 06 15/36. Устройство для распознавания случайных процессов.
3. А. с. 1057904, СССР, МКИ³ С 01 7/28. Обнаружитель сигналов.
4. А. с. 1073889, СССР, МКИ³ Н 04 В1/10. Устройство автоматического регулирования среднего числа шумовых выбросов.
5. Бакулев П. А., Степин В. М. Методы и устройства селекции движущихся целей.— М.: Радио и связь, 1986.— 288 с.
6. Белецкий А. Я., Корнильев Э. А., Хмелько Ю. М. Анализ эффективности автоматической регулировки порога квантования при нестабильной дисперсии шума // Изв вузов. Радиоэлектроника. — 1974 — Т. 17, № 1.— С. 86—89.
7. Белецкий А. Я., Прокопенко И. Г. Анализ эффективности накопителей бинарных коррелированных сигналов на малых объемах выборки // Теория и техника радиолокации, радионавигации и радиосвязи в гражданской авиации: Межвуз. сб. науч. тр.—Рига, 1979,—С. 17—21.
8. Браммер К., Зиффлинг Г. Фильтры Калмана — Бьюси: Пер. с англ.— М.: Наука, 1982.— 200 с.
9. Виноградов В. И. Информационно-вычислительные системы.— М: Энергоатомиздат, 1986.— 336 с.
10. Гаек Я., Шидак З. Теория ранговых критериев.— М.: Наука, 1971.—375 с.
11. Гильбо Е. П., Челпанов И. Б. Обработка сигналов на основе упорядоченного выбора.— М.: Сов. радио, 1975. - 344 с.
12. Градштейн И. С, Рыжик И. М. Таблицы сумм рядов и интегралов — М.: Наука, 1971.— 1108 с.
13. Гришин Ю. П., Казаринов Ю. М., Катков В. И. Динамические системы устойчивые к отказам.— М.: Радио и связь, 1985.— 176 с
14. Давлетьянц А. И., Корнильев Э. А. Основные соотношения для анализа одного класса адаптивных алгоритмов обнаружения // Теория и техника радиолокации, радионавигации и радиосвязи в гражданской авиации: Межвуз. сб. науч. тр — Рига, 1976,—С. 3—5
15. Давлетьянц А. И., Корнильев Э. А. Синтез алгоритма обнаружения сигналов на фоне аддитивной совокупности гауссова

- шума и хаотической импульсной помехи // Изв. АН ТССР. Сер. Физ.-техн., хим. и геол. наук.— 1973.— №4—С. 3—8.
16. Дэйвид Г. Порядковые статистики: Пер. с англ. / Под ред. В. В. Петрова.— М. : Наука, 1979.—336 с.
17. Заморин А. П., Мязев А. А., Селиванов Ю. П. Вычислительные машины, системы, комплексы: Справ.— М. : Энерго-атомиздат, 1985.— 279 с.
18. Кендал М. Дж., Стюарт А. Статистические выводы и связи.— М. : Наука, 1973.— 899 с.
19. Корнильев Э. А. Непараметрические методы обнаружения радиолокационных сигналов на фоне гауссовых помех с неизвестной дисперсией // Вопр. авиацион. радиотехники.— 1971 — № 6.—С 28—30.
20. Корнильев Э. А., Новиков В. С., Прокопенко И. Г. Алгоритм экстраполяции нестационарных случайных последовательностей, реализующий метод максимального правдоподобия — К., 1973.— (Укр. РФАП; № 83).
21. Корнильев Э. А., Прокопенко И. Г. Устойчивые тесты проверки статистических гипотез при негауссовых помехах // Тез. докл. Всесоюз. науч.-техн. конф. Статистические методы в теории передачи и преобразования информационных сигналов.— К. : КНИГА, 1986.—С. 154.
22. Корнильев Э. А., Хмелько Ю. М. Непараметрическое обнаружение гауссовых сигналов на фоне шума и хаотической импульсной помехи // Изв. вузов. Радиоэлектроника.— 1973.— Т. 16, № 12.—С. 91—94.
23. Корнильев Э. А., Чирков Д. В. Характеристика обнаружения адаптивного подавителя импульсных помех // Вопр. авиацион. радиотехники.— 1972.— № 7.—С. 8—10.
24. Крамер Г. Математические методы статистики.— М. : Мир, 1975.— 648 с.
25. Кузьмин С. З. Основы теории цифровой обработки радиолокационных сигналов.— М. : Сов. радио, 1974.— 380 с.
26. Лапий В. Ю., Калужный А. Я., Красный Л. Г. Устройства ранговой обработки информации.— К. : Техніка, 1986.— 120 с.
27. Левин Б. Р. Теоретические основы статистической радиотехники: В 3-х кн.— М. : Сов. радио, 1976.— Кн. 3.— 286 с.
28. Мальтийский А. Н., Подольский А. Г. Радиовещательный прием в автомобиле.— М. : Радио и связь, 1982.— 187 с.
29. Миленкий А. В. Классификация сигналов в условиях неопределенности.— М. : Сов. радио, 1975.— 328 с.

30. Новиков В. С., Корнильев Э. А., Прокопенко И. Г. Экстраполяция нестационарных случайных последовательностей по методу максимального правдоподобия // Основные вопросы теории и практики надежности.— М. : Сов. радио, 1976.— С. 87-97.
31. Оводенко А. А. Робастные локационные устройства.— Л. : Изд-во ЛГУ, 1981,— 182 с.
32. Оуэн Д. Б. Сборник статистических таблиц: Пер с англ. / Под ред. Л. Н. Большева.— М. : ВЦ АН СССР, 1966.— 586 с.
33. Прокопенко И. Г. Квазинепараметрическая процедура обнаружения радиолокационных сигналов // Теория и техника радиолокации, радионавигации и радиосвязи в гражданской авиации: Межвуз. сб. науч. тр.— Рига, 1976.— С. 3—12.
34. Прокопенко И. Г., Кириченко Е. П. Анализ эффективности реверсивных накопителей // Теория и техника радиолокации, радионавигации и радиосвязи в гражданской авиации: Межвуз. сб. науч. тр.— Рига, 1986.— С. 17—22.
35. Прокопенко И. Г., Корнильев Э. А., Кириченко Е. П. Повышение информативности наблюдения сигналов на фоне гладких шумов // Контроль и диагностирование АиРЭО воздушных судов гражданской авиации.— К. : КНИГА, 1984.— С. 57—62.
36. Прокопенко И. Г., Омельчук И. П. Анализ эффективности автоматического регулятора порога квантования в бинарных обнаружителях // Теория и техника радиолокации, радионавигации и радиосвязи в гражданской авиации. Межвуз. сб. науч. тр.—Рига, 1980.—С. 9—14.
37. Прокопенко И. Г., Сеницын Р. Б. Многомерный ранговый алгоритм обнаружения локационных сигналов // Теория и техника радиолокации, радионавигации и радиосвязи в гражданской авиации: Межвуз. сб. науч. тр.— Рига, 1985.— С. 47—51.
38. Прокопенко И. Г., Сеницын Р. Б. Оптимальный ранговый обнаружитель радиолокационных сигналов // Обработка пространственно-временных сигналов: Межвуз. сб. науч. тр.— Воронеж, 1983,—С. 96—101.
39. Проектирование специализированных информационно-измерительных систем / Г. Н. Воробьев, Е. С. Потапов, В. В. Сюзев; Ю. М. Смирнов. Под ред. Ю. М. Смирнова.— М. : Высш. шк., 1984.- 350 с.
40. Сархан и Гринберг. Введение в теорию порядковых статистик. Пер. с англ. / Под ред. А. Я. Боярского.— М. : Статистика, 1970.— 413 с.
41. Сейдж Э., Меле Дж. Теория оценивания и ее применение в связи и управлении — М. : Связь, 1976.— 493 с.

42. Семерджян М. А., Налбандян Ж. С. Матричный процессор ЕС-2345.— М. : Финансы и статистика, 1984.— 133 с.
43. Семесенко М. П. Методы обработки и анализа измерений в научных исследованиях.— К. : Донецк: Вища шк., 1983.— 240 с.
44. Серебрянников М. Г., Первозванский А. А. Выявление скрытых периодичностей.— М. : Наука, 1965.— 244 с.
45. Смирнов Н. В. Теория вероятностей и математическая статистика.— М. : Наука, 1970.— 289 с.
46. Стратонович Р. Л. Принципы адаптивного приема.— М. : Связь, 1973 — 144 с.
47. Тарасов С. А. Параллельные вычислительные системы // Радиоэлектроника: (Состояние и тенденции развития).— М. : НИИЭИР, 1985.— 120 с.
48. Теория обнаружения сигналов / П. С. Акимов, П. А. Бакут, В. А. Богданович и др.— М. : Радио и связь, 1984.— 440 с.
49. Тербер К. Дж. Архитектура высокопроизводительных вычислительных систем.— М. : Наука, 1985.— 272 с.
50. Уилкс С. Математическая статистика. Пер. с англ. / Под ред. Ю. В. Линника.— М. : Наука, 1967.— 632 с.
51. Устойчивые статистические методы оценки данных: Пер. с англ. / Под ред. Н. Г. Волкова.— М. : Машиностроение, 1984.— 232 с.
52. Фалькович С. Е., Хомяков Э. Е. Статистическая теория измерительных радиосистем.— М. : Радио и связь, 1981 — 288 с.
53. Фомин А. Ф., Новоселов О. П., Пономарев А. В. Отбраковка аномальных результатов измерений.— М. : Энергоатомиздат, 1985.— 200 с.
54. Хан Г., Шапиро С. Статистические модели в инженерных задачах.— М. : Мир, 1969.— 272 с.
55. Харрингтон Д. В. Исследования обнаружения повторяющихся сигналов при помощи двоичного накопления: Прием сигналов при наличии шумов: Пер. с англ. / Под ред. Л. С. Гуткина.— М. : Изд-во иностр. лит., 1960.— 270 с.
56. Хубер П. Дж. Робастность в статистике: Пер. с англ.— М. : Мир, 1984,— 304 с.
57. Шалыгин А. С., Палагин Ю. И. Прикладные методы статистического моделирования.— Л. : Машиностроение. Ленинград. отделение, 1986,— 320 с.
58. Шметтерер Л. Введение в математическую статистику: Пер. с нем. / Под ред. Ю. В. Линника,— М. : Наука, 1976,— 520 с.
59. Краснощеков П. С, Петров А. А. Принципы построения моделей. М., 1983.

60. Норенков И. П. Введение в автоматизированное проектирование технических устройств и систем. М., 1986.
61. Бусленко Н. П., Калашников В. В., Коваленко И. И. Лекции по теории систем. М., 1973.
62. Цветков В. Д. Системно-структурное моделирование и автоматизация проектирования технологических процессов. Минск, 1979.
63. Якимов И. М. Моделирование систем. Казань, 1980.
64. Автоматизация поискового конструирования и искусственный интеллект в машинном проектировании/А И. Головинкин, И. К. Бобков, Г. Я. Буш и др.; Под ред. А. И. Половинкина. М., 1981.
65. Максимей И. В. Имитационное моделирование на ЭВМ. М., 1988.
66. Шрайбер Г. Дж. Моделирование на GPSS. М., 1980.
67. Автоматизация проектирования в радиоэлектронике и вычислительной технике. М., 1984.
68. Н. П. Бусленко. К теории сложных систем. Известия АН СССР, «Техническая кибернетика», № 5, 1963.
69. Н. П. Бусленко. Моделирование производственных процессов. «Наука», 1965.
70. Н. П. Бусленко. Моделирование сложных систем. «Наука», 1968
71. Н. П. Бусленко. Теория больших систем. «Наука», 1969
72. Г. Х. Гуд, Р. Э. Маккол. Системотехника. Введение в проектирование больших систем. «Советское радио», 1962
73. Р. Калмаи, П. Фалб, М. Арби и б. Очерки по математической теории систем. Пер. с англ., «Мир», 1971
74. Месерович, Д. Макко, У. Такахара. Теория иерархических многоуровневых систем. Пер. с англ., ЦНИИ «Электроника», 1970
75. У. Портер. Современные Основания общей теории систем. Пер. с англ., «Наука», 1971
76. Б. С. Флейшман. Элементы теории потенциальной эффективности сложных систем. «Советское радио», 1971
77. Д. Н. Харфас. Системы и моделирование. «Мир», 1967
78. С. М. Ермаков. Метод Монте — Карло и смежные вопросы. «Наука», 1971
79. И. Е. Казаков, Б. П. Доступов. Статистическая динамика нелинейных автоматических систем. Физматгиз, 1962
80. В. И. Чернецкий. Анализ точности нелинейных систем управления. Машиностроение, 1968
81. И. Н. Коваленко. О некоторых классах сложных систем. Известия АН СССР. «Техническая кибернетика» № 6, 1964; № 1, 3 1965

82. Ю. Б. Гермейер. Введение в теорию исследования операций. «Наука», 1971
83. Ю. С. Голубев — Новожилов. Многомашинные комплексы вычислительных средств. «Советское радио», 1967
84. В. В. Налимов, Н. А. Чернова. Статистические методы планирования экстремальных экспериментов. «Наука», 1965
85. В. Роберт Миллер. Перт — система управления. «Экономика», 1965
86. Г. Ван — Три с. Синтез оптимальных нелинейных систем управления. «Мир», 1966
87. Л. А. Люстерник, В. И. Соболев. Элементы функционального анализа. «Наука», 1965
88. Д. И. Голенко. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах. «Наука», 1965.
89. В. И. Горюхи и др. Методы теории чувствительности в автоматическом управлении. «Энергия», 1971
90. D. Watson. Some Factors Involved in the Establishment Programmer Performance Standards. Computer Bulletin, vol. 13, N 6, 1969, 192—194.
91. P. R. Masteg Evaluating Programmer Performance. The Australian Computer Journal, vol. 1, N 3, 1968, 124—129.
92. А. Н. Колмогоров. Интерполирование и экстраполирование стационарных случайных последовательностей. Т. 5, № 1, Известия АН СССР.
93. Я. З. Цыпкин. Адаптация и обучение в автоматических системах. «Наука», 1968
94. Д. Блекуэлл, М. А. Гиршик. Теория игр и статистических решений. ИЛ, 1958
95. Б. П. Ван Дер Варден. Математическая статистика. ИЛ, 1960
96. Т. Андерсон. Введение в многомерный статистический анализ, Физматгиз, 1963
97. Л. П. Сыроев. Оценки параметров, обнаружение и различие сигналов. «Наука», 1969
98. Д. Худсон. Статистика для физиков. «Мир», 1970
99. С. К. Годунов, В. С. Рябенский. Введение в теорию разностных схем. Физматгиз, 1962
100. Б. П. Демидович, И. А. Марон, Э. З. Шувалова. Численные методы анализа. Физматгиз, 1962

Научно-практическое издание

Кононюк Анатолий Ефимович

Обобщенная теория моделирования

Книга 1

Начала

Часть 3

Авторская редакция

Подписано в печать 25.05.2012 г.

Формат 60x84/16.

Усл. печ. л. 26,5. Тираж 300 экз.

Издатель и изготовитель:

Издательство «Освита Украины»

04214, г. Киев, ул. Героев Днепра, 63, к. 40

Свидетельство о внесении в Государственный реестр издателей ДК №1957 от 23.04.2009 г.

Тел./факс (044) 411-4397; 237-5992

E-mail: osvita2005@ukr.net, www.rambook.ru

Издательство «Освита Украины» приглашает

авторов к сотрудничеству по выпуску изданий, касающихся вопросов управления, модернизации, инновационных процессов, технологий, методических и методологических аспектов образования и учебного процесса в высших учебных заведениях.

Предоставляем все виды издательских и полиграфических услуг