

Министерство образования и науки Российской Федерации

---

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ  
ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

---

**Приоритетный национальный проект «Образование»  
Национальный исследовательский университет**

*А.Л. ГЕЛЬГОР*

*А.И. ГОРЛОВ*

*Е.А. ПОПОВ*

## **ОБЩАЯ ТЕОРИЯ СВЯЗИ**

# **ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ОПТИМАЛЬНЫЙ ПРИЁМ СИГНАЛОВ**

*Рекомендовано Учебно-методическим объединением по университетскому  
политехническому образованию в качестве учебного пособия  
для студентов высших учебных заведений, обучающихся по направлению  
подготовки магистров “Техническая физика”*

Санкт-Петербург

Издательство политехнического университета

2013

УДК 621.391  
ББК 32.811.1  
Г 32

Рецензенты:

Кафедра радиопередающих устройств и средств подвижной связи СПбГУТ  
им. проф. М.А. Бонч-Бруевича, зав. кафедрой д.т.н., профессор М. Сиверс

Кафедра военных телекоммуникационных систем Военной академии связи  
им. С.М. Будённого, нач. кафедры к.в.н., доц. А. Боговик

*Гельгор А.Л.* **Общая теория связи. Проверка статистических гипотез. Оценивание параметров. Оптимальный приём сигналов:** учеб. пособие / Гельгор А.Л., Горлов А.И., Попов Е.А. — СПб.: Изд-во Политехн. ун-та, 2013. — 227 с.

В учебном пособии излагаются основные подходы к решению задач проверки статистических гипотез, оцениванию параметров и оптимального приёма сигналов: изучаются стратегии и критерии принятия решений; формулируются алгоритмы, реализующие выбранную стратегию, и рассчитывается их помехоустойчивость; изучаются характеристики различных статистических оценок и методы их получения. Рассматриваются ансамбли различных, в том числе, многопозиционных сигналов с амплитудно-фазовой модуляцией; предлагаются варианты синтеза алгоритмов оптимального обнаружения и различения, а также особенности практической реализации таких алгоритмов с использованием современных программно-аппаратных средств.

Пособие предназначено для студентов, обучающихся по направлению подготовки “Техническая физика”. Оно может быть также использовано при обучении студентов направлений подготовки “Инфокоммуникационные технологии и системы связи”, “Радиотехника”.

Работа выполнена в рамках реализации программы развития национального исследовательского университета “Модернизация и развитие политехнического университета как университета нового типа, интегрирующего мультидисциплинарные научные исследования и надотраслевые технологии мирового уровня с целью повышения конкурентоспособности национальной экономики”

Печатается по решению редакционно-издательского совета Санкт-Петербургского государственного политехнического университета.

© Гельгор А.Л., Горлов А.И., Попов Е.А., 2013

© Санкт-Петербургский государственный  
политехнический университет, 2013

ISBN

# ОГЛАВЛЕНИЕ

Введение .....	4
Глава 1. Проверка статистических гипотез .....	7
1.1. Стратегии, критерии и алгоритмы принятия решения .....	8
1.2. Последовательная процедура проверки гипотез .....	39
1.3. Сложные гипотезы .....	51
1.4. Многоальтернативные задачи .....	62
Вопросы и задания к Главе 1 .....	72
Глава 2. Оценивание параметров .....	77
2.1. Характеристики и свойства статистических оценок .....	77
2.2. Методы получения оценок .....	98
2.2.1. Метод моментов .....	98
2.2.2. Оценки по максимуму апостериорной вероятности ...	100
2.2.3. Оценки максимального правдоподобия .....	102
2.2.4. Байесовские оценки .....	111
2.3. Оценки по методу наименьших квадратов. регрессион- ный анализ .....	117
Вопросы и задания к Главе 2 .....	129
Глава 3. Оптимальный приём сигналов .....	131
3.1. Разложение Карунена–Лоэва .....	133
3.2. Функционал отношения правдоподобия и алгоритмы приёма для модели аддитивного белого гауссовского шума ...	149
3.2.1. Обнаружение одиночного сигнала .....	154
3.2.2. Различение сигналов .....	164
3.3. Практическая реализация алгоритмов оптимальной обра- ботки сигналов .....	203
Вопросы и задания к Главе 3 .....	225
Библиографический список .....	227

## ВВЕДЕНИЕ

Главным положением, лежащем в основе методологии теории и практики построения различных телекоммуникационных и информационных систем, является наличие в канале передачи данных случайных и неуправляемых воздействий — *помех*, негативное влияние которых приводит к частичной или, даже, полной потере информации. Конкретные причины, порождающие помехи в канале, могут быть весьма разнообразными: искажение радиоволн при их распространении в какой-либо физической среде (интерференция, вызванная спорадическими явлениями в ионосфере или переотражениями плотной городской застройки); коммутационные помехи (дребезг контактов); помехи промышленного происхождения (близость к троллейбусно-трамвайным линиям, электросварочные работы); помехи во входных цепях радиоприёмного устройства (тепловой или дробовой шум) и др.

Наличие случайных воздействий в канале передачи предопределяет вероятностный (статистический) подход к разработке и анализу функционирования системы, когда, с одной стороны, характеристики, количественно описывающие качество работы системы, являются вероятностными (например, вероятность безотказной работы устройства, вероятность правильного различения используемых сигналов, вероятность неудачного хэндовера), а с другой — также вероятностными являются сами идеи, лежащие в основе синтеза алгоритмов и устройств обработки сигналов. Так, стратегия принятия решений о передаваемых сообщениях (сигналах) формулируется в терминах задач проверки статистических гипотез; получение количественных показателей верности передаваемой информации связано с вычислением вероятностей неравенств, фигурирующих в алгоритмах оптимальной обработки; оценки параметров основаны на статистических характеристиках выборочных наблюдений.

Нет смысла много рассуждать о том, что глобальность современной теории связи не даёт возможность сколько-нибудь подробно изучать её в рамках одного — даже очень “толстого” издания<sup>1</sup>. Известные книги

---

<sup>1</sup> Неуклонное развитие науки приводит к тому, что такое утверждение вскоре можно будет отнести ко всем изучаемым дисциплинам любого образовательного направления.

Дж. Прокиса и Б. Склера [15, 16] объём которых составляет около 1000 страниц, при всех их достоинствах, в том числе, как университетских курсов, нельзя рассматривать как последовательное изложение *всех* аспектов теории связи. В этой связи, по-видимому, целесообразно представление отдельных разделов рамками относительно небольших книг, связанных между собой и логически, и структурно, и, разумеется, обозначениями.

Пособие состоит из трёх глав.

В первой главе излагаются основные подходы к решению задач проверки статистических гипотез для выборок конечного объёма: сравниваются стратегии и критерии принятия решений; формулируются алгоритмы, реализующие выбранную стратегию, и рассчитывается их помехоустойчивость; предлагаются структурные схемы для практической реализации таких алгоритмов. Наряду с “классическим” подходом анализа с фиксированным объёмом выборки рассматривается перспективный метод последовательного анализа, при котором объём выборки в процессе принятия решения может изменяться.

Вторая глава посвящена решению задач оценивания параметров на основе рассмотренной в первой главе общей байесовской стратегии принятия решений. Производится сравнение характеристик для различных статистических оценок и методов их получения (метод моментов, метод максимального правдоподобия, метод максимума апостериорной вероятности и др.). Также рассматривается методика получения оценок, основанная на минимизации суммы квадратов в модели линейной регрессии.

Третья глава содержит, пожалуй, самый важный в практическом плане материал: задачи синтеза алгоритмов оптимального обнаружения и различения, а также особенности их практической реализации с использованием современных программно-аппаратных средств. Рассматриваются ансамбли различных, в том числе, многопозиционных сигналов с амплитудно-фазовой модуляцией, изучаются подходы к вычислению их помехоустойчивости.

В конце каждой главы предложен набор заданий, решение которых должно способствовать более глубокому пониманию изучаемого материала. Предлагаемый перечень невелик, поскольку предполагается, что в процессе изучения читатель активно работает с традиционными “толстыми” задачами, например [17].

Учебное пособие написано по материалам учебных дисциплин, преподаваемых авторами в течение многих лет на радиофизическом факультете Санкт-Петербургского государственного политехнического университета и предназначено для студентов, обучающихся по направлениям подготовки “Техническая физика”, “Инфокоммуникационные технологии и системы связи”, “Радиотехника”. Также оно может оказаться полезным тем специалистам, чья профессиональная деятельность связана с разработкой алгоритмов, устройств и систем передачи информации.

## ГЛАВА 1. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Решение задач оптимального приёма основано на *априорных* (предварительных) *сведениях* о подлежащих приёму сигналах и соответствующей обработке принимаемых реализаций.

Понятно, что при полном отсутствии каких бы то ни было априорных знаний о сигнале его приём невозможен. С другой стороны, полностью известный сигнал не несёт в себе никакой информации, ибо, если заранее известны его форма и параметры, то он может быть в точности воспроизведён на приёмной стороне. Поэтому носителями полезной информации могут быть только случайные параметры сигнала — это положение лежит в основе определения количества информации, содержащегося в сообщении [1].

Априорные сведения, касающиеся как полезного сигнала, так и шумовой компоненты, могут быть получены путём предварительного статистического анализа физической среды передачи информации, в результате которого у наблюдателя увеличиваются знания об исследуемой ситуации; такие вновь сформированные знания являются *апостериорными сведениями*.

Исходным объектом наблюдений является *источник*, могущий находиться в одном из  $m$  возможных *состояний*  $Q_1, \dots, Q_m$ , каждое из которых случайным образом реализуется в определённые интервалы времени наблюдения. Предполагается, что множество  $T$  всех возможных состояний источника образуют полную группу попарно несовместных случайных событий с соответствующими априорными вероятностями состояний  $P(Q_1), \dots, P(Q_m)$ .

Результаты наблюдений за источником фиксируются либо аналоговыми устройствами в виде реализации  $z(t)$  случайного процесса, имеющей конечную длительность, либо в виде конечной совокупности чисел  $z_1, z_2, \dots, z_M$ , являющихся значениями некоторых случайных величин, в частности, отсчётными значениями реализации  $z(t)$ , взятыми в определённые моменты времени  $t_1, t_2, \dots, t_M$ . В последнем случае результаты наблюдений представляются *выборкой*  $\mathbf{z} = (z_1, \dots, z_M)$  — конечномерным вектором  $M$ -мерного *выборочного пространства* (*пространства наблюдений*)  $S$ . При этом вероятностная мера на множестве  $S$  (многомерная плотность вероятности  $w(\mathbf{z})$  или многомерная интегральная функция распределения

$F(\mathbf{z})$ ) оказывается не полностью определённой. Так например, для плотности  $w(\mathbf{z}/\boldsymbol{\eta})$  могут оказаться неизвестными один или несколько параметров  $\eta_1, \dots, \eta_l$ , но вид функциональной зависимости известен (параметрический случай). В более сложном — непараметрическом — случае неизвестна сама функциональная зависимость  $w(\mathbf{z})$ , а известны лишь некоторые её характеристики (непрерывность, одномодальность и т. п.).

Задача оптимального приёма (или, как часто говорят, оптимального синтеза) состоит в отыскании алгоритма обработки результатов наблюдения, приводящего к наилучшему (в определённом смысле) выбору решения, т. е. установлению соответствия между точками  $\mathbf{z}$  пространства наблюдений  $S$  и “точками”  $\boldsymbol{\gamma}(\mathbf{z})$  пространства решений  $\Gamma$ . Каждое решение  $\boldsymbol{\gamma}(\mathbf{z})$  является *статистикой* — некоторой функцией выборки  $(z_1, \dots, z_M)$  или же функционалом от наблюдаемой непрерывной реализации  $z(t)$ .

Методы оптимального приёма можно отнести к двум основным разновидностям прикладных задач: *проверки статистических гипотез* и *оценки параметров сигнала* [2, 3].

Проверка статистических гипотез состоит в выборе решения по результатам наблюдения, когда требуется принять одну (из конечной совокупности) гипотезу, описывающую определённое состояние изучаемого объекта (явления).

При решении задачи оценки параметров сигнала необходимо с минимальной погрешностью определить значения параметров в наблюдаемой реализации. При этом, если полезный сигнал зависит от нескольких параметров, то может быть поставлена задача о совместной оценке двух и большего числа параметров.

Результатом решения обоих типов задач являются алгоритмы обработки, структурные схемы решающих или измерительных устройств, а также оценка качества работы таких устройств, в том числе, оценка предельной точности разрешения или измерения.

## **1.1. СТРАТЕГИИ, КРИТЕРИИ И АЛГОРИТМЫ ПРИНЯТИЯ РЕШЕНИЯ**

Понятие гипотезы, с которым имеет дело математическая статистика, более узко, чем общее понятие научной гипотезы. Например, существует научная гипотеза о существовании (возможно, в прошлом) жизни на Мар-



се. Однако такие гипотезы не относятся к числу тех, что проверяются в рамках математической теории; статистические гипотезы касаются анализа наблюдаемых величин, получившихся в результате какого-либо эксперимента. Примерами статистических гипотез могут служить следующие гипотезы.

1. Наблюдаемая выборка принадлежит нормальному распределению с какими-либо математическим ожиданием и дисперсией.

2. Наблюдаемая выборка из нормального распределения имеет заданные математическое ожидание и дисперсию.

3. Две наблюдаемые выборки принадлежат одному и тому же (неизвестному) распределению.

4. В анализируемом колебании на входе приёмного устройства присутствует полезный сигнал.

Заметим, что в примере 2 распределение, из которого была получена выборка, имеет известный вид — нормальный, а гипотеза касается лишь значений параметров. Такие гипотезы называются *параметрическими*, в отличие от непараметрической гипотезы в примере 1, в котором необходимо принять решение о виде распределения. Более того, в примере 2 гипотеза определяет значения всех (в данном случае — двух) параметров распределения, тогда как в примере 1 количество заданных параметров равно нулю. Для теории такое различие является существенным, и в общем случае оно формулируется следующим образом. Если распределение имеет  $l$  параметров, а гипотеза утверждает, что  $k$  из них имеют заданные значения, то она называется *простой*, когда  $k = l$ , и *сложной*, когда  $k < l$ .

Геометрически возможные значения параметров можно представить в виде некоторой области в  $l$ -мерном пространстве. Если гипотезой выделяется единственная точка этого пространства, то такая гипотеза — простая; если же выделяется область, содержащая более одной точки, то такая гипотеза — сложная. При этом значение  $l - k$  называется *числом степеней свободы* гипотезы, а  $k$  — *числом ограничений* гипотезы.

Обратимся к простейшему случаю задачи проверки статистических гипотез — выбору между двумя простыми гипотезами, когда множество  $T$  возможных состояний источника содержит только два элемента.

Пусть имеется выборка  $\mathbf{z} = \{z_1, z_2, \dots, z_M\}$  наблюдаемых значений за состоянием источника. Требуется по этой выборке определить, какая из

двух гипотез имеет место (рис. 1.1): *основная гипотеза*<sup>1</sup>  $H_1$  о том, что источник находится в “основном состоянии”  $Q_1$ , а наблюдаемые значения описываются условной плотностью распределения  $w(\mathbf{z}/H_1)$ , или *альтернативная гипотеза*  $H_2$  о том, что источник находится в “альтернативном состоянии”  $Q_2$ , а наблюдаемые значения описываются условной плотностью распределения  $w(\mathbf{z}/H_2)$ . Поскольку каждая из двух гипотез относится только к одному состоянию источника — основному или альтернативному, такие гипотезы называются простыми, а соответствующая задача — простой проверкой гипотез.

Понятно, что после принятия решения о выборе той или иной гипотезы у наблюдателя не может быть полной уверенности в том, что такое решение правильное, поскольку любой процедуре простой проверки гипотез присущи два вида ошибок. Говорят, что происходит *ошибка первого рода*, когда основная гипотеза  $H_1$  является верной, но отвергается в результате проверки. Аналогично, *ошибка второго рода* возникает в том случае, когда верна альтернативная гипотеза  $H_2$ , но вместо неё принимается  $H_1$ . Эти ошибки характеризуются соответствующими вероятностями  $\alpha$  и  $\beta$ . Вероятность ошибки первого рода  $\alpha$  часто называется *уровнем значимости* процедуры, а вероятность  $(1 - \beta)$  — *мощностью* процедуры.

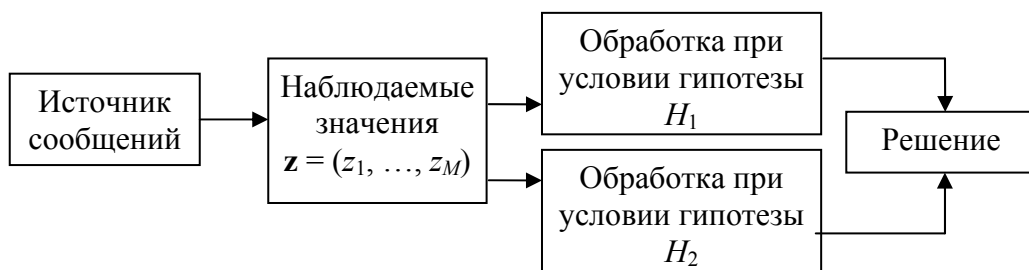


Рис. 1.1. Схема решения задачи проверки гипотез

Задача простой проверки гипотез состоит в том, чтобы разделить все

<sup>1</sup> Исторически в литературе, посвящённой статистической проверке гипотез, основная и альтернативная гипотезы обозначались  $H_0$  и  $H_1$  соответственно. К сожалению, в задачах, связанных с различением сигналов (особенно, многопозиционных), такие обозначения оказываются не совсем удобными. Поэтому авторы решились нарушить многолетнюю традицию, обозначив гипотезы индексами тех сигналов, которым они соответствуют.

множество  $S$  возможных значений источника на два непересекающихся подмножества (области)  $S_1$  и  $S_2$  так, что если выборка  $\mathbf{z}$  попадёт в подмножество  $S_1$ , то будет принято решение в пользу гипотезы  $H_1$ , а если в подмножество  $S_2$ , то — в пользу гипотезы  $H_2$ . В этом случае вероятности ошибок

$$\alpha = P\{\mathbf{z} \in S_2 / H_1\}, \beta = P\{\mathbf{z} \in S_1 / H_2\},$$

можно выразить через интегралы<sup>1</sup>, взятые по  $M$ -мерным областям  $S_1$  и  $S_2$ :

$$\alpha = \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} = 1 - \int_{S_1} w(\mathbf{z} / H_1) d\mathbf{z}, \quad (1.1.1)$$

$$\beta = \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} = 1 - \int_{S_2} w(\mathbf{z} / H_2) d\mathbf{z}. \quad (1.1.2)$$

Рассмотрим в качестве примера задачу проверки гипотезы о среднем значении нормального распределения с заданной единичной дисперсией, когда для распределения

$$w(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-(z - a)^2 / 2\right]$$

требуется проверить простую гипотезу  $H_1: a = a_1$  против простой альтернативы  $H_2: a = a_2 \neq a_1$ .

В случае выборки объёмом  $M = 2$  результаты наблюдений можно представить графически. На рис. 1.2 показана совокупность выборочных точек при  $N = 100$  числе повторений эксперимента, в котором для определённости выбраны  $a_1 = 1$ ,  $a_2 = -1$ . Нижнее скопление, в котором экспериментальные точки представлены символом “+”, соответствует выполнению гипотезы  $H_2$ , верхнее, в котором точки представлены символом “o” — гипотезы  $H_1$ . Видно, что все точки расположены вокруг центров с координатами  $(1; 1)$  и  $(-1; -1)$ .

В целях разбиения выборочного пространства (в рассматриваемом случае — всей плоскости) на непересекающиеся подмножества  $S_1$  и  $S_2$  необходимо найти такую область, в которой содержалась бы заданная доля

---

<sup>1</sup> Если выборка относится к дискретной случайной величине, то вместо интегралов следует рассматривать дискретные суммы, а вместо плотностей — дискретные вероятности. Оба случая могут быть, конечно, записаны в единой форме с помощью интеграла Стильтеса.

$\alpha$  точек “о”, соответствующих гипотезе  $H_1$ . На рис. 1.2 показана прямая  $C$

$$z_2 = -z_1,$$

разбивающая плоскость таким образом, что в нижней полуплоскости оказываются три точки “о”, т. е. при  $N = 100$  повторений вероятность ошибки первого рода составляет  $\alpha = 0,02$ . При этом, как нетрудно видеть, вероятность ошибки второго рода равна  $\beta = 0,04$  (пять точек “+” выше линии  $C$ ).

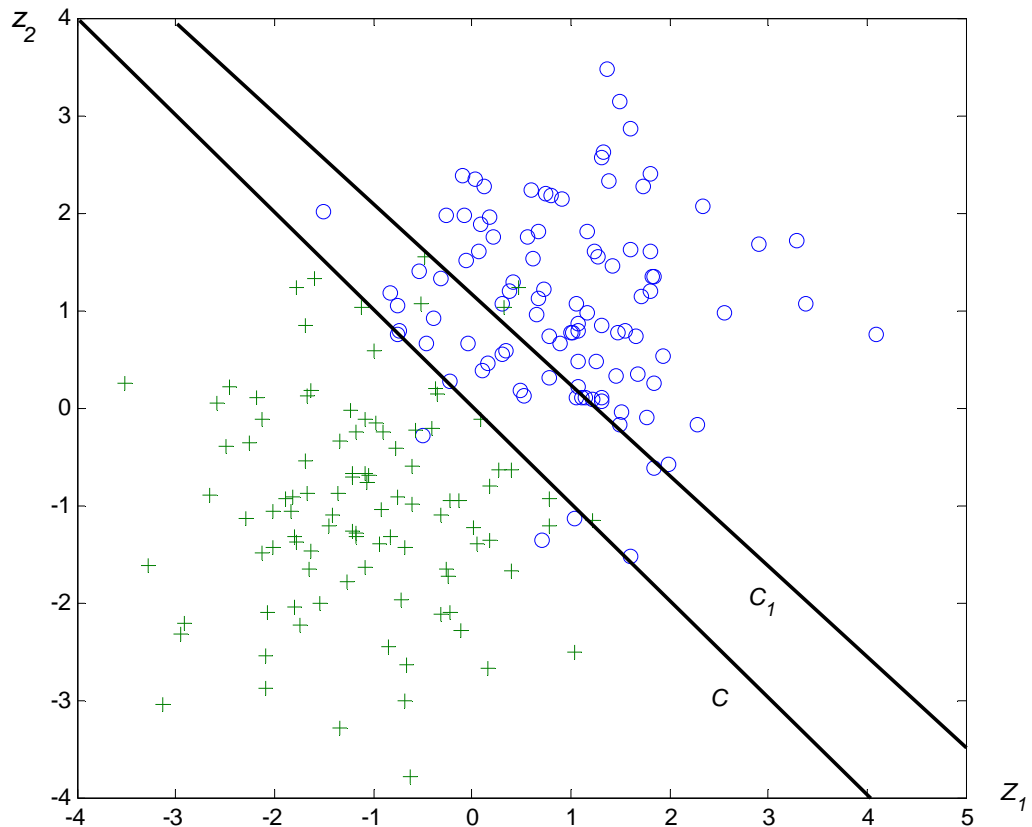


Рис. 1.2. Иллюстрация эксперимента для  $M = 2$  и  $N = 100$

Можно попытаться уменьшить одну из вероятностей ошибок, например, уменьшить  $\beta$  в два раза, перейдя к другой границе разбиения областей прямой  $C_1$

$$z_2 = 1 - z_1,$$

однако, как нетрудно видеть, при этом резко, на порядок, возрастает вероятность  $\alpha$ . Аналогичная ситуация возникает и при попытке уменьшить  $\alpha$ , смещая границу разбиения областей вниз.

Отмеченная ситуация носит общий характер. А именно, *при фиксированном объёме выборки невозможно одновременно сделать сколь угодно малыми вероятности ошибок первого и второго родов*. Например, чтобы ошибки первого рода появлялись редко, можно уменьшить до очень малого размера подмножество  $S_2$ . Но тогда подмножество  $S_1$  будет охватывать почти всё выборочное пространство, что приведёт к недопустимому росту вероятности ошибок второго рода.

Из вышеприведённых рассуждений следует, что прямая  $C$ , которая в  $M$ -мерном пространстве имеет вид

$$z_1 + z_2 + \dots + z_M = 0,$$

является “достаточно удачным” решением данной задачи проверки гипотез, однако понятно, что такой визуально-эвристический подход возможен лишь благодаря простоте поставленных в задаче условий. В более сложных ситуациях простые эвристические методы вряд ли будут давать удовлетворительные результаты, особенно, если требуется осуществить наилучшее, т. е. оптимальное разбиение пространства.

Чтобы осуществить оптимальный выбор областей  $S_1$  и  $S_2$ , необходимо задать некоторый количественный *критерий оптимальности*. Для этого, прежде всего, будем полагать, что оба состояния источника, отражаемые соответствующими гипотезами  $H_1$  и  $H_2$ , подчиняются некоторому распределению вероятностей; эти вероятности обозначим  $p(H_1)$  и  $p(H_2)$  и назовём *априорными вероятностями* гипотез.

Далее, каждому решению припишем числовую характеристику — *стоимость*, которая представляет собой потери (риски) или выигрыш — в том случае, когда стоимость оказывается отрицательной, возникающие при принятии конкретного решения. Обозначим через  $c_{ij}$  ( $i, j = 1, 2$ ) стоимость принятия гипотезы  $H_i$  в том случае, когда верной является гипотеза  $H_j$ , тогда можно ввести в рассмотрение *ценовую матрицу* (матрицу стоимостей)

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}, \quad (1.1.3)$$

в которой стоящие на главной диагонали элементы относятся к правильным решениям, а элементы, находящиеся вне главной диагонали — к ошибочным.

Рассмотрим в качестве критерия оптимальности *критерий минималь-*

ного среднего риска, определяемый следующим образом.

Очевидно, что при каждом испытании возможен один из четырёх исходов:

- верна гипотеза  $H_1$ , выбирается гипотеза  $H_1$ ;
- верна гипотеза  $H_1$ , выбирается гипотеза  $H_2$ ;
- верна гипотеза  $H_2$ , выбирается гипотеза  $H_1$ ;
- верна гипотеза  $H_2$ , выбирается гипотеза  $H_2$ .

Первый и четвёртый исходы соответствуют правильным решениям, а второй и третий — ошибочным. Тогда смысл вводимого критерия оптимальности состоит в том, чтобы *средняя* стоимость принятия решения была бы как можно меньше. При этом далее всегда будем предполагать, что стоимость ошибочного решения при фиксированной верной гипотезе больше, чем стоимость правильного решения, т. е. всегда справедливы соотношения

$$c_{21} > c_{11}, c_{12} > c_{22}. \quad (1.1.4)$$

Получение выражения для среднего риска связано с двойным усреднением: усреднение по типу принятия решения (верно / не верно) на основе условных вероятностей  $p(H_i/H_j)$  и усреднение по совокупности гипотез на основе априорных вероятностей  $p(H_1)$  и  $p(H_2)$ .

Для каждой из гипотезы  $H_i$  введём соответствующие *условные риски*  $R_i$  принятия решения:

$$R_1 = c_{11}P(H_1 / H_1) + c_{12}P(H_1 / H_2) \quad (1.1.5a)$$

и

$$R_2 = c_{21}P(H_2 / H_1) + c_{22}P(H_2 / H_2). \quad (1.1.5б)$$

Тогда средний риск  $R$  принятия решения по отношению ко всему эксперименту равен

$$R = c_{11}P(H_1 / H_1)p(H_1) + c_{12}P(H_1 / H_2)p(H_2) + c_{21}P(H_2 / H_1)p(H_1) + c_{22}P(H_2 / H_2)p(H_2). \quad (1.1.6)$$

Как уже было сказано выше, принятие решения в пользу какой-либо из гипотез означает отнесение выборки  $\mathbf{z}$  к одному из непересекающихся подмножеств  $S_1$  или  $S_2$ . Следовательно, при известных условных плотностях  $w(\mathbf{z}/H_1)$  и  $w(\mathbf{z}/H_2)$  вероятности  $p(H_i/H_j)$  могут быть выражены через интегралы в  $M$ -мерном пространстве:

$$p(H_1 / H_1) = \int_{S_1} w(\mathbf{z} / H_1) d\mathbf{z} = 1 - \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} = 1 - \alpha;$$

$$p(H_1 / H_2) = \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} = \beta;$$

$$p(H_2 / H_1) = \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} = \alpha;$$

$$p(H_2 / H_2) = \int_{S_2} w(\mathbf{z} / H_2) d\mathbf{z} = 1 - \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} = 1 - \beta.$$

Отсюда средний риск равен

$$\begin{aligned} R = & c_{11}p(H_1) \int_{S_1} w(\mathbf{z} / H_1) d\mathbf{z} + c_{12}p(H_2) \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} + \\ & + c_{21}p(H_1) \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} + c_{22}p(H_2) \int_{S_2} w(\mathbf{z} / H_2) d\mathbf{z}. \end{aligned}$$

Поскольку в результате процедуры какое-то решение все же должно быть принято (многомерная точка  $\mathbf{z}$  должна оказаться в объединении подмножеств  $S_1$  и  $S_2$ ), справедливо соотношение

$$\int_{S_1} w(\mathbf{z} / H_i) d\mathbf{z} + \int_{S_2} w(\mathbf{z} / H_i) d\mathbf{z} = 1, \quad i = 1, 2,$$

поэтому

$$\begin{aligned} R = & \int_{S_1} [(c_{12} - c_{22})w(\mathbf{z} / H_2)p(H_2) - \\ & - (c_{21} - c_{11})w(\mathbf{z} / H_1)p(H_1)] d\mathbf{z} + c_{22}p(H_2) + c_{21}p(H_1). \end{aligned} \quad (1.1.7)$$

В выражении (1.1.7) внеинтегральные слагаемые  $c_{22}p(H_2)$  и  $c_{21}p(H_1)$  отражают фиксированную составляющую среднего риска, а сам интеграл представляет собой ту стоимость, которая определяется точками получившегося при разбиении подмножества  $S_1$ . Согласно предположению (1.1.4), разности  $c_{12} - c_{22}$  и  $c_{21} - c_{11}$  положительны, поэтому в целях уменьшения среднего риска те точки  $\mathbf{z}$ , в которых оказывается справедливое неравенство

$$(c_{21} - c_{11})w(\mathbf{z} / H_1)p(H_1) > (c_{12} - c_{22})w(\mathbf{z} / H_2)p(H_2)$$

следует включить в  $S_1$ , а те точки, для которых выполняется обратное неравенство — наоборот, исключить из  $S_1$ , т. е. включить в множество  $S_2$ . Граничные точки, в которых наблюдается равенство обоих подынтегральных

ных членов, на средний риск не влияют, поэтому их можно распределять произвольно. Для определённости будем относить их в область  $S_2$ .

Итак, согласно критерию минимума среднего риска, области решений определяются следующим образом:

- подмножество  $S_1$ , соответствующее гипотезе  $H_1$ , образуют те точки  $\mathbf{z}$ , для которых

$$(c_{21} - c_{11})w(\mathbf{z} / H_1)p(H_1) > (c_{12} - c_{22})w(\mathbf{z} / H_2)p(H_2);$$

- подмножество  $S_2$ , соответствующее гипотезе  $H_2$ , образуют те точки  $\mathbf{z}$ , для которых

$$(c_{21} - c_{11})w(\mathbf{z} / H_1)p(H_1) \leq (c_{12} - c_{22})w(\mathbf{z} / H_2)p(H_2).$$

Введём константу

$$\Lambda_0 = \frac{(c_{12} - c_{22})p(H_2)}{(c_{21} - c_{11})p(H_1)}, \quad (1.1.8)$$

называемую *порогом принятия решения*. Тогда можно сформулировать стратегию принятия решений, оптимальную по критерию минимума среднего риска: *если отношение условных плотностей вероятностей  $w(\mathbf{z}/H_1)$  и  $w(\mathbf{z}/H_2)$  превышает пороговое значение  $\Lambda_0$ , то принимается гипотеза  $H_1$ ; в противном случае эта гипотеза отвергается, и принимается гипотеза  $H_2$ .*

В литературе такая стратегия известна как *оптимальная байесовская стратегия*<sup>1</sup>, а величина

$$\Lambda(\mathbf{z}) = \frac{w(\mathbf{z} / H_1)}{w(\mathbf{z} / H_2)} \quad (1.1.9)$$

называется *отношением правдоподобия* (поскольку условные плотности вероятностей  $w(\mathbf{z}/H_i)$  в статистике называются функциями правдоподобия). Она представляет собой отношение двух функций от случайных величин и, таким образом, сама является случайной величиной, причём независимо от размерности  $M$  функция  $\Lambda(\mathbf{z})$  — одномерная. Таким образом, процедура принятия решения сводится к вычислению отношения правдоподобия  $\Lambda(\mathbf{z})$

---

<sup>1</sup> Название стратегии по имени английского математика XVII века Томаса Бэйеса (Thomas Báyes) является весьма условным и связано, скорее, с возможностью использования в рамках этой стратегии соответствующей формулы (формулы Бэйеса) для вычисления обратных вероятностей.



и сравнения с пороговым значением  $\Lambda_0(C, P(H_1), P(H_2))$ , зависящим в общем случае от элементов ценовой матрицы и априорных вероятностей гипотез.

В качестве иллюстрации на рис. 1.3 показано разбиение одномерной области  $S$  (всей вещественной прямой) на непересекающиеся одномерные интервалы  $S_1$  и  $S_2$ .

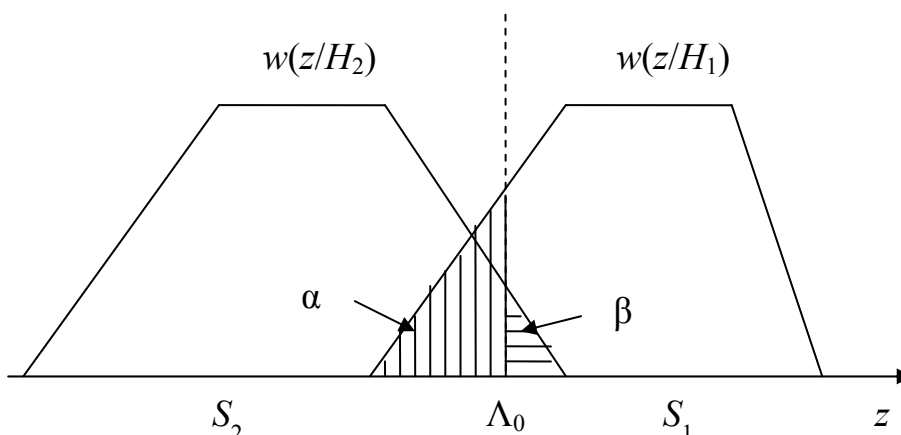


Рис. 1.3. Разбиение одномерной области на интервалы принятия решения

Пороговая точка  $\Lambda_0$ , являющаяся границей разбиения, по договорённости относится к интервалу  $S_2$ . Схематично показаны условные (в данном случае — одномерные) плотности вероятности  $w(z/H_1)$  и  $w(z/H_2)$ , соответствующие каждой из гипотез. Площадь под кривой  $w(z/H_1)$ , образуемая левее точки  $\Lambda_0$ , равна вероятности  $\alpha$  ошибки первого рода. Аналогично, площадь под кривой  $w(z/H_2)$ , образуемая правее точки  $\Lambda_0$ , равна вероятности  $\beta$  ошибки второго рода.

Зачастую оказывается полезной несколько иная запись, основанная на монотонности некоторых элементарных функций, в частности логарифмической функции<sup>1</sup>: *принимается решение в пользу гипотезы  $H_1$ , если выполняется неравенство*

$$\ln \Lambda(\mathbf{z}) > \ln \Lambda_0 \quad (1.1.10)$$

*или в пользу  $H_2$ , если неравенство не выполняется.*

Прямой расчёт вероятностей ошибок первого и второго рода посредством вычисления многомерных интегралов (1.1.1) и (1.1.2) как правило,

---

<sup>1</sup> Причиной этого является использование экспоненциальной функции в моделях, описывающих помехи в каналах передачи, например, в моделях гауссовского или пуассоновского шума.

связан со значительными техническими трудностями. Однако возможно существенное упрощение задачи посредством сведения её к однократному интегрированию.

Обозначим через  $w_\Lambda(L/H_1)$  условную плотность распределения отношения правдоподобия при условии справедливости гипотезы  $H_1$  и через  $w_\Lambda(L/H_2)$  — аналогичную плотность для гипотезы  $H_2$ . Тогда, согласно определению областей разбиения, можно записать

$$\begin{aligned}\alpha &\equiv P\{\mathbf{z} \in S_2 / H_1\} = P\{\Lambda(\mathbf{z}) \leq \Lambda_0 / H_1\} = \\ &= \int_{-\infty}^{\Lambda_0} w_\Lambda(L/H_1) dL = F_\Lambda(\Lambda_0 / H_1); \quad (1.1.11)\end{aligned}$$

$$\begin{aligned}\beta &\equiv P\{\mathbf{z} \in S_1 / H_2\} = P\{\Lambda(\mathbf{z}) > \Lambda_0 / H_2\} = \\ &= \int_{\Lambda_0}^{\infty} w_\Lambda(L/H_2) dL = 1 - F_\Lambda(\Lambda_0 / H_2), \quad (1.1.12)\end{aligned}$$

где  $F_\Lambda(\Lambda_0/H_1)$  и  $F_\Lambda(\Lambda_0/H_2)$  — интегральные функции распределения отношения правдоподобия при гипотезах  $H_1$  и  $H_2$  соответственно.

Рассмотрим следующий пример. Пусть гипотезой  $H_1$  является утверждение о том, что постоянное напряжение на выходе источника имеет значение  $a_1$ , а по гипотезе  $H_2$  — значение  $a_2 \neq a_1$ . Для определённости будем считать, что  $a_2 < a_1$ . Наблюдение за состоянием источника осуществляется посредством взятия  $M$  независимых отсчётов на фоне не зависящего от сигнала аддитивного гауссовского шума с нулевым средним и дисперсией  $\sigma^2$ .

В данной задаче отдельными результатами наблюдения являются элементы выборки

$$z_k = a_i + n_k, \quad i = 1, 2; \quad k = 1, \dots, M,$$

где распределение  $w(n_k)$  для  $k$ -го шумового отсчёта  $n_k$  есть

$$w(n_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{n_k^2}{2\sigma^2}\right).$$

Нетрудно видеть, что значение источника  $a_1$  или  $a_2$  определяет появление ненулевого математического ожидания в условных распределениях плотностей  $w(z_k/H_i)$ :

$$w(z_k / H_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z_k - a_i)^2}{2\sigma^2}\right].$$

В силу независимости отсчётов  $M$ -элементные распределения  $w(\mathbf{z}/H_i)$  факторизуются:

$$w(\mathbf{z} / H_i) = \frac{1}{(2\pi\sigma^2)^{M/2}} \prod_{k=1}^M \exp\left[-\frac{(z_k - a_i)^2}{2\sigma^2}\right],$$

так что, согласно (1.1.7), отношение правдоподобия имеет вид

$$\begin{aligned} \Lambda(\mathbf{z}) &= \frac{\prod_{k=1}^M \exp\left[-\frac{(z_k - a_1)^2}{2\sigma^2}\right]}{\prod_{k=1}^M \exp\left[-\frac{(z_k - a_2)^2}{2\sigma^2}\right]} = \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^M [(z_k - a_1)^2 - (z_k - a_2)^2]\right\} = \\ &= \exp\left[-\frac{(a_1^2 - a_2^2)M}{2\sigma^2} + \frac{a_1 - a_2}{\sigma^2} \sum_{k=1}^M z_k\right]. \end{aligned} \quad (1.1.13)$$

Воспользуемся альтернативной формулировкой (1.1.10), позволяющей использовать логарифм отношения правдоподобия:

$$\ln \Lambda(\mathbf{z}) = -\frac{(a_1^2 - a_2^2)M}{2\sigma^2} + \frac{a_1 - a_2}{\sigma^2} \sum_{k=1}^M z_k. \quad (1.1.13a)$$

Тогда можно записать правило принятия решения в пользу, например, гипотезы  $H_1$  в виде выполнения неравенства

$$-\frac{(a_1^2 - a_2^2)M}{2\sigma^2} + \frac{a_1 - a_2}{\sigma^2} \sum_{k=1}^M z_k > \ln \Lambda_0.$$

Здесь величины  $a_1$ ,  $a_2$ ,  $\sigma^2$  и  $M$  не изменяются в процессе решения задачи (наблюдения за источником), и их можно перенести в правую часть последнего неравенства, образуя модифицированный порог принятия решения. Таким образом, правило принятия гипотезы  $H_1$  окончательно можно записать в следующем виде:

$$\sum_{k=1}^M z_k > \Lambda_1 = \frac{\sigma^2 \ln \Lambda_0}{a_1 - a_2} + \frac{(a_1 + a_2)M}{2}. \quad (1.1.14)$$

На рис. 1.4 представлена структурная схема решающего устройства, содержащая устройство выборки и хранения (УВХ), сумматор ( $\Sigma$ ) и устройство сравнения, на один вход которого подаётся накопленная сумма из  $M$  отсчётов принятого колебания, а на другой — модифицированный порог  $\Lambda_1$ , зависящий, в общем случае, от параметров сигнала  $a_1$  и  $a_2$ , оцененного независимым образом параметра  $\sigma$ , а также исходного порога  $\Lambda_0$  принятия решения.

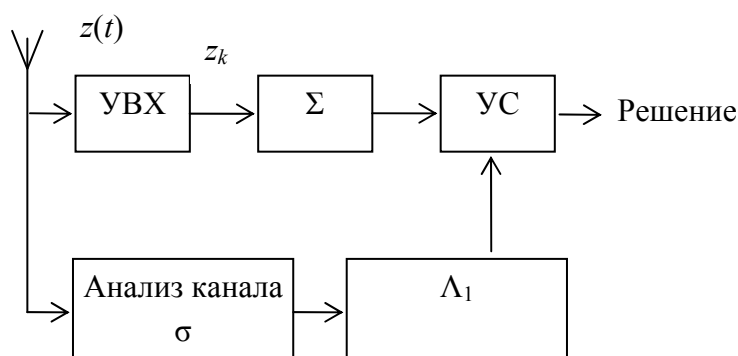


Рис. 1.4. Структурная схема решающего устройства

Необходимость независимого измерения  $\sigma$ , т. е. оценки текущего состояния канала, является достаточно серьёзной, хотя и решаемой задачей. Однако, как нетрудно видеть, если  $\Lambda_0 = 1$  (что практически всегда так), то необходимость в таких измерениях отпадает в силу обнуления соответствующего слагаемого.

Граница, разделяющая области разбиения, определяется равенством

$$\sum_{k=1}^M z_k = \Lambda_1,$$

геометрически представляющей собой  $M$ -мерный симплекс, т. е. гиперплоскость, перпендикулярную единичному вектору  $\mathbf{1} = (1, \dots, 1)$  и расположенную на расстоянии  $\Lambda_1$  от начала координат<sup>1</sup>.

Заметим, что в получившемся решающем устройстве наблюдаемые данные входят в отношение правдоподобия только в виде суммы. Это пример так называемой *достаточной статистики*. В общем случае (см. далее разд. 2.1) оценка

<sup>1</sup> Это означает, что выбранная эвристическим способом граница на рис. 1.2 является наилучшей по критерию минимума среднего риска.

$$\hat{\theta} = f(z_1, z_2, \dots, z_M)$$

называется достаточной оценкой параметра  $\theta$  в распределении  $w(\mathbf{z}/\theta)$ , если условное распределение

$$W(z_1, z_2, \dots, z_M / \hat{\theta})$$

не зависит от этого параметра  $\theta$ .

Из этого определения следует, что достаточная оценка содержит всю информацию о неизвестном параметре  $\theta$ , которую можно получить в процессе наблюдения. Иными словами, для оценки параметра  $\theta$  нет необходимости знать по отдельности каждый из элементов  $z_k$  выборки  $\mathbf{z}$ , а достаточно иметь лишь какую-нибудь одну функцию выборочных значений — достаточную оценку (достаточную статистику).

Приведённое определение не даёт простого признака достаточной оценки, так как вычисление условного распределения  $W(z_1, z_2, \dots, z_M / \hat{\theta})$  может оказаться весьма трудоёмкой задачей. Простым и удобным для практического использования является признак факторизации функции правдоподобия выборки, заключающийся в возможности представления  $w(\mathbf{z}/\theta)$  в виде произведения двух неотрицательных сомножителей:

$$W(z_1, z_2, \dots, z_M / \theta) = g(\hat{\theta} / \theta) h(z_1, z_2, \dots, z_M),$$

первый из которых зависит от достаточной статистики и оцениваемого параметра, а второй представляет собой функцию лишь выборки.

Определим вероятности ошибок для полученной решающей схемы.

Поскольку сумма нормальных величин также нормальна, распределение случайной величины  $\zeta = \sum_{k=1}^M z_k$  — это нормальное распределение с условными математическими ожиданиями

$$\mathbf{E}[\zeta / H_1] = \mathbf{E}\left[\sum_{k=1}^M z_k / H_1\right] = \sum_{k=1}^M \mathbf{E}[z_k / H_1] = M a_1;$$

$$\mathbf{E}[\zeta / H_2] = \mathbf{E}\left[\sum_{k=1}^M z_k / H_2\right] = \sum_{k=1}^M \mathbf{E}[z_k / H_2] = M a_2$$

и дисперсией (в силу независимости элементов выборки)

$$\mathbf{D}[\zeta / H_j] = \mathbf{D}\left[\sum_{k=1}^M z_k / H_j\right] = \sum_{k=1}^M \mathbf{D}[z_k / H_j] = M \sigma^2, j = 1, 2.$$

Тогда вероятность ошибки первого рода

$$\begin{aligned}
 \alpha &= \frac{1}{\sqrt{2\pi M\sigma^2}} \int_{-\infty}^{\Lambda_1} \exp\left[-\frac{(L - Ma_1)^2}{2M\sigma^2}\right] dL = \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\Lambda_1 - Ma_1}{\sqrt{M}\sigma}} \exp(-t^2 / 2) dt = F_0\left(\frac{\Lambda_1 - Ma_1}{\sqrt{M}\sigma}\right) = \\
 &= F_0\left[\frac{\sigma \ln \Lambda_0}{(a_1 - a_2)\sqrt{M}} - \frac{\sqrt{M}}{2\sigma}(a_1 - a_2)\right], \tag{1.1.15a}
 \end{aligned}$$

где  $F_0$  — табулированная функция Лапласа, являющаяся интегральной функцией стандартного (т. е. с нулевым средним и единичной дисперсией) гауссовского распределения.

Аналогичным образом вероятность ошибки второго рода

$$\begin{aligned}
 \beta &= \frac{1}{\sqrt{2\pi M\sigma^2}} \int_{\Lambda_1}^{\infty} \exp\left[-\frac{(L - Ma_2)^2}{2M\sigma^2}\right] dL = 1 - F_0\left(\frac{\Lambda_1 - Ma_2}{\sqrt{M}\sigma}\right) = \\
 &= 1 - F_0\left[\frac{\sigma \ln \Lambda_0}{(a_1 - a_2)\sqrt{M}} + \frac{\sqrt{M}}{2\sigma}(a_1 - a_2)\right]. \tag{1.1.15б}
 \end{aligned}$$

Видно, что с ростом размера выборки  $M$  вероятности ошибок стремятся к нулю (с учётом того, что  $a_2 < a_1$ ).

Полагая  $\Lambda_1 = M(a_1 + a_2)/2$  и используя свойство функции Лапласа

$$F_0(-x) = 1 - F_0(x),$$

получаем

$$\alpha = F_0\left(\frac{a_2 - a_1\sqrt{M}}{2\sigma}\right) = 1 - F_0\left(\frac{a_1 - a_2\sqrt{M}}{2\sigma}\right) = \beta,$$

т. е. одинаковые вероятности ошибок первого и второго родов. Так, при  $a_1 = -a_2$  (симметричные гипотезы) равенство вероятностей ошибок достигается при нулевом значении порога  $\Lambda_1$ . Если же при симметричных гипотезах использовать значения порога, заметно отличные от нулевого, то вероятности ошибок могут заметно различаться — тем больше, чем больше размер выборки  $M$ .

Например, устанавливая порог  $\Lambda_0 = 1$  и выбирая  $a_1 = 1,5$ ,  $a_2 = -1,5$ ,  $\sigma = 1,0$ , имеем

$$\Lambda_1 = 0, \alpha = \beta = 1,7 \cdot 10^{-2} \text{ для } M = 2;$$

и

$$\Lambda_1 = 0, \alpha = \beta = 4,0 \cdot 10^{-4} \text{ для } M = 5.$$

Если теперь при тех же  $a_1, a_2$  и  $\sigma$  выбрать  $\Lambda_0 = 0,25$ , то получим

$$\Lambda_1 = -0,46, \alpha = 0,7 \cdot 10^{-2}, \beta = 3,6 \cdot 10^{-2} \text{ для } M = 2;$$

и

$$\Lambda_1 = -0,46, \alpha = 1,8 \cdot 10^{-4}, \beta = 8,2 \cdot 10^{-4} \text{ для } M = 5.$$

Вернёмся к рассмотрению оптимальной байесовской стратегии. Может оказаться так, что по условиям решаемой задачи фигурирующие в (1.8) относительные цены  $c_{12} - c_{22}$  и  $c_{21} - c_{11}$  окажутся одинаковыми. Более того, зачастую эти значения вообще неизвестны, и тогда соотношение

$$c_{12} - c_{22} = c_{21} - c_{11} \quad (1.1.16)$$

принимается за неимением какой-либо другой априорной информации (этим достигается одинаковая нежелательность любых ошибочных решений). Тогда оптимальная стратегия описывается неравенством

$$\frac{w(\mathbf{z} / H_1) p(H_1)}{w(\mathbf{z} / H_2) p(H_2)} > 1, \quad (1.1.17)$$

т. е. следует принять гипотезу  $H_i$  ( $i = 1, 2$ ), соответствующую наибольшему из выражений  $w(\mathbf{z}/H_i)p(H_i)$ . Такую стратегию называют *стратегией идеального наблюдателя*.

Часто стратегия идеального наблюдателя определяется более конкретно: плата за все правильные решения равна нулю, плата за все ошибочные решения равна единице, т. е. элементы ценовой матрицы

$$c_{ij} = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases}; i, j = 1, 2.$$

Если в дополнение к тому, что выполняется равенство (1.1.16), для относительных цен, также оказываются одинаковыми и априорные вероятности гипотез  $p(H_i)$  (либо они предполагаются равными, что, например, имеет место в цифровых системах связи), то принятие решение основывается только на использовании условных плотностей  $w(\mathbf{z}/H_i)$  — функций правдоподобия:

$$\frac{w(\mathbf{z} / H_1)}{w(\mathbf{z} / H_2)} > 1. \quad (1.1.18)$$

Решение, состоящее в принятии гипотезы  $H_i$  ( $i = 1, 2$ ), для которой функция правдоподобия  $w(\mathbf{z}/H_i)$  принимает максимальное значение, называется *стратегией, оптимальной по критерию максимума правдоподобия* (МП-стратегией).

В том случае, когда известны априорные вероятности гипотез  $p(H_i)$ , но отсутствуют какие-либо обоснованные соображения относительно цен  $c_{ij}$ , может быть использован несколько иной подход к выработке правила принятия решения, основанный на апостериорных (“обратных”) вероятностях.

Согласно формуле Байеса, апостериорная вероятность гипотезы  $H_i$ , когда наблюдается выборка  $\mathbf{z} = \{z_1, z_2, \dots, z_M\}$ , равна

$$P\{H_i/\mathbf{z}\} = \frac{w(\mathbf{z}/H_i)P(H_i)}{w(\mathbf{z}/H_1)P(H_1) + w(\mathbf{z}/H_2)P(H_2)}, \quad i = 1, 2.$$

Эти вероятности представляют собой наиболее полную характеристику состояния источника при соответствующей гипотезе и располагаемых априорных данных. Поэтому естественно принять следующее правило: решение принимается в пользу той гипотезы, для которой максимальна апостериорная вероятность. Такую стратегию называют *стратегией, оптимальной по критерию максимума апостериорной вероятности* (МАВ-стратегией).

Заметим, что стратегии (1.1.17) и (1.1.18) являются частными случаями общей байесовской стратегии. Однако все эти стратегии могут рассматриваться и использоваться независимо друг от друга.

Перепишем выражение (1.1.6) для среднего риска с использованием вероятностей  $\alpha$  и  $\beta$  ошибок первого и второго рода:

$$R = c_{11}p(H_1)(1 - \alpha) + c_{21}p(H_1)\alpha + c_{12}p(H_2)\beta + c_{22}p(H_2)(1 - \beta).$$

Выразим средний риск как функцию какой-либо одной априорной вероятности, например  $p(H_1)$ . Учитывая, что  $p(H_2) = 1 - p(H_1)$  и обозначая для краткости  $p(H_1) = p_1$ , имеем

$$R(p_1) = c_{22} + (c_{12} - c_{22})\beta(p_1) + \\ + [c_{11} - c_{22} + (c_{21} - c_{11})\alpha(p_1) - (c_{12} - c_{22})\beta(p_1)]p_1. \quad (1.1.19)$$

Как видно, это соотношение в общем случае *не* является простой линейной зависимостью  $R$  от  $p_1$ , ибо вероятность  $p_1$  входит в (1.1.19) как явным образом, так и неявно — поскольку по мере изменения  $p_1$  изменяются



и области решений, а следовательно, и вероятности ошибок  $\alpha$  и  $\beta$ , как например, в (1.1.15) и (1.1.16).

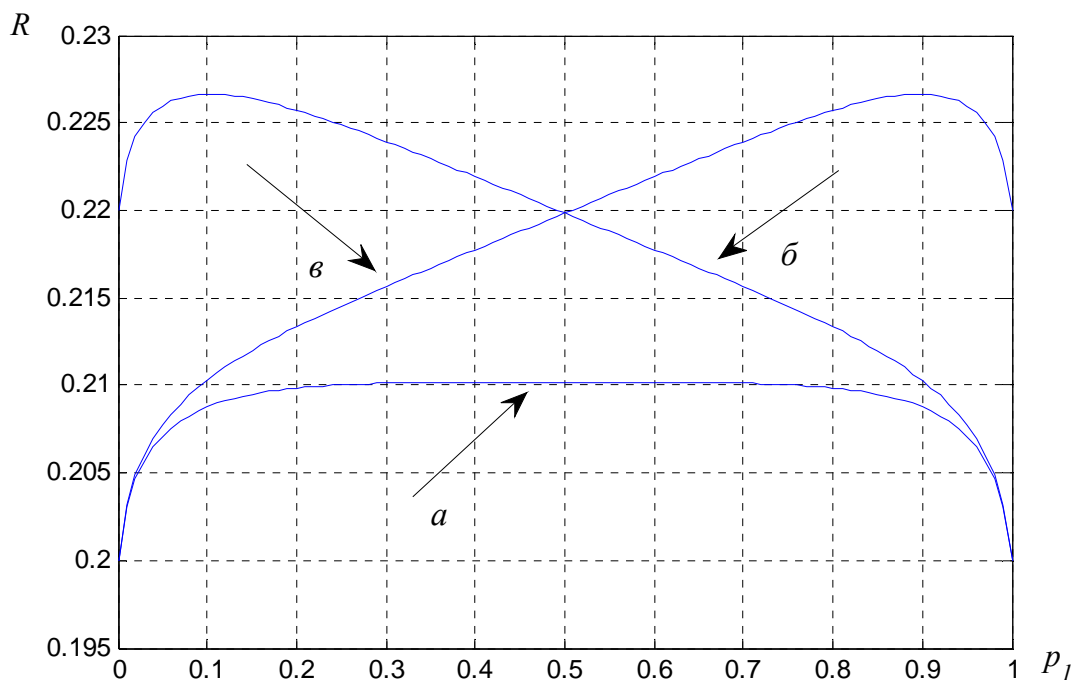


Рис. 1.5. Зависимость среднего риска от вероятности ошибки для различных ценовых матриц:  $C_1$  (*a*),  $C_2$  (*б*) и  $C_3$  (*в*)

В качестве иллюстрации на рис. 1.5 представлены зависимости  $R(p_1)$  для трёх различных ценовых матриц

$$C_1 = \begin{pmatrix} 0,2 & 0,8 \\ 0,8 & 0,2 \end{pmatrix}, C_2 = \begin{pmatrix} 0,22 & 0,78 \\ 0,8 & 0,2 \end{pmatrix} \text{ и } C_3 = \begin{pmatrix} 0,2 & 0,8 \\ 0,78 & 0,22 \end{pmatrix}$$

и независимой гауссовской выборки, т. е. когда вероятности ошибок вычисляются согласно (1.1.5). Эти зависимости могут иметь симметричный (относительно  $p_1 = 0,5$ ) или несимметричный вид, что определяется симметричностью самих элементов  $c_{ij}$  ( $i, j = 1, 2$ ) ценовой матрицы<sup>1</sup>.

Рассмотрим ситуацию, при которой наблюдателю в точности не известно значение априорных вероятностей гипотез и, следовательно, возникают трудности с определением порогового значения. Предположим, что вместо истинного значения  $p_1$  выбрано некоторое приближённое значение

<sup>1</sup> Ещё раз отметим, что важны не абсолютные значения элементов  $c_{ij}$ , а их разностное отношение, входящее в определение порога (1.1.8).

$p'_1$ , по которому рассчитываются и порог  $\Lambda'_0$ , и соответствующие вероятности ошибок  $\alpha'$ ,  $\beta'$ . При этом осуществляется некоторая, вообще говоря, неоптимальная стратегия, для которой в определении среднего риска по-прежнему фигурируют истинные априорные вероятности  $p_1$  и  $p_2 = 1 - p_1$ :

$$R' = c_{22} + (c_{12} - c_{22})\beta' + [c_{11} - c_{22} - (c_{12} - c_{22})\beta' + (c_{21} - c_{11})\alpha']p_1. \quad (1.1.20)$$

Соотношение (1.1.20) представляет собой прямую линию с отрицательным или положительным наклоном в зависимости от коэффициента перед  $p_1$ , а поскольку  $R'$  — средний риск некоторой неоптимальной стратегии, полученная прямая  $R'(p_1)$  располагается выше оптимальной байесовской зависимости  $R_{\min}(p_1)$  во всех точках, кроме одной:  $p_1 = p'_1$ , в которой происходит касание двух зависимостей (рис. 1.6).

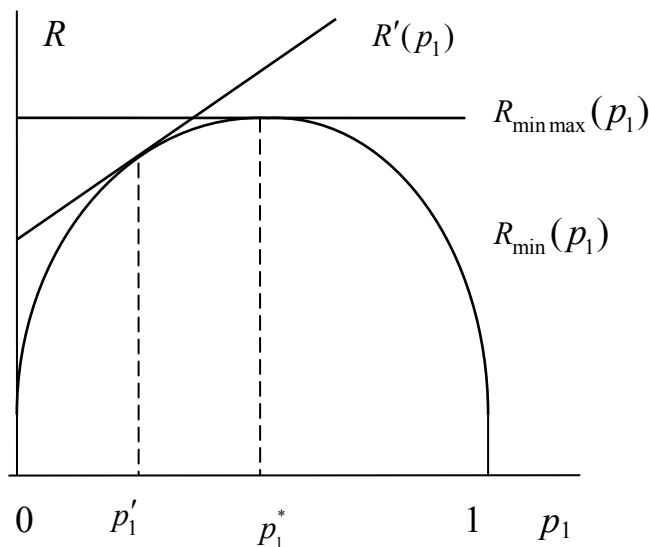


Рис 1.6. Средний риск при неточно известных априорных вероятностях

Если, к примеру,  $R'(p_1)$  — это прямая с положительным углом наклона, то при близких к 1 значениях  $p_1$  наблюдается существенное превышение реального среднего риска  $R'$  над минимально возможным байесовским средним риском  $R$ . Аналогичная ситуация возникает и при отрицательном угле наклона: для близких к нулевым значениям  $p_1$  наблюдаются большие различия между  $R'(p_1)$  и  $R(p_1)$ .

Анализ представленных на рис. 1.6 зависимостей показывает возможность использования стратегии, при которой любые возможные потери не будут превосходить наперёд заданной величины. Для реализации такой

стратегии зависимость среднего риска от априорной вероятности выбирается в виде горизонтальной прямой  $R_{\min \max}(p_1)$ , проходящей через точку  $p_1^*$  локального максимума  $R(p_1)$ . Конечно, при этом заметно возрастают потери в окрестности точки  $p_1 = p_1'$ , где различие между  $R$  и  $R'$  было минимальным, однако одновременно гарантируется, что максимальные потери (в окрестностях  $p_1 = 0$  или  $p_1 = 1$ ) не будут превосходить значения  $R^* = R(p_1^*)$ . Другими словами, при выбранной стратегии минимизируется максимальный риск, поэтому её принято называть *оптимальной минимаксной стратегией*. Разность  $R^* - R$  между минимаксным (при неизвестной  $p_1$ ) и байесовским (при известной  $p_1$ ) рисками является платой за отсутствие априорной информации о состояниях источника.

Поскольку при минимаксной стратегии угловой коэффициент зависимости  $R(p_1)$  равен нулю, из (1.1.20) получаем соотношение для элементов ценовой матрицы:

$$c_{11} - c_{22} - (c_{12} - c_{22})\beta(p_1^*) = -(c_{21} - c_{11})\alpha(p_1^*). \quad (1.1.21)$$

В частности, при  $c_{11} = c_{22} = 0$ , что определяет внутреннее по отношению к интервалу  $[0; 1]$  расположение точки максимума  $p_1^*$ , имеем

$$c_{12}\beta(p_1^*) = c_{21}\alpha(p_1^*). \quad (1.1.22)$$

Можно показать, что минимаксная стратегия является частным случаем общей байесовской стратегии для наименее благоприятного априорного распределения состояний. Такое наименее благоприятное распределение характеризуется тем, что для него минимальный средний риск имеет наибольшую величину среди всех других величин среднего риска, вычисленных для всех других распределений. К сожалению, не существует общего метода, при помощи которого можно было бы находить наименее благоприятное распределение, однако доказано, что байесовское правило, которому соответствуют одинаковые для всех состояний условные риски, является минимаксным. Это обстоятельство может быть использовано для определения наименее благоприятного распределения и, следовательно, минимаксного правила.

Так, для рассматриваемого примера нормальной выборки имеем следующее трансцендентное уравнение для определения  $p_1^*$ :

$$\frac{c_{12}}{c_{21}} \left\{ 1 - F_0 \left[ \frac{\sigma \ln \Lambda_0(p_1^*)}{(a_1 - a_2)\sqrt{M}} + \frac{\sqrt{M}}{2\sigma}(a_1 - a_2) \right] \right\} =$$

$$= F_0 \left[ \frac{\sigma \ln \Lambda_0(p_1^*)}{(a_1 - a_2)\sqrt{M}} - \frac{\sqrt{M}}{2\sigma}(a_1 - a_2) \right],$$

где

$$\Lambda_0 = \frac{(c_{12} - c_{22})(1 - p_1^*)}{(c_{21} - c_{11})p_1^*}.$$

Полагая  $c_{11} = c_{22} = 0$  (нулевой штраф за правильные решения) и  $c_{12} = c_{21}$  (одинаковый штраф за неправильные решения), получаем  $p_1^* = 1/2$ , т. е. наименее благоприятным с точки зрения байесовского риска является равновероятное появление состояний источника.

В ряде случаев оказывается крайне сложно, а иногда полностью невозможно определить даже хотя бы приблизительно априорные вероятности гипотез. Наиболее ярким примером такой ситуации являются радиолокационные задачи, связанные с обнаружением сигналов. Аналогичные сложности возникают и в ряде связных задач, например, при обнаружении начала пакета передачи служебных данных. В этих условиях более приемлемым оказывается использование *стратегии, оптимальной по критерию Неймана–Пирсона* (НП-стратегии).

С точки зрения потребителя желательно сделать значение  $\alpha$  как можно меньше, а  $1 - \beta$  как можно больше. Однако понятно, что указанные цели противоречат друг другу, и разумным выходом из этой ситуации является ограничение одного из значений и максимизация (или минимизация) другого. Согласно стратегии Неймана–Пирсона (Neuman, Pearson), фиксируется уровень значимости  $\alpha$  (т. е. вероятность отвергнуть правильную гипотезу) и используется такая процедура проверки, которая имеет наибольшую мощность  $1 - \beta$  (т. е. наименьшую вероятность ошибки второго рода) среди всех процедур, характеризующихся уровнем значимости, не большим, чем  $\alpha$ .

Оптимальность сформулированной стратегии означает следующее. Пусть  $S_2$  — область принятия гипотезы  $H_2$ , включающая все те элементы  $\mathbf{z}$ ,

для которых  $\Lambda(\mathbf{z}) \leq \Lambda_0$ , причём порог  $\Lambda_0$  выбирается из условия заданного уровня значимости:

$$P\{\Lambda(\mathbf{z}) \leq \Lambda_0 / H_1\} = \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} = \alpha.$$

Тогда утверждается, что мощность  $1 - \beta(S_2)$  данной процедуры проверки гипотез будет больше аналогичного значения мощности  $1 - \beta(\Theta)$  для любого подмножества  $\Theta$ , получаемого из всего множества  $S$  возможных значений источника, на основании удовлетворения условия

$$P\{\mathbf{z} \in \Theta / H_1\} = \int_{\Theta} w(\mathbf{z} / H_1) d\mathbf{z} \leq \alpha. \quad (1.1.23)$$

Покажем справедливость такого утверждения. Рассмотрим вначале частный случай, когда области  $\Theta$  и  $S_2$  не пересекаются (рис. 1.7, *a*). Это означает, что в область  $\Theta$  попадают те элементы  $\mathbf{z}$ , для которых наряду с (1.1.23) выполняется условие

$$\Lambda(\mathbf{z}) = \frac{w(\mathbf{z} / H_1)}{w(\mathbf{z} / H_2)} > \Lambda_0.$$

Другими словами, рассматриваются два правила принятия решения: сравнение отношения правдоподобия  $\Lambda(\mathbf{z})$  с пороговым значением  $\Lambda_0$  и принадлежность элемента  $\mathbf{z}$  к области  $\Theta$ . При этом в первом случае вероятность ошибки первого рода по определению равна  $\alpha$ , а во втором — не превосходит  $\alpha$ . Покажем, что вероятность ошибки второго рода  $\beta(S_2)$  в первом случае не превосходит аналогичного значения  $\beta(\Theta)$  во втором случае.

На основании принятия решения по отношению правдоподобия в области  $S_2$  для элементов  $\mathbf{z}$  справедливо соотношение  $\Lambda(\mathbf{z}) \leq \Lambda_0$ , при котором

$$1 - \beta(S_2) = \int_{S_2} w(\mathbf{z} / H_2) d\mathbf{z} \geq \int_{S_2} \frac{1}{\Lambda_0} w(\mathbf{z} / H_1) d\mathbf{z} = \frac{\alpha}{\Lambda_0}.$$

Далее, по определению области  $\Theta$ , т. е. согласно неравенству (1.1.23),

$$\frac{\alpha}{\Lambda_0} \geq \frac{1}{\Lambda_0} \int_{\Theta} w(\mathbf{z} / H_1) d\mathbf{z}.$$

Наконец, поскольку при реализации описываемой оптимальной стратегии в область  $\Theta$  попадут те элементы  $\mathbf{z}$ , для которых справедливо соотношение  $\Lambda(\mathbf{z}) > \Lambda_0$ , имеем

$$\frac{1}{\Lambda_0} \int_{\Theta} w(\mathbf{z} / H_1) d\mathbf{z} > \int_{\Theta} w(\mathbf{z} / H_2) d\mathbf{z} = 1 - \beta(\Theta),$$

т. е.

$$1 - \beta(S_2) > 1 - \beta(\Theta),$$

что и требовалось показать.

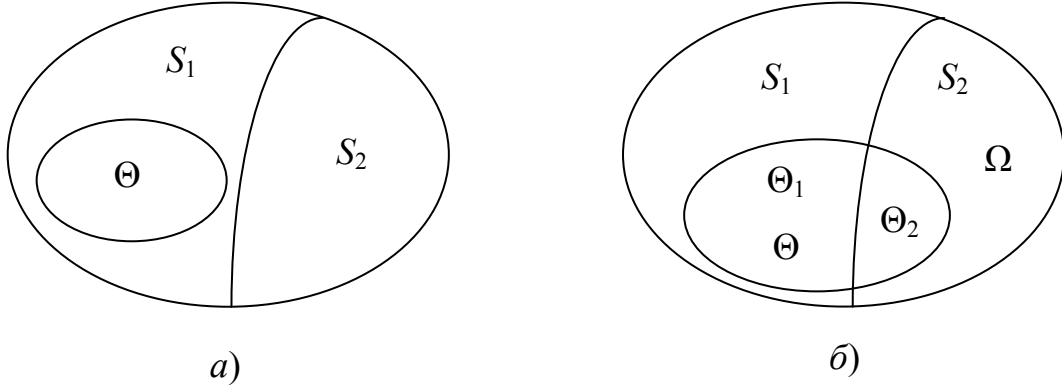


Рис. 1.7. Разбиение множества значений источника с пересечением (а) и непересечением (б) областей  $\Theta$  и  $S_1$

Обратимся теперь к более общей ситуации, при которой области  $\Theta$  и  $S_2$  имеют непустое пересечение (рис. 1.7, б), и для удобства записи введём следующие обозначения:

$\Theta_1 = S_1 \cap \Theta$  — часть области  $\Theta$ , которая не пересекается с областью  $S_2$ ;

$\Theta_2 = S_2 \cap \Theta$  — часть области  $\Theta$ , которая не пересекается с областью  $S_1$ ;

$\Omega = S_2 \setminus \Theta$  — часть области  $S_2$ , которая не пересекается с областью  $\Theta$ .

Тогда в силу свойства аддитивности интеграла по соответствующим областям и на основании правила принятия решения в области  $\Omega$

$$\begin{aligned} 1 - \beta(S_2) &= \int_{S_2} w(\mathbf{z} / H_2) d\mathbf{z} = \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \int_{\Omega} w(\mathbf{z} / H_2) d\mathbf{z} \geq \\ &\geq \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{1}{\Lambda_0} \int_{\Omega} w(\mathbf{z} / H_1) d\mathbf{z} = \\ &= \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{1}{\Lambda_0} \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} - \frac{1}{\Lambda_0} \int_{\Theta_2} w(\mathbf{z} / H_1) d\mathbf{z} = \\ &= \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{\alpha}{\Lambda_0} - \frac{1}{\Lambda_0} \int_{\Theta_2} w(\mathbf{z} / H_1) d\mathbf{z}. \end{aligned}$$

Далее, на основании неравенства (1.1.23) и конфигурации областей имеем:

$$\begin{aligned} & \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{\alpha}{\Lambda_0} - \frac{1}{\Lambda_0} \int_{\Theta_2} w(\mathbf{z} / H_1) d\mathbf{z} \geq \\ & \geq \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{1}{\Lambda_0} \int_{\Theta} w(\mathbf{z} / H_1) d\mathbf{z} - \frac{1}{\Lambda_0} \int_{\Theta_2} w(\mathbf{z} / H_1) d\mathbf{z} = \\ & = \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{1}{\Lambda_0} \int_{\Theta_1} w(\mathbf{z} / H_1) d\mathbf{z}. \end{aligned}$$

Поскольку в область  $\Theta_1$  попадут те элементы  $\mathbf{z}$ , для которых справедливо соотношение  $\Lambda(\mathbf{z}) > \Lambda_0$ ,

$$\begin{aligned} & \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \frac{1}{\Lambda_0} \int_{\Theta_1} w(\mathbf{z} / H_1) d\mathbf{z} > \int_{\Theta_2} w(\mathbf{z} / H_2) d\mathbf{z} + \int_{\Theta_1} w(\mathbf{z} / H_2) d\mathbf{z} = \\ & = \int_{\Theta} w(\mathbf{z} / H_2) d\mathbf{z} = 1 - \beta(\Theta). \end{aligned}$$

Итак, в общем случае

$$1 - \beta(S_2) < 1 - \beta(\Theta),$$

т. е. мощность НП-стратегии, больше мощности любой другой стратегии, основанной на соотношении (1.1.23).

Отметим, что стратегия Неймана–Пирсона может быть получена как решение задачи на условный экстремум (задача 1.8).

Проиллюстрируем НП-стратегию на примере уже рассмотренной задачи проверки гипотезы о среднем значении нормальной случайной величины.

Для анализируемой  $M$ -элементной выборки отношение правдоподобия по-прежнему определяется (1.1.13) или эквивалентным ему соотношением (1.1.14), однако теперь величина порога  $\Lambda_1$  рассчитывается при заданной вероятности  $\alpha$  ошибки первого рода из условия

$$P \left\{ \sum_{k=1}^M z_k \leq \Lambda_1 / H_1 \right\} = \alpha.$$

Данное условие, по виду аналогичное (1.1.16), представляет собой уравнение для нахождения значения  $\zeta_\alpha$ , являющегося  $\alpha$ -процентной точкой распределения нормальной величины  $\zeta$  с математическим ожиданием  $\mathbf{E}[\zeta / H_1] = M a_1$  и дисперсией  $\mathbf{D}[\zeta / H_1] = M \sigma^2$ :

$$\frac{\Lambda_1 - Ma_1}{\sigma\sqrt{M}} = \zeta_\alpha.$$

Отсюда величина порога равна

$$\Lambda_1 = Ma_1 + \zeta_\alpha \sigma\sqrt{M}. \quad (1.1.24)$$

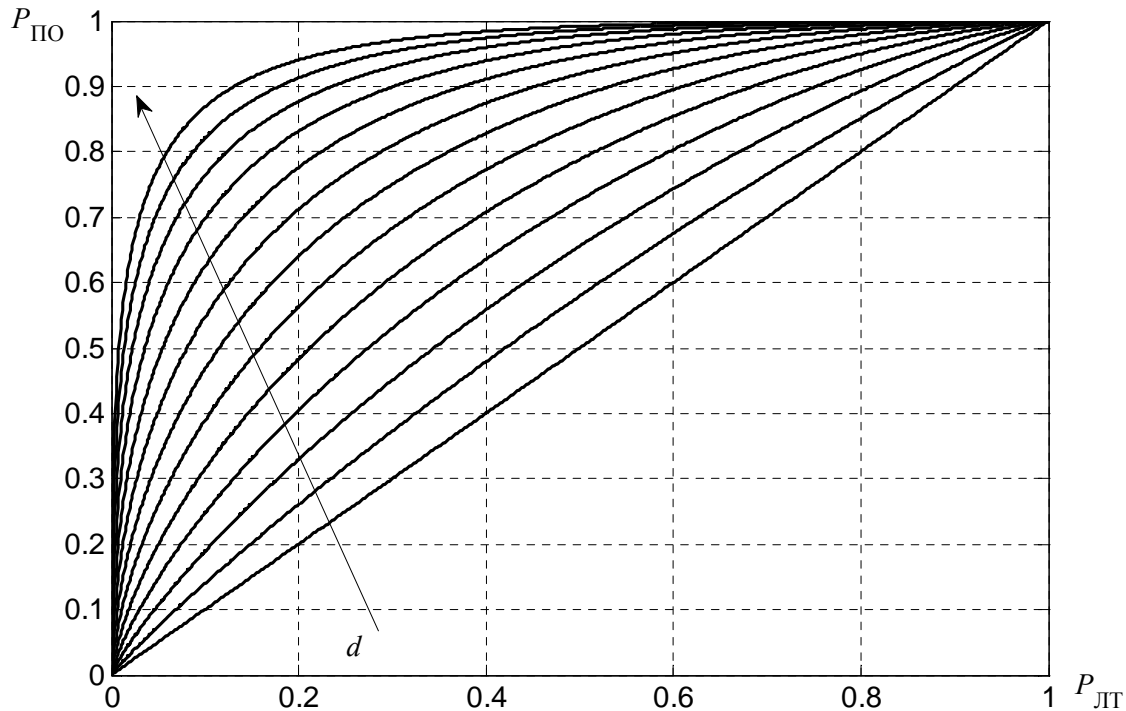


Рис. 1.8. Рабочая характеристика приёмника  $P_{\text{ПО}}$  ( $P_{\text{ЛТ}}$ ) при изменении  $d$

Вероятность ошибки  $\beta$  второго рода рассчитывается согласно (1.1.15), где в качестве порога  $\Lambda_1$  подставляется значение, вычисленное по (1.1.24):

$$\beta = 1 - F_0 \left( \frac{M(a_1 - a_2) + \zeta_\alpha \sigma\sqrt{M}}{\sqrt{M}\sigma} \right). \quad (1.1.25)$$

Соотношение (1.1.25) устанавливает связь между размером выборки, заданной вероятностью ошибки первого рода и минимально возможной вероятностью ошибки второго рода. Видно, что при  $M \rightarrow \infty$  вероятность ошибки  $\beta \rightarrow 0$  аналогично тому, как это было для байесовской стратегии.



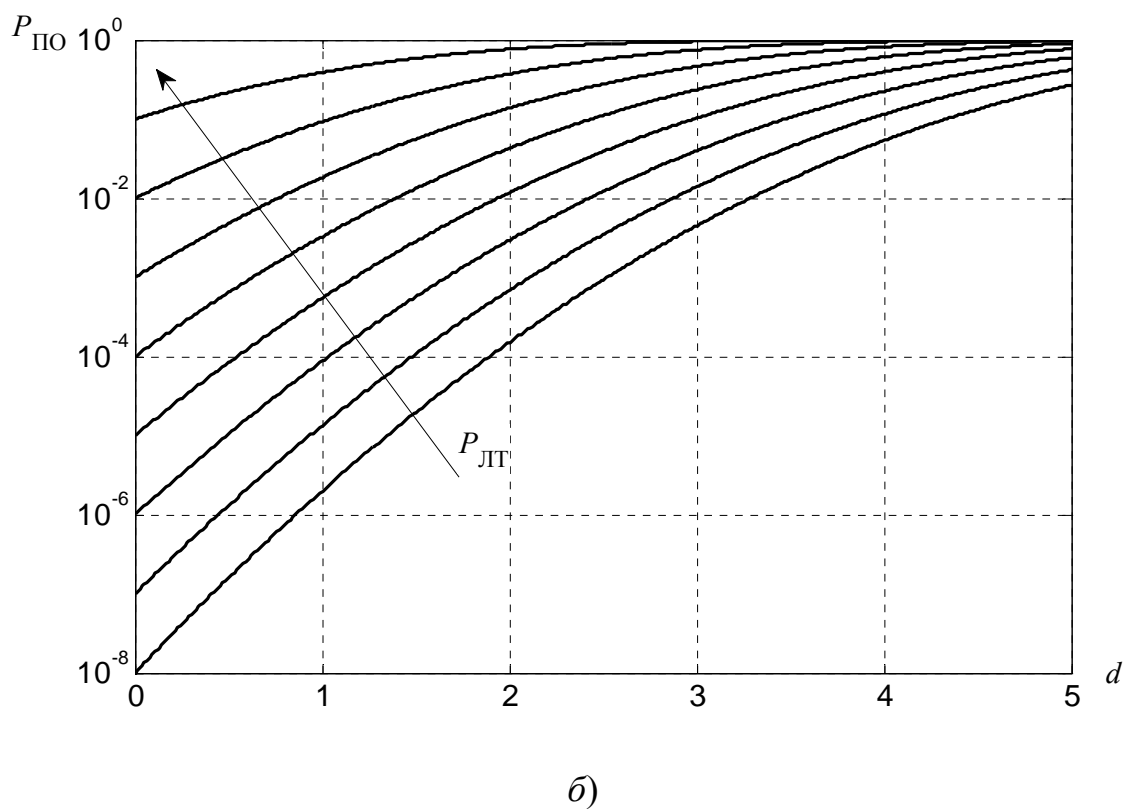
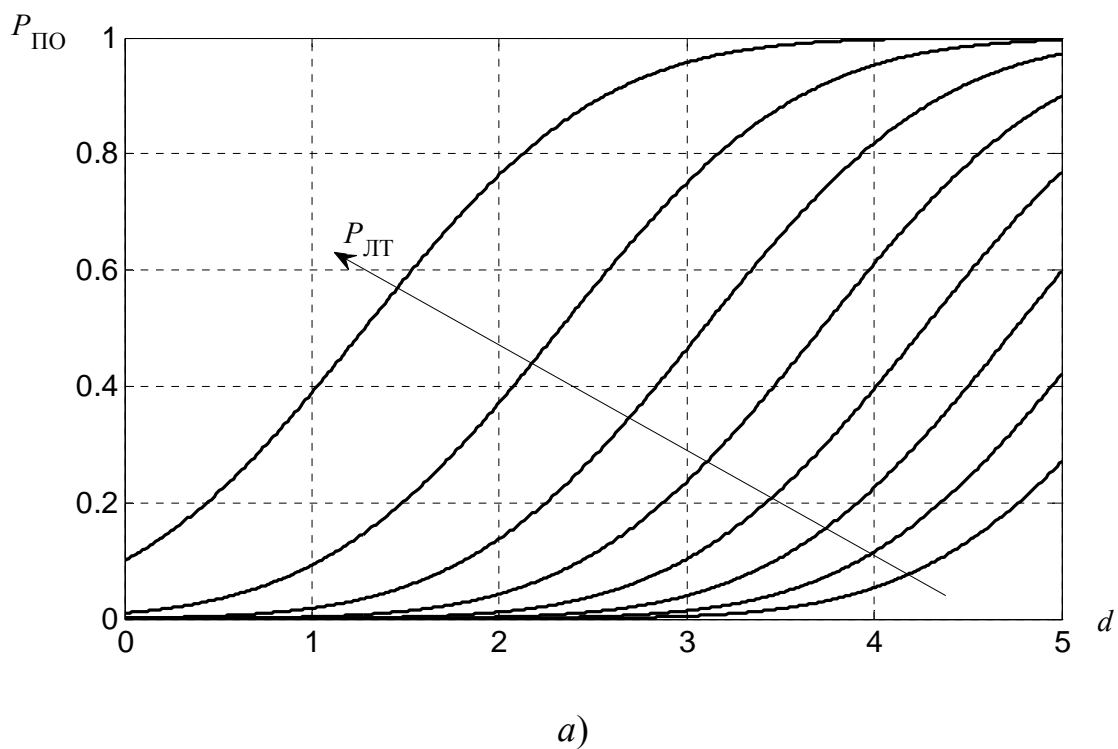


Рис. 1.9. Рабочая характеристика приёмника  $P_{\Pi O}(d)$  при изменении  $P_{ЛГ}$  в натуральном (а) и логарифмическом (б) масштабе по оси ординат

На рис. 1.8 показано семейство зависимостей вероятности правильного обнаружения от вероятности ложной тревоги при изменении в интервале  $0 \dots 2,5$  значения параметра

$$d = \frac{a_1 - a_2}{\sigma} \sqrt{M},$$

являющегося нормированным (на величину среднеквадратического отклонения) расстоянием между двумя возможными значениями математического ожидания. Как и следовало ожидать, при фиксированной  $P_{\text{ЛТ}}$  достоверность решения (т. е. вероятность правильного обнаружения) монотонно возрастает с увеличением нормированного расстояния  $d$  между возможными значениями средних. На рис. 1.9 это же семейство зависимостей — в натуральном и логарифмическом масштабах по оси ординат — представлено в другом виде: в качестве аргумента выбран  $d$ , а  $P_{\text{ЛТ}}$  является параметром, изменяющимся от  $10^{-8}$  до  $10^{-1}$ .

Выше рассматривались задачи проверки гипотез для выборки, элементы которой являются значениями одномерной случайной величины. Непосредственным обобщением таких задач являются задачи проверки гипотез для многомерных распределений, когда выборка  $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$  является векторной, т. е. каждый элемент выборки сам является вектором-столбцом, состоящим из  $n$  элементов. Иными словами, выборка является прямоугольной матрицей<sup>1</sup> размером  $n \times M$ :

$$\mathbf{z} = \begin{pmatrix} z_1^{(1)} & z_2^{(1)} & \dots & z_M^{(1)} \\ z_1^{(2)} & z_2^{(2)} & \dots & z_M^{(2)} \\ \dots & \dots & \dots & \dots \\ z_1^{(n)} & z_2^{(n)} & \dots & z_M^{(n)} \end{pmatrix}. \quad (1.1.26)$$

При этом входящие в выборку вектора  $\mathbf{z}_k$  ( $k = 1, \dots, M$ ), т. е. столбцы матрицы (1.1.26), представляют собой совокупность, вообще говоря, зависимых величин.

Проиллюстрируем теорию проверки гипотез на примере многомерного нормального распределения,  $n$ -мерная плотность вероятности которого может быть записана в матричном представлении:

---

<sup>1</sup> Верхние индексы элементов отражают позицию случайной величины в многомерном распределении, нижние — позицию вектора в выборке.

$$w^{(n)}(\mathbf{z}_k) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{K}}} \exp \left[ -\frac{1}{2} (\mathbf{z}_k - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) \right],$$

где  $\mathbf{a} = [a^{(1)}, \dots, a^{(n)}]^T$  — вектор-столбец средних значений,  $\mathbf{K}$  — корреляционная матрица, элементами которой являются ковариации  $b_{ij}$  пар элементов  $z^{(i)}$  и  $z^{(j)}$  ( $i, j = 1, \dots, n$ ), входящих в каждый из векторов  $\mathbf{z}_k$  ( $k = 1, \dots, M$ ):

$$\begin{aligned} b_{ij} &= \text{Cov}[z^{(i)}, z^{(j)}] = \int_{Z^n} (z^{(i)} - a^{(i)})(z^{(j)} - a^{(j)}) w^{(n)}(\mathbf{z}) d\mathbf{z} = \\ &= \int_{Z^n} z^{(i)} z^{(j)} w^{(n)}(\mathbf{z}) d\mathbf{z} - a^{(i)} a^{(j)} = r_{ij} - a^{(i)} a^{(j)}, \end{aligned}$$

где  $r_{ij}$  — коэффициент корреляции, совпадающий с ковариацией  $b_{ij}$  при нулевых средних, а интегрирование ведётся по всему  $n$ -мерному пространству  $Z^n$  значений вектора  $\mathbf{z}_k$ .

Сформулируем простейшую задачу: проверить гипотезу  $H_1$ , что в рассматриваемой векторной выборке вектор средних равен  $\mathbf{a}_1$  против гипотезы  $H_2$ , что вектор средних равен  $\mathbf{a}_2$  при одинаковой корреляционной матрице  $\mathbf{K}$ .

Логарифм отношения правдоподобия в этом случае имеет вид:

$$\begin{aligned} \ln \Lambda(\mathbf{z}) &= \ln \prod_{k=1}^M \frac{w^{(n)}(\mathbf{z}_k / \mathbf{a}_1)}{w^{(n)}(\mathbf{z}_k / \mathbf{a}_2)} = \sum_{k=1}^M \ln \frac{w^{(n)}(\mathbf{z}_k / \mathbf{a}_1)}{w^{(n)}(\mathbf{z}_k / \mathbf{a}_2)} = \\ &= -\frac{1}{2} \sum_{k=1}^M \left[ (\mathbf{z}_k - \mathbf{a}_1)^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}_1) - (\mathbf{z}_k - \mathbf{a}_2)^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}_2) \right] = \\ &= \sum_{k=1}^M \mathbf{z}_k^T \mathbf{K}^{-1} (\mathbf{a}_1 - \mathbf{a}_2) - \frac{M}{2} (\mathbf{a}_1 + \mathbf{a}_2)^T \mathbf{K}^{-1} (\mathbf{a}_1 - \mathbf{a}_2). \end{aligned} \quad (1.1.27)$$

Таким образом, в рассматриваемом случае оптимальное байесовское правило принятия решения формулируется следующем образом: принимается гипотеза  $H_1$ , что вектор средних равен  $\mathbf{a}_1$ , если

$$\sum_{k=1}^M \mathbf{z}_k^T \mathbf{K}^{-1} (\mathbf{a}_1 - \mathbf{a}_2) > \ln \Lambda_0 + \frac{M}{2} (\mathbf{a}_1 + \mathbf{a}_2)^T \mathbf{K}^{-1} (\mathbf{a}_1 - \mathbf{a}_2), \quad (1.1.28)$$

и принимается альтернативная гипотеза  $H_2$  (вектор средних равен  $\mathbf{a}_2$ ), если неравенство (1.1.28) не выполняется.

При  $n = 1$  векторы  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  и  $\mathbf{z}_k$  превращаются в скаляры, матрица  $\mathbf{K}^{-1}$  также содержит один скалярный элемент, равный  $1/\sigma^2$ , где  $\sigma^2$  — дисперсия

одномерного нормального распределения, и соотношение (1.1.28), как и следовало ожидать, переходит в (1.1.13a) — оптимальную байесовскую стратегию для соответствующей одномерной задачи.

Выражения (1.1.27) и (1.1.28), содержащие в себе матричные операции (умножение, транспонирование, вычисление обратной матрицы), могут быть упрощены с точки зрения вычислительных действий посредством линейного преобразования

$$\mathbf{y}_k = \mathbf{Q}\mathbf{z}_k, \quad (1.1.29)$$

позволяющего декоррелировать зависимые случайные величины, входящие в векторы  $\mathbf{z}_k$ . Если ввести вектор  $\mathbf{b}$  новых средних значений

$$\mathbf{b} = \mathbf{E}[\mathbf{y}_k] = \mathbf{E}[\mathbf{Q}\mathbf{z}_k] = \mathbf{Q}\mathbf{E}[\mathbf{z}_k] = \mathbf{Q}\mathbf{a},$$

откуда

$$\mathbf{a} = \mathbf{Q}^{-1}\mathbf{b}, \quad (1.1.30)$$

то при подстановке (1.1.29) в (1.1.28) имеем

$$\sum_{k=1}^M (\mathbf{Q}^{-1}\mathbf{y}_k)^T \mathbf{K}^{-1}\mathbf{Q}^{-1}(\mathbf{b}_1 - \mathbf{b}_2) > \ln \Lambda_0 + \frac{M}{2} [\mathbf{Q}^{-1}(\mathbf{b}_1 + \mathbf{b}_2)]^T \mathbf{K}^{-1}\mathbf{Q}^{-1}(\mathbf{b}_1 - \mathbf{b}_2).$$

Воспользуемся свойствами матричного умножения, транспонирования и обращения, справедливыми для произвольных матриц  $\mathbf{A}$  и  $\mathbf{B}$ :

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T; (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}; (\mathbf{A}^T)^T = \mathbf{A}.$$

Тогда

$$\sum_{k=1}^M \mathbf{y}_k^T (\mathbf{Q}^{-1})^T \mathbf{K}^{-1}\mathbf{Q}^{-1}(\mathbf{b}_1 - \mathbf{b}_2) > \ln \Lambda_0 + \frac{M}{2} (\mathbf{b}_1 + \mathbf{b}_2)^T (\mathbf{Q}^{-1})^T \mathbf{K}^{-1}\mathbf{Q}^{-1}(\mathbf{b}_1 - \mathbf{b}_2). \quad (1.1.31)$$

Найдём матрицу  $\mathbf{Q}$ , такую, что матричное произведение  $(\mathbf{Q}^{-1})^T \mathbf{K}^{-1}\mathbf{Q}^{-1}$  давало бы в результате диагональную матрицу  $\mathbf{D}$ , у которой ненулевыми являются лишь элементы, стоящие на главной диагонали:

$$(\mathbf{Q}^{-1})^T \mathbf{K}^{-1}\mathbf{Q}^{-1} = \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}. \quad (1.1.32)$$

При выполнении (1.1.32) все матричные в (1.1.31) операции вырождаются в скалярное произведение, и правило решения принимает значительно более простой вид:

$$\sum_{k=1}^M \mathbf{y}_k^T \mathbf{D}(\mathbf{b}_1 - \mathbf{b}_2) > \ln \Lambda_0 + \frac{M}{2} (\mathbf{b}_1 + \mathbf{b}_2)^T \mathbf{D}(\mathbf{b}_1 - \mathbf{b}_2) \quad (1.1.33)$$

или

$$\sum_{k=1}^M \sum_{i=1}^n y_k^{(i)} \lambda_i (b_1^{(i)} - b_2^{(i)}) > \ln \Lambda_0 + \frac{M}{2} \sum_{i=1}^n \lambda_i (b_1^{(i)2} - b_2^{(i)2}). \quad (1.1.33a)$$

Соотношение (1.1.32) представляет собой условие существования ортогональной матрицы [3]. При этом произведение  $\mathbf{Q}\mathbf{Q}^T$  есть единичная матрица  $\mathbf{I}$   $n$ -го порядка (откуда следует, что выполняется свойство  $\mathbf{Q}^{-1} = \mathbf{Q}^T$ ), а для определителей матриц  $\mathbf{Q}$  и  $\mathbf{K}$  справедливы соотношения:

$$|\det \mathbf{Q}| = 1, \quad \det \mathbf{K} = \lambda_1 \lambda_2 \cdots \lambda_n.$$

Нахождение ортогональной матрицы сводится к решению задачи о собственных числах и собственных векторах: *найти ортонормальный набор векторов  $\mathbf{q}_1, \dots, \mathbf{q}_n$  и набор чисел  $\lambda_1, \dots, \lambda_n$ , удовлетворяющих системе уравнений*

$$q_k \mathbf{K} = \lambda_k q_k, \quad k = 1, \dots, n, \quad (1.1.34)$$

и тогда матрица  $\mathbf{Q}$  — это матрица, строками которой являются собственные векторы матрицы  $\mathbf{K}$ .

Легко видеть, что числа  $\lambda_1, \dots, \lambda_n$  (их обычно называют *спектром* матрицы  $\mathbf{K}$ ) являются решением *характеристического уравнения*

$$\det(\mathbf{K} - \lambda \mathbf{I}) = 0, \quad (1.1.35)$$

после чего собственные векторы находятся из системы

$$(\mathbf{K} - \lambda_k \mathbf{I}) \mathbf{q}^T = 0, \quad k = 1, \dots, n. \quad (1.1.36)$$

При этом, поскольку определитель такой системы равен нулю (её ранг равен  $n - 1$ ), находятся не сами компоненты вектора  $\mathbf{q}$ , а связь между ними: какие-либо  $n - 1$  компонент вычисляются через  $n$ -й компонент, и для их однозначного нахождения необходимо использовать условие единичной нормы:

$$\|\mathbf{q}\| = \sqrt{\mathbf{q}\mathbf{q}^T} = \sqrt{\sum_{k=1}^n q_k^2} = 1. \quad (1.1.37)$$

Например, пусть  $\mathbf{z}_k = (z_k^{(1)}, z_k^{(2)})$  — двумерный нормальный вектор, задаваемый корреляционной матрицей

$$\mathbf{K} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad (1.1.38)$$

определитель которой равен

$$\det \mathbf{K} = \sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2),$$

так что обратная матрица  $\mathbf{K}^{-1}$  имеет следующий вид:

$$\mathbf{K}^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{(1 - \rho^2)} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Таким образом, выражение для плотности вероятности двумерного нормального вектора  $\mathbf{z}_k$  с математическими ожиданиями  $a^{(1)}$ ,  $a^{(2)}$  и дисперсиями  $\sigma_1^2$ ,  $\sigma_2^2$  в явном виде есть

$$w^{(2)}(z_k^{(1)}, z_k^{(2)}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \times \left[ \frac{(z_k^{(1)} - a^{(1)})^2}{\sigma_1^2} - \frac{2\rho(z_k^{(1)} - a^{(1)})(z_k^{(2)} - a^{(2)})}{\sigma_1\sigma_2} + \frac{(z_k^{(2)} - a^{(2)})^2}{\sigma_2^2} \right] \right\}.$$

Решение характеристического уравнения (1.1.35) для корреляционной матрицы (1.1.38) даёт следующие значения собственных чисел:

$$\lambda_{1,2} = \frac{1}{2} \left( \sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)} \right), \quad (1.1.39)$$

для которых системы уравнений (1.1.36), определяющие собственные векторы  $q_1$  и  $q_2$ , имеют вид

$$\frac{1}{2} \left( \sigma_1^2 - \sigma_2^2 \mp \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)} \right) q_1 + \rho\sigma_1\sigma_2 q_2 = 0,$$

$$\rho\sigma_1\sigma_2 q_1 + \frac{1}{2} \left( \sigma_1^2 - \sigma_2^2 \mp \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)} \right) q_2 = 0,$$

откуда следует связь между  $q_1$  и  $q_2$ :

$$q_2 = \frac{q_1}{2\rho\sigma_1\sigma_2} \left( \sigma_2^2 - \sigma_1^2 \mp \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)} \right).$$

Используя условие (1.1.37) единичной нормы, получаем, что

$$q_1 = \frac{\sqrt{2\rho\sigma_1\sigma_2}}{\sqrt{2\rho^2\sigma_1^2\sigma_2^2 + \sigma_1^4 + \sigma_2^4 - 2\sigma_1^2\sigma_2^2(1-\rho^2) \pm (\sigma_2^2 - \sigma_1^2)\sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)}}}, \quad (1.1.40a)$$

$$q_2 = \frac{\left( \sigma_2^2 - \sigma_1^2 + \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)} \right) / \sqrt{2}}{\sqrt{2\rho^2\sigma_1^2\sigma_2^2 + \sigma_1^4 + \sigma_2^4 - 2\sigma_1^2\sigma_2^2(1-\rho^2) \pm (\sigma_2^2 - \sigma_1^2)\sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_1^2\sigma_2^2(1-\rho^2)}}}. \quad (1.1.40б)$$

В случае  $\sigma_1 = \sigma_2 = \sigma$  формулы (1.1.39) и (1.1.40) значительно упрощаются:

$$\lambda_{1,2} = \sigma^2(1 \pm \rho), \quad (1.1.41)$$

$$\mathbf{q}_1 = \left( 1/\sqrt{2} \quad 1/\sqrt{2} \right), \quad \mathbf{q}_2 = \left( 1/\sqrt{2} \quad -1/\sqrt{2} \right), \quad (1.1.42)$$

и матрица  $\mathbf{Q}$  ортогонального преобразования имеет следующий вид:

$$\mathbf{Q} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}.$$

## 1.2. ПОСЛЕДОВАТЕЛЬНАЯ ПРОЦЕДУРА ПРОВЕРКИ ГИПОТЕЗ

Отличительной особенностью рассмотренной в предыдущем разделе процедуры принятия решения является то, что заранее, до начала наблюдений задаётся размер выборки  $M$ , а сама проверка проводится “за один шаг”, без каких-либо изменений по итогам возможных промежуточных результатов. В 1947 г. А. Вальд (A. Wald) предложил последовательную процедуру проверки гипотез, при которой отказываются от постоянного размера выборки, выбирая значение  $M$  в процессе эксперимента на основании результатов уже совершенных наблюдений. При этом вся процедура становится, вообще говоря, многошаговой.

При последовательной процедуре проверки гипотез необходимо опре-

делить два правила:

- правило остановки наблюдений;
- правило принятия решения после остановки наблюдений.

Первый шаг процедуры очевиден. Вначале формируется одноэлементная ( $M = 1$ ) выборка, содержащая единственное значение  $z_1$ . Далее, на основании этого значения по заранее определённому правилу производится либо остановка процедуры с выносом решения в пользу какой-либо гипотезы, либо отказ от принятия решения на данном шаге и продолжение наблюдений на следующем шаге — до тех пор, пока не будет принято решение в пользу какой-нибудь гипотезы.

В отличие от процедуры проверки гипотез с фиксированным размером выборки всё выборочное пространство  $S$  делится не на два, а на три подмножества: области  $S_1$  и  $S_2$ , соответствующие выносу решений в пользу гипотез  $H_1$  и  $H_2$ , а также промежуточная область  $S_{\text{пром}}$ , попадание в которую является указанием на продолжение наблюдения. При этом можно показать, что вероятность того, что число шагов в последовательной процедуре проверки гипотез не ограничено, равна нулю (задача 1.14).

Понятно, что разбить выборочное пространство на три области можно различными способами (при изначально неопределённом значении  $M$  число способов неограниченно), следовательно, необходим критерий оптимальности, по которому можно было бы сравнивать различные правила разбиений, выбирая из них наилучшее.

В предложенной Вальдом последовательной процедуре критерием оптимальности являлся *критерий минимальной средней стоимости эксперимента*. Логично считать, что стоимость эксперимента пропорциональна размеру выборки, поэтому критерий минимальной средней стоимости при фиксированных уровне значимости  $\alpha$  и мощности  $1 - \beta$  эквивалентен критерию минимума размера выборки, а точнее, *минимума среднего значения размера выборки*, ибо момент остановки процесса наблюдения — величина случайная, зависящая от предшествующих результатов наблюдений. Более того, поскольку средние значения размера выборки  $E[M/H_1]$  и  $E[M/H_2]$  являются условными — в зависимости от справедливости той или иной гипотезы и могут оказаться различными, требуется минимизация обоих средних.



Согласно Вальду, при независимых наблюдениях среди всех возможных процедур проверки гипотез, как последовательных, так и с фиксированным объёмом выборки, для которых вероятности ошибок первого и второго родов не превосходят значений  $\alpha$  и  $\beta$  соответственно, к наименьшим значениям  $E[M/H_1]$  и  $E[M/H_2]$  приводит последовательная процедура, состоящая в сравнении отношения правдоподобия с двумя пороговыми значениями  $\Lambda_{01}$  и  $\Lambda_{02}$ . При этом оптимальное разбиение выборочного пространства осуществляется следующим образом:

- область  $S_1$ , соответствующая гипотезе  $H_1$ , содержит наблюдаемые значения  $z_1, \dots, z_M$ , для которых

$$\begin{aligned} \Lambda_{01} < \Lambda(z_1, \dots, z_k) < \Lambda_{02}, \quad k = 1, \dots, M-1, \\ \Lambda(z_1, \dots, z_M) \geq \Lambda_{02}; \end{aligned} \quad (1.2.1)$$

- область  $S_2$ , соответствующая гипотезе  $H_2$ , содержит наблюдаемые значения  $z_1, \dots, z_M$ , для которых

$$\begin{aligned} \Lambda_{01} < \Lambda(z_1, \dots, z_k) < \Lambda_{02}, \quad k = 1, \dots, M-1, \\ \Lambda(z_1, \dots, z_M) \leq \Lambda_{01}; \end{aligned} \quad (1.2.2)$$

- промежуточная область  $S_{\text{пром}}$  содержит наблюдаемые значения  $z_1, \dots, z_M$ , для которых

$$\Lambda_{01} < \Lambda(z_1, \dots, z_k) < \Lambda_{02}, \quad k = 1, \dots, M. \quad (1.2.3)$$

Представленное разбиение выборочного пространства, как уже было сказано, предполагает последовательную процедуру принятия решения. Если для некоторого  $k$  отношение правдоподобия  $\Lambda(z_1, \dots, z_k)$  оказывается большим или равным  $\Lambda_{02}$ , то процедура завершается принятием решения в пользу  $H_1$ ; если  $\Lambda(z_1, \dots, z_k)$  оказывается меньшим или равным  $\Lambda_{01}$ , то процедура завершается принятием решения в пользу  $H_2$ ; наконец, если оказывается, что  $\Lambda_{01} < \Lambda(z_1, \dots, z_k) < \Lambda_{02}$ , то к набору наблюдаемых данных добавляется очередное наблюдение  $z_{k+1}$ .

Практический смысл оптимальности последовательной процедуры состоит в том, что при её выполнении в среднем требуется заметно меньше наблюдений, чем для процедуры с фиксированным объёмом выборки и такими же вероятностями ошибок.

Точное вычисление пороговых значений  $\Lambda_{01}$  и  $\Lambda_{02}$  связано с большими математическими сложностями (они, ко всему прочему, оказываются зави-

сящими от номера шага), но для них можно получить следующие оценки:

$$\Lambda_{01} \geq \frac{\alpha}{1-\beta}, \quad \Lambda_{02} \leq \frac{1-\alpha}{\beta}. \quad (1.2.4)$$

Действительно, обозначим через  $R_{1M}$  и  $R_{2M}$  множества точек  $(z_1, \dots, z_M)$ , для которых процедура проверки заканчивается на  $M$ -м шаге вынесением решения в пользу  $H_1$  или  $H_2$  соответственно:

$$\begin{aligned} R_{1M} &= \left\{ (z_1, \dots, z_M) : \Lambda_{01} < \Lambda(z_1, \dots, z_k) < \Lambda_{02}, k = 1, \dots, M-1, \right. \\ &\quad \left. \Lambda(z_1, \dots, z_M) \geq \Lambda_{02} \right\}; \\ R_{2M} &= \left\{ (z_1, \dots, z_M) : \Lambda_{01} < \Lambda(z_1, \dots, z_k) < \Lambda_{02}, k = 1, \dots, M-1, \right. \\ &\quad \left. \Lambda(z_1, \dots, z_M) \leq \Lambda_{01} \right\}. \end{aligned}$$

Пусть  $P_H(M = \mu)$  — вероятность того, что принятие решения в пользу какой-либо из гипотез заканчивается при значении  $M = \mu$ , и в силу конечности числа шагов последовательной процедуры (задача 1.14)

$$\sum_{\mu=1}^{\infty} P_H(M = \mu) = 1,$$

При этом, поскольку окончательное (без промежуточной области) принятие решения основано на разбиении  $M$ -мерного выборочного пространства на множества  $R_{1M}$  и  $R_{2M}$ , можно записать

$$\begin{aligned} \sum_{\mu=1}^{\infty} P_H(M = \mu) &= \sum_{\mu=1}^{\infty} \int_{R_{1M}} w((z_1, \dots, z_M / H_1)) d(z_1, \dots, z_M) + \\ &+ \sum_{\mu=1}^{\infty} \int_{R_{2M}} w((z_1, \dots, z_M / H_2)) d(z_1, \dots, z_M) = 1. \end{aligned}$$

Далее, учитывая, что на точках множества  $R_{1M}$  выполняется неравенство

$$\Lambda(z_1, \dots, z_M) = \frac{w((z_1, \dots, z_M / H_1))}{w((z_1, \dots, z_M / H_2))} \geq \Lambda_{02},$$

а на точках множества  $R_{2M}$  — неравенство

$$\Lambda(z_1, \dots, z_M) = \frac{w((z_1, \dots, z_M / H_1))}{w((z_1, \dots, z_M / H_2))} \leq \Lambda_{01},$$

для вероятностей ошибок  $\alpha$  и  $\beta$  имеем:

$$\begin{aligned}
\alpha &= P(H_2 / H_1) = \sum_{M=1}^{\infty} \int_{R_{2M}} w((z_1, \dots, z_M / H_1)) d(z_1, \dots, z_M) \leq \\
&\leq \Lambda_{01} \sum_{M=1}^{\infty} \int_{R_{2M}} w((z_1, \dots, z_M / H_2)) d(z_1, \dots, z_M) = \Lambda_{01} P(H_2 / H_2) = \\
&= \Lambda_{01} (1 - P(H_1 / H_2)) = \Lambda_{01} (1 - \beta); \\
\beta &= P(H_1 / H_2) = \sum_{M=1}^{\infty} \int_{R_{1M}} w((z_1, \dots, z_M / H_2)) d(z_1, \dots, z_M) \leq \\
&\leq \frac{1}{\Lambda_{02}} \sum_{M=1}^{\infty} \int_{R_{1M}} w((z_1, \dots, z_M / H_1)) d(z_1, \dots, z_M) = \frac{1}{\Lambda_{02}} P(H_1 / H_1) = \\
&= \frac{1}{\Lambda_{02}} (1 - P(H_2 / H_1)) = \frac{1 - \alpha}{\Lambda_{02}}.
\end{aligned}$$

Отсюда получаем искомые неравенства (1.2.4).

Наличие оценок (1.2.4) порождает естественное желание приближённо заменить истинные значения  $\Lambda_{01}$  и  $\Lambda_{02}$  их граничными значениями:

$$\Lambda_{01} \approx \Lambda'_{01} = \frac{\alpha}{1 - \beta}, \quad \Lambda_{02} \approx \Lambda'_{02} = \frac{1 - \alpha}{\beta}.$$

При этом, вообще говоря, возникает другая последовательная процедура, отличная от оптимальной, в которой число шагов, необходимое для принятия решения, превышает число шагов в оптимальной процедуре, хотя, как можно надеяться, незначительно. Кроме того, изменяются и значения вероятностей ошибок: вместо  $\alpha$  и  $\beta$  они становятся равными  $\alpha'$  и  $\beta'$ . На основании (1.2.4) можно утверждать, что пороговые значения и вероятности ошибок в такой процедуре связаны соотношениями

$$\Lambda'_{01} \geq \frac{\alpha'}{1 - \beta'}, \quad \Lambda'_{02} \leq \frac{1 - \alpha'}{\beta'}.$$

Тогда, поскольку  $\Lambda_{01} \geq \Lambda'_{01}$  и  $\Lambda_{02} \leq \Lambda'_{02}$ , имеем связь между вероятностями ошибок:

$$\frac{\alpha}{1 - \beta} \geq \frac{\alpha'}{1 - \beta'}, \quad \frac{1 - \alpha}{\beta} \leq \frac{1 - \alpha'}{\beta'}.$$

Отсюда получаем неравенства

$$\begin{aligned}\alpha'(1-\beta) &\leq \alpha(1-\beta'); \\ -\beta(1-\alpha') &\leq -\beta'(1-\alpha),\end{aligned}$$

Складывая которые, находим, что

$$\alpha' + \beta' \leq \alpha + \beta,$$

т. е. сумма вероятностей ошибок новой процедуры не превосходит суму вероятностей ошибок прежней процедуры. Таким образом, можно утверждать, что замена истинных порогов их приближёнными значениями практически не ухудшает вероятностные характеристики процедуры.

Единственным серьёзным опасением, связанным с заменой приближёнными значениями, является то, что число шагов в последовательной процедуре может заметно увеличиться. Однако существуют обоснованные причины надеяться на то, что такое увеличение будет весьма умеренным. Это можно увидеть, если воспользоваться логарифмическим представлением отношения правдоподобия, записав, в силу независимости элементов выборки, соотношения (1.2.1)–(1.2.3) следующим образом:

- для области  $S_1$

$$\ln \Lambda_{01} < \sum_{j=1}^k \ln \Lambda(z_j) < \ln \Lambda_{02}, \quad k = 1, \dots, M-1, \quad (1.2.5)$$

$$\sum_{j=1}^M \ln \Lambda(z_j) \geq \ln \Lambda_{02};$$

- для области  $S_2$

$$\ln \Lambda_{01} < \sum_{j=1}^k \ln \Lambda(z_j) < \ln \Lambda_{02}, \quad k = 1, \dots, M-1, \quad (1.2.6)$$

$$\sum_{j=1}^M \ln \Lambda(z_j) \leq \ln \Lambda_{01};$$

- для промежуточной области  $S_{\text{пром}}$

$$\ln \Lambda_{01} < \sum_{j=1}^k \ln \Lambda(z_j) < \ln \Lambda_{02}, \quad k = 1, \dots, M. \quad (1.2.7)$$

Например, принятие гипотезы  $H_1$  означает, что  $\ln \Lambda(z_j)$  удовлетворяют неравенствам

$$\ln \Lambda(z_1) + \dots + \ln \Lambda(z_{M-1}) < \ln \Lambda_{02} \leq \ln \Lambda(z_1) + \dots + \ln \Lambda(z_M), \quad (1.2.8)$$

и аппроксимация приближёнными значениями означает, что вместо двойного неравенства имеет место одно:

$$\ln \Lambda(z_1) + \dots + \ln \Lambda(z_{M-1}) < \ln \Lambda_{02}.$$

Это означает, что после  $M - 1$  наблюдения стоящая в левой части сумма будет всё ещё меньше порогового значения, так что увеличение числа шагов будет обусловлено единственным дополнительным наблюдением  $\ln \Lambda(z_M)$ , т. е. в худшем случае число шагов увеличится на единицу.

Определим средние значения размеров выборки  $\mathbf{E}[M/H_1]$  и  $\mathbf{E}[M/H_2]$  при использовании критерия Вальда в последовательной процедуре принятия решения, считая, что выборка однородная<sup>1</sup> и, следовательно,  $\ln \Lambda(\mathbf{z})$  представляет собой сумму независимых одинаково распределённых случайных величин с той особенностью, что количество слагаемых (число элементов в выборке) — случайная величина, не зависящая от элементов выборки.

Рассмотрим произвольную сумму  $\delta_v$  случайного числа случайных величин  $\xi_k$ :

$$\delta_v = \sum_{k=1}^v \xi_k,$$

где  $v$  — не зависящая от величин  $\xi_k$  дискретная случайная величина, принимающая значения  $1, 2, \dots$  с вероятностями  $p_r = P\{v = r\}$  ( $r = 1, 2, \dots$ ).

Интегральная функция распределения  $\delta_v$  может быть определена по формуле полной вероятности:

$$F_{\delta_v}(x) \equiv P\{\delta_v < x\} = \sum_{r=1}^{\infty} P\{\delta_r < x\} p_r = \sum_{r=1}^{\infty} F_{\delta_r}(x) p_r$$

или, в предположении о возможности почленного дифференцирования,

$$w_{\delta_v}(x) = \sum_{r=1}^{\infty} w_{\delta_r}(x) p_r,$$

где  $w_{\delta_v}(x)$  и  $w_{\delta_r}(x)$  — соответствующие плотности вероятности случайных величин  $\delta_v$  и  $\delta_r$ .

---

<sup>1</sup> Напомним, что выборка называется однородной, если все её элементы независимы и являются значениями одной случайной величины.

Применяя к обеим частям полученного соотношения преобразование Фурье, получаем аналогичное соотношение между характеристическими функциями  $\Theta_{\delta_v}(u)$  и  $\Theta_{\delta_r}(u)$ :

$$\Theta_{\delta_v}(u) = \sum_{r=1}^{\infty} \Theta_{\delta_r}(u) p_r.$$

В силу независимости составляющих сумму случайных величин характеристическая функция  $\Theta_{\delta_r}(x)$  факторизуется, так что

$$\Theta_{\delta_v}(u) = \sum_{r=1}^{\infty} p_r \prod_{l=1}^r \Theta_{\xi_l}(u),$$

где  $\Theta_{\xi_l}(u)$  представляет собой характеристическую функцию одной случайной величины  $\xi_l$ .

Продифференцируем обе части полученного соотношения и затем положим  $u = 0$ . При этом учтём, что значение характеристической функции в нуле равно единице. Тогда производная произведения, стоящего в правой части, равна

$$\begin{aligned} \frac{d}{du} \prod_{l=1}^r \Theta_{\xi_l}(u) &= \frac{d\Theta_{\xi_1}(u)}{du} \prod_{l=2}^r \Theta_{\xi_l}(u) + \Theta_{\xi_1}(u) \frac{d}{du} \prod_{l=2}^r \Theta_{\xi_l}(u) = \\ &= \frac{d\Theta_{\xi_1}(u)}{du} \prod_{l=2}^r \Theta_{\xi_l}(u) + \Theta_{\xi_1}(u) \left[ \frac{d\Theta_{\xi_2}(u)}{du} \prod_{l=3}^r \Theta_{\xi_l}(u) + \Theta_{\xi_2}(u) \frac{d}{du} \prod_{l=3}^r \Theta_{\xi_l}(u) \right], \end{aligned}$$

откуда видно, что

$$\left( \frac{d}{du} \prod_{l=1}^r \Theta_{\xi_l}(u) \right)_{u=0} = \sum_{l=1}^r \Theta'_{\xi_l}(0).$$

Теперь, используя связь между производными характеристической функции и начальными моментами распределения:

$$\left( \frac{d^k \Theta(u)}{du^k} \right)_{u=0} = j^k m_k[\xi],$$

получаем выражение для среднего значения суммы случайного числа независимых случайных величин:

$$\mathbf{E}[\delta_v] = \sum_{r=1}^{\infty} p_r \sum_{l=1}^r \mathbf{E}[\xi_l]. \quad (1.2.9)$$

Если математические ожидания величин  $\xi_k$  одинаковы и равны  $x_0$ , то из (1.2.9) следует выражение

$$\mathbf{E}[\delta_v] = x_0 \sum_{r=1}^{\infty} r p_r = x_0 \mathbf{E}[v], \quad (1.2.10)$$

известное как *тождество Вальда*.

Применим полученное выражение к задаче определения среднего размера выборки при последовательном критерии проверки гипотез. Для каждой из гипотез  $H_i$  ( $i = 1, 2$ ) имеем:

$$\begin{aligned} \mathbf{E}[\ln \Lambda(z_1, \dots, z_M) / H_i] &= \mathbf{E}\left[\sum_{k=1}^M \ln \Lambda(z_k) / H_i\right] = \\ &= \mathbf{E}[M / H_i] \mathbf{E}[\ln \Lambda(z) / H_i], \end{aligned}$$

откуда

$$\mathbf{E}[M / H_i] = \frac{\mathbf{E}[\ln \Lambda(z_1, \dots, z_M) / H_i]}{\mathbf{E}[\ln \Lambda(z) / H_i]}, \quad (1.2.11)$$

где

$$\mathbf{E}[\ln \Lambda(z) / H_i] = \int_{-\infty}^{\infty} w(z / H_i) \ln \Lambda(z) dz.$$

Если пренебречь возможным пересечением пороговых значений на заключительном шаге процедуры, т. е. считать, что в момент принятия решения в пользу какой-либо из процедур логарифм отношения правдоподобия совпадает с одним из пороговых значений, то можно трактовать  $\ln \Lambda(\mathbf{z})$  как дискретную случайную величину, принимающую значения  $\ln \Lambda_{01}$  и  $\ln \Lambda_{02}$  с вероятностями  $(1 - \alpha)$  и  $\alpha$  для гипотезы  $H_1$  или с вероятностями  $\beta$  и  $(1 - \beta)$  для гипотезы  $H_2$ . Тогда, с учётом (2.8), условные математические ожидания равны

$$\mathbf{E}[\ln \Lambda(z_1, \dots, z_M) / H_1] = (1 - \alpha) \ln \frac{\alpha}{(1 - \beta)} + \alpha \ln \frac{1 - \alpha}{\beta}; \quad (1.2.12a)$$

$$\mathbf{E}[\ln \Lambda(z_1, \dots, z_M) / H_2] = \beta \ln \frac{\alpha}{1 - \beta} + (1 - \beta) \ln \frac{1 - \alpha}{\beta}. \quad (1.2.12b)$$

Теперь, подставляя эти значения в (1.2.9) и зная одномерные условные распределения  $w(z/H_i)$ , получаем искомые средние  $\mathbf{E}[M/H_1]$  и  $\mathbf{E}[M/H_2]$ :

$$\mathbf{E}[M / H_1] = \frac{(1 - \alpha) \ln \frac{\alpha}{(1 - \beta)} + \alpha \ln \frac{1 - \alpha}{\beta}}{\int_{-\infty}^{\infty} \ln(\Lambda(z)) w(z / H_1) dz}; \quad (1.2.13a).$$

$$\mathbf{E}[M / H_2] = \frac{\beta \ln \frac{\alpha}{1 - \beta} + (1 - \beta) \ln \frac{1 - \alpha}{\beta}}{\int_{-\infty}^{\infty} \ln(\Lambda(z)) w(z / H_2) dz}. \quad (1.2.13б)$$

Итак, полученные средние значения объёма выборки оптимальной последовательной процедуры являются минимально возможными среди всех других процедур принятия решений при заданных вероятностях ошибок  $\alpha$  и  $\beta$ . Однако минимальность средних значений отнюдь не гарантирует того, что в каком-либо конкретном эксперименте процедура не окажется весьма длинной, превзойдя по размеру выборки  $M$  аналогичную — с теми же значениями  $\alpha$  и  $\beta$  — процедуру принятия решения “в целом”, т. е. по всем наблюдениям сразу.

Логичным усовершенствованием последовательных процедур проверки гипотез является привлечение механизма усечения, когда устанавливается максимально возможное значение размера выборки  $M_{\max}$ . До тех пор, пока  $M < M_{\max}$ , процедура проводится также, как и в неусечённом варианте, т. е. отношение правдоподобия сравнивается с двумя пороговыми значениями. Если же наступил момент, когда  $M = M_{\max}$ , а решение так и не было принято, то производится переход к однопороговой процедуре принятия решения “в целом”. Понятно, что чем меньше  $M_{\max}$ , тем меньшим (в среднем) будет возможный выигрыш во времени, получаемый от последовательной процедуры.

Проиллюстрируем применение последовательного критерия на примере рассмотренной в разд. 1.1 задачи проверки гипотезы о среднем значении нормальной случайной величины, для которой отношение правдоподобия, согласно (1.1.13), имеет следующий вид:

$$\ln \Lambda(\mathbf{z}) = -\frac{(a_1^2 - a_2^2)M}{2\sigma^2} + \frac{a_1 - a_2}{\sigma^2} \sum_{k=1}^M z_k, \quad a_1 > a_2.$$



Тогда на  $M$ -м шаге принимается гипотеза  $H_1$  о том, что математическое ожидание нормальной случайной величины равно  $a_1$ , если

$$\sum_{k=1}^M z_k \geq \frac{a_1 + a_2}{2} M + \frac{\sigma^2}{a_1 - a_2} \ln \frac{1 - \alpha}{\beta}; \quad (1.2.14)$$

принимается гипотеза  $H_2$  о том, что математическое ожидание нормальной случайной величины равно  $a_2$ , если

$$\sum_{k=1}^M z_k \leq \frac{a_1 + a_2}{2} M + \frac{\sigma^2}{a_1 - a_2} \ln \frac{\alpha}{1 - \beta}; \quad (1.2.15)$$

наконец, наблюдение продолжается, если не выполняется ни одно из неравенств (1.2.14) и (1.2.15).

Чтобы определить средние размеры выборки необходимо в соответствии с (1.2.13) определить средние значения логарифма отношения правдоподобия  $\ln \Lambda(z)$  для двух возможных средних  $a_1$  и  $a_2$ . Полагая в (1.1.13)  $M = 1$ , имеем:

$$\ln \Lambda(z) = \frac{a_1 - a_2}{\sigma^2} z - \frac{a_1^2 - a_2^2}{2\sigma^2},$$

где  $z$  — нормальная случайная величина с дисперсией  $\sigma^2$  и средним  $a_1$  или  $a_2$ . Тогда

$$\mathbf{E}[\ln \Lambda(z) / H_{1,2}] = -\frac{a_1^2 - a_2^2}{2\sigma^2} + \frac{a_1 - a_2}{\sigma^2} \mathbf{E}[z / H_{1,2}] = -\frac{a_1^2 - a_2^2}{2\sigma^2} + \frac{a_1 - a_2}{\sigma^2} a_{1,2},$$

т. е.

$$\mathbf{E}[\ln \Lambda(z) / H_1] = \frac{(a_1 - a_2)^2}{2\sigma^2}; \quad (1.2.16, a)$$

$$\mathbf{E}[\ln \Lambda(z) / H_2] = -\frac{(a_1 - a_2)^2}{2\sigma^2}. \quad (1.2.16, б)$$

Теперь, подставляя (1.2.16) в (1.2.13), получаем

$$\mathbf{E}[M / H_1] = \frac{2\sigma^2}{(a_1 - a_2)^2} \left[ (1 - \alpha) \ln \frac{\alpha}{1 - \beta} + \alpha \ln \frac{1 - \alpha}{\beta} \right]; \quad (1.2.15, a)$$

$$\mathbf{E}[M / H_2] = -\frac{2\sigma^2}{(a_1 - a_2)^2} \left[ \beta \ln \frac{\alpha}{1 - \beta} + (1 - \beta) \ln \frac{1 - \alpha}{\beta} \right]. \quad (1.2.15, б)$$

В частности, когда ошибки первого и второго рода одинаковы,

$$\alpha = \beta = \varepsilon < 1/2,$$

имеем

$$\bar{M} = \mathbf{E}[M / H_1] = \mathbf{E}[M / H_2] = \frac{2\sigma^2}{(a_1 - a_2)^2} (1 - 2\varepsilon) \ln \frac{1 - \varepsilon}{\varepsilon}. \quad (1.2.16)$$

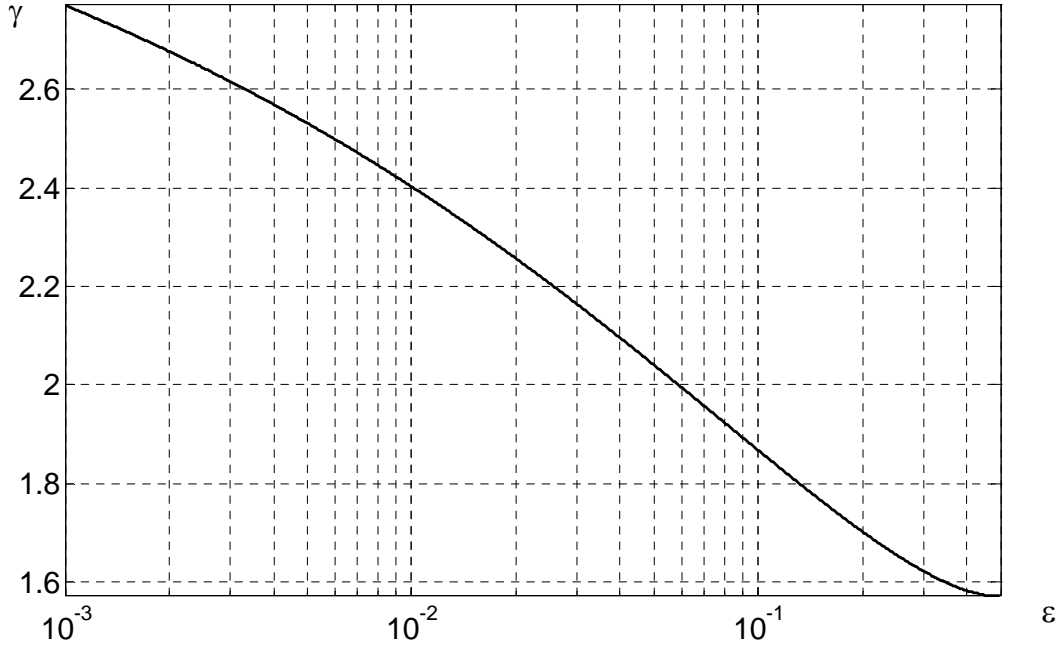


Рис. 1.10. Зависимость коэффициента выигрыша  $\gamma$  от вероятности ошибки

Сравним значения среднего размера выборки при последовательном анализе с аналогичными значениями для байесовской стратегии с фиксированным объёмом выборки, определяемым при равенстве ошибки первого и второго рода соотношением (1.1.15). Пусть  $x_\varepsilon$  — корень уравнения

$$1 - F_0(x) = \varepsilon,$$

где, как обычно,  $F_0$  — функция Лапласа, т. е. интегральная функция стандартного нормального распределения. Тогда соответствующее значение объёма выборки для байесовской стратегии равно<sup>1</sup>

$$M_\varepsilon = \frac{4\sigma^2}{(a_1 - a_2)^2} x_\varepsilon^2,$$

<sup>1</sup> Строго говоря, необходимо брать целую часть этого выражения. Однако, поскольку средние значения объёма выборки также представлены дробными значениями, это не сильно влияет на коэффициент выигрыша.

и можно ввести коэффициент выигрыша  $\gamma$  при использовании последовательного анализа:

$$\gamma(\varepsilon) = \frac{M_\varepsilon}{\bar{M}} = \frac{2}{1 - 2\varepsilon \ln[(1 - \varepsilon) / \varepsilon]} \cdot x_\varepsilon^2. \quad (1.2.17)$$

На рис. 1.10 представлена (в логарифмическом масштабе по оси абсцисс) зависимость  $\gamma(\varepsilon)$  коэффициента выигрыша при использовании последовательного анализа от значения одинаковых вероятностей ошибок первого и второго рода. Видно, что при малых (порядков  $10^{-3} \dots 10^{-2}$ ) значениях  $\varepsilon$  коэффициент выигрыша составляет  $2 \dots 3$ ; при бóльших (порядка  $10^{-1}$ ) значениях вероятностей ошибок эффективность применения последовательного анализа заметно снижается, и коэффициент выигрыша уменьшается до  $1,6$ .

### 1.3. СЛОЖНЫЕ ГИПОТЕЗЫ

В предыдущих разделах данной главы рассматривались вопросы, связанные с процедурами проверки простой основной гипотезы против простой альтернативной гипотезы, когда каждая из двух гипотез относится только к одному состоянию источника. Однако на практике часто приходится иметь дело с более сложными задачами, например, когда основная и/или альтернативная гипотезы имеют ряд параметров. Тогда возникает задача выбора между гипотезами  $H_1$  и  $H_2$  в условиях, когда они, по сути, представляют собой дискретные или даже континуальные множества гипотез. Более того, часто встречаются ситуации, при которых количество  $m$  конкурирующих гипотез больше двух — такие задачи называются многоальтернативными (см. разд. 4). Понятно, что в подобных случаях задачи оптимального разбиения при заданном критерии оптимальности могут оказаться весьма сложными.

В данном разделе будут кратко рассмотрены некоторые методы решения задач проверки сложных статистических гипотез.

Рассмотрим простейший случай — проверку двух сложных гипотез при наличии одного вещественного параметра  $\eta$  с известным распределением  $w_\eta(y)$ .

Априорная информация о любом параметре в конечном итоге содержит указание на то, что он принадлежит некоторому множеству возмож-

ных значений. Следовательно, множество  $T$  всех возможных состояний источника можно отобразить на два подмножества  $Y_1$  и  $Y_2$  значений параметра  $\eta$  так, что к гипотезе  $H_1$  относятся значения  $\eta \in Y_1$ , а к гипотезе  $H_2$  — значения  $\eta \in Y_2$ . Подмножества  $Y_1$  и  $Y_2$  могут не пересекаться, а могут иметь общие элементы — вплоть до полного совпадения.

Пусть, к примеру, по гипотезе  $H_1$  наблюдаемая величина является нормальной случайной величиной с нулевым средним и произвольной дисперсией  $\sigma^2$ , а по гипотезе  $H_2$  — нормальной случайной величиной с произвольной дисперсией  $\sigma^2$  и средним  $a$ , лежащим в заданном интервале  $[a_1; a_2]$ . Тогда гипотеза  $H_1$  — это точка  $a = 0$ , а гипотеза  $H_2$  — указанный промежуток, и распределение  $w_\eta(y)$  отображает множество значений параметров в выборочное пространство наблюдений (рис. 1.11). В такой постановке задача является *задачей проверки простой гипотезы против сложной альтернативы*.

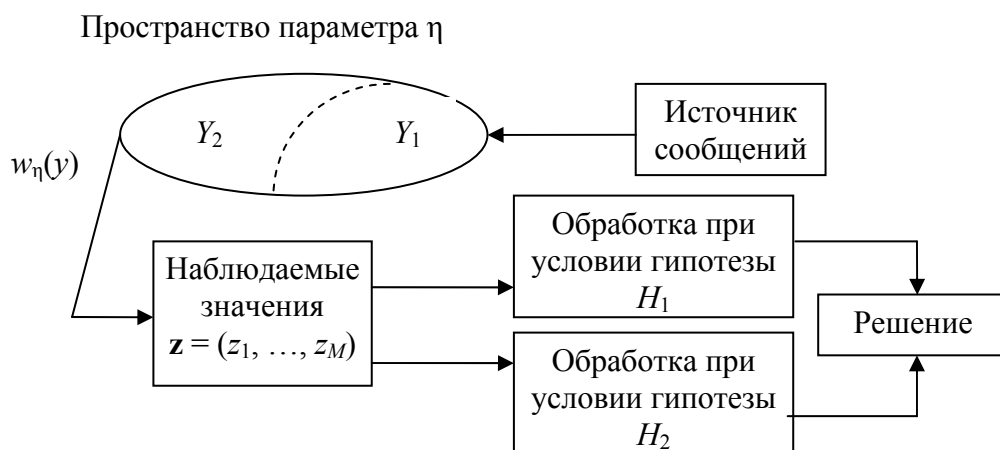


Рис. 1.11. Схема принятия решения в задаче проверки сложных гипотез

Далее, как и в случае простых гипотез, правило принятия решения делит пространство наблюдений на два непересекающихся подмножества  $S_1$  и  $S_2$ . Заметим, что в данной задаче интерес представляет только решение о выборе той или иной гипотезы, но не сами значения параметра  $a$ . По этой причине подобные параметры часто называют *несущественными* или *нежелательными*, поскольку их наличие приводит, вообще говоря, к ухудшению характеристик устройств приёма.

Идея получения алгоритма принятия решения и, следовательно, по-

строения решающего устройства принципиально не отличается от случая решения задачи проверки простых гипотез: по-прежнему, необходимо, задавшись каким-либо критерием оптимальности, обеспечить минимизацию (максимизацию) целевой функции посредством выбора конфигураций подмножеств  $S_1$  и  $S_2$ . Разница состоит лишь в том, что конструируемая целевая функция зависит от неизвестного случайного параметра, т. е. оказывается случайной величиной при каждом наблюдении за состояниями источника. Следовательно, для получения окончательного выражения необходимо провести усреднение целевой функции по этому параметру.

Будем при выборе двух конкурирующих гипотез  $H_1$  и  $H_2$  по-прежнему использовать общую байесовскую стратегию, основанную на минимизации среднего риска при заданных ценовой матрице (1.3) и априорных вероятностях гипотез  $p(H_1)$  и  $p(H_2)$ , где условные риски  $R_1$  и  $R_2$  зависят от параметра  $\eta$ :

$$R_1(\eta) = c_{11}P(H_1 / H_1, \eta) + c_{12}P(H_1 / H_2, \eta);$$

$$R_2(\eta) = c_{21}P(H_2 / H_1, \eta) + c_{22}P(H_2 / H_2, \eta).$$

Тогда средний риск принятия решений получается путём усреднения не только по гипотезам, но и по значениям параметра:

$$\begin{aligned} \bar{R} &= c_{11}p(H_1) \int_{Y_1} P(H_1 / H_1, \eta) w_\eta(y) dy + \\ &+ c_{21}p(H_1) \int_{Y_1} P(H_2 / H_1, \eta) w_\eta(y) dy + \\ &+ c_{12}p(H_2) \int_{Y_2} P(H_1 / H_2, \eta) w_\eta(y) dy + \\ &+ c_{22}p(H_2) \int_{Y_2} P(H_2 / H_2, \eta) w_\eta(y) dy = \\ &= c_{11}p(H_1) \int_{Y_1} \int_{S_1} w(\mathbf{z} / H_1, \eta) w_\eta(y) dz dy + \\ &+ c_{21}p(H_1) \int_{Y_1} \int_{S_2} w(\mathbf{z} / H_1, \eta) w_\eta(y) dz dy + \\ &+ c_{12}p(H_2) \int_{Y_2} \int_{S_1} w(\mathbf{z} / H_2, \eta) w_\eta(y) dz dy + \end{aligned}$$

$$+c_{22}p(H_2) \int_{Y_2} \int_{S_2} w(\mathbf{z} / H_2, \eta) w_\eta(y) dz dy.$$

Далее проведём действия, аналогичные тем, что проводились при получении правила решения для простых гипотез: приведём выражение к виду, содержащему интегральную (переменную) и внеинтегральную (постоянную) составляющие среднего риска; после этого постараемся минимизировать переменную составляющую посредством поиска соотношения между апостериорными плотностями, соответствующими различным гипотезам.

По-прежнему используя тот факт, что

$$\int_{S_1} w(\mathbf{z} / H_i, \eta) dz + \int_{S_2} w(\mathbf{z} / H_i, \eta) dz = 1, \quad i = 1, 2,$$

имеем:

$$\begin{aligned} \bar{R} = & c_{11}p(H_1) \int_{Y_1} \int_{S_1} w(\mathbf{z} / H_1, \eta) w_\eta(y) dz dy + \\ & + c_{21}p(H_1) \int_{Y_1} \left[ 1 - \int_{S_1} w(\mathbf{z} / H_1, \eta) dz \right] w_\eta(y) dy + \\ & + c_{12}p(H_2) \int_{Y_2} \int_{S_1} w(\mathbf{z} / H_2, \eta) w_\eta(y) dz dy + \\ & + c_{22}p(H_2) \int_{Y_2} \left[ 1 - \int_{S_1} w(\mathbf{z} / H_2, \eta) dz \right] w_\eta(y) dy. \end{aligned}$$

Изменим порядок интегрирования и введём усреднённые условные плотности вероятности

$$\begin{aligned} \bar{w}(\mathbf{z} / H_1) &= \int_{Y_1} w(\mathbf{z} / H_1, \eta) w_\eta(y) dy; \\ \bar{w}(\mathbf{z} / H_2) &= \int_{Y_2} w(\mathbf{z} / H_2, \eta) w_\eta(y) dy. \end{aligned}$$

Тогда получим выражение

$$\begin{aligned} \bar{R} = & \int_{S_1} \left[ (c_{12} - c_{22}) \bar{w}(\mathbf{z} / H_2) p(H_2) - (c_{21} - c_{11}) \bar{w}(\mathbf{z} / H_1) p(H_1) \right] dz + \\ & + c_{21}p(H_1) \int_{Y_1} w_\eta(y) dy + c_{22}p(H_2) \int_{Y_2} w_\eta(y) dy, \end{aligned} \quad (1.3.1)$$

в точности совпадающее с (1.1.7), но в котором вместо условных плотностей  $w(\mathbf{z}/H_i)$  фигурируют усреднённые условные плотности  $\bar{w}(\mathbf{z}/H_i)$ .

Далее констатируем тот факт, что входящий в (1.3.1) интеграл представляет собой переменную составляющую стоимости, и в целях её уменьшения в область  $S_1$  следует включить те точки  $\mathbf{z}$ , для которых справедливо неравенство

$$(c_{21} - c_{11})\bar{w}(\mathbf{z}/H_1)p(H_1) > (c_{12} - c_{22})\bar{w}(\mathbf{z}/H_2)p(H_2).$$

Итак, правило принятия решения в случае сложных гипотез сводится к сравнению отношения усреднённых функций правдоподобия

$$\bar{\Lambda}(\mathbf{z}) = \frac{\bar{w}(\mathbf{z}/H_1)}{\bar{w}(\mathbf{z}/H_2)} = \frac{\int_{Y_1} w(\mathbf{z}/H_1, \eta) w_\eta(y) dy}{\int_{Y_2} w(\mathbf{z}/H_2, \eta) w_\eta(y) dy} \quad (1.3.2)$$

с пороговым значением

$$\Lambda_0 = \frac{(c_{12} - c_{22})p(H_2)}{(c_{21} - c_{11})p(H_1)}.$$

Если  $\bar{\Lambda}(\mathbf{z}) > \Lambda_0$ , то принимается решение в пользу гипотезы  $H_1$ ; если  $\bar{\Lambda}(\mathbf{z}) < \Lambda_0$ , то — в пользу  $H_2$ . Граничный случай  $\bar{\Lambda}(\mathbf{z}) = \Lambda_0$ , который не влияет на средний риск, при необходимости будем относить в пользу гипотезы  $H_2$ .

Как и в случае проверки простых гипотез, зачастую целесообразен переход к эквивалентным неравенствам, основанным на монотонности логарифмической функции: сравнивается  $\ln \bar{\Lambda}(\mathbf{z})$  с пороговым значением  $\ln \Lambda_0$ .

Если в выражении (1.3.2) положить, что распределения значений параметра  $\eta$  представляют собой дельта-функции:

$$w_\eta(y) = \begin{cases} \delta(y - y_1), & y \in Y_1; \\ \delta(y - y_2), & y \in Y_2, \end{cases}$$

то задача сводится к задаче проверки простых гипотез.

Если лишь одна из плотностей равна дельта функции, то возникает задача проверки простой гипотезы против сложной альтернативы. Так, если в области  $Y_2$  параметр  $\eta$  принимает лишь одно возможное значение  $y_2$ , в

то время как в области  $Y_1$  — непрерывный набор значений, описываемый плотностью  $w_\eta(x)$ , то

$$\begin{aligned}\bar{\Lambda}(\mathbf{z}) &= \frac{1}{w(\mathbf{z} / H_2)} \int_{Y_1} w(\mathbf{z} / H_1, \eta) w_\eta(y) dy = \\ &= \int_{Y_1} \frac{w(\mathbf{z} / H_1, \eta)}{w(\mathbf{z} / H_2)} w_\eta(y) dy = \int_{Y_1} \Lambda(\mathbf{z} / \eta) w_\eta(y) dy,\end{aligned}\quad (1.3.3)$$

т. е. отношение усреднённых функций правдоподобия равно усреднённому отношению правдоподобия.

В дальнейшем отношение усреднённых функций правдоподобия будем для краткости называть усреднённым отношением правдоподобия, полагая, что это не вызовет коллизий по отношению к тем случаям, где применение данного термина не является полностью корректным.

Обратимся к уже упомянутому примеру, когда рассматривается нормальная случайная величина с известной дисперсией  $\sigma^2$ , и проверяется простая гипотеза  $H_1$  о том, что математическое ожидание равно  $a_0$  против сложной альтернативной гипотезы  $H_2$ , состоящей в том, что это математическое ожидание не равно  $a_0$ , а случайным образом распределено в некоторой области. Для определённости будем считать распределение  $a$  также нормальным со средним  $\bar{a}$  и дисперсией  $\sigma_a^2$ .

Отношение правдоподобия  $\Lambda(\mathbf{z})$  для данных гипотез согласно (1.1.13) есть

$$\Lambda(\mathbf{z}) = \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^M \left[ (z_k - a_0)^2 - (z_k - a)^2 \right] \right\}.$$

Тогда усреднённое отношение правдоподобия равно

$$\begin{aligned}\bar{\Lambda}(\mathbf{z}) &= \exp \left[ -\frac{1}{2\sigma^2} \sum_{k=1}^M (z_k - a_0)^2 \right] \times \\ &\times \int_{-\infty}^{\infty} \exp \left[ \frac{1}{2\sigma^2} \sum_{k=1}^M (z_k - a)^2 \right] \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp \left[ -\frac{1}{2\sigma_a^2} (a - \bar{a})^2 \right] da.\end{aligned}$$

После несложных преобразований записанное выражение приводится к виду



$$\bar{\Lambda}(\mathbf{z}) = \frac{1}{\sqrt{1 - M \frac{\sigma_a^2}{\sigma^2}}} \exp \left[ -\frac{\sigma_a^2 \left( \sum_{k=1}^M (z_k - a_0)^2 \right)^2}{2\sigma^2 \left( 1 - M \frac{\sigma_a^2}{\sigma^2} \right)} \right] \times \\ \times \exp \left[ \frac{a_0 - \bar{a}}{\sigma^2} \sum_{k=1}^M z_k - \frac{M}{2\sigma^2} (a_0^2 - \bar{a}^2) \right].$$

Тогда правило принятия решения при оптимальной байесовской стратегии можно сформулировать следующим образом: принимается простая гипотеза о том, что математическое ожидание анализируемой случайной величины равно  $a_0$ , если

$$-\frac{\sigma_a^2 \left( \sum_{k=1}^M (z_k - a_0)^2 \right)^2}{2\sigma^2 \left( 1 - M \frac{\sigma_a^2}{\sigma^2} \right)} + \frac{a_0 - \bar{a}}{\sigma^2} \sum_{k=1}^M z_k > \ln \left[ \sqrt{1 - M \frac{\sigma_a^2}{\sigma^2}} \Lambda_0 \right] + \frac{M}{2\sigma^2} (a_0^2 - \bar{a}^2), \quad (1.3.4)$$

и отвергается в противном случае.

Заметим, что когда  $\sigma_a/\sigma = 0$ , полученное правило решения совпадает с выражением (1.1.14), описывающим различие двух простых гипотез в гауссовском шуме. Это и понятно, поскольку стремление  $\sigma_a$  к нулю (при конечном значении  $\sigma$ ) означает вырождение распределения случайного параметра  $a$  в дельта-функцию, т. е. гипотеза из сложной становится простой.

Обобщим полученное решение на задачи проверки сложных гипотез с несколькими неизвестными параметрами. Пусть распределение случайной величины зависит от  $K$  параметров  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_K)$ , и проверяется сложная гипотеза  $H_1$  о том, что указанная совокупность параметров принадлежит  $K$ -мерной области  $Y_1$  против альтернативной сложной гипотезы  $H_2$  о том, что  $\boldsymbol{\eta}$  принадлежит  $K$ -мерной области  $Y_2$ .

Проводя такие же рассуждения, как и для скалярного параметра  $\eta$ , получим следующее правило. Принимается решение в пользу гипотезы  $H_1$ , если для наблюдаемой выборки  $\mathbf{z}$  выполняется неравенство

$$\bar{\Lambda}(\mathbf{z}) = \frac{\bar{w}(\mathbf{z} / H_1)}{\bar{w}(\mathbf{z} / H_2)} = \frac{\int_{Y_1} w(\mathbf{z} / H_1, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y}}{\int_{Y_2} w(\mathbf{z} / H_2, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y}} > \Lambda_0; \quad (1.3.5)$$

если же справедливо обратное неравенство, то принимается решение в пользу гипотезы  $H_2$ .

Как и в случае проверки простых гипотез возможны ситуации, характеризующиеся недостатком информации, необходимой для определения порогового значения  $\Lambda_0$ . В частности, могут быть известны априорные вероятности гипотез  $p(H_i)$ , но отсутствовать какие-либо обоснованные соображения относительно выбора значений  $c_{ij}$  элементов ценовой матрицы. Как было указано в разд. 1.1, в этом случае целесообразна стратегия, оптимальная по критерию максимума апостериорной вероятности (МАВ-стратегия) при дополнительном усреднении апостериорных вероятностей гипотез  $P\{H_i/\mathbf{z}\}$  по неизвестным параметрам:

$$\overline{P\{H_i/\mathbf{z}\}} = \frac{P(H_i) \int_{Y_i} w(\mathbf{z} / H_i, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y}}{P(H_1) \int_{Y_1} w(\mathbf{z} / H_1, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y} + P(H_2) \int_{Y_2} w(\mathbf{z} / H_2, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y}},$$

$$i = 1, 2. \quad (1.3.6)$$

Тогда принимается решение в пользу гипотезы  $H_1$ , если выполняется неравенство

$$\overline{P\{H_1/\mathbf{z}\}} > \overline{P\{H_2/\mathbf{z}\}},$$

и в пользу гипотезы  $H_2$ , если это неравенство не выполняется. В силу одинаковости знаменателя в (1.3.6) для обеих усреднённых апостериорных вероятностей достаточно сравнить соответствующие числители, т. е. принимать решение в пользу  $H_1$  при выполнении условия

$$P(H_1) \int_{Y_1} w(\mathbf{z} / H_1, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y} > P(H_2) \int_{Y_2} w(\mathbf{z} / H_2, \boldsymbol{\eta}) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y}. \quad (1.3.7)$$

Итак, в основе проверки сложных гипотез лежит усреднение функций правдоподобия по неизвестному параметру. Однако может оказаться так, что априорное распределение параметров неизвестно, и при вычислении усреднённого отношения правдоподобия возникают затруднения.

Одним из способов преодоления указанных затруднений является *обобщённый критерий максимального правдоподобия для сложных гипотез*, который, вообще говоря, не имеет строгого обоснования, а основывается на эвристических представлениях. Согласно этому критерию, принимается гипотеза  $H_1$ , если для наблюдаемой выборки  $\mathbf{z}$  выполняется неравенство

$$\Lambda(\mathbf{z}) = \frac{\max_{\eta \in Y_1} w(\mathbf{z} / \eta)}{\max_{\eta \in Y_2} w(\mathbf{z} / \eta)} > \Lambda_0, \quad (1.3.8)$$

и принимается решение в пользу  $H_2$ , если справедливо противоположное неравенство.

В качестве примера применения обобщённого критерия максимального правдоподобия вновь рассмотрим задачу проверки простой гипотезы о среднем (равном  $a_0$ ) значении нормальной случайной величины с известной дисперсией  $\sigma^2$  против сложной гипотезы о том, что это среднее распределено нормально с параметрами  $\bar{a}$  и  $\sigma_a^2$ .

Как известно (более подробно об этом будет сказано в следующей главе), максимум функции правдоподобия нормальной выборки с неизвестным математическим ожиданием  $a$

$$w(\mathbf{z} / a) = (2\pi\sigma^2)^{-M/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^M (z_k - a)^2\right]$$

достигается в том случае, когда в качестве оценки  $a$  берётся среднее арифметическое:

$$\hat{a} = \frac{1}{M} \sum_{k=1}^M z_k.$$

Тогда, в соответствии с (1.3.8),

$$\begin{aligned} \Lambda(\mathbf{z}) &= \frac{\max_a w(\mathbf{z} / a)}{w(\mathbf{z} / a_0)} = \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^M [\hat{a}^2 - 2z_k(\hat{a} - a_0) - a_0^2]\right\} = \\ &= \exp\left[\frac{M}{2\sigma^2} (\hat{a} - a_0)^2\right]. \end{aligned}$$

Теперь, переходя к логарифму отношения правдоподобия, можно записать следующее правило принятия решения в пользу гипотезы  $H_1$ :

$$\left| \frac{1}{M} \sum_{k=1}^M z_k - a_0 \right| > \sqrt{\frac{2\sigma^2}{M} \ln \Lambda_0}. \quad (1.3.9)$$

Следуя концепции, изложенной в разд. 1.1, рассмотрим далее два метода проверки сложных гипотез, применение которых оправдано в условиях априорной неопределённости: минимаксной стратегии и стратегии, оптимальной по критерию Неймана – Пирсона.

При использовании минимаксной стратегии по-прежнему производится усреднение условного риска  $R(\eta)$  по неизвестным параметрам и представление его в виде, аналогичном (1.1.19):

$$\begin{aligned} \bar{R}(p_1) = & c_{22} + (c_{12} - c_{22})\bar{\beta}(p_1) + \\ & + [c_{11} - c_{22} + (c_{21} - c_{11})\bar{\alpha}(p_1) - (c_{12} - c_{22})\bar{\beta}(p_1)]p_1, \end{aligned} \quad (1.3.10)$$

где

$$\bar{\alpha} = \int_{S_2} \bar{w}(\mathbf{z} / H_1) d\mathbf{z}$$

и

$$\bar{\beta} = \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z}$$

усреднённые вероятности ошибок первого и второго рода соответственно.

Теперь для определения точки  $p_1^*$ , через которую будет проходить горизонтальная линия, обеспечивающая минимум максимальных потерь (рис. 1.6), необходимо обеспечить равенство нулю углового коэффициента:

$$c_{11} - c_{22} - (c_{12} - c_{22})\bar{\beta}(p_1^*) = -(c_{21} - c_{11})\bar{\alpha}(p_1^*), \quad (1.3.11)$$

что приводит к решению, как правило, трансцендентного уравнения.

Критерий Неймана–Пирсона также может быть применён для решения задач проверки сложных гипотез, но, в отличие от рассмотренных выше методов, где основным действием (и, как правило, основной трудностью) является усреднение условных вероятностей или других вероятностных характеристик, стратегия, оптимальная по критерию Неймана–Пирсона, изобилует рядом нюансов, связанных со спецификой решаемой задачи.

Рассмотрим вначале применение критерия Неймана–Пирсона к задаче проверки простой гипотезы  $H_1$  против сложной альтернативы  $H_2$ , когда фиксируется вероятность  $\alpha$  ошибки первого рода, соответствующая неко-

тому фиксированному значению  $\eta = \eta_0$ , и минимизируется зависящая от параметра  $\eta$  вероятность  $\beta(\eta)$  ошибки второго рода (или, что эквивалентно, максимизируется мощность процедуры  $1 - \beta(\eta)$ ) для всех простых гипотез, содержащихся в сложной альтернативе  $H_2$ . Такое правило называют *равномерно наиболее мощным* (РНМ-правилом).

Использование РНМ-правила означает, что его рабочая характеристика, как например, кривая вероятности правильного обнаружения, показанная на рис. 1.8, при прочих одинаковых параметрах окажется лучше (выше), чем аналогичные характеристики для любых других правил на всём множестве значений параметра  $\eta$ .

К сожалению, существование РНМ-правила при проверке простой гипотезы против сложной альтернативы является скорее исключением, поскольку правило, максимизирующее мощность процедуры при какой-то фиксированной рассматриваемой альтернативе, обычно оказывается зависящим от этой альтернативы, и соответствующая экстремальная задача имеет решение только в некоторых специальных случаях. Поэтому часто сужают множество рассматриваемых правил, ограничиваясь *классом несмещённых правил*.

Несмещённое правило удовлетворяет дополнительному условию, что вероятность отвергнуть ложную гипотезу не меньше вероятности отвергнуть правильную гипотезу:

$$P\{z \in S_2 / H_2, \eta\} = 1 - \beta(\eta) \geq \alpha(\eta_0).$$

Другими словами, вероятность  $\alpha$  ошибки первого рода при некотором значении параметра  $\eta = \eta_0$  является нижней границей значений мощности критерия  $1 - \beta(\eta)$  для всех значений параметра  $\eta$ . Если  $\beta(\eta)$  — непрерывная функция, то минимальное значение  $1 - \beta(\eta)$  достигается при  $\eta = \eta_0$ .

Отметим, что оптимальное по НП-критерию правило решения при проверке простой гипотезы против сложной альтернативы, вообще говоря, не имеет структуру байесовского правила, как это имело место при проверке двух простых гипотез.

Пусть теперь проверяется сложная гипотеза  $H_1$  против сложной альтернативы  $H_2$ , и вероятность  $\alpha$  ошибки первого рода также зависит от параметра  $\eta$ , значения которого принадлежат области  $Y_1$ . Тогда следует попытаться найти такую совокупность правил выбора решения, которую

можно удовлетворить одним, одинаковым для всех  $\eta$  из области  $Y_1$  уровнем значимости  $\alpha$ , а после этого — отыскать среди них равномерно наиболее мощное для сложной альтернативы  $H_2$ . Конечно, как и при проверке простой гипотезы против сложной альтернативы может оказаться, что такого РНМ-правила не существует, тогда придётся сужать класс правил, вводя дополнительные ограничения.

Так, несмещённость правила для рассматриваемой задачи означает, что все функции мощности  $1 - \beta(\eta)$  в этом классе ограничены значением  $\pi(\eta)$ , равным вероятности попадания выборки  $\mathbf{z}$  в критическую область:

$$\pi(\eta) = P\{\mathbf{z} \in S_2 / \eta\}.$$

Отсюда следует, что функция мощности равна вероятности правильного принятия альтернативы  $H_2$  в случае, когда  $\eta \in Y_1$ , и равна вероятности ошибки первого рода, когда  $\eta \in Y_2$ . Несмещённость требует, чтобы при проверке сложной гипотезы против сложной альтернативы одновременно выполнялись следующие неравенства

$$\begin{aligned} \pi(\eta) &\geq \alpha, & \eta \in Y_1; \\ \pi(\eta) &\leq \alpha, & \eta \in Y_2. \end{aligned}$$

## 1.4. МНОГОАЛЬТЕРНАТИВНЫЕ ЗАДАЧИ

Во всех предыдущих разделах рассматривались задачи проверки либо простых, либо сложных, но, в любом случае, *двух* конкурирующих гипотез. Однако большинство практических вопросов, лежащих в области разработки информационных и телекоммуникационных систем, связано с решением именно многоальтернативных задач, что обусловлено использованием многопозиционных сигнально-кодовых ансамблей.

Методика решения многоальтернативных задач проверки гипотез принципиально не отличается от двоичного случая с тем лишь отличием, что выборочное пространство  $S$  разбивается не на два, а на  $m > 2$  непересекающихся подмножеств, каждое из которых соответствует принятию одной из гипотез:

$$S = \bigcup_{i=1}^m S_i, \quad S_i \cap S_j = \emptyset; \quad i, j = 1, \dots, m. \quad (1.4.1)$$

В общем случае ценовая матрица имеет следующий вид:

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix}, \quad (1.4.2)$$

где  $c_{ij}$  ( $i, j = 1, \dots, m$ ) — плата за принятие гипотезы  $H_i$ , в то время, как в действительности верна гипотеза  $H_j$ . Как обычно, предполагается, что плата за ошибочное решение больше, чем за правильное, т. е. выполняются условия  $c_{ij} > c_{jj}$  и  $c_{ij} > c_{ii}$ .

В данном разделе для простоты ограничимся рассмотрением задачи проверки  $m$  простых гипотез; её обобщение на случай сложных гипотез очевидно и не вызывает каких-либо существенных затруднений.

Запишем частные риски  $R_i$ , соответствующие принятию  $i$ -й ( $i = 1, \dots, m$ ) гипотезы:

$$\begin{aligned} R_1 &= c_{11}P(H_1 / H_1) + c_{12}P(H_1 / H_2) + \dots + c_{1m}P(H_1 / H_m); \\ &\dots \\ R_i &= c_{i1}P(H_i / H_1) + c_{i2}P(H_i / H_2) + \dots + c_{im}P(H_i / H_m); \\ &\dots \\ R_m &= c_{m1}P(H_m / H_1) + c_{m2}P(H_m / H_2) + \dots + c_{mm}P(H_m / H_m). \end{aligned}$$

Тогда средний риск получается посредством усреднения частных рисков по априорным вероятностям гипотез:

$$\begin{aligned} R &= c_{11}P(H_1 / H_1)p(H_1) + \dots + c_{i1}P(H_i / H_1)p(H_1) + \dots + \\ &\quad + c_{m1}P(H_m / H_1)p(H_1) + c_{12}P(H_1 / H_2)p(H_2) + \dots + \\ &\quad + c_{i2}P(H_i / H_2)p(H_2) + \dots + c_{m2}P(H_m / H_2)p(H_2) + \dots + \\ &\quad + c_{1m}P(H_1 / H_m)p(H_m) + \dots + c_{im}P(H_i / H_m)p(H_m) + \dots + \\ &\quad + c_{mm}P(H_m / H_m)p(H_m) = \\ &= \sum_{j=1}^m \sum_{i=1}^m c_{ij}P(H_j)P(H_i / H_j). \end{aligned} \quad (1.4.3)$$

Фигурирующие в (1.4.3) условные вероятности можно представить в виде  $M$ -мерных интегралов от условных плотностей вероятностей  $w(\mathbf{z}/H_i)$  по соответствующим областям  $S_j$ :

$$P(H_i / H_j) = \int_{S_i} w(\mathbf{z} / H_j) d\mathbf{z}; \quad i, j = 1, \dots, m.$$

При  $i = j$  эти вероятности представляют собой вероятности правильного решения; при  $i \neq j$  — вероятности ошибочного решения.

Далее, аналогично тому, как это делалось для двоичного случая, необходимо таким образом выбрать конфигурацию подмножеств  $S_i$ , чтобы обеспечить минимизацию среднего риска. Прежде чем получить решение в общем виде рассмотрим сначала какой-либо частный случай, а затем обобщим полученные результаты.

Пусть, к примеру<sup>1</sup>,  $m = 3$ , т. е. выборочное пространство представляет собой объединение трёх непересекающихся подмножеств  $S_1, S_2$  и  $S_3$ :

$$S = S_1 \cup S_2 \cup S_3.$$

Представим средний риск в виде, аналогичном (1.1.7), т. е. в виде суммы внеинтегральной составляющей (фиксированной стоимости) и переменной интегральной составляющей, значение которой может изменяться посредством изменения конфигурации подмножеств  $S_i$ . При этом вероятности правильных решений выразим через комбинации вероятностей ошибочных решений:

$$\begin{aligned} \int_{S_1} w(\mathbf{z} / H_1) d\mathbf{z} &= \int_{S \setminus (S_2 \cup S_3)} w(\mathbf{z} / H_1) d\mathbf{z} = \int_S w(\mathbf{z} / H_1) d\mathbf{z} - \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} - \\ &- \int_{S_3} w(\mathbf{z} / H_1) d\mathbf{z} = 1 - \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} - \int_{S_3} w(\mathbf{z} / H_1) d\mathbf{z} \end{aligned}$$

и аналогично

$$\begin{aligned} \int_{S_2} w(\mathbf{z} / H_2) d\mathbf{z} &= 1 - \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} - \int_{S_3} w(\mathbf{z} / H_2) d\mathbf{z}; \\ \int_{S_3} w(\mathbf{z} / H_3) d\mathbf{z} &= 1 - \int_{S_1} w(\mathbf{z} / H_3) d\mathbf{z} - \int_{S_2} w(\mathbf{z} / H_3) d\mathbf{z}. \end{aligned}$$

Тогда

$$\begin{aligned} R &= c_{11}p(H_1) - c_{11}p(H_1) \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} - c_{11}p(H_1) \int_{S_3} w(\mathbf{z} / H_1) d\mathbf{z} + \\ &+ c_{21}p(H_1) \int_{S_2} w(\mathbf{z} / H_1) d\mathbf{z} + c_{31}p(H_1) \int_{S_3} w(\mathbf{z} / H_1) d\mathbf{z} + c_{12}p(H_2) \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} + \end{aligned}$$

---

<sup>1</sup> Данный пример взят в сугубо иллюстративных целях; он не имеет какого-либо практического применения, поскольку объём реально используемых ансамблей сигналов представляет собой, как правило, целую степень двойки.



$$\begin{aligned}
& +c_{22}p(H_2) - c_{22}p(H_2) \int_{S_1} w(\mathbf{z} / H_2) d\mathbf{z} - c_{22}p(H_2) \int_{S_3} w(\mathbf{z} / H_2) d\mathbf{z} + \\
& +c_{32}p(H_2) \int_{S_3} w(\mathbf{z} / H_2) d\mathbf{z} + c_{13}p(H_3) \int_{S_1} w(\mathbf{z} / H_3) d\mathbf{z} + c_{23}p(H_3) \int_{S_2} w(\mathbf{z} / H_3) d\mathbf{z} + \\
& +c_{33}p(H_3) - c_{33}p(H_3) \int_{S_1} w(\mathbf{z} / H_3) d\mathbf{z} - c_{33}p(H_3) \int_{S_2} w(\mathbf{z} / H_3) d\mathbf{z}.
\end{aligned}$$

Объединив интегралы по одинаковым областям, имеем:

$$\begin{aligned}
R = & c_{11}p(H_1) + c_{22}p(H_2) + c_{33}p(H_3) + \\
& + \int_{S_1} [(c_{12} - c_{22})p(H_2)w(\mathbf{z} / H_2) + (c_{13} - c_{33})p(H_3)w(\mathbf{z} / H_3)] d\mathbf{z} + \\
& + \int_{S_2} [(c_{21} - c_{11})p(H_1)w(\mathbf{z} / H_1) + (c_{23} - c_{33})p(H_3)w(\mathbf{z} / H_3)] d\mathbf{z} + \\
& + \int_{S_3} [(c_{31} - c_{11})p(H_1)w(\mathbf{z} / H_1) + (c_{32} - c_{22})p(H_2)w(\mathbf{z} / H_2)] d\mathbf{z}. \quad (1.4.4)
\end{aligned}$$

Как нетрудно видеть, в отличие от двоичного случая, *минимизация среднего риска не может быть обеспечена выполнением определённых неравенств внутри подынтегральных функций*, поскольку они представляет собой сумму положительных слагаемых. Следовательно, анализируемую выборку  $\mathbf{z}$  необходимо ставить в соответствие тому подмножеству, в котором подынтегральная функция меньше двух других.

Обозначим для удобства подынтегральные функции:

$$\begin{aligned}
f_1(\mathbf{z}) &= (c_{12} - c_{22})p(H_2)w(\mathbf{z} / H_2) + (c_{13} - c_{33})p(H_3)w(\mathbf{z} / H_3); \\
f_2(\mathbf{z}) &= (c_{21} - c_{11})p(H_1)w(\mathbf{z} / H_1) + (c_{23} - c_{33})p(H_3)w(\mathbf{z} / H_3); \\
f_3(\mathbf{z}) &= (c_{31} - c_{11})p(H_1)w(\mathbf{z} / H_1) + (c_{32} - c_{22})p(H_2)w(\mathbf{z} / H_2).
\end{aligned}$$

Тогда можно записать следующую оптимальную стратегию:

принимается решение в пользу гипотезы  $H_1$ , если

$$f_1(\mathbf{z}) < f_2(\mathbf{z}) \text{ и } f_1(\mathbf{z}) < f_3(\mathbf{z}); \quad (1.4.5a)$$

принимается решение в пользу гипотезы  $H_2$ , если

$$f_2(\mathbf{z}) < f_1(\mathbf{z}) \text{ и } f_2(\mathbf{z}) < f_3(\mathbf{z}); \quad (1.4.5б)$$

принимается решение в пользу гипотезы  $H_3$ , если

$$f_3(\mathbf{z}) < f_2(\mathbf{z}) \text{ и } f_3(\mathbf{z}) < f_1(\mathbf{z}). \quad (1.4.5в)$$

После упрощения и деления обеих частей на величину  $w(\mathbf{z}/H_1)p(H_1)$  неравенства (1.4.5) принимают следующий вид:

для гипотезы  $H_1$

$$(c_{21} - c_{11}) + (c_{22} - c_{12}) \frac{p(H_2) w(\mathbf{z}/H_2)}{p(H_1) w(\mathbf{z}/H_1)} + (c_{23} - c_{13}) \frac{p(H_3) w(\mathbf{z}/H_3)}{p(H_1) w(\mathbf{z}/H_1)} > 0;$$

$$(c_{31} - c_{11}) + (c_{32} - c_{12}) \frac{p(H_2) w(\mathbf{z}/H_2)}{p(H_1) w(\mathbf{z}/H_1)} + (c_{33} - c_{13}) \frac{p(H_3) w(\mathbf{z}/H_3)}{p(H_1) w(\mathbf{z}/H_1)} > 0;$$

для гипотезы  $H_2$

$$(c_{11} - c_{21}) + (c_{12} - c_{22}) \frac{p(H_2) w(\mathbf{z}/H_2)}{p(H_1) w(\mathbf{z}/H_1)} + (c_{13} - c_{23}) \frac{p(H_3) w(\mathbf{z}/H_3)}{p(H_1) w(\mathbf{z}/H_1)} > 0;$$

$$(c_{31} - c_{21}) + (c_{32} - c_{22}) \frac{p(H_2) w(\mathbf{z}/H_2)}{p(H_1) w(\mathbf{z}/H_1)} + (c_{33} - c_{23}) \frac{p(H_3) w(\mathbf{z}/H_3)}{p(H_1) w(\mathbf{z}/H_1)} > 0;$$

для гипотезы  $H_3$

$$(c_{21} - c_{31}) + (c_{22} - c_{32}) \frac{p(H_2) w(\mathbf{z}/H_2)}{p(H_1) w(\mathbf{z}/H_1)} + (c_{23} - c_{33}) \frac{p(H_3) w(\mathbf{z}/H_3)}{p(H_1) w(\mathbf{z}/H_1)} > 0;$$

$$(c_{11} - c_{31}) + (c_{12} - c_{32}) \frac{p(H_2) w(\mathbf{z}/H_2)}{p(H_1) w(\mathbf{z}/H_1)} + (c_{13} - c_{33}) \frac{p(H_3) w(\mathbf{z}/H_3)}{p(H_1) w(\mathbf{z}/H_1)} > 0.$$

Видно, что в получившихся неравенствах, представляющих собой три пары взаимообратных неравенств, фигурируют два отношения правдоподобия: гипотезы  $H_2$  по отношению к гипотезе  $H_1$

$$\Lambda_{21}(\mathbf{z}) = \frac{w(\mathbf{z}/H_2)}{w(\mathbf{z}/H_1)}$$

и гипотезы  $H_3$  по отношению к гипотезе  $H_1$

$$\Lambda_{31}(\mathbf{z}) = \frac{w(\mathbf{z}/H_3)}{w(\mathbf{z}/H_1)}.$$

При рассмотрении двоичной задачи подчёркивалось, что введение отношения правдоподобия, являющегося одномерной функцией выборки  $\mathbf{z}$ , позволяет для вычисления вероятностей ошибок перейти от многомерного интегрирования условных плотностей  $w(\mathbf{z}/H_i)$  по областям  $M$ -мерного выборочного пространства  $S$  к одномерному интегрированию отношения правдоподобия  $\Lambda(\mathbf{z})$  по вещественной оси. Другими словами, в двоичном случае отношение правдоподобия отображает многомерное выборочное

пространство на вещественную ось. Аналогичная ситуация имеет место и для задачи со многими гипотезами с той лишь разницей, что множество решений, на которое отображается выборочное пространство, является в общем случае также многомерным, хотя, конечно, размерность такого множества много меньше размерности  $M$  выборочного пространства.

В рассматриваемом случае (при  $m = 3$ ) выборочное пространство отображается на двумерную область решений, представляющее собой первый квадрант координатной плоскости  $(\Lambda_{21}(\mathbf{z}), \Lambda_{31}(\mathbf{z}))$  (рис. 1.12, *a*).

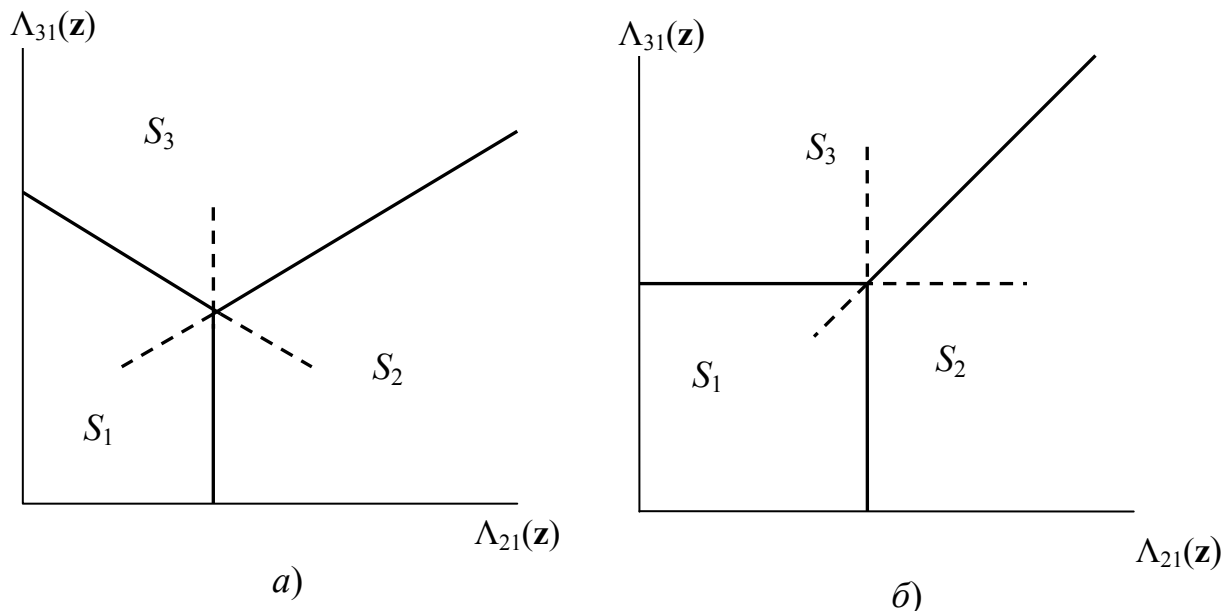


Рис. 1.12. Области решений при  $m = 3$  в общем случае (*a*) и для “идеального” наблюдателя (*б*)

Перепишем условия принятия решения в терминах отношений правдоподобия.

Для гипотезы  $H_1$ :

$$(c_{21} - c_{11}) + (c_{22} - c_{12}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{23} - c_{13}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) > 0;$$

$$(c_{31} - c_{11}) + (c_{32} - c_{12}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{33} - c_{13}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) > 0;$$

(1.4.6a)

для гипотезы  $H_2$

$$\begin{aligned}
& (c_{11} - c_{21}) + (c_{12} - c_{22}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{13} - c_{23}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) > 0; \\
& (c_{31} - c_{21}) + (c_{32} - c_{22}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{33} - c_{23}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) > 0;
\end{aligned}
\tag{1.4.6б}$$

для гипотезы  $H_3$

$$\begin{aligned}
& (c_{21} - c_{31}) + (c_{22} - c_{32}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{23} - c_{33}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) > 0; \\
& (c_{11} - c_{31}) + (c_{12} - c_{32}) \frac{p(H_2)}{p(H_1)} \Lambda_{31}(\mathbf{z}) + (c_{13} - c_{33}) \frac{p(H_3)}{p(H_1)} \Lambda_{21}(\mathbf{z}) > 0.
\end{aligned}
\tag{1.4.6в}$$

Из соотношений (1.4.6) следует, что область принятия решений разбивается на подмножества  $S_1$ ,  $S_2$  и  $S_3$  прямыми линиями, уравнения которых получаются путём замены пары взаимнообратных неравенств в соответствующие равенства:

$$\begin{aligned}
& (c_{21} - c_{11}) + (c_{22} - c_{12}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{23} - c_{13}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) = 0; \\
& (c_{31} - c_{11}) + (c_{32} - c_{12}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{33} - c_{13}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) = 0; \\
& (c_{31} - c_{21}) + (c_{32} - c_{22}) \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) + (c_{33} - c_{23}) \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) = 0.
\end{aligned}
\tag{1.4.7}$$

Нетрудно показать, что эти прямые пересекаются в одной точке и, следовательно, однозначно определяют три области решений.

В частности, рассматривая условия “идеального” наблюдателя

$$c_{ij} = \begin{cases} 1, & i \neq j; \\ 0, & i = j; \end{cases} \quad i, j = 1, \dots, m,$$

из (1.4.7) получаем следующие уравнения:

$$1 - \frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) = 0;$$

$$1 - \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) = 0;$$

$$\frac{p(H_2)}{p(H_1)} \Lambda_{21}(\mathbf{z}) - \frac{p(H_3)}{p(H_1)} \Lambda_{31}(\mathbf{z}) = 0, \quad (1.4.8)$$

представляющие собой в координатах  $(\Lambda_{21}(\mathbf{z}), \Lambda_{31}(\mathbf{z}))$  соответственно вертикальную прямую, горизонтальную прямую и наклонную прямую, которая при равенстве априорных вероятностей является биссектрисой первого квадранта (рис. 1.12, б).

Вернёмся в (1.4.8) к записям  $\Lambda_{21}(\mathbf{z})$  и  $\Lambda_{31}(\mathbf{z})$  через условные вероятности. Имеем:

$$p(H_1)w(\mathbf{z}/H_1) = p(H_2)w(\mathbf{z}/H_2);$$

$$p(H_1)w(\mathbf{z}/H_1) = p(H_3)w(\mathbf{z}/H_3);$$

$$p(H_2)w(\mathbf{z}/H_2) = p(H_3)w(\mathbf{z}/H_3).$$

Если теперь поделить эти соотношения на безусловное распределение выборки

$$w(\mathbf{z}) = \sum_{j=1}^m w(\mathbf{z}/H_j)P(H_j),$$

то образуются апостериорные вероятности гипотез

$$P\{H_i/\mathbf{z}\} = \frac{w(\mathbf{z}/H_i)P(H_i)}{w(\mathbf{z}/H_1)P(H_1) + w(\mathbf{z}/H_2)P(H_2) + w(\mathbf{z}/H_3)P(H_3)}, \quad i = 1, \dots, 3.$$

Таким образом, стратегия идеального наблюдателя, заключающаяся, как нетрудно видеть, в нахождении максимального значения  $p(H_i)w(\mathbf{z}/H_i)$ , эквивалентна стратегии максимума апостериорной вероятности. Как уже было сказано в разд. 1.1, применение МАВ-стратегии оправдано в тех случаях, когда отсутствует информация об априорных вероятностях гипотез.

После того, как рассмотрен частный пример, обратимся к общему случаю многоальтернативной задачи с  $m$  конкурирующими гипотезами. Перепишем (1.4.3) с использованием  $M$ -мерного интегрирования условных плотностей и изменим порядок интегрирования и суммирования:

$$R = \sum_{j=1}^m \sum_{i=1}^m c_{ij} p(H_j) \int_{S_i} w(\mathbf{z}/H_j) d\mathbf{z} = \sum_{i=1}^m \int_{S_i} \sum_{j=1}^m c_{ij} w(\mathbf{z}/H_j) p(H_j) d\mathbf{z}.$$

Тогда можно утверждать, что средний риск минимизируется при отнесе-

нии наблюдаемой выборки  $\mathbf{z}$  в определённое подмножество  $S_k$  ( $k = 1, \dots, m$ ), если сумма

$$\sum_{j=1}^m c_{kj} w(\mathbf{z} / H_j) p(H_j)$$

меньше аналогичной суммы для всех других значений  $i$ , отличных от  $k$ :

$$\sum_{j=1}^m c_{kj} w(\mathbf{z} / H_j) p(H_j) < \sum_{j=1}^m c_{ij} w(\mathbf{z} / H_j) p(H_j), \text{ для всех } i \neq k$$

или

$$\sum_{j=1}^m (c_{ij} - c_{kj}) w(\mathbf{z} / H_j) p(H_j) > 0. \quad (1.4.9)$$

Как обычно, возникают трудности применения (1.4.9), когда неизвестны либо стоимости  $c_{ij}$ , либо априорные вероятности  $p(H_j)$ , либо и то, и другое. В таких случаях возможно и целесообразно использование частных стратегий, рассмотренных в разд. 1.1. Например, при использовании стратегии идеального наблюдателя и равенстве априорных вероятностей гипотез в сумме

$$(c_{i1} - c_{ki}) w(\mathbf{z} / H_1) + \dots + (c_{ii} - c_{ki}) w(\mathbf{z} / H_i) + \dots + (c_{ik} - c_{kk}) w(\mathbf{z} / H_k) + \dots + (c_{im} - c_{km}) w(\mathbf{z} / H_m)$$

отличными от нуля являются только слагаемые

$$(c_{ii} - c_{ki}) w(\mathbf{z} / H_i) = -w(\mathbf{z} / H_i)$$

и

$$(c_{ik} - c_{kk}) w(\mathbf{z} / H_k) = w(\mathbf{z} / H_k).$$

Следовательно, соотношение (1.4.9) вырождается в условие

$$w(\mathbf{z} / H_k) > w(\mathbf{z} / H_i), \text{ для всех } i \neq k,$$

т. е. в условие

$$\Lambda_{ki}(\mathbf{z}) > 1, \text{ для всех } i \neq k, \quad (1.4.10),$$

выбора максимального отношения правдоподобия, что является непосредственным обобщением двоичной задачи проверки гипотез.

Рассмотрим возможности нахождения вероятностей ошибок при решении многоальтернативных задач проверки гипотез.

Обозначим через  $\pi_k$  случайную величину  $w(\mathbf{z}/H_k)p(H_k)$ . Вероятность  $p_{\text{ош}}(k)$  — вероятность ошибки при принятии решения в пользу гипотезы  $H_k$  можно выразить следующим образом:

$$\begin{aligned} p_{\text{ош}}(k) &= 1 - P\left\{\pi_k > \max_{i \neq k} \pi_i / H_k\right\} = \\ &= 1 - \int_0^{\infty} \int_0^{\pi_k} \cdots \int_0^{\pi_k} w(p_1, p_2, \dots, p_m / H_k) dp_1 dp_2 \dots dp_m. \end{aligned} \quad (1.4.11)$$

Тогда средняя вероятность ошибки  $p_{\text{ош}}$  всей процедуры равна

$$p_{\text{ош}} = \sum_{k=1}^m p_{\text{ош}}(k). \quad (1.4.12)$$

Подынтегральным выражением в (1.4.11) является условная совместная плотность распределения вероятностей случайных величин  $\pi_k$ , которые, как правило, не являются независимыми, что делает точное вычисление  $p_{\text{ош}}(k)$  затруднительным. На практике применяют различные оценки вероятностей  $p_{\text{ош}}(k)$ . Одну из таких оценок в виде двухсторонней границы можно получить, если учесть, что

$$\max_{i \neq k} P\{\pi_i - \pi_k \geq 0 / H_k\} \leq p_{\text{ош}}(k) \leq \sum_{\substack{i=1 \\ i \neq k}}^m P\{\pi_i - \pi_k \geq 0 / H_k\}. \quad (1.4.13)$$

Поскольку  $\pi_k$  являются вероятностями (плотностями) и, следовательно, не могут быть отрицательными, справедливо соотношение

$$\begin{aligned} P\{\pi_i - \pi_k \geq 0 / H_k\} &= P\left\{\frac{\pi_k}{\pi_i} \leq 1 / H_k\right\} = P\left\{\frac{w(\mathbf{z} / H_k) p(H_k)}{w(\mathbf{z} / H_i) p(H_i)} \leq 1 / H_k\right\} = \\ &= P\left\{\Lambda_{ki}(\mathbf{z}) \leq \frac{p(H_i)}{p(H_k)} / H_k\right\}, \end{aligned}$$

т. е. одномерная вероятностная характеристика для отношения правдоподобия.

Методика и примеры вычисления вероятностей ошибок на основе отношений правдоподобия были подробно рассмотрены в разд. 1.1. Соотношения типа (1.4.13), часто в литературе называемые *аддитивными границами*, используются для оценки вероятностей ошибок различных многопозиционных ансамблей сигналов.

## ВОПРОСЫ И ЗАДАНИЯ К ГЛАВЕ 1

1.1. Что представляет собой понятие “статистическая гипотеза”? Какие виды статистических гипотез вам известны? Перечислите основные этапы проверки статистических гипотез.

1.2. Что такое статистический критерий? Какие виды статистических критериев вам известны?

1.3. Рассматривается двоичный источник, выход которого представляет собой последовательность независимых нулей ‘0’ и единиц ‘1’ с вероятностью появления  $P('0') = 1 - P('1') = p$ . Определить, используя шесть последовательных выходных символов, находится ли он в состоянии  $H_1$ , при котором  $p = 2/3$  или в состоянии  $H_2$ , при котором  $p = 1/2$ .

а) Показать, что в каждой из следующих ситуаций стратегия приводит к принятию решения в пользу гипотезы  $H_1$ , когда число появившихся нулей больше некоторого числа  $n_0$ , и определить  $n_0$  в каждом случае:

- $P(H_1) = P(H_2) = 1/2$ ;  $c_{12} - c_{22} = c_{21} - c_{11}$ ;
- $P(H_1) = P(H_2) = 1/2$ ;  $c_{12} - c_{22} = 1$ ;  $c_{21} - c_{11} = 10$ ;
- $P(H_1) = P(H_2) = 1/2$ ;  $c_{12} - c_{22} = 10$ ;  $c_{21} - c_{11} = 1$ ;
- вероятность  $\alpha$  ошибки первого рода не превосходит  $1/9$ ;
- априорные вероятности гипотез неизвестны;  $c_{22} = c_{11} = 0$ ;  $c_{12} = 1$ ;  $c_{21} = 10$ .

Для всех соотношений найти вероятности ошибок обоих родов.

б) Указать последовательную процедуру определения состояний источника, найти среднее и дисперсию числа выходных символов источника, необходимых для последовательного решения при  $\alpha = \beta = 1/3$ . (Обратить внимание, что из-за небольшого числа испытаний влияние “перескоков границ” существенно, поэтому ответы будут приближёнными).

в) Предположим, что при использовании последовательной процедуры источник на самом деле выдаёт нули с вероятностью  $p = 5/12$ . Найти вероятность принятия гипотезы  $H_1$  и определить математическое ожидание числа наблюдений, необходимых для завершения процедуры.

г) Какие условия накладывает минимаксный критерий на пороговые значения в последовательной процедуре?

1.4. Пусть  $H_1$  — простая гипотеза, состоящая в том, что выборка  $\mathbf{z} = \{z_1, z_2, \dots, z_M\}$  получена из одностороннего экспоненциального распределения



$$w_1(z) = \exp(-z), z > 0,$$

а  $H_2$  — простая альтернатива, состоящая в том, что выборка получена из одностороннего нормального распределения

$$w_2(z) = \sqrt{2/\pi} \exp(-z^2/2), z > 0.$$

Получить оптимальное правило принятия решения, найти вероятности ошибок первого и второго рода, а также асимптотическое поведение вероятностей ошибок при  $M \rightarrow \infty$ .

Ответ: гипотеза  $H_1$  отвергается, если

$$\frac{1}{M} \sum_{k=1}^M (z_k - 1)^2 \leq 1 + \ln \frac{2}{\pi} - \frac{2 \ln \Lambda_0}{M},$$

где  $\Lambda_0$  — константа, характеризующая порог принятия решений. При  $M \rightarrow \infty$  имеют место следующие асимптотические формулы для вероятностей ошибок первого и второго рода:

$$\alpha \sim 1 - F \left[ \frac{1}{\sqrt{2M}} \left( \ln \Lambda_0 + \frac{M}{2} \ln \frac{\pi}{2} \right) \right],$$

$$\beta \sim F \left[ \frac{\ln \Lambda_0 - (2\sqrt{2/\pi} + \ln(2/\pi) - 1)M/2}{\sqrt{M(3/2 - 2/\pi - \sqrt{2/\pi})}} \right],$$

где  $F$  — функция Лапласа.

1.5. Пусть  $H_1$  — простая гипотеза, состоящая в том, что целочисленная выборка  $\mathbf{z}$  получена из распределения Пуассона

$$p_k \equiv P(z = k) = \frac{\lambda^k}{k!} \exp(-\lambda); \lambda > 0; k = 0, 1, \dots$$

с параметром  $\lambda = \lambda_1$ , а  $H_2$  — простая альтернатива, состоящая в том, что  $\lambda = \lambda_2 > \lambda_1$ . Получить оптимальное байесовское правило принятия решения. Найти минимальную вероятность  $\beta$  ошибки второго рода при фиксированной вероятности  $\alpha$  ошибки первого рода.

Ответ: гипотеза  $H_1$  отвергается, если

$$\frac{1}{M} \sum_{k=1}^M z_k \geq \frac{\lambda_2 - \lambda_1 + (1/M) \ln \Lambda_0}{\ln(\lambda_2/\lambda_1)}.$$

При фиксированном  $\alpha$  минимальное значение  $\beta$  равно

$$\beta = 1 - \frac{\Gamma([\Lambda_0] + 1, \lambda_2)}{\Gamma([\Lambda_0] + 1)}.$$

Здесь  $\Gamma$  — гамма-функция (полная и неполная), а  $[\Lambda_0]$  — целая часть константы  $\Lambda_0$ , определяемой из условия

$$\frac{\Gamma([\Lambda_0] + 1, \lambda_1)}{\Gamma([\Lambda_0] + 1)} = \alpha.$$

1.6. Пусть  $H_1$  — простая гипотеза, состоящая в том, что выборка  $\mathbf{z}$  получена из релеевского распределения

$$w(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad x > 0$$

с параметром  $\sigma = \sigma_1$ , а  $H_2$  — простая альтернатива, состоящая в том, что выборка получена из того же распределения с параметром  $\sigma = \sigma_2 > \sigma_1$ . Получить оптимальное правило принятия решения, найти вероятности ошибок первого и второго рода.

Ответ: гипотеза  $H_1$  отвергается, если

$$\sum_{k=1}^M z_k^2 \geq Z_0,$$

где

$$Z_0 = \frac{2\sigma_1^2\sigma_2^2}{\sigma_2^2 - \sigma_1^2} \ln\left[\Lambda_0 (\sigma_2 / \sigma_1)^{2M}\right].$$

Вероятности  $\alpha$  и  $\beta$  ошибок первого и второго рода соответственно равны

$$\alpha = \frac{\Gamma(M, Z_0 / \sigma_1^2)}{\Gamma(M)}; \quad \beta = 1 - \frac{\Gamma(M, Z_0 / \sigma_2^2)}{\Gamma(M)}.$$

1.7. Пусть  $H_1$  — простая гипотеза, состоящая в том, что выборочное значение  $\mathbf{z}$  получено из нормального распределения

$$w_1(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2 / 2),$$

а  $H_2$  — простая альтернатива, состоящая в том, что значение  $\mathbf{z}$  получено из бимодального распределения вида

$$w_2(z) = \frac{1}{2\sqrt{\pi}} \left[ \exp(-(z - 2a)^2 / 2) + \exp(-(z + 2a)^2 / 2) \right].$$

Получить оптимальное правило принятия решения.

Ответ: гипотеза  $H_1$  отвергается, если

$$\frac{w_1(z)}{w_2(z)} = \frac{\exp(a^2)}{2} \left[ \exp(-(z - 2a)^2 / 2) + \exp(-(z + 2a)^2 / 2) \right] \geq \Lambda_0.$$

1.8. Получить критерий Неймана–Пирсона, исходя из следующей задачи на условный экстремум.

При фиксированной вероятности  $\alpha$  ошибки первого рода минимизировать вероятность  $\beta$  ошибки второго рода посредством поиска такой области  $S_1$ , в которой выражение

$$\lambda\alpha + \beta = \lambda \left[ 1 - \int_{S_1} w(\mathbf{z} / H_0) d\mathbf{z} \right] + \int_{S_1} w(\mathbf{z} / H_1) d\mathbf{z},$$

где  $\lambda$  — неизвестный множитель, принимает максимальное значение.

1.9. Пусть  $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$  — векторная выборка (т. е. каждый элемент выборки сам является вектором) из  $N$ -мерного нормального распределения с корреляционной матрицей  $\mathbf{K} = \sigma^2 \mathbf{I}_N$ , где  $\mathbf{I}_N$  — единичная матрица размерности  $N \times N$ . Выдвигается гипотеза  $H_1$  о том, что векторное среднее равно  $\mathbf{a}_1 = \{a_1^{(1)}, a_1^{(2)}, \dots, a_1^{(M)}\}$  против альтернативной гипотезы  $H_2$ , что это среднее равно  $\mathbf{a}_2 = \{a_2^{(1)}, a_2^{(2)}, \dots, a_2^{(M)}\}$ .

Показать, что оптимальное по критерию максимального правдоподобия разбиение выборочного пространства производится гиперплоскостью, перпендикулярной линии, соединяющей точки  $\mathbf{z} = \mathbf{a}_1$  и  $\mathbf{z} = \mathbf{a}_2$  и делящей эту линию пополам.

1.10. Пусть  $H_1$  — гипотеза, состоящая в том, что векторная выборка  $\mathbf{z}$  получена из  $N$ -мерного нормального распределения с векторным средним  $\mathbf{a}$  и корреляционной матрицей  $\mathbf{K}_1$  против альтернативной гипотезы  $H_2$ , что распределение имеет такое же среднее  $\mathbf{a}$ , но корреляционную матрицу  $\mathbf{K}_2$ . Получить оптимальное байесовское правило принятия решения.

Ответ: принимается гипотеза  $H_1$ , если

$$(\mathbf{z} - \mathbf{a})^T (\mathbf{K}_2^{-1} - \mathbf{K}_1^{-1})(\mathbf{z} - \mathbf{a}) < 2 \ln \Lambda_0 + \ln \frac{\det \mathbf{K}_1}{\det \mathbf{K}_2}.$$

Совершая замену переменных  $\mathbf{y} = \mathbf{f}'(\mathbf{z} - \mathbf{a})$ , где  $j$ -й ( $j = 1, \dots, N$ ) вектор-столбец  $\mathbf{f}_j$  матрицы  $\mathbf{f}$  определяется из уравнения

$$\mathbf{K}_2 \mathbf{f}_j = \lambda_j \mathbf{K}_1 \mathbf{f}_j,$$

а коэффициенты  $\lambda_1, \dots, \lambda_N$  являются корнями характеристического уравнения

$$\det[\mathbf{K}_2 - \lambda \mathbf{I}_N] = 0,$$

условие принятия решения можно привести к виду

$$\sum_{j=1}^N (\lambda_j - 1) y_j^2 < 2 \ln \Lambda_0 + \sum_{j=1}^N \ln \lambda_j.$$

1.11. Выборка  $\mathbf{z}$  взята из нормального распределения с нулевым средним. Найти оптимальное по критерию Неймана–Пирсона правило проверки простой гипотезы  $H_1$ , что дисперсия равна  $\sigma_1^2$  против сложной альтернативы  $H_2$ , что дисперсия  $\sigma^2 \neq \sigma_1^2$ . Является ли это правило равномерно наиболее мощным?

Ответ: гипотеза  $H_1$  отвергается, если

$$\sum_{k=1}^M z_k^2 \geq \sigma_0^2 \chi_\alpha^2,$$

где  $\chi_\alpha^2$  — процентная точка  $\chi^2$ -распределения с  $M$  степенями свободы. Это РМН-правило при  $\sigma^2 > \sigma_1^2$ .

1.12. Выборка  $\mathbf{z}$  взята из одностороннего экспоненциального распределения. Найти оптимальное по критерию Неймана–Пирсона правило проверки простой гипотезы  $H_1$ , что параметр распределения равен  $\lambda_1$  против сложной альтернативы  $H_2$ , что параметр распределения  $\lambda \neq \lambda_1$ . Является ли это правило равномерно наиболее мощным?

Ответ: гипотеза  $H_1$  отвергается, если

$$\sum_{k=1}^M z_k \leq \frac{1}{2\lambda_1} \chi_{1-\alpha}^2,$$

где  $\chi_{1-\alpha}^2$  — процентная точка  $\chi^2$ -распределения с  $2M$  степенями свободы. Это РМН-правило при  $\lambda > \lambda_1$ .

1.13. В чём суть метода последовательного анализа проверки статистических гипотез? Каковы критерии проведения такого анализа?

1.14. Покажите, что последовательная процедура проверки гипотез всегда заканчивается за конечное число шагов, т. е. если  $P_H(M > \mu)$  — вероятность того, что решение в пользу какой-либо из гипотез принимается более, чем за  $\mu$  шагов, то

$$\lim_{\mu \rightarrow \infty} P_H(M > \mu) = 0.$$

## ГЛАВА 2. ОЦЕНИВАНИЕ ПАРАМЕТРОВ

В данной главе будут рассмотрены основные аспекты второго направления теории статистических выводов — оценивании параметров и векторов состояния систем. Методы решения таких задач базируются на рассмотренной в предыдущей главе общей байесовской теории принятия решений, когда при выборе одной из  $m$  возможных гипотез необходимо обеспечить минимальное значение среднего риска (1.4.4). Также рассматривается методика получения оценок, основанная на минимизации суммы квадратов в модели линейной регрессии.

### 2.1. ХАРАКТЕРИСТИКИ И СВОЙСТВА СТАТИСТИЧЕСКИХ ОЦЕНОК

При всём своём разнообразии статистические оценки могут быть разделены на два класса: *точечные* и *интервальные*.

Точечные оценки  $\hat{\eta}_M$  — это функции выборочных значений, называемые *статистиками*, которые являются в определённом смысле наилучшими оценками неизвестного параметра  $\eta$ :

$$\hat{\eta}_M = g(\mathbf{z}) = g(z_1, \dots, z_M). \quad (2.1.1)$$

Под интервальной оценкой понимают некоторый интервал  $[\hat{\eta}_M^{(1)}; \hat{\eta}_M^{(2)}]$  значений неизвестного параметра, в котором оцениваемый параметр содержится с заданной вероятностью  $\gamma$ :

$$P\{\hat{\eta}_M^{(1)} < \eta < \hat{\eta}_M^{(2)}\} = \gamma, \quad (2.1.2)$$

где нижняя  $\hat{\eta}_M^{(1)}$  и верхняя  $\hat{\eta}_M^{(2)}$  оценки являются, в свою очередь, точечными оценками параметра  $\eta$ .

Рассмотрим вначале основные характеристики точечных оценок на примере скалярного параметра  $\eta$ . При этом будем различать *условную оценку*

$$\hat{\eta}_M = g(z_1, \dots, z_M / \eta), \quad (2.1.3)$$

когда оцениваемый параметр является неизвестной константой, либо *безусловную оценку*

$$\hat{\eta}_M = \int_Y g(z_1, \dots, z_M / \eta) w_\eta(y) dy, \quad (2.1.4)$$

при которой оцениваемый параметр является случайной величиной с известным априорным распределением  $w_\eta(y)$ . Интегрирование в (2.1.4) производится по всему множеству  $Y$  значений параметра  $\eta$ ; по умолчанию будем считать, что это множество — вся вещественная ось от  $-\infty$  до  $+\infty$ . В выражениях (2.1.3) и (2.1.4) условная и безусловная оценки обозначаются одинаково; как правило, различие в понятиях выявляется из контекста и не приводит к путанице.

И условная, и безусловная точечные оценки как функции выборки являются случайными величинами, распределение которых определяется соответственно либо условной  $w(\mathbf{z} / \eta)$ , либо безусловной

$$w(\mathbf{z}) = \int_Y w(\mathbf{z} / \eta) w_\eta(y) dy$$

функциями правдоподобия<sup>1</sup> выборки.

Понятно, что для одной и той же выборки можно построить бесконечно много точечных оценок, и необходимо иметь признаки (характеристики), по которым можно было отобрать хорошие или, даже, в определённом смысле наилучшие оценки. Важнейшими из таких признаков являются *состоятельность*, *достаточность*, *несмещённость* и *эффективность*.

Условная оценка  $\hat{\eta}_M$  параметра  $\eta$  называется состоятельной, если она сходится по вероятности<sup>2</sup> к оцениваемому параметру при неограниченном увеличении размера выборки, т. е. при произвольном  $\varepsilon > 0$

$$\lim_{M \rightarrow \infty} P\{|\hat{\eta}_M - \eta| \geq \varepsilon\} = 0. \quad (2.1.5a)$$

Аналогично, безусловная оценка  $\hat{\eta}_M$  параметра  $\eta$  называется состоя-

---

<sup>1</sup> Один из основоположников математической статистики Р. А. Фишер предлагал называть  $w(\mathbf{z} / \eta)$  функцией правдоподобия выборки, когда она рассматривается как функция оцениваемого параметра  $\eta$ , и обычной условной плотностью вероятности при рассмотрении её как функции  $\mathbf{z}$  при фиксированном значении  $\eta$ .

<sup>2</sup> Некоторые авторы определяют состоятельность оценки через сходимость в среднеквадратическом:

$$\lim_{M \rightarrow \infty} \mathbf{E} \left[ |\hat{\eta}_M - \eta|^2 \right] = 0.$$

Как известно, из сходимости в среднеквадратическом следует сходимость по вероятности.

тельной, если она сходится по вероятности к среднему значению  $\mathbf{E}[\eta]$  оцениваемого параметра при неограниченном увеличении размера выборки, т. е. при произвольном  $\varepsilon > 0$

$$\lim_{M \rightarrow \infty} P\{|\hat{\eta}_M - \mathbf{E}[\eta]| \geq \varepsilon\} = 0. \quad (2.1.5b)$$

Условная оценка  $\hat{\eta}_M$  называется несмещённой, если её среднее, вычисляемое по выборочному пространству  $Z^M$ , совпадает с самим оцениваемым параметром:

$$\mathbf{E}[\hat{\eta}_M] = \int_{Z^M} \hat{\eta}_M(\mathbf{z}) w(\mathbf{z} / \eta) d\mathbf{z} = \eta. \quad (2.1.6a)$$

Аналогично, безусловная оценка  $\hat{\eta}_M$  называется несмещённой, если её среднее по выборочному пространству совпадает со средним значением оцениваемого параметра:

$$\mathbf{E}[\hat{\eta}_M] = \int_{Z^M} \hat{\eta}_M(\mathbf{z}) w(\mathbf{z}) d\mathbf{z} = \mathbf{E}[\eta]. \quad (2.1.6b)$$

Разность

$$d_M(\eta) = \mathbf{E}[\hat{\eta}_M] - \eta \quad (2.1.7a)$$

для условной или

$$d_M(\eta) = \mathbf{E}[\hat{\eta}_M] - \mathbf{E}[\eta] \quad (2.1.7b)$$

для безусловной оценки называется *смещением оценки*. С практической точки зрения смещение оценки является систематической ошибкой при проведении эксперимента.

Если смещение является линейной функцией оцениваемого параметра:

$$d_M(\eta) = k\eta + b,$$

то оно может быть устранено посредством перехода к модифицированной оценке

$$\hat{\theta}_M = \frac{\hat{\eta}_M - b}{k + 1}.$$

Действительно, в этом случае из соотношения

$$\mathbf{E}[\hat{\eta}_M] - \eta = k\eta + b$$

следует, что

$$\mathbf{E}[\hat{\eta}_M] = (k + 1)\eta + b,$$

тогда разность

$$\mathbf{E}[\hat{\theta}_M] - \eta = \frac{\mathbf{E}[\hat{\eta}_M] - b}{k+1} - \eta = \frac{(k+1)\eta + b - b}{k+1} - \eta = 0.$$

Иногда рассматриваются оценки, для которых условия (2.1.6) выполняются не при любых размерах выборки, а лишь при неограниченном увеличении  $M$ . В этом случае говорят про *асимптотическую несмещённость* оценки.

Условная оценка  $\hat{\eta}_M$  называется достаточной, если условная функция правдоподобия  $w(\mathbf{z} / \hat{\eta}_M)$  не зависит от оцениваемого параметра  $\eta$ . Другими словами, достаточная оценка содержит всю информацию о неизвестном параметре, которую можно получить при наблюдении выборочных значений. При этом нет необходимости знать каждый из элементов выборки в отдельности, а достаточно иметь лишь некоторую функцию совокупности выборочных значений — так называемую *достаточную статистику*.

В реальности прямое определение свойства достаточности может оказаться затруднительным ввиду сложности вычисления  $w(\mathbf{z} / \hat{\eta}_M)$ , и на практике обычно применяют *признак факторизации функции правдоподобия*, заключающийся в возможности представления  $w(\mathbf{z} / \hat{\eta}_M)$  в виде произведения двух неотрицательных сомножителей:

$$w(\mathbf{z} / \hat{\eta}_M) = f_1(\hat{\eta}_M / \eta) f_2(\mathbf{z}), \quad (2.1.8)$$

один из которых зависит в явном виде от достаточной статистики и оцениваемого параметра, а второй — только от выборочных значений.

Заметим также, что если  $\hat{\eta}_M$  — достаточная оценка параметра  $\eta$ , то для параметра  $\theta = \varphi(\eta)$ , где  $\varphi$  — некоторая функция, достаточной оценкой будет  $\varphi(\hat{\eta}_M)$ .

Несмещённая оценка  $\hat{\eta}_{M \text{ eff}}$  называется эффективной, если она имеет наименьшую дисперсию среди других оценок:

$$\mathbf{D}[\hat{\eta}_M] \geq \mathbf{D}[\hat{\eta}_{M \text{ eff}}]. \quad (2.1.9a)$$

Для смещённых оценок условие наименьшей дисперсии заменяется условием наименьшей среднеквадратической ошибки:

$$\mathbf{E}[(\hat{\eta}_M - \eta)^2] \geq \mathbf{E}[(\hat{\eta}_{M \text{ eff}} - \eta)^2]. \quad (2.1.9б)$$

На практике работать с эффективными оценками, как правило, не удаётся; часто в условиях решаемых задач они вообще могут не существовать.



Поэтому вводится понятие *относительной эффективности* среди набора используемых оценок  $\hat{\eta}_M^{(1)}, \dots, \hat{\eta}_M^{(n)}$ . При этом показатель эффективности  $\delta_M$  определяет отношение дисперсий для несмещённых оценок:

$$\delta_M = \frac{\mathbf{D}[\hat{\eta}_M^{(k)}]}{\mathbf{D}[\hat{\eta}_M^{(1)}]}, k = 1, \dots, n \quad (2.1.10a)$$

или отношение среднеквадратических ошибок для смещённых оценок:

$$\delta_M = \frac{\mathbf{E}[(\hat{\eta}_M^{(k)} - \eta)^2]}{\mathbf{E}[(\hat{\eta}_M^{(1)} - \eta)^2]}, k = 1, \dots, n. \quad (2.1.10б)$$

Существует *неравенство Рао–Крамера*, с помощью которого можно определить нижнюю границу дисперсии (среднеквадратической ошибки) при использовании любых оценок [3].

Пусть  $w(\mathbf{z}/\eta)$  — условная функция правдоподобия выборки. Неотрицательная величина

$$I_M(\eta) = \mathbf{D}\left[\frac{\partial}{\partial \eta} \ln w(\mathbf{z}/\eta)\right] \quad (2.1.11)$$

называется *фишеровской информацией* (информационной единицей Фишера), содержащейся в выборке.

Отметим, что случайная величина  $\partial[\ln w(\mathbf{z}/\eta)]/\partial \eta$ , имеет нулевое среднее, что можно показать, дифференцируя по параметру  $\eta$  условие нормировки для функции правдоподобия:

$$\int_{Z^M} w(\mathbf{z}/\eta) d\mathbf{z} = 1.$$

Действительно, имеем:

$$0 = \frac{d}{d\eta} \int_{Z^M} w(\mathbf{z}/\eta) d\mathbf{z} = \int_{Z^M} \frac{\partial \ln[w(\mathbf{z}/\eta)]}{\partial \eta} w(\mathbf{z}/\eta) d\mathbf{z} = \mathbf{E}\left[\frac{\partial \ln[w(\mathbf{z}/\eta)]}{\partial \eta}\right].$$

Следовательно, вычисление дисперсии в (2.1.11) эквивалентно вычислению математического ожидания от квадрата  $\partial[\ln w(\mathbf{z}/\eta)]/\partial \eta$ :

$$I_M(\eta) = \mathbf{E}\left[\left(\frac{\partial}{\partial \eta} \ln w(\mathbf{z}/\eta)\right)^2\right]. \quad (2.1.11a)$$

Используя свойство производной логарифма

$$\frac{df(x)}{dx} = f(x) \frac{d \ln f(x)}{dx},$$

можно получить ещё одно выражение для вычисления фишеровской информации:

$$I_M(\eta) = -\mathbf{E} \left[ \frac{\partial^2}{\partial \eta^2} \ln w(\mathbf{z} / \eta) \right]. \quad (2.1.11\delta)$$

Действительно, дифференцируя второй раз по  $\eta$  соотношение

$$\frac{\partial}{\partial \eta} \ln w(\mathbf{z} / \eta) = \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial w(\mathbf{z} / \eta)}{\partial \eta},$$

имеем

$$\frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} = - \left( \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial w(\mathbf{z} / \eta)}{\partial \eta} \right)^2 + \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2}.$$

Тогда, согласно (2.1.11а),

$$\begin{aligned} I_M(\eta) &= \mathbf{E} \left[ \left( \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial w(\mathbf{z} / \eta)}{\partial \eta} \right)^2 \right] = -\mathbf{E} \left[ \frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \right] + \\ &+ \mathbf{E} \left[ \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} \right]. \end{aligned}$$

Но

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} \right] &= \int_{Z^M} \frac{1}{w(\mathbf{z} / \eta)} \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} w(\mathbf{z} / \eta) d\mathbf{z} = \\ &= \int_{Z^M} \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} d\mathbf{z} = \frac{d^2}{d\eta^2} \int_{Z^M} w(\mathbf{z} / \eta) d\mathbf{z} = \frac{d^2}{d\eta^2} (1) = 0, \end{aligned}$$

отсюда получаем (2.1.11б).

Для выборки с независимыми элементами величина  $I_M(\eta)$  может быть выражена через одномерное (для любого  $z_k = z$ ,  $k = 1, \dots, M$ ) условное распределение  $w(z / \eta)$  и объём выборки  $M$ :

$$I(\eta) = \int_{-\infty}^{\infty} \left[ \frac{\partial}{\partial \eta} \ln w(z / \eta) \right]^2 w(z / \eta) dz \quad (2.1.12a)$$

для непрерывного распределения или

$$I(\eta) = \sum_{(i)} \left[ \frac{\partial}{\partial \eta} \ln p_i(z / \eta) \right]^2 p_i(z / \eta) \quad (2.1.12\delta)$$

для дискретного распределения, при этом

$$I_M(\eta) = MI(\eta).$$

Согласно неравенству Рао–Крамера, дисперсия оценки  $\mathbf{D}[\hat{\eta}_M]$  со смещением  $d_M(\eta)$  ограничена снизу следующим образом:

$$\mathbf{D}[\hat{\eta}_M] \geq \frac{[1 + d'_M(\eta)]^2}{I_M(\eta)}. \quad (2.1.13)$$

Для несмещённых оценок  $d_M(\eta) = 0$ , и

$$\mathbf{D}[\hat{\eta}_M] \geq \frac{1}{I_M(\eta)}, \quad (2.1.13a)$$

т. е. нижней границей дисперсии является величина, обратная фишеровской информации, содержащейся в выборке.

Исходя из вышесказанного, эффективность часто определяют из условия близости к нижней границе в неравенстве Рао–Крамера. В тех случаях, когда минимально возможная дисперсия оценки совпадает с правой частью (2.1.13), оба определения эффективности совпадают, однако для некоторых распределений граница Рао–Крамера может оказаться слишком грубой.

Как и для случая асимптотической несмещённости, эффективность оценки (определяемая различным образом) может проявляться не при любых размерах выборки, а лишь при неограниченном увеличении  $M$ . В этом случае говорят про *асимптотическую эффективность* оценки.

Рассмотрим основные особенности интервального оценивания. В выражении (2.1.2) вероятность  $\gamma$  принято называть *доверительной вероятностью* (иногда в литературе эту вероятность называют *коэффициентом доверия*), а сам интервал  $[\hat{\eta}_M^{(1)}; \hat{\eta}_M^{(2)}]$  — *доверительным интервалом*.

Ширина  $\hat{\eta}_M^{(2)} - \hat{\eta}_M^{(1)}$  доверительного интервала является случайной величиной — в зависимости от того, каким образом строятся точечные оценки для нижней и верхней границ интервала. Можно попытаться сделать длину доверительного интервала постоянной, выбирая

$$\hat{\eta}_M^{(1)} = \hat{\eta}_M - \varepsilon\eta; \quad \hat{\eta}_M^{(2)} = \hat{\eta}_M + \varepsilon\eta,$$

где  $\hat{\eta}_M$  — точечная оценка параметра  $\eta$ , а  $\varepsilon$  — некоторое положительное число. Тогда длина доверительного интервала постоянна и равна  $2\varepsilon\eta$ .

Значение  $\varepsilon$  можно найти различными способами. Обычно его определяют из распределения самой точечной оценки

$$P\{(1-\varepsilon)\eta < \hat{\eta}_M < (1+\varepsilon)\eta\} = \gamma, \quad (2.1.14)$$

либо другой связанной с ней случайной величиной с известным распределением, например

$$P\left\{-\varepsilon < \frac{\hat{\eta}_M}{\eta} - 1 < \varepsilon\right\} = \gamma. \quad (2.1.14a)$$

Если, например, известна плотность распределения  $w_{\hat{\eta}_M}(t)$  точечной оценки  $\hat{\eta}_M$ , то  $\varepsilon$  определяется из следующего интегрального соотношения:

$$\int_{(1-\varepsilon)\eta}^{(1+\varepsilon)\eta} w_{\hat{\eta}_M}(t) dt = \gamma, \quad (2.1.15)$$

связывающего доверительную вероятность  $\gamma$ , относительную длину  $2\varepsilon$  доверительного интервала и объём выборки  $M$ .

При заданном  $\varepsilon$  доверительная вероятность будет возрастать с ростом объёма выборки, приближаясь к единице (если, конечно, используемая точечная оценка является состоятельной). С другой стороны, при фиксированном объёме выборки доверительная вероятность будет уменьшаться с уменьшением доверительного интервала. Это означает, что невозможно увеличивать доверительную вероятность без неизбежного увеличения доверительного интервала.

Пусть, к примеру, рассматривается задача оценки параметра  $\eta$  случайной величины  $\zeta$ , распределённой по одностороннему экспоненциальному закону:

$$w_\zeta(z/\eta) = (1/\eta)\exp(-z/\eta), \quad z > 0, \eta > 0.$$

Как известно, математическое ожидание такой случайной величины есть  $\eta$ , а её дисперсия равна  $\eta^2$ .

Поскольку оцениваемый параметр равен математическому ожиданию, выберем в качестве его условной точечной оценки выборочное среднее

$$\hat{\eta}_M = \frac{1}{M} \sum_{k=1}^M z_k.$$

Нетрудно видеть, что такая оценка является несмещённой, а её дисперсия составляет

$$\mathbf{D}[\hat{\eta}_M] = \frac{1}{M^2} \sum_{k=1}^M \mathbf{D}[z_k] = \frac{\eta^2}{M}.$$

Из закона больших чисел следует, что вероятность того, что выборочное среднее, т. е. оценка  $\hat{\eta}_M$  отличается от априорного среднего, т. е. оцениваемого параметра  $\eta$ , по модулю на величину, большую, чем произвольное  $\varepsilon > 0$ , стремится к нулю, когда объём выборки  $M$  неограниченно возрастает. Другими словами,  $\hat{\eta}_M$  сходится по вероятности к  $\eta$ , что определяет состоятельность оценки.

Условная функция правдоподобия  $M$ -элементной выборки, в предположении, что её элементы независимы, равна

$$w^{(M)}(\mathbf{z} / \eta) = \prod_{k=1}^M w(z_k / \eta) = \frac{1}{\eta^M} \exp\left(-\frac{1}{\eta} \sum_{k=1}^M z_k\right) = \frac{1}{\eta^M} \exp\left(-M \frac{\hat{\eta}_M}{\eta}\right).$$

Видно, что условное распределение

$$w^{(M)}(\mathbf{z} / \hat{\eta}_M) = \frac{1}{(\hat{\eta}_M)^M} \exp(-M) = \left(\frac{1}{M} \sum_{k=1}^M z_k\right)^{-M} \exp(-M)$$

не зависит от оцениваемого параметра, следовательно, согласно признаку факторизации (2.1.8), оценка является достаточной ( $f_1$  — это сама функция правдоподобия  $w^{(M)}(\mathbf{z} / \eta)$ , а множитель  $f_2$  тождественно равен единице).

Найдём фишеровскую информацию, содержащуюся в выборке. Из (2.1.12) имеем:

$$I_M(\eta) = M \int_0^{\infty} \left(\frac{1}{\eta} - \frac{z}{\eta^2}\right)^2 \frac{1}{\eta} \exp(-z/\eta) dz = \frac{M}{\eta^2} = \left(\frac{\eta^2}{M}\right)^{-1} = (\mathbf{D}[\hat{\eta}_M])^{-1},$$

откуда следует, что нижней границей дисперсии в неравенстве Рао–Крамера является величина, обратная фишеровской информации, содержащейся в выборке. Следовательно, выборочное среднее является эффективной оценкой параметра  $\eta$  экспоненциального распределения.

Получим интервальную оценку параметра  $\eta$ . Из (2.1.14) имеем следующее соотношение для определения параметров интервального оценивания:

$$P \left\{ 2M(1-\varepsilon) < \frac{2M\hat{\eta}_M}{\eta} < 2M(1+\varepsilon) \right\} = \gamma.$$

Как известно [3], нормированная на величину  $\eta/2$  сумма из  $M$  независимых экспоненциально распределённых случайных величин

$$\frac{2}{\eta} \sum_{k=1}^M z_k = \frac{2M \hat{\eta}_M}{\eta}$$

имеет  $\chi^2$ -распределение с  $2M$  степенями свободы. Отсюда

$$(1 - \varepsilon) = \frac{1}{2M} \chi_{(1-\gamma)/2}^2; \quad (1 + \varepsilon) = \frac{1}{2M} \chi_{(1+\gamma)/2}^2,$$

где  $\chi_\alpha^2$  — это  $\alpha$ -процентная точка  $\chi^2$ -распределения с  $2M$  степенями свободы.

Итак, при заданной доверительной вероятности  $\gamma$  значения оцениваемого параметра  $\eta$  лежат в следующем интервале:

$$\frac{2M \hat{\eta}_M}{\chi_{(1+\gamma)/2}^2} < \eta < \frac{2M \hat{\eta}_M}{\chi_{(1-\gamma)/2}^2}. \quad (2.1.16)$$

Обратимся теперь к случаю, когда функция распределения зависит не от одного, а от нескольких неизвестных параметров  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$ , описываемых априорной совместной плотностью распределения  $w_{\boldsymbol{\eta}}(\mathbf{y})$ , и по выборке  $\mathbf{z}$  формируются  $L$  условных оценок

$$\hat{\eta}_M^{(j)} = g(z_1, \dots, z_M / \eta_j), \quad j = 1, \dots, L \quad (2.1.17)$$

или, посредством усреднения по совместному распределению параметров,  $L$  безусловных оценок

$$\hat{\eta}_M^{(j)} = \int_Y g(z_1, \dots, z_M / \eta_j) w_{\boldsymbol{\eta}}(\mathbf{y}) d\mathbf{y}, \quad j = 1, \dots, L. \quad (2.1.18)$$

Интегрирование в (2.1.18) производится по всему  $L$ -мерному множеству  $Y$  значений вектора параметров  $\boldsymbol{\eta}$ .

Характеристики оценок, введённые применительно к скалярному параметру  $\eta$ , теперь применяются как компонентные к векторной оценке

$$\hat{\boldsymbol{\eta}}_M = (\hat{\eta}_M^{(1)}, \hat{\eta}_M^{(2)}, \dots, \hat{\eta}_M^{(L)}).$$

Оценка  $\hat{\boldsymbol{\eta}}_M$  называется состоятельной, если состоятельными являются все её компоненты.

Оценка  $\hat{\boldsymbol{\eta}}_M$  называется несмещённой, если несмещёнными являются все её компоненты.

Оценка  $\hat{\boldsymbol{\eta}}_M$  называется достаточной, если все её компоненты являются *взаимно достаточными*, т. е. если функцию правдоподобия  $w(\mathbf{z} / \mathbf{y})$ , согласно признаку факторизации, можно представить в виде произведения двух сомножителей

$$w(\mathbf{z} / \hat{\boldsymbol{\eta}}_M) = f_1(\hat{\boldsymbol{\eta}}_M / \boldsymbol{\eta}) f_2(\mathbf{z}), \quad (2.1.19)$$

один из которых зависит в явном виде от вектора оценки и вектора оцениваемых параметров, а второй — только от выборочных значений.

Понятие фишеровской информации обобщается для векторного параметра посредством введения *информационной матрицы Фишера*  $\mathbf{I}_M$ , элементы которой равны

$$I_M^{(i,j)}(\boldsymbol{\eta}) = \mathbf{E} \left[ \frac{\partial}{\partial \eta_i} \ln w(\mathbf{z} / \boldsymbol{\eta}) \frac{\partial}{\partial \eta_j} \ln w(\mathbf{z} / \boldsymbol{\eta}) \right], \quad i, j = 1, \dots, L. \quad (2.1.20)$$

Нетрудно показать, по аналогии с (2.1.11), что математическое ожидание элементов информационной матрицы Фишера равно нулю. Как и в (2.1.12),  $I_M^{(i,j)}(\boldsymbol{\eta})$  могут быть выражены через объём выборки  $M$  и одномерное условное распределение  $w(z / \boldsymbol{\eta})$ :

$$I_M^{(i,j)}(\boldsymbol{\eta}) = M \mathbf{E} \left[ \frac{\partial}{\partial \eta_i} \ln w(z / \boldsymbol{\eta}) \frac{\partial}{\partial \eta_j} \ln w(z / \boldsymbol{\eta}) \right], \quad i, j = 1, \dots, L. \quad (2.1.21)$$

Обобщённое неравенство Рао—Крамера формулируется следующим образом. Если определитель информационной матрицы Фишера отличен от нуля (т. е. если она невырождена), то для любого ненулевого вектора строки  $\mathbf{u} = (u_1, \dots, u_L)$  справедливо неравенство

$$\sum_{i=1}^L \sum_{j=1}^L \left\{ \mathbf{E} \left[ (\hat{\eta}_M^{(i)} - \eta_i)(\hat{\eta}_M^{(j)} - \eta_j) \right] - v_M^{(i,j)} \right\} u_i u_j \geq 0, \quad (2.1.22)$$

где  $v_M^{(i,j)}$  — элементы матрицы  $\mathbf{I}_M^{-1}$ , обратной информационной матрице Фишера. Соотношение (2.1.22) можно переписать в матричной форме:

$$\mathbf{u}(\mathbf{M} - \mathbf{I}_M^{-1})\mathbf{u}^T \geq 0. \quad (2.1.22a)$$

Здесь  $\mathbf{M}$  — корреляционная матрица ошибок, т. е. матрица, элементами которой являются ковариации разности между оценкой параметра и его истинным значением:

$$\mu_{ij} = \mathbf{E} \left[ (\hat{\eta}_M^{(i)} - \eta_i)(\hat{\eta}_M^{(j)} - \eta_j) \right], \quad i, j = 1, \dots, L. \quad (2.1.23)$$

Тогда обобщённое неравенство Рао—Крамера — это условие неотрицательной знакоопределённости матричного умножения в (2.1.22a).

Компоненты векторной оценки называются *совместно эффективными*, если в (2.1.22) достигается равенство для произвольного объёма вы-

борки. Если же это равенство имеет место лишь при  $M \rightarrow \infty$ , то компоненты векторной оценки называются *совместно асимптотически эффективными*.

В качестве примера векторной оценки параметров одномерного распределения рассмотрим совместную оценку математического ожидания  $a$  и дисперсии  $\sigma^2$  гауссовского распределения

$$w(z / a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-a)^2}{2\sigma^2}\right],$$

когда в качестве оценки  $a$  используется выборочное среднее

$$\hat{a}_M = \frac{1}{M} \sum_{k=1}^M z_k,$$

а в качестве оценки  $\sigma^2$  — модифицированная выборочная дисперсия, т. е. выборочная дисперсия, умноженная на  $M/(M-1)$ :

$$\hat{\sigma}_M^2 = \frac{1}{M-1} \sum_{k=1}^M (z_k - \hat{a}_M)^2.$$

Данные оценки являются несмещёнными и состоятельными для любых распределений, в том числе, и для рассматриваемого гауссовского.

Условная функция правдоподобия выборки объёма  $M$ , равная

$$\begin{aligned} w(\mathbf{z} / a, \sigma^2) &= (2\pi\sigma^2)^{-M/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^M (z_k - a)^2\right] = \\ &= (2\pi\sigma^2)^{-M/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^M z_k^2 + \frac{a}{\sigma^2} \sum_{k=1}^M z_k - \frac{Ma^2}{2\sigma^2}\right], \end{aligned} \quad (2.1.24)$$

содержит две статистики:

$$\sum_{k=1}^M z_k = M\hat{a}_M$$

и

$$\sum_{k=1}^M z_k^2 = (M-1)\hat{\sigma}_M^2 + M\hat{a}_M^2.$$

Тогда, подставляя их в (2.1.24) и сравнивая полученное выражение с (2.1.19), убеждаемся в выполнении признака факторизации, т. е. обе статистики являются совместно достаточными и, таким образом, векторная оценка  $(\hat{a}_M, \hat{\sigma}_M^2)$  — достаточная.



Покажем, что компоненты векторной оценки — независимые случайные величины. Рассмотрим следующее преобразование (в литературе его часто называют *преобразованием Хельмерта*) элементов выборки:

$$\begin{aligned} u_1 &= \frac{1}{\sigma\sqrt{2}}(z_1 - z_2); \\ u_2 &= \frac{1}{\sigma\sqrt{6}}(z_1 + z_2 - 2z_3); \\ u_3 &= \frac{1}{\sigma\sqrt{12}}(z_1 + z_2 + z_3 - 3z_4); \\ &\dots \\ u_{M-1} &= \frac{1}{\sigma\sqrt{M(M-1)}}(z_1 + z_2 + \dots + z_{M-1} - (M-1)z_M); \\ u_M &= \frac{1}{\sigma\sqrt{M}}(z_1 + z_2 + \dots + z_M). \end{aligned}$$

Каждую из случайных величин  $u_k$  можно трактовать как  $M$ -мерный вектор, построенный в базисе элементов выборки:

$$u_k = \sum_{j=1}^M c_{kj} z_j, \quad (k = 1, \dots, M),$$

где

$$c_{kj} = \frac{1}{\sigma\sqrt{k+k^2}} \begin{cases} 1, & j = 1, \dots, k; \\ -k, & j = k+1; \\ 0, & j = k+2, \dots, M; \end{cases} \quad k = 1, \dots, M-1,$$

и

$$c_{mj} = \frac{1}{\sigma\sqrt{M}}.$$

Нетрудно видеть, что  $u_k$  имеют нулевое среднее, единичную дисперсию и ортогональные коэффициенты  $c_k$ , поскольку их скалярное произведение

$$\sum_{i=1}^M c_{ki} c_{ji} = \begin{cases} 0, & k \neq j; \\ 1/\sigma^2, & k = j. \end{cases}$$

Следовательно, если элементы  $z_k$  независимы и нормально распределены, то также независимы и нормально распределены элементы  $u_k$ .

Кроме того, преобразованию Хельмерта присуще ещё одно свойство

ортогональности коэффициентов  $c_k$  (ортогональность “по столбцу” матрицы преобразования):

$$\sum_{k=1}^M c_{kj} c_{ki} = \delta_{ji} / \sigma^2 = \begin{cases} 0, & j \neq i; \\ 1 / \sigma^2, & j = i; \end{cases} \quad j, i = 1, \dots, M.$$

Тогда

$$\begin{aligned} \sum_{k=1}^M u_k^2 &= \sum_{k=1}^M \left( \sum_{j=1}^M c_{kj} z_j \right)^2 = \sum_{k=1}^M \sum_{j=1}^M \sum_{i=1}^M c_{ki} c_{kj} z_j z_i = \\ &= \sum_{j=1}^M \sum_{i=1}^M z_j z_i \sum_{k=1}^M c_{ki} c_{kj} = \sum_{j=1}^M \sum_{i=1}^M z_j z_i \delta_{ij} / \sigma^2 = \sum_{j=1}^M (z_j / \sigma)^2, \end{aligned}$$

и можно записать следующее представление оценки  $\hat{\sigma}_M^2$  через величины  $u_k$ :

$$\hat{\sigma}_M^2 = \frac{1}{M-1} \sum_{k=1}^M z_k^2 - \frac{M}{M-1} \hat{a}_M^2 = \frac{\sigma^2}{M-1} \sum_{k=1}^M u_k^2 - \frac{\sigma^2}{M-1} u_M^2 = \frac{\sigma^2}{M-1} \sum_{k=1}^{M-1} u_k^2.$$

Таким образом, оценка  $\hat{\sigma}_M^2$  распределена также, как нормированная сумма из  $(M-1)$  квадратов гауссовских случайных величин с нулевым средним и единичной дисперсией, т. е. имеет  $\chi^2$ -распределение с  $M-1$  степенями свободы:

$$w_{\hat{\sigma}_M^2}(s) \sim \exp\left(-\frac{s}{2}\right) s^{\frac{M-1}{2}-1}.$$

Символ “ $\sim$ ” означает функциональное представление с точностью до числовых коэффициентов. Видно, что распределение оценки  $\hat{\sigma}_M^2$ , содержащей в себе выборочное среднее  $\hat{a}_M$ , на самом деле не зависит от значений  $\hat{a}_M$ , что говорит о независимости компонент векторной оценки  $(\hat{a}_M, \hat{\sigma}_M^2)$ .

Как будет показано далее, независимость выборочных среднего и дисперсии существенно облегчает вычисление элементов информационной матрицы Фишера и матрицы ошибок.

Распределение  $\hat{a}_M$  — это распределение суммы независимых гауссовских случайных величин суть также гауссовское распределение. Центрируя и нормируя, получаем стандартную (с нулевым средним и единичной дисперсией) гауссовскую величину:

$$\alpha_M = \frac{(\hat{a}_M - a)}{\sigma} \sqrt{M} \in N(0, 1).$$

Сумма

$$\sum_{k=1}^M \left( \frac{z_k - \hat{a}}{\sigma} \right)^2 = \frac{M-1}{\sigma^2} \hat{\sigma}_M^2 \in \chi^2(M-1)$$

квадратов нормированных гауссовских величин имеет  $\chi^2$ -распределение с  $M-1$  степенями свободы (одна степень свободы “расходуется” как связь между элементами выборки в выборочном среднем).

Определим элементы информационной матрицы Фишера.

$$I_M^{(1,1)}(a, \sigma^2) = M \mathbf{E} \left[ \left( \frac{\partial}{\partial a} \ln w(z/a, \sigma^2) \right)^2 \right] = M \mathbf{E} \left[ \frac{(z-a)^2}{\sigma^4} \right] = \frac{M}{\sigma^4} \sigma^2 = \frac{M}{\sigma^2}.$$

При нахождении элементов побочной диагонали учтём независимость выборочных средних и дисперсии, что позволит факторизовать математическое ожидание произведения:

$$\begin{aligned} I_M^{(1,2)}(a, \sigma^2) &= I_M^{(2,1)}(a, \sigma^2) = M \mathbf{E} \left[ \frac{\partial}{\partial a} \ln w(z/a, \sigma^2) \frac{\partial}{\partial \sigma^2} \ln w(z/a, \sigma^2) \right] = \\ &= M \mathbf{E} \left[ \frac{\partial}{\partial a} \ln w(z/a, \sigma^2) \right] \mathbf{E} \left[ \frac{\partial}{\partial \sigma^2} \ln w(z/a, \sigma^2) \right] = \\ &= M \mathbf{E} \left[ \frac{(z-a)}{\sigma^2} \right] \mathbf{E} \left[ \frac{(z-a)^2}{2\sigma^4} - \frac{1}{2\sigma^2} \right] = M \cdot 0 \cdot 0 = 0. \end{aligned}$$

Отсюда следует, что в рассматриваемом примере информационная матрица Фишера является диагональной.

$$I_M^{(2,2)}(a, \sigma^2) = M \mathbf{E} \left[ \left( \frac{\partial}{\partial \sigma^2} \ln w(z/a, \sigma^2) \right)^2 \right] = M \mathbf{E} \left[ \left( \frac{(z-a)^2}{2\sigma^4} - \frac{1}{2\sigma^2} \right)^2 \right].$$

Используя значения различных моментов гауссовского распределения, получаем, что

$$I_M^{(2,2)}(a, \sigma^2) = \frac{M}{2\sigma^4}.$$

Итак, информационная матрица Фишера имеет следующий вид:

$$\mathbf{I}_M = \begin{pmatrix} M/\sigma^2 & 0 \\ 0 & M/(2\sigma^4) \end{pmatrix}. \quad (2.1.25)$$

Нетрудно видеть, что её определитель

$$\det \mathbf{I}_M = \frac{M^2}{2\sigma^6} > 0,$$

следовательно, элементы обратной матрицы равны

$$v_M^{(1,1)} = \sigma^2 / M; \quad v_M^{(1,2)} = 0; \quad v_M^{(2,1)} = 0; \quad v_M^{(2,2)} = 2\sigma^4 / M,$$

т. е. обратная матрица есть

$$\mathbf{I}_M^{-1} = \begin{pmatrix} \sigma^2 / M & 0 \\ 0 & 2\sigma^4 / M \end{pmatrix}.$$

Найдём элементы матрицы ошибок согласно (2.1.23):

$$\mu_{11} = \mathbf{E}[(\hat{a} - a)^2] = \mathbf{D}[\hat{a}] = \frac{\sigma^2}{M};$$

$$\mu_{12} = \mu_{21} = \mathbf{E}[(\hat{a} - a)(\hat{\sigma}^2 - \sigma^2)] = \mathbf{E}[(\hat{a} - a)]\mathbf{E}[(\hat{\sigma}^2 - \sigma^2)] = 0;$$

$$\mu_{22} = \mathbf{E}[(\hat{\sigma}^2 - \sigma^2)^2] = \frac{2\sigma^4}{M-1}.$$

Тогда

$$\mathbf{M} = \begin{pmatrix} \sigma^2 / M & 0 \\ 0 & 2\sigma^4 / (M-1) \end{pmatrix}. \quad (2.1.26)$$

Обобщённое неравенство Рао–Крамера (2.1.22) для произвольного ненулевого двухкомпонентного вектора-строки  $\mathbf{u} = (u_1, u_2)$  имеет следующий вид:

$$\mathbf{u}(\mathbf{M} - \mathbf{I}_M^{-1})\mathbf{u}^T = (u_1 \quad u_2) \begin{pmatrix} \frac{\sigma^2}{M} - \frac{\sigma^2}{M} & 0 \\ 0 & \frac{2\sigma^4}{M-1} - \frac{2\sigma^4}{M} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \frac{2\sigma^4 u_2^2}{M(M-1)}.$$

Видно, что для любого конечного  $M$  матричное произведение  $\mathbf{u}(\mathbf{M} - \mathbf{I}_M^{-1})\mathbf{u}^T$  всегда положительно, и, следовательно, выборочные среднее и дисперсия *не* являются совместно эффективными оценками (равенство нулю в (2.1.22) не достигается). При  $M \rightarrow \infty$  матричное произведение  $\mathbf{u}(\mathbf{M} - \mathbf{I}_M^{-1})\mathbf{u}^T \rightarrow 0$ , что говорит об асимптотической эффективности используемой векторной оценки.

Построим интервальные оценки для математического ожидания и дисперсии гауссовского распределения, причём в качестве точечных со-

ставляющих по-прежнему будем использовать несмещённые состоятельные оценки в виде выборочных среднего и дисперсии.

Введём нормированную ошибку оценки:

$$\tau_M = \frac{(\hat{a}_M - a)}{\hat{\sigma}} \sqrt{M} = \frac{(\hat{a}_M - a)}{\sigma} \sqrt{M} \sqrt{\sigma^2 / \hat{\sigma}^2} = \alpha_M \sqrt{\sigma^2 / \hat{\sigma}^2},$$

которая, представляет собой отношение двух независимых случайных величин: стандартной гауссовской случайной величины  $\alpha_M$  и случайной величины  $\sqrt{\hat{\sigma}^2 / \sigma^2}$ , распределённой как корень квадратный из  $\chi^2$ -распределения с  $M - 1$  степенями свободы. Как известно [3], такая случайная величина  $\tau_M$  имеет распределение Стьюдента с  $M - 1$  степенями свободы:

$$w_{\tau_M}(t) = \frac{1}{\sqrt{\pi(M-1)}} \frac{\Gamma(M/2)}{\Gamma((M-1)/2)} \left(1 + \frac{t^2}{M-1}\right)^{-M/2}.$$

Таким образом, вероятность  $\gamma$  того, что абсолютное значение нормированной ошибки  $\tau_M$  не превосходит заданного значения  $\tau_0$ , равна

$$P\{|\tau_M| \leq \tau_0\} = \int_{-\tau_0}^{\tau_0} w_{\tau_M}(t) dt = \gamma. \quad (2.1.27)$$

Теперь, располагая, например, табулированными значениями  $\tau_0$ , удовлетворяющими соотношению (2.1.27) при различных  $M$  и  $\gamma$ , построим следующую интервальную оценку математического ожидания гауссовского распределения:

$$\hat{a} - \tau_0 \sqrt{\frac{\hat{\sigma}^2}{M}} < a < \hat{a} + \tau_0 \sqrt{\frac{\hat{\sigma}^2}{M}}. \quad (2.1.28)$$

При построении интервальной оценки выберем в качестве границ значения  $\varepsilon_1 \hat{\sigma}_M^2$  и  $\varepsilon_2 \hat{\sigma}_M^2$ :

$$P\left\{\varepsilon_1 \hat{\sigma}_M^2 < \sigma^2 < \varepsilon_2 \hat{\sigma}_M^2\right\} = P\left\{\frac{M-1}{\varepsilon_2} < (M-1) \frac{\sigma^2}{\hat{\sigma}_M^2} < \frac{M-1}{\varepsilon_1}\right\} = \gamma,$$

здесь  $\varepsilon_1$  и  $\varepsilon_2$  — некоторые положительные числа. Точечная оценка  $(M-1) \hat{\sigma}_M^2 / \sigma^2$ , как уже говорилось, имеет  $\chi^2$ -распределения с  $M - 1$  степенями свободы. Тогда положим

$$\varepsilon_1 = \frac{M-1}{\chi^{2(1)}}; \quad \varepsilon_2 = \frac{M-1}{\chi^{2(2)}},$$

где  $\chi^{2(1)}$  и  $\chi^{2(2)}$  — процентные точки  $\chi^2$ -распределения (рис. 2.1), определяемые из условия (заштрихованные области)

$$P\{\chi^2 > \chi_1^2\} = P\{\chi^2 < \chi_2^2\} = \frac{1-\gamma}{2},$$

т. е.

$$\chi^{2(1)} = \chi_{(1-\gamma)/2}^2; \quad \chi^{2(2)} = \chi_{(1+\gamma)/2}^2.$$

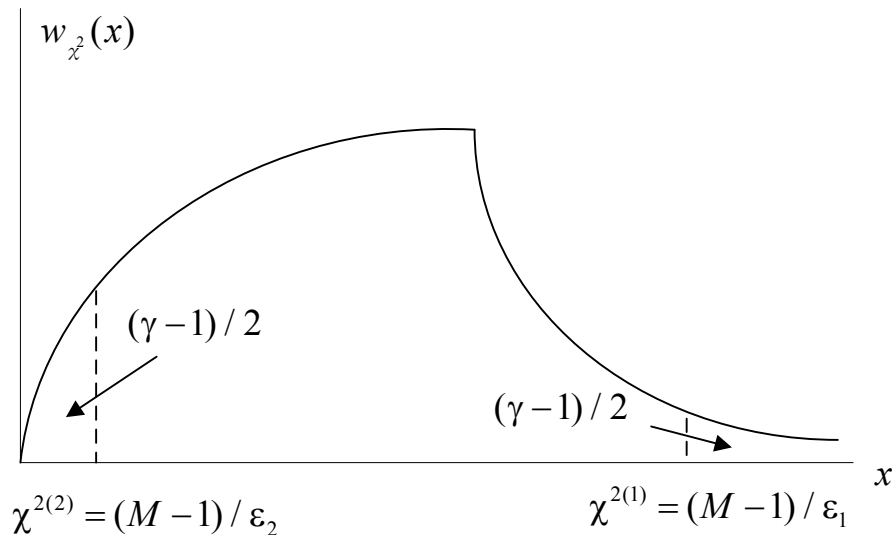


Рис. 2.1. Интервальное оценивание дисперсии гауссовского распределения

Итак, интервальная оценка дисперсии гауссовского распределения определяется следующим неравенством:

$$\frac{M-1}{\chi_{(1-\gamma)/2}^2} \hat{\sigma}_M^2 < \sigma^2 < \frac{M-1}{\chi_{(1+\gamma)/2}^2} \hat{\sigma}_M^2. \quad (2.1.29)$$

Предыдущее изложение было связано с оцениванием (скалярным или векторным) параметров одномерного распределения. Обобщим основные положения теории оценивания для совокупности из  $n$ , вообще говоря, зависимых случайных величин. Как и в задачах проверки гипотез, в данном случае выборка является матрицей вида (1.1.26):

$$\mathbf{z} = \begin{pmatrix} z_1^{(1)} & z_2^{(1)} & \dots & z_M^{(1)} \\ z_1^{(2)} & z_2^{(2)} & \dots & z_M^{(2)} \\ \dots & \dots & \dots & \dots \\ z_1^{(n)} & z_2^{(n)} & \dots & z_M^{(n)} \end{pmatrix}.$$

Аналогичным образом, функция правдоподобия выборки  $w(\mathbf{z} / \boldsymbol{\eta})$  зависит от  $M$  векторных аргументов  $\mathbf{z}_k$  ( $k = 1, \dots, M$ ), а её неизвестные пара-

метры представляются в виде векторов (когда оценивается один параметр) или матриц (в случае нескольких оцениваемых параметров). Для независимых элементов выборки функция правдоподобия факторизуется:

$$w(\mathbf{z} / \boldsymbol{\eta}) \equiv w^{(M,n)}(\mathbf{z} / \boldsymbol{\eta}) = \prod_{k=1}^M w^{(n)}(\mathbf{z}_k / \boldsymbol{\eta}). \quad (2.1.30)$$

В выражении (2.1.30) обозначение  $w^{(M,n)}(\mathbf{z} / \boldsymbol{\eta})$  подчёркивает, что данное распределение относится к  $M$ -элементной выборке, взятой из  $n$ -мерного распределения  $w^{(n)}(\mathbf{z}_k / \boldsymbol{\eta})$ , содержащего вектор  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$  неизвестных параметров.

По каждой выборке  $\mathbf{z}$  формируются условные (точечные) оценки неизвестных параметров, а если известно совместное распределение  $w_{\boldsymbol{\eta}}(\mathbf{y})$  вектора параметров, то могут быть определены и безусловные оценки. Свойства оценок (несмещённость, состоятельность, эффективность, достаточность) формулируются также, как и в одномерном случае — посредством покомпонентного рассмотрения. Так, под средним значением случайной матрицы понимается матрица, элементы которой равны средним значениям элементов случайной матрицы.

Как и в главе 1, проиллюстрируем рассмотренные выше обобщения на примере совместной оценки среднего и корреляционной матрицы  $n$ -мерного нормального распределения

$$w^{(n)}(\mathbf{z}_k / \mathbf{a}, \mathbf{K}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{K}}} \exp \left[ -\frac{1}{2} (\mathbf{z}_k - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) \right],$$

где  $\mathbf{a} = [a^{(1)}, \dots, a^{(n)}]^T$  — вектор-столбец средних значений,  $\mathbf{K}^{-1}$  — матрица, обратная корреляционной.

Пусть  $\mathbf{z}_1, \dots, \mathbf{z}_k$  — независимые выборочные  $n$ -элементные векторы, для которых функция правдоподобия, согласно (2.1.30), равна

$$w^{(M,n)}(\mathbf{z}_k / \mathbf{a}) = (2\pi)^{-Mn/2} (\det \mathbf{K})^{-M/2} \exp \left[ -\frac{1}{2} \sum_{k=1}^M (\mathbf{z}_k - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) \right],$$

Введём вектор-столбец  $\hat{\mathbf{a}}_M = [\hat{a}_M^{(1)}, \dots, \hat{a}_M^{(n)}]^T$  выборочных средних

$$\hat{\mathbf{a}}_M = \frac{1}{M} \sum_{k=1}^M \mathbf{z}_k \quad (2.1.31)$$

и выборочную корреляционную матрицу

$$\hat{\mathbf{K}} = \frac{1}{M} \sum_{k=1}^M (\mathbf{z}_k - \hat{\mathbf{a}}_M)(\mathbf{z}_k - \hat{\mathbf{a}}_M)^T, \quad (2.1.32)$$

элементами которой являются выборочные коэффициенты корреляции  $\hat{b}_{ij}$ , рассчитываемые для  $i$ -й и  $j$ -й строк выборочной матрицы ( $i, j = 1, \dots, n$ ):

$$\hat{b}_{ij} = \frac{1}{M} \sum_{k=1}^M (z_k^{(i)} - \hat{a}_M^{(i)})(z_k^{(j)} - \hat{a}_M^{(j)}) = \frac{1}{M} \sum_{k=1}^M z_k^{(i)} z_k^{(j)} - \hat{a}_M^{(i)} \hat{a}_M^{(j)}. \quad (2.1.32a)$$

Математическое ожидание  $\hat{b}_{ij}$  равно

$$\begin{aligned} \mathbf{E}[\hat{b}_{ij}] &= \frac{1}{M} \sum_{k=1}^M \mathbf{E}[z_k^{(i)} z_k^{(j)}] - \mathbf{E}[\hat{a}_M^{(i)} \hat{a}_M^{(j)}] = r_{ij} - \frac{1}{M^2} \sum_{k=1}^M \sum_{l=1}^M \mathbf{E}[x_k^{(i)} x_l^{(j)}] = \\ &= r_{ij} - \frac{1}{M^2} \sum_{k=1}^M \mathbf{E}[x_k^{(i)} x_k^{(j)}] - \sum_{k \neq l} \mathbf{E}[x_k^{(i)}] \mathbf{E}[x_l^{(j)}] = \\ &= r_{ij} - \frac{1}{M} r_{ij} - \frac{M^2 - M}{M^2} a^{(i)} a^{(j)} = \left(1 - \frac{1}{M}\right) (r_{ij} - a^{(i)} a^{(j)}) = \left(1 - \frac{1}{M}\right) b_{ij}. \end{aligned}$$

Отсюда следует, что  $\hat{b}_{ij}$  является смещённой оценкой корреляционной матрицы:

$$\mathbf{E}[\hat{\mathbf{K}}] = (1 - 1/M) \mathbf{K},$$

так что в качестве несмещённой оценки может быть выбрана следующая:

$$\hat{\mathbf{K}}' = \frac{M}{M-1} \hat{\mathbf{K}} = \frac{1}{M-1} \sum_{k=1}^M (\mathbf{z}_k - \hat{\mathbf{a}}_M)(\mathbf{z}_k - \hat{\mathbf{a}}_M)^T.$$

Оценка вектора средних, очевидно, является несмещённой и эффективной. Состоятельность рассматриваемых оценок вытекает из состоятельности выборочных точечных оценок математического ожидания и корреляционной (ковариационной) функции.

Выясним достаточность оценок вектора средних и корреляционной матрицы, т. е. проверим признак факторизации функции правдоподобия многомерного нормального распределения. Для этого используем известные свойства следа матрицы.

Если  $\mathbf{Y} = (y_{il})$ ,  $i, j = 1, \dots, N$  — произвольная прямоугольная матрица размером  $N \times N$ , то её следом<sup>1</sup>  $\text{Tr } \mathbf{Y}$  называют сумму диагональных элементов:

<sup>1</sup> Иногда след матрицы обозначают символом  $\text{Sp}$ .



$$\text{Tr } \mathbf{Y} = \sum_{i=1}^N y_{ij}.$$

След скалярной величины, очевидно, совпадает с ней самой. Для любых матриц  $\mathbf{Y}_1$  и  $\mathbf{Y}_2$  одинаковой размерности справедливы следующие соотношения:

$$\begin{aligned}\text{Tr}(\mathbf{Y}_1 + \mathbf{Y}_2) &= \text{Tr}(\mathbf{Y}_1) + \text{Tr}(\mathbf{Y}_2); \\ \text{Tr}(\mathbf{Y}_1 \cdot \mathbf{Y}_2) &= \text{Tr}(\mathbf{Y}_2 \cdot \mathbf{Y}_1).\end{aligned}$$

На основании этого имеем:

$$\begin{aligned}& \sum_{k=1}^M (\mathbf{z}_k - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) \equiv \text{Tr} \sum_{k=1}^M (\mathbf{z}_k - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) = \\ &= \sum_{k=1}^M \text{Tr} (\mathbf{z}_k - \mathbf{a})^T \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) = \sum_{k=1}^M \text{Tr} \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) (\mathbf{z}_k - \mathbf{a})^T = \\ &= \text{Tr} \sum_{k=1}^M \mathbf{K}^{-1} (\mathbf{z}_k - \mathbf{a}) (\mathbf{z}_k - \mathbf{a})^T = \text{Tr} \mathbf{K}^{-1} \sum_{k=1}^M (\mathbf{z}_k - \mathbf{a}) (\mathbf{z}_k - \mathbf{a})^T.\end{aligned}$$

Нетрудно видеть, что

$$\frac{1}{M} \sum_{k=1}^M (\mathbf{z}_k - \mathbf{a}) (\mathbf{z}_k - \mathbf{a})^T = \hat{\mathbf{K}} + (\hat{\mathbf{a}}_M - \mathbf{a}) (\hat{\mathbf{a}}_M - \mathbf{a})^T.$$

Тогда

$$\begin{aligned}\text{Tr} \mathbf{K}^{-1} \sum_{k=1}^M (\mathbf{z}_k - \mathbf{a}) (\mathbf{z}_k - \mathbf{a})^T &= \text{Tr} \left( M \mathbf{K}^{-1} \hat{\mathbf{K}} \right) + \text{Tr} \left( M \mathbf{K}^{-1} (\hat{\mathbf{a}}_M - \mathbf{a}) (\hat{\mathbf{a}}_M - \mathbf{a})^T \right) = \\ &= M \text{Tr} \left( \mathbf{K}^{-1} \hat{\mathbf{K}} \right) + M \text{Tr} \left( (\hat{\mathbf{a}}_M - \mathbf{a})^T \mathbf{K}^{-1} (\hat{\mathbf{a}}_M - \mathbf{a}) \right).\end{aligned}$$

Теперь, подставляя это соотношение под экспоненту в выражении для многомерного нормального распределения, в (2.1.30) имеем:

$$\begin{aligned}w(\mathbf{z} / \mathbf{a}, \mathbf{K}) &= (2\pi)^{-Mn/2} (\det \mathbf{K})^{-M/2} \times \\ &\times \exp \left[ -\frac{1}{2} M \text{Tr} \left( \mathbf{K}^{-1} \hat{\mathbf{K}} \right) \right] \exp \left[ -\frac{1}{2} M \text{Tr} \left( (\hat{\mathbf{a}}_M - \mathbf{a})^T \mathbf{K}^{-1} (\hat{\mathbf{a}}_M - \mathbf{a}) \right) \right].\end{aligned}\quad (2.1.33)$$

Видно, что признак факторизации функции правдоподобия выполняется, и можно утверждать, что  $\hat{\mathbf{a}}_M$  и  $\hat{\mathbf{K}}$  являются совместно достаточными оценками вектора средних и корреляционной матрицы многомерного нормального распределения.

## 2.2. МЕТОДЫ ПОЛУЧЕНИЯ ОЦЕНОК

В предыдущем разделе изучались общие свойства оценок. Рассмотрим теперь различные методы их получения.

### 2.2.1. Метод моментов

Исторически первым методом получения точечных оценок следует считать предложенный К. Пирсоном *метод моментов*. Кратко изложим его суть.

Пусть  $\mathbf{z} = (z^{(1)}, \dots, z^{(M)})$  — выборка из распределения  $w_\zeta(z, \eta)$ , содержащего один или  $n > 1$  неизвестных параметров. Будем считать, что распределение случайной величины  $\zeta$  достаточно “традиционное”, так что у него существуют<sup>1</sup>  $n$  первых моментов  $m_k(\eta) = E[\zeta^k]$ ,  $k = 1, \dots, n$ , которые также зависят от неизвестных параметров  $\eta$ .

Рассмотрим соответствующие выборочные моменты

$$m_k^* = \frac{1}{M} \sum_{i=1}^M (z^{(i)})^k, \quad k = 1, \dots, n \quad (2.2.1)$$

и составим систему уравнений относительно параметров  $\eta$ , приравняв теоретические и выборочные моменты:

$$m_k(\eta_1, \dots, \eta_n) = m_k^*, \quad k = 1, \dots, n. \quad (2.2.2)$$

Предположим, что система уравнений (2.2.2) однозначно разрешима и обозначим её решения

$$\hat{\eta}_k(\mathbf{z}) = \varphi_k(m_1^*, \dots, m_n^*), \quad k = 1, \dots, n. \quad (2.2.3)$$

Тогда  $\hat{\eta}_k$  и будут оценками неизвестных параметров распределения, полученные по методу моментов.

Если функции  $\varphi_k$  непрерывны, то, поскольку выборочные моменты являются состоятельными статистиками, также состоятельными будут и оценки (2.2.3).

Например, пусть случайная величина  $\zeta$  распределена равномерно на интервале  $[a; b]$ . В данном случае неизвестны два параметра распределения — границы  $a$  и  $b$  интервала равномерности, и для их нахождения необходимо использовать два теоретических момента распределения:

---

<sup>1</sup> Отсюда следует, что метод моментов неприменим к тем распределениям, у которых не существуют некоторые первые моменты, например, к распределению Коши.

$$m_1 \equiv \mathbf{E}[\zeta] = (a + b) / 2;$$

$$m_2 \equiv \mathbf{E}[\zeta^2] = (a^2 + ab + b^2) / 3$$

и два статистических момента:

$$m_1^* = \frac{1}{M} \sum_{i=1}^M z^{(i)};$$

$$m_2^* = \frac{1}{M} \sum_{i=1}^M (z^{(i)})^2.$$

Система уравнений (2.2.2) имеет следующий вид:

$$(a + b) / 2 = m_1^*;$$

$$(a^2 + ab + b^2) / 3 = m_2^*,$$

и её решением является

$$\hat{a} = m_1^* - \sqrt{3} \sqrt{m_2^* - m_1^{*2}};$$

$$\hat{b} = m_1^* + \sqrt{3} \sqrt{m_2^* - m_1^{*2}}.$$

В терминах выборки  $m_1^*$  — это выборочное среднее  $\bar{z}$ , а  $\sqrt{m_2^* - m_1^{*2}}$  — выборочное стандартное отклонение  $s$ , так что решение можно записать как

$$\hat{a} = \bar{z} - \sqrt{3}s;$$

$$\hat{b} = \bar{z} + \sqrt{3}s.$$

Рассмотрим оценивание методом моментов параметров гауссовского распределения  $a = \mathbf{E}[\zeta]$  и  $\sigma^2 = \mathbf{D}[\zeta]$ . В данном случае система уравнений (2.2.2) имеет очень простой вид:

$$a = m_1^* \equiv \bar{z};$$

$$\sigma^2 = m_2^* - (m_1^*)^2 \equiv s^2$$

Видно, оценкой математического ожидания является выборочное среднее, а оценкой дисперсии — выборочная дисперсия.

Говоря в целом о методе моментов, можно отметить, что система уравнений (2.2.2) для некоторых широко используемых распределений оказывается достаточно простой, и получение оценок не составляет особых сложностей. Однако эффективность таких оценок оказывается, как правило, не очень высока; в выборках достаточно большого объёма они имеют сравнительно большую дисперсию. В этой связи оценки, получен-

ные методом моментов, если и используют, то — в качестве первого приближения, опираясь на которое можно затем находить более эффективные последующие приближения. Выборочные моменты больших порядков использовать нецелесообразно, так как качество оценок при этом значительно снижается.

### 2.2.2. Оценки по максимуму апостериорной вероятности

Метод оценки по максимуму апостериорной вероятности основан на предположении, что оцениваемый параметр  $\eta$  является случайной величиной с известным априорным распределением  $w_\eta(y)$ .

Если  $w(\mathbf{z} / \eta)$  — функция правдоподобия выборки, то можно найти безусловную (усреднённую по априорному распределению параметра) функцию правдоподобия:

$$w(\mathbf{z}) = \int_{-\infty}^{\infty} w(\mathbf{z} / \eta) w_\eta(y) dy.$$

Далее,

$$w(\mathbf{z}, \eta) = w(\mathbf{z} / \eta) w_\eta(y)$$

является совместным распределением выборки и оцениваемого параметра, тогда по формуле Байеса

$$w(\eta / \mathbf{z}) = \frac{w(\mathbf{z} / \eta) w_\eta(y)}{w(\mathbf{z})} = \frac{w(\mathbf{z} / \eta) w_\eta(y)}{\int_{-\infty}^{\infty} w(\mathbf{z} / \eta) w_\eta(y) dy} \quad (2.2.4)$$

является апостериорным (послеопытным) распределением параметра  $\eta$ , дающим полные сведения об интересующем параметре на основе анализа выборки  $\mathbf{z}$ .

Если апостериорное распределение имеет достаточно “хороший” вид: унимодальность, малую асимметрию и др., то представляется естественным выбрать такую оценку  $\hat{\eta}$  параметра  $\eta$ , которая максимизирует апостериорное распределение. Необходимым условием такой оценки является уравнение

$$\frac{\partial w(\eta / \mathbf{z})}{\partial \eta} = 0. \quad (2.2.5)$$

Строго говоря, дополнительно к необходимому условию (2.2.5) следует удостовериться в выполнении достаточного условия — знакопостоян-

ства второй производной. Однако в статистических задачах проверки гипотез и оценки параметров наличие экстремума, как правило, вытекает из существа изучаемых вопросов

Часто вместо уравнения (2.2.5) используется его модификация, при которой производная вычисляется не от самой  $w(\mathbf{z})$ , а от  $f[w(\mathbf{z})]$ , где  $f$  — монотонная функция:

$$\frac{\partial f[w(\eta / \mathbf{z})]}{\partial \eta} = 0. \quad (2.2.5a)$$

Корни уравнения (2.2.5a) являются корнями уравнения (2.2.5).

Для гауссовской выборки в качестве функции  $f$  используется логарифм, тогда уравнение для оценки по максимуму апостериорной вероятности (МАН-оценки) имеет следующий вид:

$$\frac{\partial \ln w(\eta / \mathbf{z})}{\partial \eta} = 0. \quad (2.2.6)$$

Из (2.2.4) видно, что стоящее в знаменателе безусловное распределение  $w(\mathbf{z})$  не зависит от оцениваемого параметра, поэтому эквивалентным условием МАН-оценки является уравнение

$$\frac{\partial f[w(\mathbf{z} / \eta)w_{\eta}(y)]}{\partial \eta} = 0. \quad (2.2.7)$$

В частности, для гауссовской выборки с независимыми элементами получаем уравнение

$$\frac{\partial \ln w_{\eta}(y)}{\partial \eta} + \sum_{k=1}^M \frac{\partial \ln w(z_k / \eta)}{\partial \eta} = 0. \quad (2.2.8)$$

Во многих практических задачах априорное распределение  $w_{\eta}(y)$  параметра  $\eta$  оказывается неизвестным. В этой ситуации для того, чтобы иметь возможность применить метод максимума апостериорной вероятности целесообразно задаться некоторым “подходящим” априорным распределением, выбираемым из эвристических соображений. Например, если ничего неизвестно о природе возникновения тех или иных значений  $\eta$ , то разумно предположить его равномерное распределение в определённом интервале  $[\eta_1; \eta_2]$ . Тогда апостериорное распределение  $w(\eta / \mathbf{z})$  оказывается пропорциональным функции правдоподобия выборки, и оценивание по максимуму апостериорной вероятности эквивалентно оцениванию по максимуму функции правдоподобия (см. разд. 2.2.3).

Если оцениванию подлежат несколько параметров  $\eta_1, \dots, \eta_L$ , то совместные МАВ-оценки определяются как решения системы уравнений

$$\frac{\partial f[w(\eta_1, \dots, \eta_L / \mathbf{z})]}{\partial \eta_j} = 0, j = 1, \dots, L. \quad (2.2.9)$$

Можно показать [3], что МАВ-оценки состоятельны и асимптотически эффективны.

### 2.2.3. Оценки максимального правдоподобия

Оценка максимального правдоподобия (МП-оценка) параметра  $\eta$  — это такая условная оценка  $\hat{\eta}$ , для которой функция правдоподобия выборки  $w(\mathbf{z} / \eta)$  имеет максимальное значение. Оценивание методом максимального правдоподобия получило распространение в 20-е годы XX века благодаря работам Р. Фишера (хотя, как обычно, основные идеи такого подхода были известны и раньше), и в настоящее время такое оценивание является одним из наиболее универсальных и эффективных методов.

Необходимым условием МП-оценки является выполнение равенства

$$\frac{\partial f[w(\mathbf{z} / \eta)]}{\partial \eta} = 0, \quad (2.2.10)$$

где  $f$  — монотонная функция, в качестве которой для гауссовской выборки используется логарифм, и для выборки с независимыми элементами

$$\sum_{k=1}^M \frac{\partial \ln[w(z_k / \eta)]}{\partial \eta} = 0. \quad (2.2.11)$$

Формально условие (2.2.10) является частным случаем необходимого условия (2.2.7) МАВ-оценки, когда априорное распределение параметра — равномерное. Однако подход к получению МП-оценок может рассматриваться как вполне самостоятельная стратегия.

Отметим, что в (2.2.10) фигурирует многомерная условная плотность непрерывного распределения. Однако таким же образом можно составить функцию правдоподобия и для дискретного распределения, образуя произведение условных дискретных вероятностей.

Уравнение (2.2.11) в общем случае является нелинейным или, даже, трансцендентным и может иметь несколько решений; в этом случае каждое решение даёт свою оценку максимального правдоподобия, и необхо-

димо из всех таких оценок выбрать максимальную, “абсолютную” оценку максимального правдоподобия.

Если решение уравнения (2.2.11) не удаётся получить простыми приёмами, то возможно его решение методом последовательных приближений на основе степенного разложения логарифма функции правдоподобия по оцениваемому параметру в окрестности некоторого значения  $\eta_0$ :

$$\begin{aligned} \ln w(\mathbf{z} / \eta) = & \ln w(\mathbf{z} / \eta_0) + (\eta - \eta_0) \left( \frac{\partial \ln w(\mathbf{z} / \eta)}{\partial \eta} \right)_{\eta=\eta_0} + \\ & + \frac{1}{2} (\eta - \eta_0)^2 \left( \frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \right)_{\eta=\eta_0} + \dots \end{aligned}$$

Рассмотрим вначале параболическую аппроксимацию (удерживание трёх членов в степенном разложении) функции правдоподобия и, кроме того, заменим вторую производную её средним значением:

$$\frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \approx \mathbf{E} \left[ \frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \right].$$

Нетрудно видеть, что среднее значение второй производной с точностью до знака совпадает с фишеровской информацией. Действительно:

$$\begin{aligned} \mathbf{E} \left[ \frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \right] &= \int_{Z^M} \frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \ln w(\mathbf{z} / \eta) d\mathbf{z} = \\ &= \int_{Z^M} \left[ \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} - \frac{1}{w(\mathbf{z} / \eta)} \left( \frac{\partial w(\mathbf{z} / \eta)}{\partial \eta} \right)^2 \right] d\mathbf{z} = \\ &= \int_{Z^M} \left[ \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} - \left( \frac{\partial \ln w(\mathbf{z} / \eta)}{\partial \eta} \right)^2 w(\mathbf{z} / \eta) \right] d\mathbf{z}. \end{aligned}$$

Так как

$$\int_{Z^M} \frac{\partial^2 w(\mathbf{z} / \eta)}{\partial \eta^2} d\mathbf{z} = \frac{\partial^2}{\partial \eta^2} \int_{Z^M} w(\mathbf{z} / \eta) d\mathbf{z} = \frac{\partial^2}{\partial \eta^2} 1 = 0,$$

получаем

$$\mathbf{E} \left[ \frac{\partial^2 \ln w(\mathbf{z} / \eta)}{\partial \eta^2} \right] = - \int_{Z^M} \left( \frac{\partial w(\mathbf{z} / \eta)}{\partial \eta} \right)^2 w(\mathbf{z} / \eta) d\mathbf{z} = -I_M(\eta).$$

Таким образом, в условиях параболической аппроксимации и указанной замены имеем следующее приближённое уравнение максимального правдоподобия:

$$\left( \frac{\partial \ln w(\mathbf{z} / \eta)}{\partial \eta} \right)_{\eta=\eta_0} - (\eta - \eta_0) I_M(\eta_0) = 0, \quad (2.2.12)$$

откуда

$$\hat{\eta}_1 = \eta_0 + \frac{1}{I_M(\eta_0)} \left( \frac{\partial \ln w(\mathbf{z} / \eta)}{\partial \eta} \right)_{\eta=\eta_0}. \quad (2.2.13)$$

Уравнение (2.2.12) и его решение (2.2.13) будем трактовать как первое приближение. Тогда получение второго приближения  $\hat{\eta}_2$  связано с подстановкой в правую часть (2.2.13) вместо значения  $\eta_0$  величины  $\hat{\eta}_1$ :

$$\hat{\eta}_2 = \hat{\eta}_1 + \frac{1}{I_M(\hat{\eta}_1)} \left( \frac{\partial \ln w(\mathbf{z} / \eta)}{\partial \eta} \right)_{\eta=\hat{\eta}_1}.$$

Аналогично получают дальнейшие приближения: если найдено решение  $\hat{\eta}_k$  на  $k$ -м шаге, то его дальнейшее уточнение  $\hat{\eta}_{k+1}$  вычисляется из уравнения

$$\hat{\eta}_{k+1} = \hat{\eta}_k + \frac{1}{I_M(\hat{\eta}_k)} \left( \frac{\partial \ln w(\mathbf{z} / \eta)}{\partial \eta} \right)_{\eta=\hat{\eta}_k}, \quad k = 0, 1, \dots \quad (2.2.14)$$

Процесс вычисления осуществляют рекуррентно до достижения желаемой точности, например, по критерию

$$|\hat{\eta}_{k+1} - \hat{\eta}_k| < \varepsilon,$$

где  $\varepsilon$  — заданная, достаточно малая величина.

Важным в этом методе является выбор начальной точки  $\eta_0$ , в качестве которой обычно берут какую-нибудь легко вычисляемую состоятельную оценку, например, по методу моментов.

Одним из полезных свойств МП-оценок является инвариантность относительно преобразований параметров. Это означает, что если  $\hat{\eta}_1, \dots, \hat{\eta}_L$  — МП-оценки, то при взаимно-однозначном отображении

$$\hat{\mu}_1 = \varphi_1(\eta_1, \dots, \eta_L);$$

...

$$\hat{\mu}_L = \varphi_L(\eta_1, \dots, \eta_L);$$



значения  $\hat{\mu}_1, \dots, \hat{\mu}_L$  также будут МП-оценками. Это означает, что для каждой конкретной задачи можно сначала выбрать “удачную”, например, с точки зрения простоты уравнения правдоподобия, параметризацию, а затем получать МП-оценки с помощью обратных преобразований.

Рассмотрим некоторые примеры вычисления МП-оценок.

### Гауссовское распределение

Получим МП-оценки математического ожидания  $a$  и дисперсии  $\sigma^2$  гауссовского распределения  $N(a, \sigma^2)$  с функцией правдоподобия (2.1.24) и логарифмом функции правдоподобия

$$\ln w(\mathbf{z} / a, \sigma^2) - \frac{M}{2} \ln(2\pi) - \frac{M}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^M (z_k - a)^2.$$

Тогда уравнения правдоподобия (2.2.11) имеют следующий вид:

$$\frac{\partial \ln w(\mathbf{z} / a, \sigma^2)}{\partial a} = \frac{1}{\sigma^2} \sum_{k=1}^M (z_k - a) = 0;$$

$$\frac{\partial \ln w(\mathbf{z} / a, \sigma^2)}{\partial \sigma^2} = -\frac{M}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{k=1}^M (z_k - a)^2 = 0.$$

Решением этих уравнений является выборочное среднее

$$\hat{a} = \bar{z} = \frac{1}{M} \sum_{k=1}^M z_k$$

и выборочная дисперсия

$$\hat{\sigma}^2 = s^2 = \frac{1}{M} \sum_{k=1}^M (z_k - \bar{z})^2.$$

Таким образом, для гауссовского распределения МП-оценки совпадают с оценками, полученными по методу моментов.

Как следует из соотношений (2.1.24)–(2.1.26), выборочные среднее и дисперсия не являются совместно эффективными оценками.

### Равномерное распределение

Пусть случайная величина  $\zeta$  распределена равномерно на интервале  $[0; c]$ :

$$w_\zeta(z / c) = (1 / c) \text{rect}(z, 0, c),$$

где для удобства записи введена функция прямоугольного окна:

$$\text{rect}(x, a, b) = \begin{cases} 1, & x \in [a; b]; \\ 0, & x \notin [a; b]. \end{cases}$$

Тогда функция правдоподобия выборки равна

$$w_{\zeta}(\mathbf{z} / c) = \frac{1}{c^M} \prod_{k=1}^M \text{rect}(z_k, 0, c).$$

Введём ещё одну “служебную” функцию, часто используемую в теории вероятностей и её приложениях — *индикатор*  $I(A)$  произвольного события  $A$ , определяемый как  $I(A) = 1$  при наступлении  $A$  и  $I(A) = 0$  в противном случае. В этих обозначениях можно записать

$$w_{\zeta}(\mathbf{z} / c) = (1 / c^M) I(z_{\max} \leq c) I(z_{\min} \geq 0), \quad (2.2.15)$$

где  $z_{\max}$  и  $z_{\min}$  — достаточные статистики, являющиеся, соответственно, максимальным и минимальным значениями выборки.

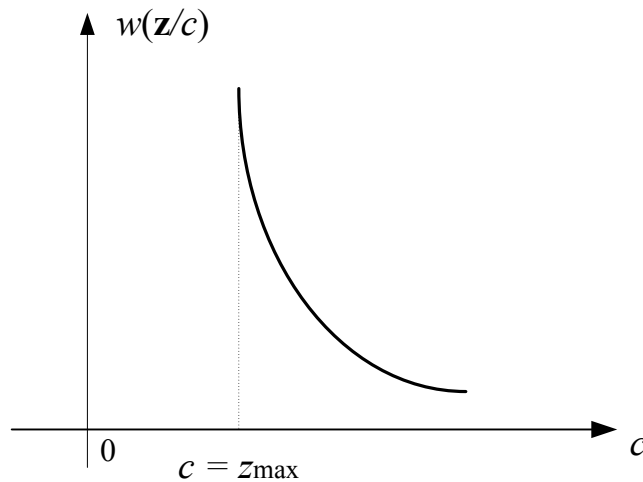


Рис. 2.2. Функция правдоподобия равномерного распределения

На рис. 2.2 показана зависимость (2.2.15) от параметра  $c$  — монотонное убывание при  $c \geq z_{\max}$ ; в точке  $c = z_{\max}$  функция правдоподобия имеет максимальное значение, возникающее скачкообразно, поэтому производной  $w_{\zeta}(\mathbf{z} / c)$  по  $c$  в этой точке не существует. Таким образом, в данном случае МП-оценка  $\hat{c} = z_{\max}$  не является решением уравнением правдоподобия равномерного распределения. Такая ситуация является характерной, когда выборочное пространство зависит от оцениваемого параметра (в данном случае выборочное пространство —  $M$ -мерный куб со стороной  $c$ ).

Пусть теперь равномерное распределение имеет единичный интервал и границы, определяемые неизвестным параметром  $c$ :

$$w_{\zeta}(z / c) = \text{rect}(z, c, c + 1).$$

Тогда функцию правдоподобия, равную

$$w_{\zeta}(\mathbf{z} / c) = I(c \leq z_{\min} \leq z_{\max} \leq c + 1), \quad (2.2.16)$$

максимизирует любое число в диапазоне от  $z_{\max} - 1$  до  $z_{\min}$ . Например, в качестве решения можно выбрать среднее значение

$$\hat{c} = (z_{\max} - 1 + z_{\min}) / 2.$$

### Двумерное гауссовское распределение

Рассмотрим выборку

$$\mathbf{z} = \left( (z_1^{(1)}, z_2^{(1)}), (z_1^{(2)}, z_2^{(2)}), \dots, (z_1^{(M)}, z_2^{(M)}) \right),$$

взятую из двумерного гауссовского распределения с нулевыми средними и корреляционной матрицей вида

$$\mathbf{K} = \begin{pmatrix} \sigma^2 & r\sigma^2 \\ r\sigma^2 & \sigma^2 \end{pmatrix},$$

в которой дисперсия  $\sigma^2$  и коэффициент корреляции  $r$  являются неизвестными параметрами.

Уравнение правдоподобия для двумерной плотности

$$w(z_1, z_2 / \sigma^2, r) = \frac{1}{2\pi\sqrt{1-r^2}\sigma^2} \exp \left[ -\frac{z_1^2 - 2rz_1z_2 + z_2^2}{2\sigma^2(1-r^2)} \right]$$

будет иметь достаточно сложный вид, поэтому, пользуясь свойством инвариантности МП-оценок, перейдём к новым параметрам:

$$\begin{aligned} \mu_1(\sigma^2, r) &= -(2\sigma^2(1-r^2))^{-1}; \\ \mu_2(\sigma^2, r) &= r(\sigma^2(1-r^2))^{-1}, \end{aligned} \quad (2.2.17)$$

для которых, с учётом того, что

$$\sigma^2 = -\frac{2\mu_1}{4\mu_1^2 - \mu_2^2}, \quad r = -\frac{\mu_2}{2\mu_1} \quad (2.2.18)$$

и, следовательно,

$$\sigma^2 \sqrt{1-r^2} = -(4\mu_1^2 - \mu_2^2)^{-1/2},$$

двумерная плотность равна

$$w(z_1, z_2 / \mu_1, \mu_2) = -\frac{\sqrt{4\mu_1^2 - \mu_2^2}}{2\pi} \exp\left[\mu_1(z_1^2 + z_2^2) + \mu_2 z_1 z_2\right] =$$

$$= -\frac{1}{2\pi} \exp\left[\mu_1(z_1^2 + z_2^2) + \mu_2 z_1 z_2 + \frac{1}{2} \ln(4\mu_1^2 - \mu_2^2)\right],$$

и логарифм функции правдоподобия имеет следующий вид:

$$\ln(\mathbf{z}_1, \mathbf{z}_2 / \mu_1, \mu_2) = -M \ln 2\pi + \sum_{k=1}^M \left[\mu_1(z_{1k}^2 + z_{2k}^2) + \mu_2 z_{1k} z_{2k}\right] + \frac{M}{2} \ln(4\mu_1^2 - \mu_2^2).$$

Отсюда уравнения правдоподобия:

$$\frac{\partial \ln(\mathbf{z}_1, \mathbf{z}_2 / \mu_1, \mu_2)}{\partial \mu_1} = \sum_{k=1}^M (z_{1k}^2 + z_{2k}^2) + \frac{4M\mu_1}{4\mu_1^2 - \mu_2^2} = 0;$$

$$\frac{\partial \ln(\mathbf{z}_1, \mathbf{z}_2 / \mu_1, \mu_2)}{\partial \mu_2} = \sum_{k=1}^M z_{1k} z_{2k} + \frac{M\mu_2}{4\mu_1^2 - \mu_2^2} = 0,$$

решения которых есть

$$\frac{1}{M} \sum_{k=1}^M (z_{1k}^2 + z_{2k}^2) = -\frac{4M\mu_1}{4\mu_1^2 - \mu_2^2};$$

$$\frac{1}{M} \sum_{k=1}^M z_{1k} z_{2k} = -\frac{M\mu_2}{4\mu_1^2 - \mu_2^2}.$$

Теперь, чтобы получить МП-оценки, необходимо выполнить обратные преобразования. На основании (2.2.17) и (2.2.18) находим:

$$\hat{\sigma}^2 = \frac{1}{2M} \sum_{k=1}^M (z_{1k}^2 + z_{2k}^2); \quad \hat{r} = \frac{2 \sum_{k=1}^M z_{1k} z_{2k}}{\sum_{k=1}^M (z_{1k}^2 + z_{2k}^2)}. \quad (2.2.19)$$

### Распределение Коши

Пусть  $M$ -элементная выборка  $\mathbf{z}$  соответствует случайной величине  $\zeta$ , подчиняющейся распределению Коши:

$$w_\zeta(z / \eta) = \frac{1}{\pi} \cdot \frac{1}{1 + (z - \eta)^2},$$

где  $\eta$  — неизвестный параметр, имеющий смысл коэффициента сдвига относительно центра.

Найдём, прежде всего, фишеровскую информацию, содержащуюся в выборке. Поскольку

$$\ln w_{\zeta}(\mathbf{z} / \eta) = -M \ln \pi - \sum_{k=1}^M \ln [1 + (z_k - \eta)^2],$$

на основании (2.1.11) имеем:

$$\frac{\partial \ln(\mathbf{z} / \eta)}{\partial \eta} = 2 \sum_{k=1}^M \frac{z_k - \eta}{1 + (z_k - \eta)^2};$$

$$\frac{\partial^2 \ln(\mathbf{z} / \eta)}{\partial \eta^2} = 2 \sum_{k=1}^M \frac{(z_k - \eta)^2 - 1}{[1 + (z_k - \eta)^2]^2}$$

и

$$I_M(\eta) = -\mathbf{E} \left[ \frac{\partial^2 \ln(\mathbf{z} / \eta)}{\partial \eta^2} \right] = -\frac{1}{\pi} \int_{-\infty}^{\infty} \left[ \frac{\partial^2 \ln(\mathbf{z} / \eta)}{\partial \eta^2} \right] \frac{dz_k}{1 + (z_k - \eta)^2} =$$

$$= -\frac{2}{\pi} \sum_{k=1}^M \int_{-\infty}^{\infty} \frac{(z_k - \eta)^2 - 1}{[1 + (z_k - \eta)^2]^3} dz_k = -\frac{2}{\pi} \sum_{k=1}^M \int_{-\infty}^{\infty} \frac{x_k^2 - 1}{(1 + x_k^2)^2} dx_k = \frac{M}{2}.$$

Решение уравнения правдоподобия

$$\frac{\partial \ln(\mathbf{z} / \eta)}{\partial \eta} = 2 \sum_{k=1}^M \frac{z_k - \eta}{1 + (z_k - \eta)^2} = 0$$

в замкнутом виде не представляется возможным, поэтому одним из способов решения данной задачи является метод последовательных приближений, основанный на разложении логарифма функции правдоподобия в степенной ряд. Согласно (2.2.14),

$$\hat{\eta}_{i+1} = \hat{\eta}_i + \frac{4}{M} \sum_{k=1}^M \frac{z_k - \hat{\eta}_i}{1 + (z_k - \hat{\eta}_i)^2}, \quad i = 0, 1, \dots \quad (2.2.20)$$

Это рекуррентное соотношение легко вычисляется при задании начального условия  $\hat{\eta}_0$ , в качестве которого можно, взять, например, выборочную медиану, являющуюся состоятельной оценкой теоретической медианы.

Приведённые примеры МП-оценок касались непрерывных распределений. Рассмотрим получение оценок максимального правдоподобия для некоторых дискретных распределений.

### Пуассоновское распределение

Оценим параметр  $\lambda$  в пуассоновском распределении

$$p(z = i / \lambda) = \frac{\exp(-\lambda) \lambda^z}{z!}, \quad i = 0, 1, \dots$$

Функция правдоподобия выборки равна

$$p(\mathbf{z} / \lambda) = \exp(-M\lambda) \frac{\lambda^{z_1+z_2+\dots+z_M}}{z_1!z_2!\dots z_M!},$$

тогда уравнение правдоподобия есть

$$\frac{\partial p(\mathbf{z} / \lambda)}{\partial \lambda} = -M + \frac{1}{\lambda} \sum_{k=1}^M z_k = 0,$$

из которого следует, что МП-оценкой параметра пуассоновского распределения является выборочное среднее:

$$\hat{\lambda} = \bar{z} = \frac{1}{M} \sum_{k=1}^M z_k.$$

На основании (2.1.12б) фишеровская информация

$$I(\lambda) = -\mathbf{E} \left[ \partial^2 p \ln(z / \lambda) / \partial \lambda^2 \right] = \frac{1}{\lambda^2} \sum_{i=0}^{\infty} i p(z = i / \lambda) = \frac{1}{\lambda},$$

так как фигурирующая бесконечная сумма представляет собой математическое ожидание пуассоновского распределения, равное  $\lambda$ . Отсюда

$$I_M(\eta) = \frac{M}{\lambda^2} \lambda = \frac{M}{\lambda}.$$

Оценим дисперсию полученной оценки на основании нижней границы в неравенстве Рао—Крамера. Согласно (2.1.13), имеем

$$\mathbf{D}[\hat{\lambda}] \geq \hat{\lambda} / M.$$

### **Последовательность независимых испытаний (схема Бернулли)**

Пусть для случайной величины  $\zeta$ , принимающей значения 1 и 0 с соответствующими вероятностями  $q$  и  $1 - q$  в результате  $M$  испытаний исход  $z = 1$  осуществился  $K$  раз, а исход  $z = 0$  осуществился  $M - K$  раз. Тогда функция правдоподобия определяется биномиальным распределением

$$p(\mathbf{z} / q) = C_M^K q^K (1 - q)^{M-K},$$

и уравнение правдоподобия имеет вид

$$\frac{\partial \ln p(\mathbf{z} / q)}{\partial q} = \frac{K}{q} - \frac{M - K}{1 - q} = 0.$$

Отсюда МП-оценкой вероятности элементарного исхода является относительная частота появления исхода  $z = 1$ :

$$\hat{q} = K / M.$$

Оценим дисперсию такой оценки. С учётом того, что математическое ожидание биномиального распределения равно

$$\sum_{k=0}^M k C_M^K q^K (1-q)^{M-K} = Mq,$$

фишеровская информация составляет

$$I_M(q) = \frac{M}{q(1-q)}.$$

Отсюда

$$\mathbf{D}[\hat{q}] \geq \frac{\hat{q}(1-\hat{q})}{M}.$$

#### 2.2.4. Байесовские оценки

Рассмотренные выше методы получения статистических оценок, связанные с вычислением апостериорной вероятности или функции правдоподобия, можно рассматривать как частные случаи более общего, *байесовского метода*, в основе которого лежит использование *функции потерь*.

Пусть  $\eta$  — подлежащий оценке скалярный параметр, являющийся случайной величиной с известной априорной плотностью распределения<sup>1</sup>  $w(\eta)$ , а  $\hat{\eta} = \hat{\eta}(\mathbf{z})$  — некоторая его точечная оценка, полученная из выборки  $\mathbf{z}$ . То, что априорное распределение  $w(\eta)$  предполагается известным можно трактовать следующим образом: до начала “основного” эксперимента, связанного с анализом выборки  $\mathbf{z}$ , проводится “вспомогательный” эксперимент, направленный на изучение вероятностных свойств параметра  $\eta$ , и по его итогам формируется закон распределения  $w(\eta)$ .

Байесовский метод оценивания является аналогом байесовской стратегии в задачах теории проверки статистических гипотез, изложенной в первой главе. Значение параметра играет роль номера гипотезы, однако теперь  $\eta$  является, вообще говоря, непрерывной величиной, принимающей

---

<sup>1</sup> В предыдущих разделах по возможности сохранялся общепринятый подход к записи законов распределений: греческими символами обозначать случайные величины, а их значения — соответствующими латинскими символами, например,  $w_\eta(y)$ . В данном разделе в силу специфики используемых выражений одним и тем же греческим символом обозначается и сама случайная величина, и её значение.

значения из некоторого континуального множества  $Y$ . Следовательно, непрерывной также является *ошибка оценивания*  $\Delta\eta$ , определяемая<sup>1</sup> как

$$\Delta\eta = \eta - \hat{\eta}.$$

Аналогично тому, как в задачах проверки гипотез вводились частные риски, обусловленные выбором конкретной гипотезы, в задачах оценивания вводится неотрицательная *функция потерь*

$$c(\Delta\eta) = c(\eta - \hat{\eta}) \quad (2.2.21)$$

функциональный вид которой выбирается из эвристических соображений; желательно иметь минимум функции потерь при  $\eta = \hat{\eta}$  и её возрастание при отклонении оцениваемого параметра от его точечной оценки.

В приложениях наиболее распространены следующие простые функции потерь, содержащие в явном виде ошибку оценивания.

- Дельта-образная функция потерь<sup>2</sup> (рис. 2.3, а):

$$c(\Delta\eta) = a - \delta(\eta - \hat{\eta}), \quad a > 0; \quad (2.2.22)$$

- модульная функция потерь (рис. 2.3, б):

$$c(\Delta\eta) = |\eta - \hat{\eta}|; \quad (2.2.23)$$

- прямоугольная функция потерь (рис. 2.3, в):

$$c(\Delta\eta) = \begin{cases} 1, & |\eta - \hat{\eta}| \leq \varepsilon; \\ 0, & |\eta - \hat{\eta}| > \varepsilon; \end{cases} \quad (2.2.24)$$

- квадратичная функция потерь (рис. 2.3, г):

$$c(\Delta\eta) = (\eta - \hat{\eta})^2; \quad (2.2.25)$$

Все приведённые функции потерь являются чётными неубывающими функциями ошибки оценивания. Возможны и другие, более сложные формы  $c(\Delta\eta)$ , например, трапецеидальная, колоколообразная или многоинтервальная ломаная линия.

---

<sup>1</sup> Во многих работах, особенно, в области классической математической статистики ошибка оценивания вводится с обратным знаком, т. е.  $\Delta\eta = \hat{\eta} - \eta$ . В большинстве случаев это оказывается несущественным в силу использования квадратичных мер, однако всё же это может привести к разночтению в получаемых соотношениях.

<sup>2</sup> Для этой функции в точке  $\hat{\eta} = \eta$  формально теряется свойство неотрицательности.



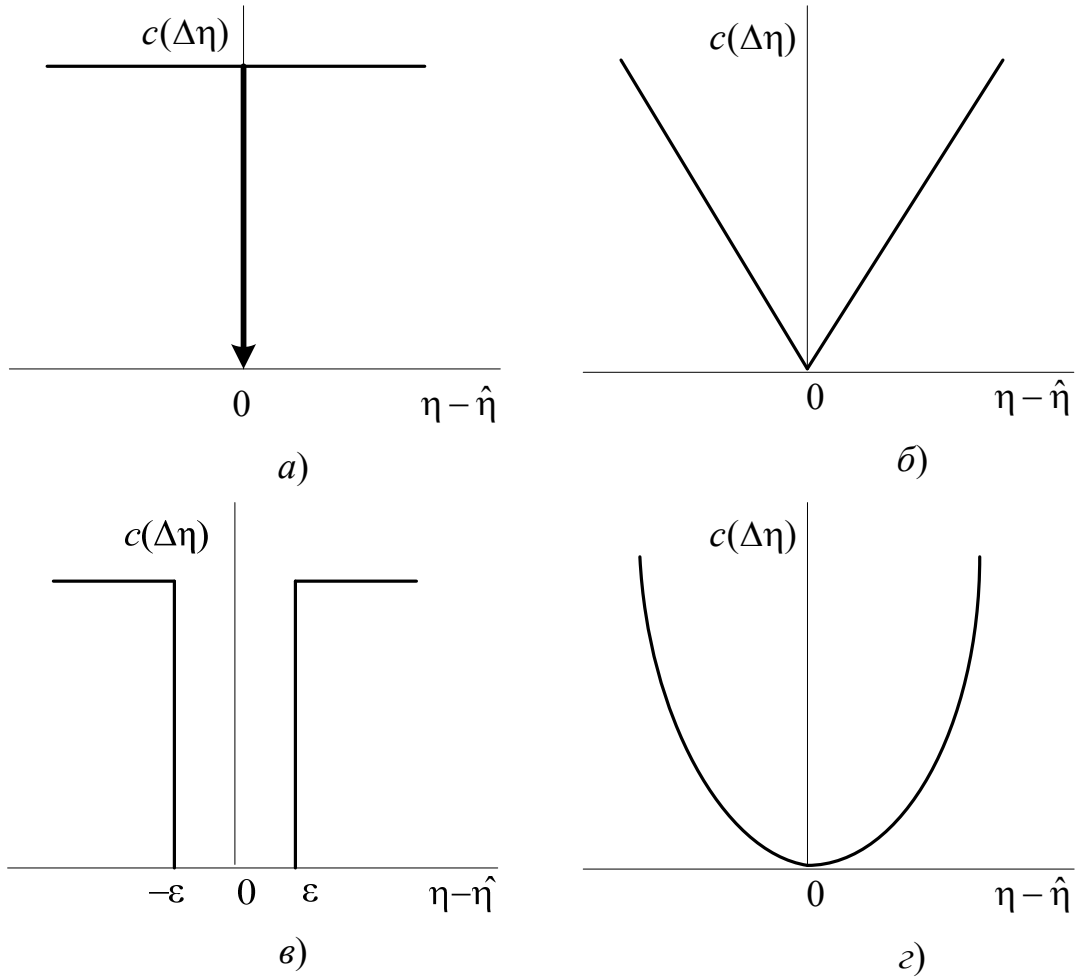


Рис. 2.3. Различные формы функции потерь

Как известно (разд. 1.1, 1.4), для получения алгоритмов принятия решения при решении задач проверки статистических гипотез в рамках общей байесовской стратегии производится суммирование по всем элементам ценовой матрицы и интегрирование по подмножествам  $S_i$  ( $i = 1, \dots, m$ ), связанным с принятием определённой гипотезы. В рассматриваемом случае деление выборочного пространства на дискретное количество подмножеств теряет смысл, и средний (байесовский) риск  $R$ , соответствующий всем возможным значениям параметра  $\theta$ , конструируется на основе совместной плотности  $w(\eta, \mathbf{z})$  распределения значений параметра  $\eta$  и выборки  $\mathbf{z}$  посредством совместного интегрирования по всему выборочному пространству  $Z^M$  и по всему множеству  $Y$  значений параметра  $\eta$ :

$$R = R(\hat{\eta}) = \int_Y \int_{Z^M} c(\Delta\eta) w(\eta, \mathbf{z}) d\eta d\mathbf{z}. \quad (2.2.26)$$

Итак, байесовский риск является усреднённым интегральным значением некоторой функции от ошибки оценивания, и необходимо обеспечить минимизацию (2.2.26) путём надлежащего выбора оценки  $\hat{\eta}$ . Такую оценку будем называть *оптимальной байесовской оценкой*.

Оценка, являющаяся оптимальной для какой-то одной формы функции потерь, может не оказаться оптимальной для других форм. К счастью, для большого класса задач оценивания оценки, полученные для выбранной формы функции, сохраняют свойство оптимальности для большинства используемых простых функций потерь.

Записывая совместное распределение  $w(\eta, \mathbf{z})$  в виде

$$w(\eta, \mathbf{z}) = w(\eta / \mathbf{z}) w(\mathbf{z}),$$

где  $w(\eta / \mathbf{z})$  — апостериорная плотность вероятности параметра  $\eta$  для данной выборки  $\mathbf{z}$ , средний риск можно представить в виде среднего (математического ожидания)

$$R = \mathbf{E}[\rho(\hat{\eta} / \mathbf{z})] = \int_{Z^M} \rho(\hat{\eta} / \mathbf{z}) w(\mathbf{z}) d\mathbf{z} \quad (2.2.27)$$

по выборке  $\mathbf{z}$  от некоторой величины  $\rho(\hat{\eta} / \mathbf{z})$ , представляющей собой усреднённую по апостериорному распределению  $w(\eta / \mathbf{z})$  параметра  $\eta$  функцию потерь, называемую *апостериорным (байесовским) риском*:

$$\rho(\hat{\eta} / \mathbf{z}) = \int_Y c(\Delta\eta) w(\eta / \mathbf{z}) d\eta. \quad (2.2.28)$$

Тогда байесовская оценка находится из условия минимума функционала (2.2.27), зависящего от распределения выборки  $w(\mathbf{z})$  и апостериорного риска  $\rho(\hat{\eta} / \mathbf{z})$ . Из неотрицательности  $w(\mathbf{z})$  следует, что для минимизации среднего риска достаточно найти минимальное значение  $\rho(\hat{\eta} / \mathbf{z})$ , т. е. приравнять к нулю частную производную апостериорного риска по искомому параметру:

$$\frac{\partial \rho(\hat{\eta} / \mathbf{z})}{\partial \hat{\eta}} = 0. \quad (2.2.29)$$

Получим выражение, определяющее байесовскую оценку для наиболее часто используемой — квадратичной функции потерь. Подставляя в (2.2.29) выражение (2.2.24), получаем апостериорный риск

$$\rho(\hat{\eta} / \mathbf{z}) = \int_Y (\eta - \hat{\eta})^2 w(\eta / \mathbf{z}) d\eta = \int_Y \eta^2 w(\eta / \mathbf{z}) d\eta - 2\hat{\eta} \int_Y \eta w(\eta / \mathbf{z}) d\eta + \hat{\eta}^2,$$

совпадающий со среднеквадратической ошибкой при точечном оценивании параметра, т. е. в данном случае условие минимума риска сводится к поиску оценки с наименьшей среднеквадратической ошибкой. Уравнение (2.2.29) в данном случае имеет следующее решение:

$$\hat{\eta} = \int_Y \eta w(\eta / \mathbf{z}) d\eta \equiv \mathbf{E}[\eta / \mathbf{z}]. \quad (2.2.30)$$

Из (2.2.30) следует, что байесовская оценка при квадратичной функции потерь совпадает с условным (апостериорным) средним значением оцениваемого параметра при заданных значениях выборки  $\mathbf{z}$ .

Вычисление байесовской оценки согласно (2.2.30) предполагает знание апостериорной вероятности  $w(\eta / \mathbf{z})$  оцениваемого параметра, которая может быть выражена через функцию правдоподобия  $w(\mathbf{z} / \eta)$  и априорное распределение параметра  $w(\eta)$  посредством формулы Байеса:

$$w(\eta / \mathbf{z}) = \frac{w(\mathbf{z} / \eta)w(\eta)}{w(\mathbf{z})} = \frac{w(\mathbf{z} / \eta)w(\eta)}{\int_Y w(\mathbf{z} / \eta)w(\eta) d\eta}.$$

Тогда

$$\hat{\eta} = \frac{\int_Y \eta w(\mathbf{z} / \eta)w(\eta) d\eta}{\int_Y w(\mathbf{z} / \eta)w(\eta) d\eta}. \quad (2.2.30a)$$

При этом может оказаться так, что априорное распределение  $w(\eta)$  неизвестно. В этом случае вид  $w(\eta)$  следует определять из дополнительных, например, эвристических соображений.

### Гауссовское распределение

Пусть, к примеру, требуется получить байесовскую оценку математического ожидания  $a$  гауссовского распределения  $N(a, \sigma^2)$  при известной дисперсии  $\sigma^2$ , если распределение  $a$  также гауссовское с известными параметрами  $a_0$  и  $\sigma_0^2$ :

$$w(a) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[ -\frac{(a - a_0)^2}{2\sigma_0^2} \right].$$

Апостериорное распределение параметра  $a$  при известной выборке  $\mathbf{z}$  есть

$$w(a / \mathbf{z}) = \frac{w(\mathbf{z} / a)w(a)}{\int_{-\infty}^{\infty} w(\mathbf{z} / a)w(a)da} = \left\{ \int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2} \left( \frac{(a' - a_0)^2}{\sigma_0^2} - \frac{(a - a_0)^2}{\sigma_0^2} + \frac{1}{\sigma^2} \sum_{k=1}^M (z_k - a')^2 - \frac{1}{\sigma^2} \sum_{k=1}^M (z_k - a)^2 \right) \right] da' \right\}^{-1}.$$

Обозначим

$$\gamma = \frac{\sigma^2}{M\sigma_0^2}, \quad A = \frac{1}{1 + \gamma^2} \left[ \frac{1}{M} \sum_{k=1}^M z_k + \gamma^2 a_0 \right],$$

тогда

$$\begin{aligned} w(a / \mathbf{z}) &= \exp \left[ -\frac{M}{2\sigma^2} (1 + \gamma^2) (a^2 - 2aA) \right] \times \\ &\times \left\{ \int_{-\infty}^{\infty} \exp \left[ -\frac{M}{2\sigma^2} (1 + \gamma^2) (a'^2 - 2a'A) \right] da' \right\}^{-1} = \\ &= \left[ \frac{M}{2\pi\sigma^2} (1 + \gamma^2) \right]^{1/2} \exp \left[ -\frac{M}{2\sigma^2} (1 + \gamma^2) (a - A)^2 \right], \end{aligned}$$

и байесовская оценка согласно (2.2.30) равна

$$\hat{a} = \left[ \frac{M}{2\pi\sigma^2} (1 + \gamma^2) \right]^{1/2} \int_{-\infty}^{\infty} a \exp \left[ -\frac{M}{2\sigma^2} (1 + \gamma^2) (a - A)^2 \right] da = A.$$

Полученная оценка представляет собой взвешенное отношение двух величин: априорного среднего значения  $a_0$  оцениваемого параметра и среднего выборочного

$$\bar{a} = \frac{1}{M} \sum_{k=1}^M z_k,$$

являющегося (см. разд. 2.2.3) оценкой максимального правдоподобия. Когда  $M \rightarrow \infty$  (при этом апостериорное распределение стремится к дельта-функции  $\delta(a - \bar{a})$ ),  $\gamma^2 \rightarrow 0$ , и байесовская оценка стремится к МП-оценке.

Обобщим метод байесовских оценок на многомерный случай, когда оцениванию подлежат  $L$  параметров  $\eta_1, \dots, \eta_L$ .

Совокупное оценивание  $L$  неизвестных параметров приводит к появлению  $L$  ошибок оценивания  $\Delta\eta_1, \dots, \Delta\eta_L$ , которым необходимо поставить в соответствие  $L$ -компонентную функцию потерь. Если использовать аддитивную, по отношению к компонентам, функцию потерь, то

$$c(\Delta\eta) = \sum_{j=1}^L c_j(\eta - \hat{\eta}).$$

В частности, при квадратичной функции потерь

$$c(\Delta\eta) = \sum_{j=1}^L (\eta_j - \hat{\eta}_j)^2.$$

Тогда апостериорный риск представляется в виде суммы из  $L$  слагаемых:

$$\rho(\hat{\eta} / \mathbf{z}) = \sum_{j=1}^L \int_Y (\eta_j - \hat{\eta}_j)^2 w(\eta_j / \mathbf{z}) d\eta_j,$$

и для его минимизации необходимо решить систему уравнений

$$\frac{\partial \rho(\hat{\eta}_j / \mathbf{z})}{\partial \hat{\eta}_j} = 0, j = 1, \dots, L.$$

Отсюда получаем набор оценок:

$$\hat{\eta}_j = \int_Y \eta_j w(\eta_j / \mathbf{z}) d\eta_j. \quad (2.2.30б)$$

## 2.3. ОЦЕНКИ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ. РЕГРЕССИОННЫЙ АНАЛИЗ

Рассмотренные в разд. 2.2 методы получения статистических оценок были связаны с обработкой “собственной” выборки оцениваемых параметров. Представляет особый интерес подход, при котором оценка одной — ненаблюдаемой — случайной величины  $\eta$ , в том числе, векторной, когда  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$  является  $L$ -элементным набором, производится на основе анализа другой — наблюдаемой — случайной величины  $\xi$ , функционально связанной с первой.

Задачи, связанные с изучением вероятностных связей между наблюдаемыми и изучаемыми величинами, относятся к разделу математической статистики, который принято называть *математической теорией актив-*

ного эксперимента, основу которой составляют такие разделы, как *дисперсионный анализ, корреляционный анализ, регрессионный анализ*.

Методами дисперсионного анализа устанавливается наличие влияния определённых факторов на изучаемый процесс, отображаемый наблюдаемой выборкой. Корреляционный анализ позволяет оценить “силу” такой связи, а методами регрессионного анализа можно выбрать конкретную математическую модель и оценить её адекватность при установлении взаимосвязи между случайными величинами.

В общем случае предполагается, что при оценке ненаблюдаемого  $L$ -элементного вектора  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_L)$  посредством какого-либо эксперимента имеет место влияние  $L$  неслучайных параметров  $\mathbf{z} = (z_1, \dots, z_L)$  — *факторов*, значения которых могут меняться от опыта к опыту; эти параметры находятся под контролем экспериментатора и измеряются с пренебрежимо малой ошибкой. Такими параметрами могут быть входные характеристики технического устройства, вид и значения которых определяют процесс на выходе. Факторы могут быть обусловлены и не зависящими от исследователя причинами, например, внутренними шумовыми процессами. Но в любом случае статистические данные состоят из наблюдаемых значений  $x_k$  “доступной” случайной величины и соответствующих значений факторов:

$$(x_k, z_1^{(k)}, \dots, z_L^{(k)}), k = 1, \dots, M.$$

При этом предполагается, что  $L < M$ .

Наиболее простой, но, тем не менее, важной в практическом отношении является *модель линейной регрессии*, в которой значения наблюдаемой случайной величины представляют собой линейную зависимость от ненаблюдаемых величин:

$$x_k = \sum_{j=1}^L z_j^{(k)} \eta_j + e_k, k = 1, \dots, M, \quad (2.3.1)$$

где  $e_k$  — значения “источника погрешности”, т. е. не зависящей от  $\boldsymbol{\eta}$  случайной величины  $\varepsilon$ , по отношению к которым из физических соображений логично предположить, что они некоррелированы, имеют нулевое математическое ожидание  $\mathbf{E}[e_k] = 0$  и одинаковую дисперсию  $\mathbf{D}[e_k] = \sigma^2$ .

По терминологии регрессионного анализа управляемые параметры (факторы)  $\mathbf{z}$  принято называть *регрессорами* или *предикторами*.

Соотношение (2.3.1) является линейной моделью по отношению к величинам  $\eta$ , но требование линейности никак не относится к параметрам  $z$ . Например, все  $z_j$  ( $j = 1, \dots, L$ ) могут быть  $j$ -й степенью одного параметра; для определённости обозначим его через  $c$ :

$$x_k = \eta_1 c_k + \eta_2 c_k^2 + \eta_3 c_k^3 + \dots + \eta_L c_k^L + e_k.$$

В противоположность этому, соотношение вида

$$x_k = \eta_1 + \eta_2 c_k + \exp(\eta_3) c_k^2 + e_k,$$

не является моделью линейной регрессии, так как она не линейна по параметру  $\eta_3$ .

Исторически первыми задачами, для решения которых ещё на рубеже XVIII–XIX столетий привлекались методы регрессионного анализа (А. Лежандр, К. Гаусс, А. Марков и др.), были задачи “наилучшей” теоретической аппроксимации экспериментальных зависимостей в области астрономии и геодезии.

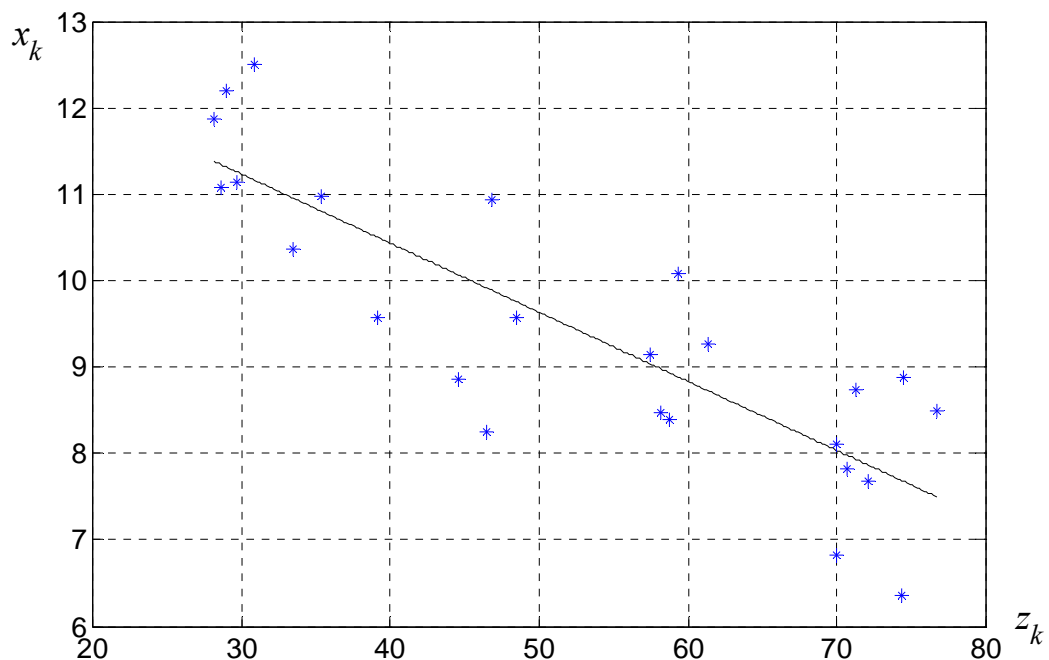


Рис. 2.4. К оцениванию методом наименьших квадратов

В классической постановке задача оценивания неизвестных параметров формулируется следующим образом. Задан набор точек, образующих некоторое “облако” на плоскости (рис. 2.4). Предполагается, что это обла-

ко в действительности представляет собой “размытую” линию, а размытость обусловлена различными погрешностями при проведении эксперимента. Требуется подобрать некоторым оптимальным образом параметры линии: угол наклона и коэффициент смещения.

В рамках общей модели линейной регрессии (2.3.1) эта задача формулируется как задача оптимального оценивания двух параметров, скажем  $\eta_1$  и  $\eta_2$ , при наличии одного<sup>1</sup> управляемого фактора  $z$ :

$$x_k = \eta_1 z_k + \eta_2 + e_k, \quad k = 1, \dots, M. \quad (2.3.2)$$

В качестве критерия оптимальности предлагается минимум суммы квадратов разностей выборочных значений  $x_k$  и “истинных” значений — значений, лежащих на прямой

$$x_k^* = \eta_1 z_k + \eta_2,$$

т. е.

$$\Delta(\eta_1, \eta_2) = \sum_{k=1}^M e_k^2 = \sum_{k=1}^M (x_k - \eta_1 z_k - \eta_2)^2 \rightarrow \min. \quad (2.3.3)$$

Для нахождения минимума неотрицательной функции двух переменных достаточно приравнять к нулю частные производные по каждому из неизвестных параметров:

$$\frac{\partial \Delta(\eta_1, \eta_2)}{\partial \eta_1} = -2 \sum_{k=1}^M (x_k - \eta_1 z_k - \eta_2) z_k = 0;$$

$$\frac{\partial \Delta(\eta_1, \eta_2)}{\partial \eta_2} = -2 \sum_{k=1}^M (x_k - \eta_1 z_k - \eta_2) = 0.$$

Полученные уравнения в литературе называют *нормальными*. Введём стандартные обозначения для выборочного среднего и выборочной дисперсии:

$$\bar{x} = \frac{1}{M} \sum_{k=1}^M x_k; \quad \bar{z} = \frac{1}{M} \sum_{k=1}^M z_k; \quad \overline{z^2} = \frac{1}{M} \sum_{k=1}^M z_k^2; \quad \bar{z} - \overline{z^2} = s_z^2,$$

тогда получим следующие выражения для *оптимальных по методу минимума суммы квадратов оценок* (МНК-оценок) неизвестных параметров:

---

<sup>1</sup> Можно говорить о двух факторах, один из которых принимает значения в заданной области, а второй — тождественно равен единице.



$$\hat{\eta}_1 = \frac{1}{Ms_z^2} \sum_{k=1}^M (z_k - \bar{z}) x_k = \frac{1}{s_z^2} \left( \frac{1}{M} \sum_{k=1}^M z_k x_k - \bar{z} \bar{x} \right); \quad (2.3.4a)$$

$$\hat{\eta}_2 = \frac{1}{Ms_z^2} \sum_{k=1}^M (\bar{z}^2 - \bar{z} z_k) x_k = \bar{x} - \bar{z} \hat{\eta}_1. \quad (2.3.4b)$$

Видно, что МНК-оценки являются линейными по отношению к выборочным значениям  $x_k$ .

Найдём характеристики полученных оценок. Математическое ожидание и дисперсия фактически являются условными, так как вычисляются при фиксированных (управляемых) значениях  $z_k$ . Поэтому выразим в (2.3.4) наблюдаемые значения  $x_k$  через фиксированные значения  $z_k$  и значения погрешностей  $e_k$ , которые, как уже было сказано, имеют нулевое среднее и одинаковую дисперсию  $\sigma^2$ :

$$\begin{aligned} \hat{\eta}_1 &= \frac{1}{Ms_z^2} \sum_{k=1}^M (z_k - \bar{z})(\eta_1 z_k + \eta_2 + e_k) = \\ &= \eta_1 + \frac{1}{Ms_z^2} \sum_{k=1}^M z_k e_k - \frac{\bar{z}}{Ms_z^2} \sum_{k=1}^M e_k; \end{aligned} \quad (2.3.5a)$$

$$\begin{aligned} \hat{\eta}_2 &= \frac{1}{M} \sum_{k=1}^M (\eta_1 z_k + \eta_2 + e_k) - \bar{z} \hat{\eta}_1 = \\ &= \eta_2 - \frac{\bar{z}}{Ms_z^2} \sum_{k=1}^M z_k e_k + \frac{(1 - \bar{z}^2 / s_z^2)}{M} \sum_{k=1}^M e_k. \end{aligned} \quad (2.3.5b)$$

Тогда операции усреднения при вычислении математического ожидания и дисперсии будут относиться только к значениям погрешностей  $e_k$ .

Нетрудно видеть, что полученные оценки являются несмещёнными. Действительно:

$$\begin{aligned} \mathbf{E}[\hat{\eta}_1] &= \eta_1 + \frac{1}{Ms_z^2} \sum_{k=1}^M z_k \mathbf{E}[e_k] - \frac{\bar{z}}{Ms_z^2} \sum_{k=1}^M z_k \mathbf{E}[e_k] = \eta_1; \\ \mathbf{E}[\hat{\eta}_2] &= \eta_2 - \frac{\bar{z}}{Ms_z^2} \sum_{k=1}^M z_k \mathbf{E}[e_k] - \frac{(1 - \bar{z}^2 / s_z^2)}{M} \sum_{k=1}^M \mathbf{E}[e_k] = \eta_2. \end{aligned}$$

Вычислим дисперсии оценок. Имеем:

$$\mathbf{E}[\hat{\eta}_1^2] = \eta_1^2 + \frac{1}{M^2 (s_z^2)^2} \sum_{k=1}^M \sum_{n=1}^M z_k z_n \mathbf{E}[e_k e_n] + \frac{\bar{z}^2}{M^2 (s_z^2)^2} \sum_{k=1}^M \sum_{n=1}^M \mathbf{E}[e_k e_n] +$$

$$+ \frac{2\eta_1}{Ms_z^2} \sum_{k=1}^M z_k \mathbf{E}[e_k] - \frac{2\eta_1 \bar{z}}{Ms_z^2} \sum_{k=1}^M \mathbf{E}[e_k] - \frac{2\bar{z}}{M^2(s_z^2)^2} \sum_{k=1}^M \sum_{n=1}^M z_k \mathbf{E}[e_k e_n].$$

В силу независимости значений  $e_k$

$$\mathbf{E}[e_k e_n] = \delta_{kn} \sigma^2 = \begin{cases} \sigma^2, & m = n; \\ 0, & m \neq n; \end{cases}$$

следовательно,

$$\begin{aligned} \mathbf{E}[\hat{\eta}_1^2] &= \eta_1^2 + \frac{\sigma^2}{M^2(s_z^2)^2} \sum_{k=1}^M z_k^2 + \frac{\bar{z}^2}{M^2(s_z^2)^2} M \sigma^2 - \frac{2\sigma^2 \bar{z}^2}{M^2(s_z^2)^2} \sum_{k=1}^M z_k = \\ &= \eta_1^2 + \frac{\sigma^2}{M(s_z^2)^2} (\bar{z}^2 - \bar{z}^2) = \eta_1^2 + \frac{\sigma^2}{Ms_z^2}, \end{aligned}$$

и

$$\mathbf{D}[\hat{\eta}_1] = \mathbf{E}[\hat{\eta}_1^2] - \mathbf{E}^2[\eta_1] = \eta_1^2 + \frac{\sigma^2}{Ms_z^2} - \eta_1^2 = \frac{\sigma^2}{Ms_z^2}. \quad (2.3.6a)$$

Аналогично

$$\begin{aligned} \mathbf{E}[\hat{\eta}_2^2] &= \eta_2^2 + \frac{\bar{z}^2}{M^2(s_z^2)^2} \sum_{k=1}^M \sum_{n=1}^M z_k z_n \mathbf{E}[e_k e_n] + \frac{(1 + \bar{z}^2 / s_z^2)}{M^2} \sum_{k=1}^M \sum_{n=1}^M \mathbf{E}[e_k e_n] - \\ &- \frac{2\eta_2 \bar{z}}{Ms_z^2} \sum_{k=1}^M z_k \mathbf{E}[e_k] + \frac{2\eta_2 (1 + \bar{z}^2 / s_z^2)}{M} \sum_{k=1}^M \mathbf{E}[e_k] - \frac{2\bar{z} (1 + \bar{z}^2 / s_z^2)}{M^2 s_z^2} \sum_{k=1}^M \sum_{n=1}^M z_k \mathbf{E}[e_k e_n] = \\ &= \eta_2^2 + \frac{\bar{z}^2 \sigma^2 \bar{z}^2}{M(s_z^2)^2} + \frac{\sigma^2}{M} (1 + \bar{z}^2 / s_z^2) (1 - \bar{z}^2 / s_z^2) = \eta_2^2 + \frac{\sigma^2 \bar{z}^2}{Ms_z^2}, \end{aligned}$$

и

$$\mathbf{D}[\hat{\eta}_2] = \eta_2^2 + \frac{\sigma^2 \bar{z}^2}{Ms_z^2} - \eta_2^2 = \frac{\sigma^2 \bar{z}^2}{Ms_z^2}. \quad (2.3.6a)$$

Найдём ковариацию оцениваемых параметров

$$\text{Cov}[\hat{\eta}_1, \hat{\eta}_2] = \mathbf{E}[\hat{\eta}_1 \hat{\eta}_2] - \eta_1 \eta_2 = \mathbf{E}[\hat{\eta}_1 \bar{x}] - \bar{z} \mathbf{E}[\hat{\eta}_1^2] - \eta_1 \eta_2.$$

Поскольку

$$\begin{aligned} \mathbf{E}[\hat{\eta}_1 \bar{x}] &= \frac{1}{M^2 s_z^2} \sum_{k=1}^M \sum_{n=1}^M (z_k - \bar{z}) \mathbf{E}[x_k x_n] = \frac{1}{M^2 s_z^2} \sum_{k=1}^M \sum_{n=1}^M (z_k - \bar{z}) \times \\ &\times ((\eta_1 z_k + \eta_2)(\eta_1 z_n + \eta_2) + (\eta_1 z_k + \eta_2) \mathbf{E}[e_n] + (\eta_1 z_n + \eta_2) \mathbf{E}[e_k] + \mathbf{E}[e_k e_n]) = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{M^2 s_z^2} \sum_{k=1}^M \sum_{n=1}^M (z_k - \bar{z}) ((\eta_1 z_k + \eta_2)(\eta_1 z_n + \eta_2) + \delta_{kn} \sigma^2) = \\
&= (\eta_1 \bar{z} + \eta_2) \frac{\eta_1 \bar{z}^2 - \eta_1 (\bar{z})^2}{s_z^2} = (\eta_1 \bar{z} + \eta_2) \eta_1, \\
\text{Cov}[\hat{\eta}_1, \hat{\eta}_2] &= \bar{z} (\eta_1^2 - \mathbf{E}[\hat{\eta}_1^2]) = -\bar{z} \mathbf{D}[\hat{\eta}_1] = \frac{-\bar{z} \sigma^2}{M s_z^2}. \quad (2.3.7)
\end{aligned}$$

Из (2.3.7) следует, что при равенстве нулю выборочного среднего  $\bar{z}$  МНК-оценки параметров  $\eta_1$  и  $\eta_2$  некоррелированы.

Нормируя (2.3.7), получаем коэффициент корреляции  $r$  МНК-параметров:

$$r[\hat{\eta}_1, \hat{\eta}_2] = \frac{\text{Cov}[\hat{\eta}_1, \hat{\eta}_2]}{\sqrt{\mathbf{D}[\hat{\eta}_1] \mathbf{D}[\hat{\eta}_2]}} = -\frac{\bar{z}}{\sqrt{\bar{z}^2}}. \quad (2.3.8)$$

Одним из главных преимуществ МНК-оценок является то, что такие оценки имеют наименьшую дисперсию среди всех линейных несмещённых оценок — это будет показано далее при рассмотрении более общего (многофакторного) случая метода наименьших квадратов.

В качестве иллюстрации рассмотрим задачу оценивания параметров аппроксимирующей прямой для экспериментальных данных, представленных в табл. 2.1 и на рис. 2.4.

Таблица 2.1

**Экспериментальные данные для МНК-аппроксимации в модели линейной регрессии с одним предиктором**

$x_k$	$z_k$	$x_k$	$z_k$	$x_k$	$z_k$	$x_k$	$z_k$
35,3	11,0	74,4	6,4	28,1	11,9	70,0	6,8
29,7	11,1	76,7	8,5	39,1	9,6	74,5	8,9
30,8	12,5	70,7	7,8	46,8	10,0	72,1	7,7
58,8	8,4	57,5	9,1	48,5	9,6	58,1	8,5
61,4	9,3	46,4	8,2	59,3	10,1	44,6	8,7
71,3	8,7	28,9	12,2	70,0	8,1	33,4	10,4

Применение соотношений (2.3.4) даёт следующие значения параметров:

$$\hat{\eta}_1 = -0,1; \hat{\eta}_2 = 13,6,$$

так что аппроксимирующая прямая имеет вид

$$x = -0,1z + 13,6.$$

Полученные выше результаты касались простейшего случая применения линейной регрессии — МНК-оценки одного параметра. Рассмотрим общую модель линейной регрессии (2.3.1), в которой участвует более одного предиктора.

Для удобства запишем соотношение (2.3.1) в матричном виде. Для этого введём матрицу размерности  $M \times L$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \dots \\ \mathbf{z}^{(M)} \end{pmatrix} = \begin{pmatrix} z_1^{(1)} & z_2^{(1)} & \dots & z_L^{(1)} \\ z_1^{(2)} & z_2^{(2)} & \dots & z_L^{(1)} \\ \dots & \dots & \dots & \dots \\ z_1^{(M)} & z_2^{(M)} & \dots & z_L^{(M)} \end{pmatrix}, \quad (2.3.9)$$

называемую *регрессионной матрицей*. Обычно элементы матрицы  $\mathbf{Z}$  выбирают таким образом, чтобы её столбцы были линейно независимы, т. е. чтобы её ранг равнялся  $L$ . Однако в некоторых случаях при планировании и проведении эксперимента элементы матрицы  $\mathbf{Z}$  выбираются из нулей и единиц; столбцы такой матрицы, называемой в таком случае *матрицей планирования*, могут оказать линейно зависимыми.

Далее будем считать, что  $L$ -элементный набор  $\boldsymbol{\eta}$  оцениваемых параметров,  $M$ -элементный набор  $\mathbf{e}$  значений ошибок и  $M$ -элементный набор  $\mathbf{x}$  наблюдаемых значений являются вектор-столбцами. Тогда запишем следующее матричное соотношение:

$$\mathbf{x} = \mathbf{Z}\boldsymbol{\eta} + \mathbf{e}. \quad (2.3.10)$$

Получение МНК-оценок в общей модели линейной регрессии принципиально не отличается от рассмотренного частного случая и связано с минимизацией квадратичной формы

$$\mathbf{e}^T \mathbf{e} = (\mathbf{x} - \mathbf{Z}\boldsymbol{\eta})^T (\mathbf{x} - \mathbf{Z}\boldsymbol{\eta}) = \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{Z}\boldsymbol{\eta} - (\mathbf{Z}\boldsymbol{\eta})^T \mathbf{x} + (\mathbf{Z}\boldsymbol{\eta})^T \mathbf{Z}\boldsymbol{\eta}.$$

С учётом того, что для матриц  $\mathbf{A}$  и  $\mathbf{B}$  при транспонировании справедливы соотношения

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T, \quad (\mathbf{A}^T)^T = \mathbf{A},$$

имеем

$$(\mathbf{Z}\boldsymbol{\eta})^T \mathbf{x} = ((\mathbf{Z}\boldsymbol{\eta})^T \mathbf{x})^T = \mathbf{x}^T \boldsymbol{\eta} \mathbf{Z},$$

так что

$$\mathbf{e}^T \mathbf{e} = \mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{Z}\boldsymbol{\eta} + (\mathbf{Z}\boldsymbol{\eta})^T \mathbf{Z}\boldsymbol{\eta}.$$

Далее получившуюся квадратичную форму дифференцируем по<sup>1</sup>  $\boldsymbol{\eta}$ , приравняем нулю и приходим к уравнению

$$(\mathbf{Z}^T \mathbf{Z})\boldsymbol{\eta} = \mathbf{Z}^T \mathbf{x}, \quad (2.3.11)$$

которое, как нетрудно показать, имеет единственное решение

$$\boldsymbol{\eta}_{\text{МНК}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{x}. \quad (2.3.12)$$

Минимальное значение квадратичной формы  $\mathbf{e}^T \mathbf{e}$ , равное

$$\mathbf{e}^T \mathbf{e} = (\mathbf{x} - \mathbf{Z}\boldsymbol{\eta}_{\text{МНК}})^T (\mathbf{x} - \mathbf{Z}\boldsymbol{\eta}_{\text{МНК}}) = \mathbf{x}^T \mathbf{x} - (\boldsymbol{\eta}_{\text{МНК}})^T \mathbf{Z}^T \mathbf{Z}\boldsymbol{\eta}_{\text{МНК}} \quad (2.3.13)$$

называется *остаточной суммой квадратов* и в литературе по регрессионному анализу (в том числе, русскоязычной) обычно обозначается буквами  $\text{RSS}^2$ .

В матричной записи достаточно легко найти статистические характеристики МНК-оценок.

Несмещённость:

$$\begin{aligned} E[\boldsymbol{\eta}_{\text{МНК}}] &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T E[\mathbf{x}] = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T E[\mathbf{Z}\boldsymbol{\eta} + \mathbf{e}] = \\ &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T (\mathbf{Z}\boldsymbol{\eta} + E[\mathbf{e}]) = (\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{Z}^T \mathbf{Z})\boldsymbol{\eta} = \boldsymbol{\eta}. \end{aligned}$$

При вычислении дисперсии МНК-оценок представим дисперсию матрицы ошибок в виде

$$\mathbf{D}[\mathbf{e}] = \sigma^2 \mathbf{I}_M,$$

где  $\mathbf{I}_M$  — единичная матрица размерности  $M \times M$ , а также воспользуемся свойством дисперсии от произведения неслучайной  $\mathbf{A}$  и случайной  $\mathbf{B}$  матриц:

$$\mathbf{D}[\mathbf{A}\mathbf{B}] = \mathbf{A}\mathbf{D}[\mathbf{B}]\mathbf{A}^T.$$

Тогда

$$\begin{aligned} \mathbf{D}[\boldsymbol{\eta}_{\text{МНК}}] &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{D}[\mathbf{x}] ((\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T)^T = \\ &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{D}[\mathbf{e}] \mathbf{Z} ((\mathbf{Z}^T \mathbf{Z})^{-1})^T. \end{aligned}$$

Учитывая, что для вещественных матриц

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1},$$

имеем

$$((\mathbf{Z}^T \mathbf{Z})^{-1})^T = ((\mathbf{Z}^T \mathbf{Z})^T)^{-1} = (\mathbf{Z}^T \mathbf{Z})^{-1},$$

и

$$\begin{aligned} \mathbf{D}[\boldsymbol{\eta}_{\text{МНК}}] &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \sigma^2 \mathbf{I}_M \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} = \\ &= \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{Z}^T \mathbf{Z}) (\mathbf{Z}^T \mathbf{Z})^{-1} = \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1}, \end{aligned} \quad (2.3.14)$$

<sup>1</sup> Матричная производная означает, что дифференцированию по соответствующему скалярному параметру подвергаются все строки (столбцы) матрицы.

<sup>2</sup>  $\text{RSS}$  — аббревиатура от *residual sum of squares*.

что является обобщением (2.3.6).

МНК-оценки  $\boldsymbol{\eta}_{\text{МНК}}$  обладают свойством<sup>1</sup> оптимальности, заключающимся в том, что *в классе любых других линейных несмещённых оценок они обладают минимальной дисперсией*. Покажем это.

Пусть  $\mathbf{P}$  — матрица размерности  $L \times M$ , так что  $\mathbf{q} = \mathbf{P}\mathbf{x}$  — произвольная линейная несмещённая оценка параметров  $\boldsymbol{\eta}$ , полученная на основе линейных операций с наблюдаемыми значениями  $\mathbf{x}$ . Требование несмещённости оценки означает, что для произвольных значений  $\boldsymbol{\eta}$  выполняется равенство

$$\mathbf{E}[\mathbf{q}] = \mathbf{P}\mathbf{E}[\mathbf{x}] = \mathbf{P}\mathbf{Z}\boldsymbol{\eta} = \boldsymbol{\eta},$$

откуда следует, что

$$\mathbf{P}\mathbf{Z} = \mathbf{I}_L. \quad (2.3.17)$$

(Заметим, что в данном случае  $\mathbf{P} \neq \mathbf{Z}^{-1}$ , поскольку эти матрицы не являются квадратными.)

Дисперсия такой оценки равна

$$\mathbf{D}[\mathbf{q}] = \mathbf{P}\mathbf{D}[\mathbf{x}]\mathbf{P}^T = \mathbf{P}(\sigma^2\mathbf{I}_M)\mathbf{P}^T = \sigma^2\mathbf{P}\mathbf{P}^T.$$

Запишем для  $\mathbf{P}\mathbf{P}^T$  с учётом (2.3.17) следующее тождество:

$$\begin{aligned} \mathbf{P}\mathbf{P}^T &\equiv (\mathbf{Z}^T\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{Z})^{-1} + \mathbf{P}\mathbf{P}^T + (\mathbf{Z}^T\mathbf{Z})^{-1} = \\ &= (\mathbf{Z}^T\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{I}_L^T - ((\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{I}_L^T)^T + \\ &\quad + \mathbf{P}\mathbf{P}^T + (\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{Z})(\mathbf{Z}^T\mathbf{Z})^{-1})^T = \\ &= (\mathbf{Z}^T\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{P}^T - \mathbf{P}\mathbf{Z}((\mathbf{Z}^T\mathbf{Z})^{-1})^T - \\ &\quad + \mathbf{P}\mathbf{P}^T + (\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{Z}^T\mathbf{Z})(\mathbf{Z}^T\mathbf{Z})^{-1})^T = \\ &= (\mathbf{Z}^T\mathbf{Z})^{-1} + (\mathbf{P} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)(\mathbf{P} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)^T. \end{aligned}$$

Тогда

$$\begin{aligned} \mathbf{D}[\mathbf{q}] &= \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1} + \sigma^2(\mathbf{P} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)(\mathbf{P} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)^T = \\ &= \mathbf{D}[\boldsymbol{\eta}_{\text{МНК}}] + \sigma^2(\mathbf{P} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)(\mathbf{P} - (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)^T. \end{aligned}$$

Из общих свойств матриц следует, что произведение произвольной матрицы на её транспонированный вид является<sup>2</sup> неотрицательно-определённой матрицей, отсюда

<sup>1</sup> В литературе данное свойство известно как теорема Гаусса–Маркова

<sup>2</sup> Матрица  $\mathbf{A}$  называется неотрицательно-определённой, если для произвольного ненулевого вектора  $\mathbf{y}$  справедливо соотношение  $\mathbf{y}^*\mathbf{A}\mathbf{y} \geq 0$  (\* — знак комплексного сопряжения). Данное свойство означает, что при вычислении различных квадратичных форм, например, дисперсии отдельных параметров, в результате будут получаться неотрицательные значения.

$$D[\mathbf{q}] \geq D[\boldsymbol{\eta}_{\text{МНК}}],$$

причём равенство достигается только в том случае, когда

$$\mathbf{P} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T,$$

т. е. когда матрица  $\mathbf{P}$  линейного преобразования совпадает с МНК-оценкой, получаемой из соотношения (2.3.12).

Рассмотрим пример линейной регрессии с двумя предикторами  $z_1$  и  $z_2$  и тремя неизвестными параметрами  $\eta_0$ ,  $\eta_1$  и  $\eta_2$ , в которой наблюдаемыми значениями являются

$$x_k = \eta_0 + \eta_1 z_{1k} + \eta_2 z_{2k} + e_k, \quad k = 1, \dots, M. \quad (2.3.18)$$

Экспериментальные данные для модели (2.3.18) представлены в табл. 2.2.

Таблица 2.2

**Экспериментальные данные для модели линейной регрессии с двумя предикторами**

$x_k$	$z_{1k}$	$z_{2k}$	$x_k$	$z_{1k}$	$z_{2k}$
11,0	35,3	7,4	11,9	28,1	6,6
11,1	29,7	8,0	9,6	39,1	7,3
12,5	30,8	7,4	10,0	46,8	8,1
8,4	58,8	7,5	9,6	48,5	8,4
9,3	61,4	5,5	10,1	59,3	6,1
8,7	71,3	8,9	8,1	70,0	4,9
6,4	74,4	4,1	6,8	70,0	4,6
8,5	76,7	4,1	8,9	74,5	3,7
7,8	70,7	4,1	7,7	72,1	4,7
9,1	57,5	4,5	8,5	58,1	5,3
8,2	46,4	10,3	8,7	44,6	6,2
12,2	28,9	6,9	10,4	33,4	6,8

Вектор-столбец наблюдаемых значений  $\mathbf{X}$  и регрессионная матрица  $\mathbf{Z}$  размерностью  $24 \times 3$  для рассматриваемых экспериментальных данных имеют следующий вид:

$$\mathbf{X} = \begin{pmatrix} 11,0 \\ 11,1 \\ 12,5 \\ \dots \\ 8,7 \\ 10,4 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 1 & 35,5 & 7,4 \\ 1 & 29,7 & 8,0 \\ 1 & 30,8 & 7,4 \\ \dots & \dots & \dots \\ 1 & 44,6 & 6,2 \\ 1 & 33,4 & 6,8 \end{pmatrix}.$$

Единичный первый столбец в регрессионной матрице отражает тот факт, что соответствующий предиктор ( $z_0$ ) является формальным; он никак не влияет на проведение эксперимента.

Как видно из рассмотренных выше теоретических выкладок, одним из ключевых во всём регрессионном анализе объектов является матрица, представляющая собой произведение регрессионной матрицы на себя транспонированную. В данном случае эта матрица равна

$$\mathbf{Z}^T \mathbf{Z} = \begin{pmatrix} 24 & 1286 & 151 \\ 1286 & 75517 & 7696 \\ 151 & 7696 & 1028 \end{pmatrix},$$

а обратная ей матрица

$$(\mathbf{Z}^T \mathbf{Z})^{-1} = \begin{pmatrix} 2,5582 & -0,0218 & -0,2134 \\ -0,0218 & 0,0002 & -0,2134 \\ -0,2134 & -0,2134 & 0,0219 \end{pmatrix}.$$

Отметим, однако, что матрица  $\mathbf{Z}^T \mathbf{Z}$  имеет не очень хорошую обусловленность — соответствующее число обусловленности имеет порядок  $10^5$ .

Вычисление МНК-оценки согласно (2.3.12) приводит к следующему результату:

$$\boldsymbol{\eta}_{\text{МНК}} = \begin{pmatrix} \eta_{0\text{МНК}} \\ \eta_{1\text{МНК}} \\ \eta_{2\text{МНК}} \end{pmatrix} = \begin{pmatrix} 14,4 \\ -0,1 \\ -0,1 \end{pmatrix}.$$

При этом остаточная сумма квадратов равна

$$\text{RSS} = 17,9.$$

Таким образом, окончательно получаем, что модель линейной регрессии имеет следующий вид:

$$x = 14,4 - 0,1z_1 - 0,1z_2. \quad (2.3.19)$$



## ВОПРОСЫ И ЗАДАНИЯ К ГЛАВЕ 2

- 2.1. Что такое точечная и интервальная статистические оценки?
- 2.2. Какие характеристики точечных оценок вам известны?
- 2.3. Что такое фишеровская информация и каков её статистический смысл?
- 2.4. Вычислить фишеровскую информацию для следующих распределений:
- а) гауссовского;
  - б) бернуллиевского;
  - в) пуассоновского;
  - г) Коши.
- 2.5. Какие вам известны методы получения точечных оценок? Каковы статистические характеристики получаемых в этих методах оценок?
- 2.6. Оценить методом моментов параметры гамма-распределения

$$w(z/a, \mu) = \begin{cases} \frac{z^{\mu-1} \exp(-z/a)}{\Gamma(\mu)}, & z \geq 0; \\ 0, & z < 0. \end{cases}$$

- 2.7. Найти апостериорный риск и выражение для байесовской оценки при использовании функций потерь, рассмотренных в разд. 2.2.4: дельта-образной, модульной, прямоугольной, экспоненциальной.
- 2.8. Наблюдается выборка из пуассоновского распределения с параметром  $\lambda$ . Найти апостериорный риск и байесовскую оценку для квадратичной функции потерь.
- 2.9. Найти байесовскую оценку при квадратичной функции потерь:
- а) дисперсии гауссовского распределения, если математическое ожидание известно и равно нулю, а априорное распределение дисперсии имеет экспоненциальный вид:

$$w(\sigma^2) = \frac{1}{\sigma_0^2} \exp(-\sigma^2 / \sigma_0^2);$$

- б) параметра экспоненциального распределения

$$w(z / \eta) = \frac{1}{\eta} \exp(-z / \eta), z > 0, \eta > 0,$$

когда априорное распределение параметра также имеет экспоненциальный вид:

$$w(\eta) = \frac{1}{\eta_0} \exp(-\eta / \eta_0).$$

Ответ. а)  $\hat{\eta} = -\frac{d}{d(1/\sigma_0^2)} \left[ \ln K_{M/2-1} \left( 2\sqrt{Ms^2 / \sigma_0^2} \right) \right]$ , где  $s^2$  — выборочная дисперсия; б)  $\hat{\eta} = -\frac{d}{d(1/\eta_0)} \left[ \ln K_{M-1} \left( 2\sqrt{M\bar{z} / \eta_0} \right) \right]$ , где  $\bar{z}$  — выборочное среднее. Здесь  $K_m(y)$  — функция Бесселя второго рода  $m$ -го порядка от мнимого аргумента.

2.10. Дисперсия  $\sigma^2$  гауссовского распределения известна, а математическое ожидание  $a$  распределено по экспоненциальному закону

$$w(a) = \frac{1}{a_0} \exp(-a / a_0), a > 0, a_0 > 0.$$

Найти байесовскую оценку математического ожидания для дельта-образной и квадратичной функций потерь.

2.11. Рассматривается линейная регрессия с тремя неизвестными параметрами  $\eta_0, \eta_1$  и  $\eta_2$ , и наблюдаются значения

$$x_k = \eta_0 + \eta_1 (z_{1k} - \bar{z}_1) + \eta_2 (z_{2k} - \bar{z}_2) + e_k, k = 1, \dots, M,$$

где

$$\bar{z}_1 = \sum_{k=1}^M z_{1k}, \bar{z}_2 = \sum_{k=1}^M z_{2k}.$$

Найти дисперсию МНК-оценок параметров  $\eta_0, \eta_1$  и  $\eta_2$ , если  $\mathbf{E}[e] = 0$  и  $\mathbf{D}[e] = \sigma^2 \mathbf{I}_M$ .

## ГЛАВА 3. ОПТИМАЛЬНЫЙ ПРИЁМ СИГНАЛОВ

Как следует из результатов, полученных в первой главе, использование любой из стратегий принятия решений предполагает вычисление отношения правдоподобия  $\Lambda(\mathbf{z})$  для однородной выборки  $\mathbf{z} = \{z_1, z_2, \dots, z_M\}$  конечного объёма и сравнение его с некоторым пороговым значением  $\Lambda_0$ . Знание закона распределения  $w(z_k)$ , ( $k = 1, \dots, M$ ) входящих в выборку элементов позволяет найти функции правдоподобия, соответствующие различным гипотезам, и, следовательно, определить искомые отношения правдоподобия.

Решение практических задач оптимального приёма, как правило, связано с обработкой набора сигналов  $s_1(t), \dots, s_m(t)$  — непрерывных функций времени<sup>1</sup>, т. е. случайных процессов. Как известно [2], в настоящее время используются два основных подхода к представлению случайных процессов.

При первом подходе процесс  $\zeta(t)$  описывается набором многомерных функций распределения

$$w_{\zeta}^{(M)}(\mathbf{z}, \mathbf{t}) = w_{\zeta}^{(M)}(z_1, \dots, z_M; t_1, \dots, t_M), M = 1, 2, \dots,$$

получаемых в различных временных сечениях процесса, т. е. в моменты времени наблюдения  $t_1, \dots, t_M$ . Другими словами, случайный процесс, по существу, отождествляется с совокупностью из  $M$ , вообще говоря, зависимых случайных величин  $\zeta_1 = \zeta(t_1), \dots, \zeta_M = \zeta(t_M)$ . При этом выбор моментов времени произволен. Например, если задан интервал наблюдения  $[0; T]$ , то в нём можно выбрать  $M$  равноотстоящих временных точек, полагая

$$t_k = \frac{k-1}{M-1}T, k = 1, \dots, M.$$

Однако в общем случае получаемые таким способом значения  $\zeta_k$  оказываются коррелированными между собой, а многомерное распределение, по которому вычисляется отношение правдоподобия, представляет собой весьма сложное нефакторизованное выражение.

Альтернативным и, как будет показано далее, конструктивным способом представления процессов набором дискретных величин является раз-

---

<sup>1</sup> Пока что не будем принимать во внимание операции дискретизации во времени и квантования по уровню, лежащие в основе методов цифровой обработки сигналов.

ложение Карунена–Лоева (Karhunen, Loeve), которое представляет собой обобщение идеи представления функций в ортогональном базисе, когда ортогональному разложению подвергаются случайные процессы. Результатом такого представления является набор некоррелированных координат в соответствующем базисе, что вкупе с дополнительным предположением о гауссовском характере действующих в канале помех позволяет факторизовать многомерные распределения.

Другой подход к представлению  $\zeta(t)$  состоит в рассмотрении всех реализаций  $\{z(t)\}$  случайного процесса как функций времени  $t$ , когда вероятностной мерой, заданной на множестве реализаций, является характеристический функционал

$$\Theta[v(t)] = \mathbf{E} \left\{ \exp \left[ j \int_{-\infty}^{\infty} v(t) \zeta(t) dt \right] \right\},$$

где  $v(t)$  — некоторая непрерывная функция, а стоящий под экспонентой интеграл сходится в среднеквадратическом. При таком подходе результат наблюдения на конечном интервале  $[0; T]$  представляется усечённой реализацией

$$z_T(t) = \begin{cases} z(t), & 0 \leq t \leq T; \\ 0, & t < 0, t > T. \end{cases}$$

Возникает вопрос о возможности объединения двух подходов при решении задач оптимального приёма сигналов, используя как механизм получения некоррелированных координат в ортогональном базисе, так и вероятностные характеристики непрерывной наблюдаемой реализации  $z_T(t)$ . Эта возможность базируется на том фундаментальном факте, что при определённых, достаточно общих и, как правило, выполняющихся на практике условиях существует предел по вероятности при  $M \rightarrow \infty$  отношения правдоподобия  $\Lambda^{(M)}(\mathbf{z})$  независимо от того, какому условному распределению  $w^{(M)}(\mathbf{z}/s_r(t))$ ,  $r = 1, \dots, m$ , принадлежит  $M$ -мерная выборка  $\mathbf{z}$ :

$$\Lambda_{rl}[z(t)] \equiv \lim_{M \rightarrow \infty} \Lambda_{rl}^{(M)}(\mathbf{z}) = \lim_{M \rightarrow \infty} \frac{w^{(M)}(\mathbf{z}/s_r(t))}{w^{(M)}(\mathbf{z}/s_l(t))}, \quad r, l = 1, \dots, m.$$

Этот предел называется функционалом отношения правдоподобия.

### 3.1. РАЗЛОЖЕНИЕ КАРУНЕНА — ЛОЭВА

Пусть  $\{f_k(t)\}$  — система ортонормированных на интервале наблюдения  $[t_1; t_2]$  функций, т. е. их скалярное произведение равно символу Кронекера:

$$\langle f_k, f_l \rangle = \int_{t_1}^{t_2} f_k(t) f_l(t) dt = \delta_{kl} = \begin{cases} 1, & k = l; \\ 0, & k \neq l. \end{cases}$$

Запишем формальное представление случайного процесса  $\zeta(t)$  в виде ряда

$$\zeta(t) = \sum_{k=1}^{\infty} \zeta_k f_k(t). \quad (3.1.1)$$

Равенство (3.1.1) понимается в среднеквадратическом<sup>1</sup>, когда

$$\lim_{M \rightarrow \infty} \mathbf{E} \left[ \zeta(t) - \sum_{k=1}^M \zeta_k f_k(t) \right]^2 = 0. \quad (3.1.2)$$

Случайные величины  $\{\zeta_k\}$  представляют собой координаты случайного процесса при его разложении в ортогональном базисе  $\{f_k(t)\}$ , т. е. являются скалярным произведением процесса  $\zeta(t)$  (в смысле всех его реализаций) и  $k$ -й базисной функции:

$$\zeta_k = \int_{t_1}^{t_2} \zeta(t) f_k(t) dt. \quad (3.1.3)$$

Отметим, что базисная система функций  $\{f_k(t)\}$  изначально не задана и, следовательно, она может быть выбрана так, чтобы наиболее эффективно способствовать разрешению возникающих проблем. При рассмотрении в главе 1 различных задач, связанных с проверкой статистических гипотез (когда анализу подвергались конечномерные выборки, т. е. пространство наблюдений имело конечное число измерений), было установлено, что многие задачи решаются значительно проще, если перейти в новую систему координат, в которой случайные величины были бы некоррелированы (для нормальных величин некоррелированность означает статистическую независимость).

Пусть  $a_k = \mathbf{E}[\zeta_k]$  ( $k = 1, 2, \dots$ ) — математические ожидания координат

---

<sup>1</sup> Часто для обозначения сходимости в среднеквадратическом используется обозначение l.i.m. — от английского limit in the mean.

ортогонального разложения. Тогда, учитывая сказанное, для ковариаций величин  $\zeta_1, \zeta_2, \dots$  желательно иметь свойство

$$\mathbf{E}[(\zeta_k - a_k)(\zeta_l - a_l)] = \mathbf{E}[\zeta_k \zeta_l] - a_k a_l = \lambda_k \delta_{kl}, \quad (3.1.4)$$

где  $\{\lambda_k\}$  ( $k = 1, 2, \dots$ ) — некоторые числа, значения которых определяются корреляционными свойствами процесса  $\zeta(t)$ .

Для простоты предположим, что величины  $\zeta_1, \zeta_2, \dots$  имеют нулевые средние, что имеет место в большинстве практически интересных случаях. С учётом (3.1.3) имеем:

$$\begin{aligned} \mathbf{E}[\zeta_k \zeta_l] &= \int_{t_1}^{t_2} \int_{t_1}^{t_2} f_k(t') f_l(t'') \mathbf{E}[\zeta(t') \zeta(t'')] dt' dt'' = \\ &= \int_{t_1}^{t_2} f_k(t') \left( \int_{t_1}^{t_2} f_l(t'') \mathbf{E}[\zeta(t') \zeta(t'')] dt'' \right) dt'. \end{aligned}$$

По определению

$$\mathbf{E}[\zeta(t') \zeta(t'')] = b(t', t'')$$

является корреляционной функцией процесса  $\zeta(t)$ . Поэтому для того, чтобы соотношение (3.1.4) выполнялось при всех значениях  $k$  и  $l$ , необходимо и достаточно, чтобы внутренний интеграл был равен  $\lambda_l f_l(t')$ :

$$\int_{t_1}^{t_2} f_l(t'') b(t', t'') dt'' = \lambda_l f_l(t'), \quad (3.1.5)$$

ибо при этом коэффициент корреляции

$$\mathbf{E}[\zeta_k \zeta_l] = \int_{t_1}^{t_2} f_k(t') \lambda_l f_l(t') dt' = \lambda_l \delta_{kl},$$

т. е. имеет требуемый вид (3.1.4).

Уравнение (3.1.5) представляет собой *интегральное уравнение Фредгольма 2-го рода с ядром  $b(t', t'')$* .

По методам решения таких уравнений существует обширная литература [7]. Не вдаваясь в математические подробности, перечислим основные свойства интегральных уравнений вида

$$\int_{t_1}^{t_2} f(t) b(s, t) dt = \lambda f(s),$$

когда в качестве ядра выступает симметричная и неотрицательно определённая корреляционная функция  $b(s, t)$ .

Базовым положением теории интегральных уравнений является то, что существование нетривиальных решений возможно лишь при определённых значениях  $\lambda_k$  ( $k = 1, 2, \dots$ ) — *собственных (характеристических) чисел* интегрального уравнения. При этом всякое ненулевое ядро имеет хотя бы одно собственное число. Соответствующие решения интегрального уравнения  $f_k(s)$  называются *собственными функциями* интегрального уравнения. Для симметричных ядер собственные числа вещественны и образуют дискретный набор.

Двум различным собственным числам соответствуют ортогональные собственные функции, и собственные функции образуют полную ортонормированную систему, в которой может быть разложен любой интегрируемый в квадрате детерминированный процесс.

Может оказаться так, что одному собственному числу  $\lambda_k$  соответствует несколько собственных функций  $f_{k1}(s), \dots, f_{kN}(s)$ . При этом  $N$  — его называют *рангом* собственного числа — заведомо конечно. Тогда линейная комбинация собственных функций

$$c_{k1}f_{k1}(s) + \dots + c_{kN}f_{kN}(s)$$

также является собственной функцией.

Если собственные числа образуют конечный набор, то ядро интегрального уравнения оказывается *вырожденным*, т. е. представимо в виде

$$b(s, t) = \sum_{m=1}^n \rho_m(s) \sigma_m(t),$$

где функции  $\sigma_1(t), \dots, \sigma_n(t)$  — линейно независимы.

Итак, если найдены наборы собственных чисел  $\{\lambda_k\}$  и собственных функций  $\{f_k(t)\}$  ( $k = 1, 2, \dots$ ), удовлетворяющих уравнению (3.1.5), то обеспечивается некоррелированность координат случайного процесса в разложении Карунена–Лозва. Это свойство приобретает важную практическую значимость применительно к гауссовским процессам, для которых некоррелированность величин  $\zeta_1, \zeta_2, \dots$  означает их независимость. В этом случае совместное распределение  $M$ -элементной совокупности таких величин факторизуется и может быть записано в следующем виде:

$$w_{\zeta}^{(M)}(z_1, \dots, z_M) = \frac{1}{(2\pi)^{M/2} \prod_{k=1}^M \sqrt{\lambda_k}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^M \frac{z_k^2}{\lambda_k} \right\}. \quad (3.1.6)$$

Найдём решение интегрального уравнения (3.1.5) для некоторых модельных случаев спектральных (корреляционных) характеристик процессов.

### Процессы с рациональным спектром

Рассмотрим *стационарные процессы с рациональным спектром*, т. е. процессы, энергетический спектр которых  $G(\omega)$  может быть представлен в виде отношения двух полиномов  $P(\omega)$  и  $Q(\omega)$ :

$$G(\omega) = \frac{P(\omega)}{Q(\omega)},$$

где степень полинома  $P(\omega)$  меньше степени полинома  $Q(\omega)$ .

Стандартный метод решения в этом случае достаточно прост, хотя несколько утомителен. Интегральное уравнение преобразуется в обыкновенное дифференциальное, решение которого несложно найти; затем найденное решение подставляется в исходное интегральное уравнение для удовлетворения граничным условиям.

Например, пусть на интервале времени  $[-T/2; T/2]$  (при симметричном интервале выкладки значительно упрощаются) рассматривается стационарный процесс с энергетическим спектром

$$G(\omega) = \frac{G_0}{\omega^2 + \alpha^2}, \quad -\infty < \omega < \infty,$$

или, что эквивалентно, корреляционная функция  $b(\tau)$  имеет вид

$$b(\tau) = b_0 \exp(-\alpha |\tau|), \quad -\infty < \tau < \infty.$$

Тогда имеем следующее интегральное уравнение:

$$\int_{-T/2}^{T/2} b_0 \exp(-\alpha |s - t|) f(s) ds = \lambda f(t), \quad -T/2 \leq t \leq T/2. \quad (3.1.7)$$

Для того, чтобы избавиться от модуля в подынтегральной функции, запишем

$$\int_{-T/2}^{T/2} b_0 \exp(-\alpha |s - t|) f(s) ds = \int_{-T/2}^t b_0 \exp(-\alpha |s - t|) f(s) ds + \int_t^{T/2} b_0 \exp(-\alpha |s - t|) f(s) ds.$$

При этом  $s < t$  на интервале  $[-T/2; t]$  и  $s \geq t$  на интервале  $[t; T/2]$ , так что

$$\int_{-T/2}^t b_0 \exp(\alpha(s - t)) f(s) ds + \int_t^{T/2} b_0 \exp(-\alpha(s - t)) f(s) ds = \lambda f(t). \quad (3.1.7a)$$

Возьмём первую и вторую производную от обеих частей этого урав-



нения. Для первой производной имеем<sup>1</sup>

$$-b_0\alpha \int_{-T/2}^t \exp(\alpha(s-t))f(s)ds + b_0\alpha \int_t^{T/2} \exp(-\alpha(s-t))f(s)ds = \lambda f'(t). \quad (3.1.8)$$

Для второй производной имеем

$$b_0\alpha^2 \int_{-T/2}^t \exp(\alpha(s-t))f(s)ds - b_0\alpha f(t) + \\ + b_0\alpha^2 \int_t^{T/2} \exp(-\alpha(s-t))f(s)ds - b_0\alpha f(t) = \lambda f''(t),$$

т. е.

$$b_0\alpha^2 \int_{-T/2}^{T/2} \exp(-\alpha|s-t|)f(s)ds - 2b_0\alpha f(t) = \lambda f''(t).$$

Интегральное слагаемое в левой части, согласно исходному интегральному уравнению (3.1.7), есть просто  $\alpha^2\lambda f(t)$ , отсюда получаем следующее дифференциальное уравнение для нахождения  $f(t)$ :

$$\lambda f''(t) + (2b_0\alpha - \alpha^2\lambda)f(t) = 0, \quad (3.1.9)$$

имеющее различную структуру в зависимости от соотношений между параметрами  $\lambda$ ,  $\alpha$  и  $b_0$ .

При  $\lambda = 0$  получаем тривиальное решение в виде тождественно равной нулю собственной функции.

Если  $\lambda = 2b_0/\alpha$ , то дифференциальное уравнение имеет вид  $f''(t) = 0$ , и его общее решение с двумя неизвестными константами  $c_1$  и  $c_2$  есть

$$f(t) = c_1 t + c_2. \quad (3.1.10)$$

Когда  $\lambda > 2b_0/\alpha$ , имеет место линейное дифференциальное уравнение второго порядка вида

$$f''(t) - \mu^2 f(t) = 0, \quad \mu^2 = \frac{\alpha}{\lambda}(\lambda\alpha - 2b_0) > 0, \quad (3.1.11)$$

---

<sup>1</sup> Напомним, что производная зависящего от параметра интеграла  $I(y) = \int_{\alpha(y)}^{\beta(y)} f(x, y)dx$  равна

$$I'(y) = \int_{\alpha(y)}^{\beta(y)} f'_y(x, y)dx + \beta'(y)f(\beta(y), y) - \alpha'(y)f(\alpha(y), y).$$

с общим решением

$$f(t) = c_1 \exp(-\mu t) + c_2 \exp(\mu t). \quad (3.1.12)$$

Наконец, в тех случаях, при которых выполняется соотношение  $0 < \lambda < 2b_0 / \alpha$ , получаем также линейное дифференциальное уравнение второго порядка

$$f''(t) + \mu^2 f(t) = 0, \quad \mu^2 = \frac{\alpha}{\lambda}(2b_0 - \lambda\alpha) > 0, \quad (3.1.13)$$

с общим решением

$$f(t) = c_1 \exp(-j\mu t) + c_2 \exp(j\mu t). \quad (3.1.14)$$

Можно убедиться, что при подстановке (3.1.10), (3.1.11) и (3.1.12) в (3.1.7a) получающиеся “остаточные” члены не могут быть обнулены (ввиду различных функциональных проявлений в правой и левой частях), т. е. интегральное уравнение не может быть удовлетворено и, следовательно, оно не имеет решений.

Подставляя в (3.1.7a) решение (3.1.14) и выполняя интегрирование, имеем следующее соотношение между различными параметрами:

$$\exp(-\alpha t) \left[ \frac{c_1}{\alpha + j\mu} \exp[-(\alpha + j\mu)T] + \frac{c_2}{\alpha - j\mu} \exp[-(\alpha - j\mu)T] \right] - \\ - \exp(\alpha t) \left[ \frac{c_1}{-\alpha + j\mu} \exp[-(\alpha - j\mu)T] + \frac{c_2}{-\alpha - j\mu} \exp[-(\alpha + j\mu)T] \right] = 0, \quad (3.1.15)$$

анализ которого показывает, что оно может быть удовлетворено лишь когда  $c_1 = \pm c_2$ .

В том случае, когда  $c_1 = c_2$ , из (3.1.15) имеем уравнение

$$\frac{\exp(-j\mu T)}{\alpha + j\mu} + \frac{\exp(j\mu T)}{\alpha - j\mu} = 0,$$

приводящееся к виду

$$\operatorname{tg}(\mu T) = \alpha / \mu. \quad (3.1.16a)$$

Аналогично, когда  $c_1 = -c_2$ , получаем уравнение

$$\frac{\exp(-j\mu T)}{\alpha + j\mu} - \frac{\exp(j\mu T)}{\alpha - j\mu} = 0,$$

приводящееся к виду

$$\operatorname{tg}(\mu T) = -\alpha / \mu. \quad (3.1.16b)$$

Трансцендентные уравнения (3.1.16) могут быть решены численно,

например, итеративным алгоритмом Ньютона, или графически (табличные значения функции  $x \operatorname{tg} x$  представлены, например, в [3]). В результате решения получается упорядоченный набор значений  $\{\mu_k\}$  ( $k=1, 2, \dots$ ) и, следовательно, собственных чисел  $\{\lambda_k\}$  интегрального уравнения:

$$\lambda_k = \frac{2b_0\alpha}{\alpha^2 + \mu_k^2}. \quad (3.1.17)$$

При этом случаю  $c_1 = c_2$  соответствуют нечётные, а случаю  $c_1 = -c_2$  — чётные значения номера  $k$ .

Из (3.1.17) следует, что собственные значения  $\lambda_k$  равны значениям  $G(\omega_k)$  энергетического спектра процесса на частотах  $\omega_k$ .

Итак, согласно (3.1.14), решения интегрального уравнения — набор собственных функций  $\{f_k\}$  — есть

$$f_k(t) = \begin{cases} 2C \cos \mu_k t, & k \text{ нечётное} \\ 2jC \sin \mu_k t, & k \text{ чётное} \end{cases}, \quad t \in \left[-\frac{T}{2}; \frac{T}{2}\right].$$

Для нахождения константы  $C$  воспользуемся условием нормировки собственных функций (напомним, что  $\{f_k\}$  — ортонормированный базис):

$$\|f_k(t)\|^2 = \int_{-T/2}^{T/2} f_k^2(t) dt = 1.$$

Имеем

$$C = \begin{cases} \pm \frac{1}{\sqrt{2[T + (1/\mu_k)\sin(\mu_k T)]}}, & k \text{ нечётное;} \\ \pm \frac{1}{j\sqrt{2[T - (1/\mu_k)\sin(\mu_k T)]}}, & k \text{ чётное.} \end{cases}$$

Наличие двух знаков означает наличие двух “зеркальных” наборов собственных функций, что, по существу, никак не сказывается на сути ортогонального разложения. Для определённости выберем положительную константу, тогда окончательно имеем:

$$f_k(t) = \begin{cases} \frac{\sqrt{2} \cos \mu_k t}{\sqrt{T + (1/\mu_k)\sin(\mu_k T)}}, & k \text{ нечётное;} \\ \frac{\sqrt{2} \sin \mu_k t}{\sqrt{T - (1/\mu_k)\sin(\mu_k T)}}, & k \text{ чётное;} \end{cases} \quad t \in \left[-\frac{T}{2}; \frac{T}{2}\right]. \quad (3.1.18)$$

Анализируя полученный результат, можно сделать ряд интересных наблюдений.

Как следует из решения трансцендентных уравнений (3.1.16), при увеличении интервала наблюдения  $T$  уменьшаются значения параметра  $\mu_k$  и, следовательно, возрастают значения  $\lambda_k$ .

Собственные функции согласно (3.1.18) — гармонические, но с некратными частотами. Однако по мере увеличения значения  $\mu T$  точки  $\mu_k$  оказываются близки к значениям  $(k - 1)\pi/2$  (с учётом чётности и нечётности нижнего и верхнего рядов), так что собственные функции с большими индексами асимптотически образуют набор колебаний с кратными частотами:

$$f_k(t) \approx \begin{cases} \frac{\sqrt{2} \cos \frac{(k-1)\pi t}{2T}}{\sqrt{T + (1/\mu_k) \sin(\mu_k T)}}, & k \text{ нечётное;} \\ \frac{\sqrt{2} \sin \frac{(k-1)\pi t}{2T}}{\sqrt{T - (1/\mu_k) \sin(\mu_k T)}}, & k \text{ чётное.} \end{cases} \quad (3.1.19)$$

Когда спектр процесса моделируется функцией, не являющейся отношением двух полиномов, дифференциальное уравнение, возникающее из исходного интегрального уравнения, имеет, как правило, изменяющиеся во времени коэффициенты, и их решение представляет собой достаточно сложную математическую задачу. Тем не менее, в некоторых случаях всё же удаётся выразить решение через известные табулированные функции.

### Белый процесс в конечной полосе

Рассмотрим стационарный процесс, имеющий постоянное значение энергетического спектра  $G(\omega) = N_0/2$  в конечной полосе частот  $[-2\pi F; 2\pi F]$  и, следовательно, корреляционную функцию

$$b(\tau) = 2\pi F N_0 \frac{\sin(2\pi F \tau)}{2\pi F \tau}.$$

Тогда соответствующее интегральное уравнение имеет вид

$$\int_{-T/2}^{T/2} 2\pi F N_0 \frac{\sin[2\pi F(s-t)]}{2\pi F(s-t)} f(s) ds = \lambda f(t), \quad -T/2 \leq t \leq T/2, \quad (3.1.20)$$

и его решением являются так называемые *волновые функции вытянутого сфероида*.

Волновые функции вытянутого сфероида (иногда их называют *сжатыми сфероидами*) возникают как решение уравнения Гельмгольца

$$\Delta\Psi + k^2\Psi = 0$$

в “вытянутых сферических координатах”  $(\xi, \eta, \varphi)$ , переход в которые из декартовых координат  $(x, y, z)$  осуществляется по формулам

$$\begin{aligned} x &= \xi\eta; \\ y &= \sqrt{(\xi^2 - 1)(1 - \eta^2)} \cos \varphi; \\ z &= \sqrt{(\xi^2 - 1)(1 - \eta^2)} \sin \varphi. \end{aligned}$$

Собственные функции  $\Psi_{mn}$  для данного решения записывается в виде произведения радиальной компоненты  $R_{mn}(k/2, \xi)$ , угловой компоненты  $S_{mn}(k/2, \eta)$ , а также фазовой компоненты  $\exp(jm\varphi)$ :

$$\Psi_{mn}(\xi, \eta, \varphi) = R_{mn}(k/2, \xi)S_{mn}(k/2, \eta)\exp(jm\varphi), \quad m, n = 0, 1, 2, \dots$$

Радиальная и угловая компоненты удовлетворяют одному и тому же уравнению, но в разных областях:  $|\xi| > 1$  и  $|\eta| < 1$ :

$$\begin{aligned} (\xi^2 - 1)\frac{d^2 R_{mn}(k/2, \xi)}{d\xi^2} + 2\xi\frac{dR_{mn}(k/2, \xi)}{d\xi} - \left(\lambda_{mn}(k/2) - \frac{k^2}{4}\xi^2 + \frac{m^2}{\xi^2 - 1}\right)R_{mn}(k/2, \xi) &= 0; \\ (\eta^2 - 1)\frac{d^2 R_{mn}(k/2, \eta)}{d\eta^2} + 2\xi\frac{dR_{mn}(k/2, \eta)}{d\eta} - \left(\lambda_{mn}(k/2) - \frac{k^2}{4}\eta^2 + \frac{m^2}{\eta^2 - 1}\right)R_{mn}(k/2, \eta) &= 0, \end{aligned}$$

где  $\lambda_{mn}$  — собственное число, соответствующее собственной функции.

Итак, можно показать, что на интервале  $[-T/2; T/2]$  уравнению (3.1.20) удовлетворяют волновые функции с  $m = 0$ :

$$\begin{aligned} \int_{-T/2}^{T/2} 2\pi FN_0 \frac{\sin[2\pi F(s-t)]}{2\pi F(s-t)} S_{0n}\left(\pi FT, \frac{2s}{T}\right) ds = \\ = 2\pi N_0 FT \left[ R_{0n}\left(\pi FT, \frac{2s}{T}\right) \right]^2 S_{0n}\left(\pi FT, \frac{2s}{T}\right), \end{aligned} \quad (3.1.21)$$

откуда следует, что собственные значения рассматриваемого интегрального уравнения равны

$$\lambda_k = 2\pi N_0 FT \left[ R_{0n}\left(\pi FT, \frac{2s}{T}\right) \right]^2, \quad k = 0, 1, 2, \dots \quad (3.1.22)$$

В табл. 3.1 представлен ряд нормированных  $\lambda_k$  для двух значений параметра  $2FT$ : 2,55 и 5,10.

Таблица 3.1

**Собственные значения интегрального уравнения**

$\lambda_k/(\pi N_0)$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$2FT=2,55$	0,996	0,912	0,519	0,110	0,009	0,0004	–	–	–
$2FT=5,10$	1,000	0,999	0,997	0,961	0,748	0,321	0,061	0,006	0,0004

Из анализа представленных в табл. 3.1 данных следует, что с ростом  $FT$  собственные значения начинают концентрироваться в двух областях: вблизи единицы (с учётом нормировки), когда  $k \ll 2FT + 1$  и вблизи нуля, когда  $k \gg 2FT + 1$ .

Если ввести число  $\alpha$ , определяемое равенством

$$k = 2FT + 1 + \frac{\alpha}{\pi^2} \ln(4\pi FT),$$

то можно показать, что

$$\lim_{FT \rightarrow \infty} \left| \lambda_k(FT) - \frac{1}{1 + \exp(\alpha)} \right| = 0.$$

Другими словами, переходная область между значениями 0 и 1 имеет ширину, пропорциональную  $\ln(4\pi FT)$ . Таким образом, можно утверждать, что если процесс с ограниченным на интервале  $[-2\pi F; 2\pi F]$  спектром наблюдается на интервале  $T$ , то имеет смысл учитывать только  $2FT + 1$  собственных значений.

Рассмотренный выше процесс, имеющий постоянный спектр в полосе частот  $[-2\pi F; 2\pi F]$ , принято называть *белым процессом в конечной полосе*. Термин “белый” здесь подчёркивает тот факт, что большинство ненулевых собственных чисел (при достаточно большом  $FT$ ) имеют приблизительно одинаковые, близкие к единице значения.

**Белый процесс**

Предельная ситуация при  $F \rightarrow \infty$  приводит к модели *белого процесса в бесконечной полосе*, для которой корреляционная функция равна

$$b(\tau) = \frac{N_0}{2} \lim_{F \rightarrow \infty} \frac{\sin(2\pi F \tau)}{\pi \tau}. \quad (3.1.23)$$

Нетрудно видеть, что соотношение (3.1.23) является одним из представлений дельта-функции:

$$\lim_{F \rightarrow \infty} \frac{\sin(2\pi F \tau)}{\pi \tau} = \delta(\tau).$$

Напомним, что дельта-функция<sup>1</sup>  $\delta(x)$  является *обобщённой функцией*, возникающей как предельный элемент множества — дельтаобразующей последовательности  $\{\delta(x, \sigma)\}$ :

$$\delta(x) = \lim_{\sigma \rightarrow \sigma_0} \{\delta(x, \sigma)\},$$

где  $\sigma$  — параметр (возможно, векторный), по которому осуществляется предельный переход.

Основные формальные свойства, определяющие дельта-функцию, состоят в следующем. Дельта-функция  $\delta(x - x_0)$  для любого вещественного  $x_0$  равна нулю при  $x \neq x_0$  и неограничена при  $x = x_0$ :

$$\delta(x - x_0) = \begin{cases} 0, & x \neq x_0 \\ \infty, & x = x_0 \end{cases}.$$

Интеграл от этой функции по промежутку  $[a; b]$ , содержащему значение  $x_0$ , конечен и равен единице:

$$\int_a^b \delta(x - x_0) dx = 1.$$

Таким образом, дельта-функция описывает импульс, имеющий бесконечно малую продолжительность, бесконечно большую амплитуду и площадь, равную единице.

Кроме основных свойств дельта-функции часто придают дополнительные свойства, проистекающие из физических соображений: чётность  $\delta(x) = \delta(-x)$  и единичная средняя мощность

$$\lim_{c \rightarrow \infty} \frac{1}{c} \int_{-c/2}^{c/2} \delta^2(x) dx = 1.$$

Далее приведён ряд наиболее часто используемых дельтаобразующих последовательностей.

Последовательность прямоугольных импульсов единичной площади:

$$\delta(x, \tau) = \begin{cases} 1/\tau, & x_0 \leq x \leq x_0 + \tau; \\ 0, & x < x_0, x > x_0 + \tau; \end{cases} \text{ при } \tau \rightarrow 0.$$

Последовательность гауссовских импульсов:

$$\delta(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \text{ при } \sigma \rightarrow 0.$$

Последовательность функций вида  $\sin(x)/x$ :

---

<sup>1</sup> Дельта-функцию часто связывают с именем П. Дирака

$$\delta(x, \lambda) = \frac{\sin \lambda x}{\pi x} \text{ при } \lambda \rightarrow \infty.$$

Последовательность колоколообразных импульсов:

$$\delta(x, \lambda) = \frac{1}{\pi \lambda} \frac{1}{1 + (x / \lambda)^2} \text{ при } \lambda \rightarrow 0.$$

Производные от дельта-функции определяются как пределы соответствующих производных от дельтаобразующих последовательностей. Так, например, используя преобразование Фурье последовательности гауссовских импульсов, можно записать следующее выражение для  $m$ -й производной дельта-функции:

$$\delta^{(m)}(x) = \frac{(-jx)^m}{2\pi} \int_{-\infty}^{\infty} \exp(-jtx) dt, \quad m = 0, 1, \dots$$

При этом выполняется “фильтрующее” свойство дельта-функции и её производных:

$$\int_a^b g(x) \delta^{(m)}(x - x_0) dx = (-1)^m g^{(m)}(x_0).$$

Нетрудно видеть, что интегральное уравнение для дельтообразной корреляционной функции

$$\int_{-T/2}^{T/2} \frac{N_0}{2} \delta(s - t) f(s) ds = \lambda f(t), \quad -T/2 \leq t \leq T/2 \quad (3.1.24)$$

выполняется автоматически для любой собственной функций  $f(t)$ . При этом все собственные числа оказываются одинаковыми и равными  $N_0/2$ . Таким образом, ортогональное разложение случайного процесса (3.1.1), имеющего постоянное значение энергетического спектра во всей полосе частот или, эквивалентно, дельтообразную корреляционную функцию (по этой причине такой процесс часто называют *дельтакоррелированным*), может быть осуществлено в произвольном ортонормированном базисе  $\{f_k(t)\}$ , например, в тригонометрическом (экспоненциальном) базисе Фурье-разложения. Этот факт имеет скорее принципиальный, нежели конкретный конструктивный характер, поскольку непосредственное формирование базисных функций не понадобится.

Гораздо более значимым является то, что все собственные числа одинаковы, т. е. коэффициент корреляции двух произвольных координат

$$\mathbf{E}[\zeta_k \zeta_l] = \begin{cases} 0, & k \neq l; \\ N_0 / 2, & k = l; \end{cases}$$



не зависит от индексов базисных функций. Другими словами, дисперсия случайных величин  $\zeta_1, \zeta_2, \dots$  одинакова:

$$\mathbf{D}[\zeta_k] = \mathbf{E}[\zeta_k^2] = N_0 / 2, \quad (k = 1, 2, \dots).$$

Ещё раз подчеркнём, что случайные величины  $\zeta_k$  — это не отсчётные значения  $\zeta(t_k)$  процесса  $\zeta(t)$  в моменты времени  $t_k$ , дисперсия которых имеет размерность квадрата размерности физических значений процесса, а координаты процесса в ортогональном разложении, размерность которых может отличаться от размерности  $\zeta(t_k)$ .

Если модель белого процесса в конечной полосе может в некотором приближении служить адекватной моделью реальных физических процессов, то модель белого процесса в бесконечной полосе является теоретической абстракцией, поскольку определяет процесс с бесконечной энергией (как сумму квадратов собственных чисел). Кроме того, ясно, что достаточно близкие значения физических процессов всегда будут иметь отличную от нуля корреляцию. Несмотря на это, именно такая модель — модель белого процесса используется в большинстве теоретических рассуждений, поскольку удобство свойств собственных функций и чисел компенсирует неудобства, вызванные некоторой неадекватностью физическим реалиям.

В действительности, корректное использование понятия белого процесса не приводит к сколь-нибудь существенным противоречиям.

Белый процесс, как и дельта-функция, принадлежит к классу так называемых *обобщённых функций* и получает реальный физический смысл только в подынтегральных выражениях функционалов. Если  $\xi(t)$  — стационарный процесс с постоянным (во всей полосе) спектром, то для произвольной функции  $g(t)$ , интегрируемой в квадрате на интервале  $[t_1; t_2]$ , интеграл

$$\int_{t_1}^{t_2} g(t)\xi(t)dt$$

по определению полагается нормальной случайной величиной с нулевым математическим ожиданием и дисперсией, равной

$$\frac{N_0}{2} \int_{t_1}^{t_2} g^2(t)dt.$$

Отсюда следует, что процесс  $\eta(t)$ , определяемый интегральным преобразованием

$$\eta(t) = \int_{t_1}^{t_2} g(t - \tau)\xi(\tau)dt,$$

является “обычным” стационарным нормальным процессом с конечной дисперсией и корреляционной функцией

$$\mathbf{E}[\eta(t)\eta(t + \tau)] = \frac{N_0}{2} \int_{t_1}^{t_2} g(t)g(t + \tau)dt.$$

Таким образом, стационарный белый процесс, пройдя через линейную цепь с постоянными параметрами, превращается в стационарный нормальный процесс.

“Нестационарный белый процесс” (обобщённый процесс) проще всего определить как произведение стационарного белого шума на некоторую детерминированную функцию. Тогда процесс  $\eta(t)$  будет иметь зависящие от времени математическое ожидание и дисперсию.

Можно определить и другие обобщённые процессы, имеющие бесконечную дисперсию и дельтаобразную корреляционную функцию и отличающиеся от белого процесса своим поведением в подынтегральных выражениях. Но в современной теории связи успешно используется только определённая выше конструкция.

В математических моделях с белым процессом нужно обращаться осторожно. Его нельзя возводить в квадрат, нельзя перемножать два белых процесса от одного аргумента и т. д. Такие операции лишены смысла также, как, например, возведение в квадрат дельта-функции.

### **Винеровский процесс**

В заключение данного раздела рассмотрим ортогональное разложение Карунена — Лоэва применительно к нестационарным процессам, в частности, к винеровскому процессу.

Винеровский процесс  $\gamma(t)$ , являющийся моделью броуновского движения, представляет собой нормальный однородный процесс с независимыми приращениями, т. е. образованные в моменты времени  $0 < t_1 < t_2 \leq t_3 < t_4$  приращения  $\gamma(t_4) - \gamma(t_3)$  и  $\gamma(t_2) - \gamma(t_1)$  являются независимыми нормальными величинами, а распределение приращений  $\gamma(t_2) - \gamma(t_1)$  (и, следовательно, моментные характеристики) зависит лишь от разности временных аргументов. При этом предполагается<sup>1</sup>, что  $\gamma(t)$  имеет нулевое среднее  $\mathbf{E}[\gamma(t)] = 0$  и, вообще говоря, зависящую от времени дисперсию  $\mathbf{D}[\gamma(t)] = \sigma^2(t)$ .

Для моментов времени  $t_2 > t_1$  дисперсия процесса с независимыми приращениями равна

---

<sup>1</sup> Обычно также считают, что  $\gamma(0) = 0$ , т. е. процесс в начальный момент времени выходит из начала координат.

$$\begin{aligned}\mathbf{D}[\gamma(t_1 + t_2)] &= \mathbf{D}[(\gamma(t_1) - \gamma(0)) + (\gamma(t_1 + t_2) - \gamma(t_1))] = \\ &= \mathbf{D}[\gamma(t_1)] + \mathbf{D}[(\gamma(t_1 + t_2) - \gamma(t_1))] = \sigma^2(t_1) + \sigma^2(t_2).\end{aligned}$$

Отсюда следует, что дисперсия является линейной функцией от  $t$ :

$$\sigma^2(t) = \sigma_0^2 t, \quad t \geq 0, \quad (3.1.25)$$

где константа  $\sigma_0^2$  называется *коэффициентом диффузии*.

Определим корреляционную функцию  $b(t_1, t_2)$  винеровского процесса. Имеем при  $t_2 > t_1$ :

$$\begin{aligned}b(t_1, t_2) &= \mathbf{E}[\gamma(t_1)\gamma(t_2)] \equiv \mathbf{E}[\gamma(t_1)(\gamma(t_2) - \gamma(t_1) + \gamma(t_1))] = \\ &= \mathbf{E}[\gamma(t_1)(\gamma(t_2) - \gamma(t_1)) + \gamma^2(t_1)] = \mathbf{E}[\gamma(t_1)(\gamma(t_2) - \gamma(t_1))] + \mathbf{E}[\gamma^2(t_1)].\end{aligned}$$

Поскольку приращения<sup>1</sup>  $\gamma(t_2) - \gamma(t_1)$  и  $\gamma(t_1)$  независимы,

$$\mathbf{E}[\gamma(t_1)(\gamma(t_2) - \gamma(t_1))] = \mathbf{E}[\gamma(t_1)]\mathbf{E}[(\gamma(t_2) - \gamma(t_1))] = 0,$$

отсюда

$$b(t_1, t_2) = \mathbf{E}[\gamma^2(t_1)] = \sigma_0^2 t_1.$$

Аналогично при  $t_2 < t_1$

$$b(t_1, t_2) = \mathbf{E}[\gamma^2(t_2)] = \sigma_0^2 t_2.$$

Таким образом, обобщая оба случая, можно записать:

$$b(t_2, t_1) = \mathbf{E}[\gamma^2(t_2)] = \sigma_0^2 \min[t_1, t_2]. \quad (3.1.26)$$

Подставляя (3.1.26) в (3.1.5), на интервале  $[-T/2; T/2]$  получаем следующее интегральное уравнение:

$$\begin{aligned}\lambda f(t_1) &= \int_{-T/2}^{T/2} \sigma_0^2 \min[t_2, t_1] f(t_2) dt_2 = \sigma_0^2 \int_{-T/2}^{t_1} t_2 f(t_2) dt_2 + \\ &+ \sigma_0^2 \int_{t_1}^{T/2} t_1 f(t_2) dt_2 = \sigma_0^2 \int_{-T/2}^{t_1} t_2 f(t_2) dt_2 + t_1 \sigma_0^2 \int_{t_1}^{T/2} f(t_2) dt_2.\end{aligned} \quad (3.1.27)$$

Аналогично тому, как это делалось для случая процесса с рациональным спектром, продифференцируем (3.1.27) дважды. Имеем:

$$\lambda f'(t_1) = \sigma_0^2 \int_{t_1}^{T/2} f(t_2) dt_2, \quad (3.1.28)$$

---

<sup>1</sup>  $\gamma(t_1)$  можно трактовать как приращение  $\gamma(t_1) - 0 \equiv \gamma(t_1) - \gamma(0)$ .

$$\lambda f''(t_1) = -\sigma_0^2 f(t_1), \quad (3.1.29)$$

отсюда получаем дифференциальное уравнение

$$\lambda f''(t) + \sigma_0^2 f(t) = 0 \quad (3.1.30)$$

с общим решением

$$f(t) = c_1 \exp\left(j \frac{\sigma_0}{\sqrt{\lambda}} t\right) + c_2 \exp\left(-j \frac{\sigma_0}{\sqrt{\lambda}} t\right). \quad (3.1.31)$$

Подстановка (3.1.31) в уравнение (3.1.28) для первой производной приводит к простому комплексному соотношению

$$c_1 \exp\left(j \frac{\sigma_0 T}{2\sqrt{\lambda}}\right) = c_2 \exp\left(-j \frac{\sigma_0 T}{2\sqrt{\lambda}}\right),$$

эквивалентному двум вещественным соотношениям

$$\begin{aligned} c_1 \cos \frac{\sigma_0 T}{2\sqrt{\lambda}} &= c_2 \cos \frac{\sigma_0 T}{2\sqrt{\lambda}}; \\ c_1 \sin \frac{\sigma_0 T}{2\sqrt{\lambda}} &= -c_2 \sin \frac{\sigma_0 T}{2\sqrt{\lambda}}. \end{aligned}$$

Поскольку  $c_1$  и  $c_2$  одновременно не могут равняться нулю, возможны два варианта одновременного выполнения соотношений, дающие связь между  $c_1$  и  $c_2$  и определяющие условия на собственные числа:

$$c_1 = c_2, \quad \sin \frac{\sigma_0 T}{2\sqrt{\lambda}} = 0,$$

и

$$c_1 = -c_2, \quad \cos \frac{\sigma_0 T}{2\sqrt{\lambda}} = 0.$$

Отсюда, используя, как обычно, для нахождения неизвестной константы условие нормировки, получаем два набора собственных чисел и собственных функций интегрального уравнения:

$$\lambda_k = \frac{\sigma_0^2 T^2}{4\pi^2 k^2}, \quad f_k(t) = \sqrt{2/T} \cos(2\pi k t / T), \quad k = 0, 1, \dots, \quad (3.1.32a)$$

и

$$\lambda_k = \frac{\sigma_0^2 T^2}{4\pi^2 (k + 1/2)^2}, \quad f_k(t) = \sqrt{2/T} \sin(2\pi(k + 1/2)t / T), \quad k = 0, 1, \dots, \quad (3.1.32б)$$

Отметим, что формальное дифференцирование корреляционной функции винеровского процесса путём вычисления частных производных по  $t_1$  и  $t_2$  приводит к появлению дельта-функции:

$$\frac{\partial^2 b(t_1, t_2)}{\partial t_1 \partial t_2} = \sigma_0^2 \frac{\partial^2}{\partial t_1 \partial t_2} \min(t_1, t_2) = \sigma_0^2 \delta(t_1 - t_2)$$

Из теории случайных процессов известно, что смешанная производная корреляционной функции некоторого процесса определяет корреляционную функцию процесса, представляющего собой производную (в среднеквадратическом) от такого процесса. Отсюда следует, что производная винеровского процесса — это белый шум.

### 3.2. ФУНКЦИОНАЛ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ И АЛГОРИТМЫ ПРИЁМА ДЛЯ МОДЕЛИ АДДИТИВНОГО БЕЛОГО ГАУССОВСКОГО ШУМА

Вычисление отношения правдоподобия (1.1.9) с последующим предельным переходом и получение функционала отношения правдоподобия связано с факторизацией многомерного условного распределения.

Получим вид функционала отношения правдоподобия и соответствующие алгоритмы обработки сигнала  $z(t)$  для модели АБГШ:

$$z(t) = \mu s_r(t) + n(t), \quad r = 1, \dots, m,$$

когда действующий в канале шум  $n(t)$  представим в виде набора гауссовских некоррелированных, а, следовательно, и независимых случайных величин  $n_1, n_2, \dots$ , являющихся координатами ортогонального разложения Карунена–Лозва процесса  $n(t)$  и описываемые совместной плотностью распределения

$$w_\zeta^{(M)}(n_1, \dots, n_M) = \frac{1}{(\pi N_0)^{M/2}} \exp \left\{ -\frac{1}{N_0} \sum_{k=1}^M n_k^2 \right\}. \quad (3.2.1)$$

При этом коэффициент передачи канала  $\mu$  считается постоянным на интервале анализа  $[0; T_a]$ . Отметим, что в зависимости от используемых алгоритмов обработки длительность интервала анализа может как совпадать с интервалом  $[0; T]$  формирования сигналов  $s_r(t)$ , так и отличаться от него. В

дальнейшем, если это особо не оговаривается, будем считать, что интервал анализа совпадает с интервалом формирования.

Результатом работы алгоритма должно являться решение относительно того, какой из множества возможных сигналов  $s_1(t), \dots, s_m(t)$  присутствует на входе приёмника — это формулировка общей задачи различения сигналов. Полагая  $m = 2$  и тождественно равным нулю один из двух сигналов, получаем задачу обнаружения сигналов. Несмотря на то, что задача обнаружения формально может трактоваться как частный случай задачи различения, между этими задачами всё же есть принципиальное отличие, заключающееся в используемой стратегии оптимальности.

В задачах различения рассматривается, как правило, равновероятный ансамбль сигналов, что предопределяет использование какой-либо байесовской стратегии с единичным порогом, например, стратегии, оптимальной по критерию максимума правдоподобия (МП-стратегии).

В задачах обнаружения обычно имеет место несимметричная ситуация: априорная вероятность появления полезного сигнала существенно меньше вероятности его отсутствия. В этой связи целесообразно использование стратегии, оптимальной по критерию Неймана — Пирсона (НП-стратегии), когда задаётся (достаточно малая) вероятность ложной тревоги и максимизируется вероятность правильного обнаружения.

Тем не менее, и в том, и в другом случае в основе решения задачи лежит вычисление функционала отношения правдоподобия, что позволяет их решать с общих позиций, конкретизируя, при необходимости, какие-либо особенности.

Итак, функционал отношения правдоподобия  $\Lambda_{rl}(\mathbf{z})$  для конкурирующих гипотез по отношению к сигналам  $s_r(t)$  и  $s_l(t)$  при факторизуемых распределениях имеет вид

$$\Lambda_{rl}[z(t)] = \lim_{M \rightarrow \infty} \frac{w^{(M)}(\mathbf{z} / s_r(t))}{w^{(M)}(\mathbf{z} / s_l(t))} = \lim_{M \rightarrow \infty} \frac{\prod_{k=1}^M w(z_k / s_r(t))}{\prod_{k=1}^M w(z_k / s_l(t))}, \quad r, l = 1, \dots, m. \quad (3.2.2)$$

Разложим на интервале анализа  $[0; T]$  и передаваемый сигнал  $s_r(t)$ , и наблюдаемое колебание  $z(t)$ , и действующий в канале шум  $n(t)$  в одном и

том же ортогональном базисе  $\{f_k(t)\}$ , получаемом из решения интегрального уравнения (3.1.7), причём для модели АБГШ в качестве базисного может выступать любой ортонормированный набор функций:

$$n(t) = \sum_{k=1}^M n_k f_k(t), \quad z(t) = \sum_{k=1}^M z_k f_k(t), \quad s_r(t) = \sum_{k=1}^M s_{rk} f_k(t), \quad (3.2.3)$$

где, соответственно,

$$n_k = \int_0^T n(t) f_k(t) dt, \quad z_k = \int_0^T z(t) f_k(t) dt, \quad s_{rk} = \int_0^T s_r(t) f_k(t) dt \quad (3.2.4)$$

являются координатами разложения, т. е. скалярным произведением процесса и базисной функции. При этом  $\{s_{rk}\}$  — детерминированные, а  $\{z_k\}$  и  $\{n_k\}$  — случайные величины, связанные между собой соотношением

$$z_k = \mu s_{rk} + n_k, \quad r = 1, \dots, m. \quad (3.2.5)$$

Ключевым моментом при вычислении (3.2.2) является переход от условных плотностей  $w(z_k / s_r(t))$  к условным плотностям  $w(n_k / s_r(t))$  на основании того, что при фиксированном сигнале  $s_r(t)$  функциональный вид распределения случайных величин  $z_k$  и  $n_k$ , как это видно из (3.2.5), совпадает, различаясь, возможно, в параметрах. Так, если  $n_k$  — это гауссовские величины с нулевым средним и дисперсией  $N_0/2$ , то  $z_k$  также имеют гауссовское распределение с той же дисперсией, но с математическим ожиданием, равным  $\mu s_{rk}$ . Таким образом, можно записать следующее выражение для функционала отношения правдоподобия:

$$\begin{aligned} \Lambda_{rl}[z(t)] &= \lim_{M \rightarrow \infty} \frac{\prod_{k=1}^M \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(z_k - \mu s_{rk})^2}{N_0}\right]}{\prod_{k=1}^M \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(z_k - \mu s_{lk})^2}{N_0}\right]} = \\ &= \frac{\prod_{k=1}^{\infty} \exp\left[-\frac{(z_k - \mu s_{rk})^2}{N_0}\right]}{\prod_{k=1}^{\infty} \exp\left[-\frac{(z_k - \mu s_{lk})^2}{N_0}\right]} = \frac{\exp\left[-\frac{1}{N_0} \sum_{k=1}^{\infty} (z_k - \mu s_{rk})^2\right]}{\exp\left[-\frac{1}{N_0} \sum_{k=1}^{\infty} (z_k - \mu s_{lk})^2\right]}, \quad r, l = 1, \dots, m. \end{aligned} \quad (3.2.6)$$

Учитывая наличие в (3.2.6) экспонент, целесообразно, аналогично то-

му, как это делалось для  $M$ -элементной выборки, перейти к логарифму функционала отношения правдоподобия:

$$\ln \Lambda_{rl}[z(t)] = -\frac{1}{N_0} \sum_{k=1}^{\infty} (z_k - \mu s_{rk})^2 + \frac{1}{N_0} \sum_{k=1}^{\infty} (z_k - \mu s_{lk})^2,$$

или, после упрощения,

$$\ln \Lambda_{rl}[z(t)] = \frac{2\mu}{N_0} \sum_{k=1}^{\infty} z_k (s_{rk} - s_{lk}) - \frac{\mu^2}{N_0} \sum_{k=1}^{\infty} (s_{rk}^2 - s_{lk}^2) \quad r, l = 1, \dots, m. \quad (3.2.7)$$

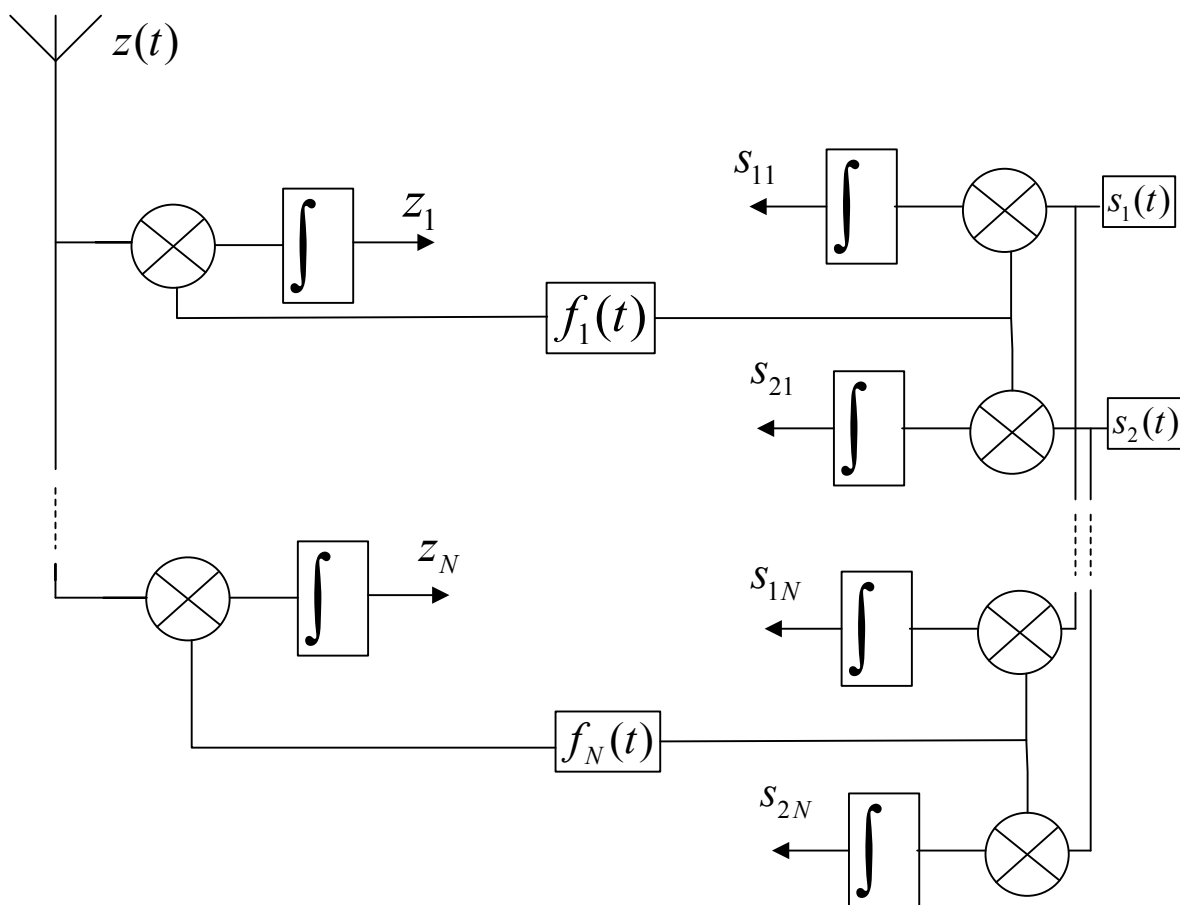


Рис. 3.1. Схема практической реализации коэффициентов ортогонального разложения для двоичных сигналов

Практическая реализация функционала  $\Lambda_{rl}[z(t)]$  согласно (3.2.6) или (3.2.7), предполагает вычисление, вообще говоря, бесконечного числа коэффициентов  $\{z_k\}$ ,  $\{s_{rk}\}$  и  $\{s_{lk}\}$ . В принципе можно ограничиться конечным, достаточно большим числом  $N$  таких коэффициентов, вычисляя их на ос-



нове умножителей, интеграторов и произвольного (конечного) набора базисных функций  $\{f_k(t)\}$  (рис. 3.1). Однако имеется возможность существенного упрощения реализации  $\Lambda_{rl}[z(t)]$  на основании того, что дискретные суммы, содержащие  $\{z_k\}$ ,  $\{s_{rk}\}$  и  $\{s_{lk}\}$ , могут быть представлены в эквивалентной интегральной форме на основании соотношений

$$\begin{aligned} \sum_{k=1}^{\infty} z_k s_{rk} &= \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} z_k s_{rn} \delta_{kn} = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} z_k s_{rn} \int_0^T f_k(t) f_n(t) dt = \\ &= \int_0^T \left( \sum_{k=1}^{\infty} z_k f_k(t) \right) \left( \sum_{n=1}^{\infty} s_{rn} f_n(t) \right) dt = \int_0^T z(t) s_r(t) dt \end{aligned}$$

и, аналогично,

$$\sum_{k=1}^{\infty} s_{rk}^2 = \int_0^T s_r^2(t) dt.$$

Тогда имеем следующее выражение для функционала отношения правдоподобия:

$$\begin{aligned} \ln \Lambda_{rl}[z(t)] &= \frac{2\mu}{N_0} \int_0^T z(t) [s_r(t) - s_l(t)] dt - \frac{\mu^2}{N_0} \int_0^T [s_r^2(t) - s_l^2(t)] dt, \\ &r, l = 1, \dots, m, \end{aligned} \quad (3.2.8)$$

которое в зависимости от используемой стратегии должно быть сравнено со значением  $\ln \Lambda_0$ , где  $\Lambda_0$  — порог принятия решения.

На рис. 3.2 показана структурная схема, реализующая логарифм функционала отношения правдоподобия на основе (3.2.8) для двоичных сигналов. Видно, что в отличие от аналогичной схемы на рис. 3.1, в которой основная “тяжесть” лежит на вычислении коэффициентов разложения сигналов  $z(t)$  и  $s_r(t)$  по некоторому ортогональному базису, основу данной схемы составляет так называемый *корреляционный интеграл* вида

$$\int_0^T z(t) s_r(t) dt,$$

где в качестве подынтегрального выражения фигурирует произведение анализируемого колебания  $z(t)$  на “опорный” сигнал  $s_r(t)$ .

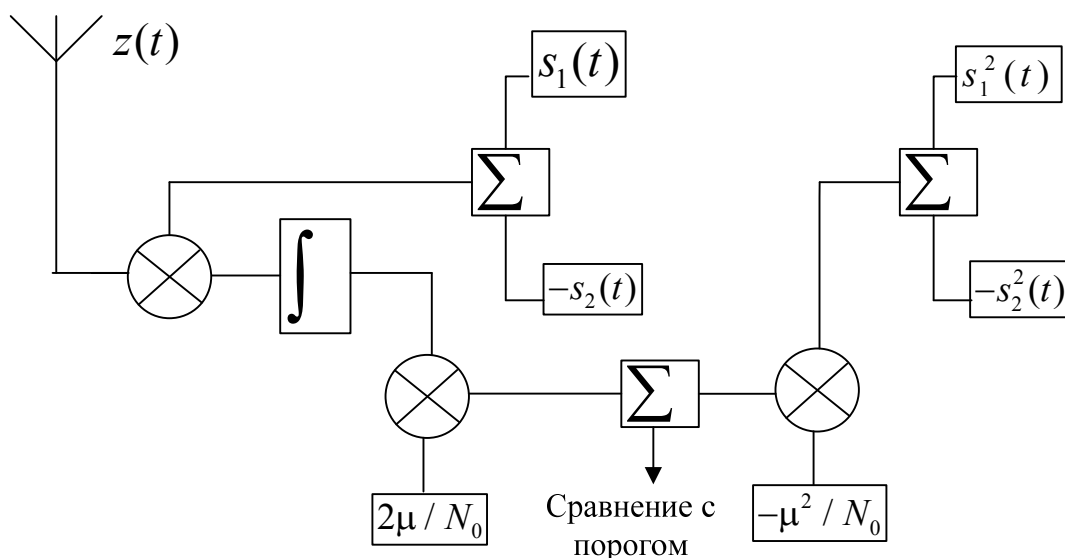


Рис. 3.2. Структурная схема, реализующая логарифм функционала отношения правдоподобия для двоичных сигналов

В общем случае реализация функционала отношения правдоподобия предполагает вычисление, например, посредством предварительного анализа, коэффициента передачи канала  $\mu$  и спектральной плотности средней мощности шума  $N_0$ .

Нетрудно видеть, что функционал (3.2.8) для двоичных сигналов является аналогом отношения правдоподобия (1.1.13), полученного в задаче распознавания сигналов на основе независимых отсчётов с той лишь разницей, что вместо “локальных” параметров — амплитуд полезных сигналов и дисперсии шума фигурируют “интегральные” параметры: мощности полезных сигналов и спектральная плотность средней мощности шума.

Рассмотрим далее некоторые типичные задачи, связанные с оптимальным приёмом сигналов.

### 3.2.1. Обнаружение одиночного сигнала

Как уже говорилось в разд. 1.1, для таких задач характерна определённая асимметрия, заключающаяся в том, что априорные вероятности наличия или отсутствия полезного сигнала на входе приёмного устройства заметно различаются между собой. Более того, определить или, хотя бы, оценить априорные вероятности, как правило, оказывается весьма затруднительно. В этих условиях целесообразно использование стратегии, оптимальной по критерию Неймана–Пирсона, когда фиксируется достаточно

малая, (например,  $10^{-3}$ ) вероятность ложной тревоги  $P_{ЛТ}$  и максимизируется вероятность правильного обнаружения.

Положим, что рассматривается функционал отношения правдоподобия (3.2.8) для двоичного случая, и один из сигналов, скажем,  $s_2(t)$  тождественно равен нулю. Требуется определить на основании наблюдения на интервале  $[0; T]$  присутствует ли на входе приёмника смесь полезного сигнала и аддитивного шума

$$z(t) = \mu s_1(t) + n(t)$$

(гипотеза  $H_1$ ) или же анализируемое колебание представляет собой “чистый” шумовой процесс

$$z(t) = n(t)$$

(гипотеза  $H_2$ ).

Алгоритм оптимального обнаружения имеет следующий вид: *выносятся решение в пользу гипотезы  $H_1$  о наличии полезного сигнала  $s_1(t) \equiv s(t)$ , если выполняется неравенство*

$$\ln \Lambda_{12}[z(t)] = \frac{2\mu}{N_0} \int_0^T z(t)s(t)dt - \frac{\mu^2}{N_0} \int_0^T s^2 dt > \ln \Lambda_{\text{НП}} , \quad (3.2.9)$$

где пороговое значение  $\Lambda_{\text{НП}}$  определяется интегральным соотношением

$$\int_{\Lambda_{\text{НП}}}^{\infty} w_{\zeta}(z / H_2) dz = P_{\text{ЛТ}} , \quad (3.2.10)$$

в котором  $w_{\zeta}(z / H_2)$  — плотность распределения анализируемого колебания на входе приёмника при условии отсутствия в нём полезного сигнала.

Заметим, что входящее в (3.2.9) слагаемое  $\int_0^T s^2(t)dt$  — это энергия  $E$  сигнала  $s(t)$ , т. е. величина, известная на приёмной стороне. Тогда в алгоритме обнаружения можно перейти к эквивалентному неравенству

$$\int_0^T z(t)s(t)dt > \Lambda'_{\text{НП}} , \quad (3.2.9a)$$

где модифицированный порог принятия решения

$$\Lambda'_{\text{НП}} = \frac{N_0}{2\mu} \ln \Lambda_{\text{НП}} + \frac{\mu}{2} E . \quad (3.2.11)$$

На рис. 3.3 показана структурная схема обнаружителя сигнала, содержащая генератор опорного сигнала  $s(t)$ , перемножитель, интегратор, а так-

же решающее устройство, на один вход которого подаётся пороговое значение, а на другой — значение с выхода интегратора.

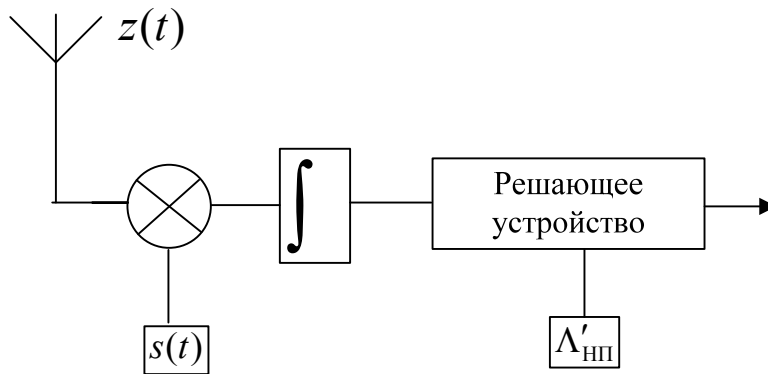


Рис. 3.3. Структурная схема обнаружителя одиночного сигнала

Рассмотрим помехоустойчивость полученного алгоритма обнаружения, определив рабочие характеристики, аналогичные тем, что были получены в разд. 1.1.

Пусть на входе приёмника отсутствует полезный сигнал:  $z(t) = n(t)$ . Тогда вероятность выполнения неравенства

$$\int_0^T n(t)s(t)dt > \Lambda'_{\text{НП}}$$

представляет собой вероятность ложной тревоги (выносится решение в пользу наличия сигнала, когда в действительности его нет). При этом левая часть неравенства является случайной величиной

$$\xi = \int_0^T n(t)s(t)dt,$$

в виде интеграла от гауссовского процесса, имеющего нулевое среднее и одностороннюю спектральную плотность средней мощности  $N_0/2$ . Поскольку интегральное преобразование — это линейная операция, не изменяющая закона распределения, можно утверждать, что  $\xi$  — гауссовская величина с нулевым математическим ожиданием:

$$\mathbf{E}[\xi] = \int_0^T \mathbf{E}[n(t)]s(t)dt = \int_0^T 0 \cdot s(t)dt = 0.$$

Найдём дисперсию  $\xi$ . Имеем:

$$\begin{aligned} \mathbf{D}[\xi] = \mathbf{E}[\xi^2] &= \mathbf{E} \left[ \left( \int_0^T n(t)s(t)dt \right)^2 \right] = \mathbf{E} \left[ \int_0^T n(t_1)s(t_1)dt_1 \int_0^T n(t_2)s(t_2)dt_2 \right] = \\ &= \int_0^T \int_0^T \mathbf{E}[n(t_1)n(t_2)]s(t_1)s(t_2)dt_1dt_2. \end{aligned}$$

Как было показано в разд. 3.1, корреляционная функция белого гауссовского шума представляет собой дельта-функцию:

$$\mathbf{E}[n(t_1)n(t_2)] \equiv b(t_1, t_2) = \frac{N_0}{2} \delta(t_1 - t_2).$$

Тогда

$$\mathbf{D}[\xi] = \frac{N_0}{2} \int_0^T s^2(t)dt = \frac{N_0 E}{2}. \quad (3.2.12)$$

Таким образом, фигурирующая в (3.2.10) подынтегральная функция  $w_\zeta(z/H_2)$  — это плотность распределения гауссовской случайной величины с нулевым средним и дисперсией (3.2.12). Отсюда получаем связь между вероятностью ложной тревоги и порогом принятия решения:

$$\begin{aligned} P_{\text{лт}} &= \frac{1}{\sqrt{2\pi(N_0 E / 2)}} \int_{\Lambda'_{\text{лт}}}^{\infty} \exp\left(-\frac{z^2}{2(N_0 E / 2)}\right) dz = \\ &= \frac{1}{\sqrt{2\pi}} \int_{\Lambda'_{\text{лт}} \sqrt{2/(N_0 E)}}^{\infty} \exp(-y^2 / 2) dy = \\ &= 1 - F\left(\Lambda'_{\text{лт}} \sqrt{\frac{2}{N_0 E}}\right) = 1 - F\left(\ln \Lambda_{\text{лт}} \sqrt{\frac{N_0}{2\mu^2 E}} + \sqrt{\frac{\mu^2 E}{2N_0}}\right), \end{aligned}$$

где, как обычно,  $F(x)$  — функция Лапласа, т. е. интегральная функция стандартного (с нулевым средним и единичной дисперсией) гауссовского распределения.

Введём параметр

$$h^2 = \frac{\mu^2 E}{N_0}, \quad (3.2.13)$$

представляющий собой отношение энергии сигнала (с учётом коэффициента передачи канала) на входе приёмного устройства к спектральной плотности шума. Его называют<sup>1</sup> *отношением сигнал-шум на входе приёмного устройства* или, короче, просто *сигнал-шум*. Тогда можно записать связь между вероятностью ложной тревоги и порогом принятия решения:

$$P_{\text{ЛТ}} = 1 - F\left(\frac{\ln \Lambda_{\text{НП}}}{h\sqrt{2}} + \frac{h}{\sqrt{2}}\right)$$

или, выражая  $\ln \Lambda_{\text{НП}}$  через  $P_{\text{ЛТ}}$ :

$$\ln \Lambda_{\text{НП}} = h\sqrt{2}\left[F^{-1}(1 - P_{\text{ЛТ}}) - h/\sqrt{2}\right], \quad (3.2.14a)$$

где  $F^{-1}$  — функция, обратная функции Лапласа.

В практических задачах, связанных с оптимальной обработкой сигналов, вероятностные характеристики помехоустойчивости обычно выражают не через функцию Лапласа, а через родственные ей функции, такие, как *функция Крампа*

$$\Phi(x) = \sqrt{\frac{2}{\pi}} \int_0^x \exp(-y^2/2) dy$$

или так называемая *функция ошибок*<sup>2</sup>

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) dy.$$

Переход от этих функций к функции Лапласа осуществляется по следующим формулам:

$$\Phi(x) = 2F(x) - 1;$$

$$\text{erf}(x) = 2F(x\sqrt{2}) - 1.$$

Выбор либо функции Крампа, либо функции ошибок для количественных расчётов целесообразно осуществлять, исходя из того, какая из них (возможно, обе) реализована в используемом для расчётов программ-

<sup>1</sup> Часто в литературе можно встретить другие определения отношения сигнал-шум на входе приёмного устройства, например, в [9] отношение сигнал-шум определяется как  $h^2 = 2E/N_0$ .

<sup>2</sup> Название функции исторически произошло от статистических задач обработки ошибок при различных измерениях, Применительно к обработке сигналов оно является не совсем удачным, однако прочно укоренилось в литературе.

ном обеспечении. В настоящее время де-факто универсальной программной средой, причём, как для учебных, так и для профессиональных инженерных целей является система MATLAB<sup>1</sup>, в которой в качестве функции, родственной функции Лапласа, реализована функция erf(x). Связь между вероятностью ложной тревоги и порогом принятия решения при использовании функции ошибок имеет вид

$$\ln \Lambda_{\text{НП}} = 2h \left[ \text{erf}^{-1}(1 - 2P_{\text{ЛТ}}) - h/2 \right], \quad (3.2.14б)$$

где erf<sup>-1</sup> — функция, обратная функции ошибок (функции, обратные erf(x) и Φ(x), также реализованы в соответствующих пакетах).

Зависимости логарифма порогового значения от вероятности ложной тревоги при различных значениях отношения сигнал-шум представлена на рис. 3.3.

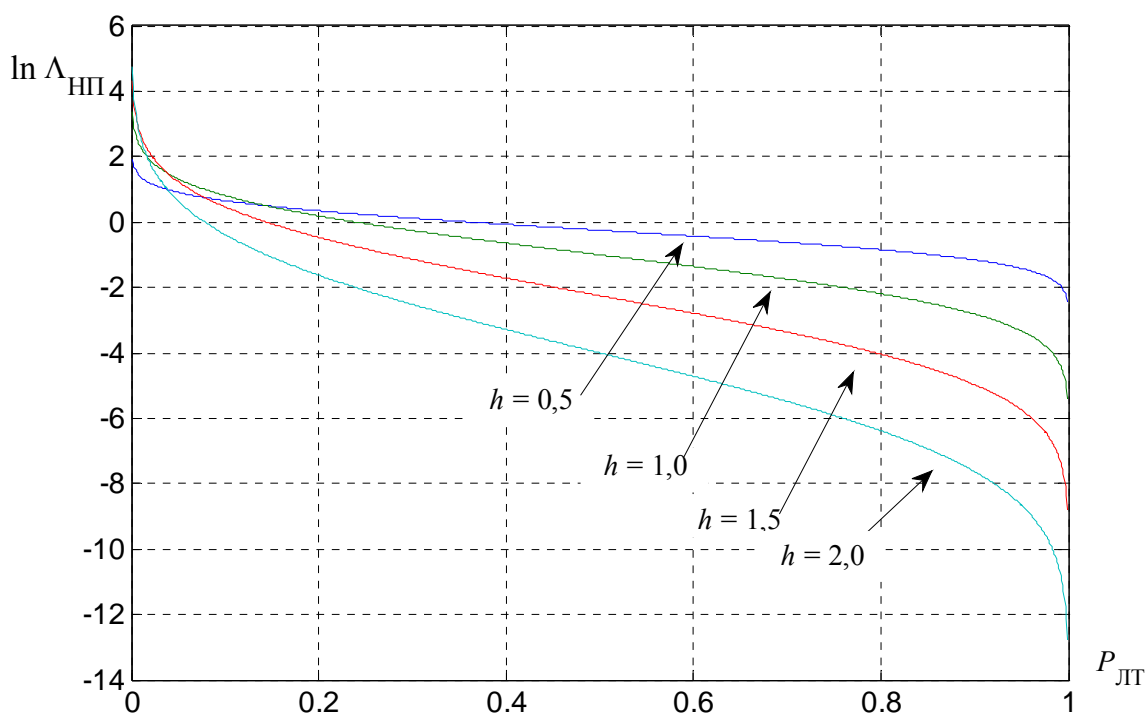


Рис. 3.3. Зависимость логарифма порогового значения от вероятности ложной тревоги

<sup>1</sup> MATLAB (MATrix LABoratory) — высокоуровневый интерпретируемый язык программирования, разработанный Кливом Моулером (Cleve Moler), а также одноимённый пакет прикладных программ, предназначенный для решения инженерных задач в различных научно-технических областях.

В практических задачах, связанных с обнаружением сигналов, реальные значения вероятности ложной тревоги лежат в диапазоне  $10^{-3} \dots 10^{-1}$ , так что рабочим участком является левый край характеристики на рис. 3.3.

Для вычисления вероятности  $P_{\text{ПО}}$  правильного обнаружения положим, что  $z(t) = s(t) + n(t)$ , т. е. полезный сигнал присутствует на входе приёмника. Тогда искомая вероятность представляет собой вероятность выполнения неравенства

$$\int_0^T [\mu s(t) + n(t)] s(t) dt > \Lambda'_{\text{НП}},$$

или, перенося в правую часть энергию,

$$\int_0^T n(t) s(t) dt > \Lambda'_{\text{НП}} - \mu E.$$

В левой части по-прежнему находится гауссовская величина  $\xi$ , имеющая нулевое среднее и дисперсию  $N_0 E / 2$ . Следовательно,

$$\begin{aligned} P_{\text{ПО}} &= \frac{1}{\sqrt{2\pi(N_0 E / 2)}} \int_{\Lambda'_{\text{НП}} - \mu E}^{\infty} \exp\left(-\frac{z^2}{2(N_0 E / 2)}\right) dz = \\ &= 1 - F\left((\Lambda'_{\text{НП}} - \mu E) \sqrt{\frac{2}{N_0 E}}\right) = 1 - F\left(\ln \Lambda_{\text{НП}} \sqrt{\frac{N_0}{2\mu^2 E}} - \frac{1}{\sqrt{2}} \sqrt{\frac{\mu^2 E}{N_0}}\right). \end{aligned}$$

В обозначениях отношения сигнал-шум окончательно имеем

$$P_{\text{ПО}} = 1 - F\left(\frac{\ln \Lambda_{\text{НП}}}{h\sqrt{2}} - h / \sqrt{2}\right) \quad (3.2.15a)$$

или

$$P_{\text{ПО}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{\ln \Lambda_{\text{НП}}}{2h} - h / 2\right). \quad (3.2.15b)$$

Как следует из полученных соотношений, помехоустойчивость оптимального обнаружителя в условиях модели АБГШ не зависит от формы полезного сигнала, а определяется лишь отношением порогового уровня к значению сигнал-шум. По заданной вероятности ложной тревоги на основании (3.2.14) определяется пороговый уровень, а затем по (3.2.15) — вероятность правильного обнаружения.

На рис. 3.4 представлены зависимости вероятности правильного обнаружения от отношения сигнал-шум (их часто называют *кривыми обнару-*



жения) при различных значениях вероятности ложной тревоги.

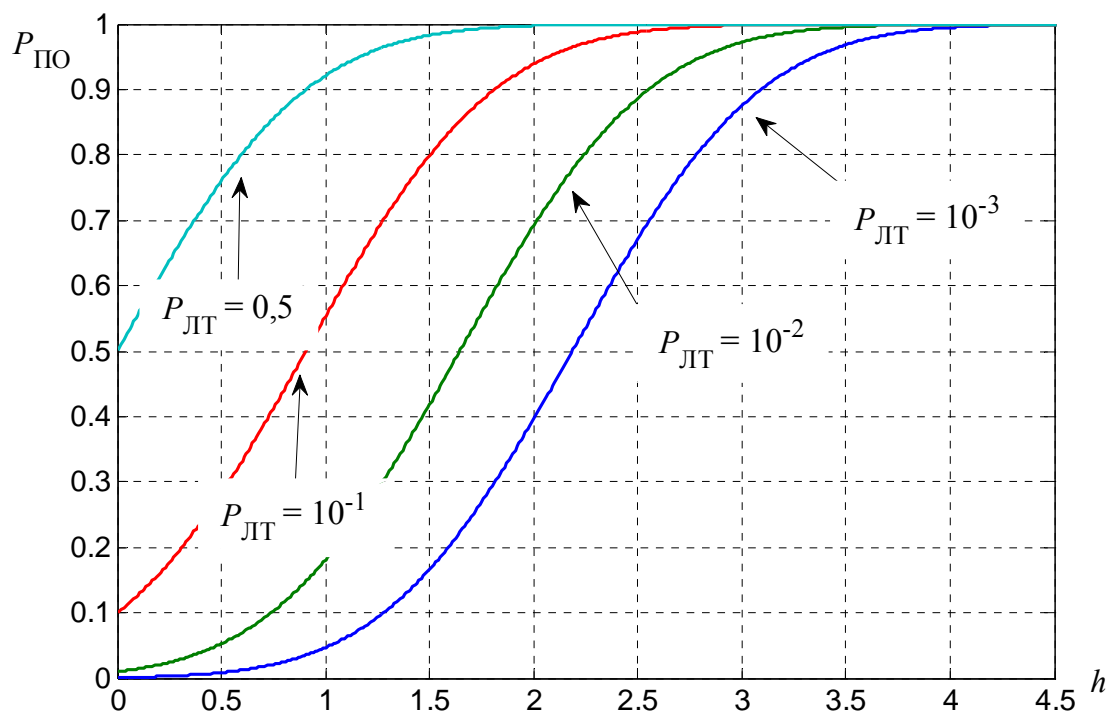


Рис. 3.4. Зависимости вероятности правильного обнаружения от отношения сигнал-шум

В (3.2.14) и (3.2.15) вероятности  $P_{ЛТ}$  и  $P_{ПО}$  выражены через безразмерные величины (порог принятия решения, отношение сигнал-шум). На практике, когда речь идёт о реальной радиоэлектронной аппаратуре, устанавливают “физический” порог на основе мощности (дисперсии) шума *на выходе* приёмника (напомним, что в модели АБГШ дисперсия шума на входе бесконечна). Для того чтобы выразить вероятностные характеристики обнаружения через дисперсию шума  $\sigma_{\text{ВЫХ}}^2$  на выходе рассмотрим алгоритм оптимального обнаружения с точки зрения его возможной схемотехнической реализации.

Как уже было сказано, в основе оптимального приёма лежит вычисление корреляционного интеграла

$$\int_0^T z(t)s(t)dt,$$

определяющего на интервале анализа  $[0; T]$  функцию взаимной корреляции

принимаемого колебания  $z(t)$  и опорного сигнала  $s(t)$ . В многочисленной литературе, посвящённой реализации устройств оптимальной обработки сигналов [10, 11], указывается на два основных способа реализации такого интеграла: на основе *коррелятора* и с помощью *согласованного фильтра*.

Коррелятор — это устройство, формирующее на своём выходе в различные моменты времени сигнал  $u_{\text{ВЫХ}}(t)$  с выхода интегратора:

$$u_{\text{ВЫХ}}(t) = \int_0^t z(\tau)s(\tau)d\tau.$$

Тогда значение  $u_{\text{ВЫХ}}(t)$  в момент времени  $t = T$  даёт искомый корреляционный интеграл. По существу, представленная на рис. 3.3 часть структурной схемы приёмника (без решающего устройства) является коррелятором, поэтому зачастую такой приёмник называют корреляционным.

Согласованный фильтр определяется как линейное устройство, обеспечивающее на своём выходе максимальное значение сигнал-шум, когда на входе присутствует аддитивная смесь полезного сигнала и помехи. В случае модели АБГШ импульсная характеристика  $h(t)$  такого устройства пропорциональна зеркалированной форме входного сигнала  $s_{\text{ВХ}}(t)$ :

$$h(t) = ks_{\text{ВХ}}(t_0 - t) \quad (3.2.16a)$$

(отсюда и название фильтра, поскольку он “согласован” по форме с входным сигналом), а частотная характеристика  $H(\omega)$  пропорциональна комплексно-сопряжённому спектру входного сигнала  $S_{\text{ВХ}}(\omega)$ :

$$H(\omega) = kS_{\text{ВХ}}^*(\omega)\exp(-j\omega t_0). \quad (3.2.16b)$$

Параметр  $t_0$  определяет момент времени, в который фиксируется значение сигнала  $s_{\text{ВЫХ}}(t)$  на выходе согласованного фильтра. Для финитных сигналов  $t_0$  принимается равным длительности входного сигнала  $s_{\text{ВХ}}(t)$ .

Сигнал на выходе фильтра определяется свёрткой входного сигнала с импульсной характеристикой:

$$\begin{aligned} u_{\text{ВЫХ}}(t) &= z(t) * h(t) = \int_0^t z(t - \tau)h(\tau)d\tau = k \int_0^t z(t - \tau)s_{\text{ВХ}}(t_0 - \tau)d\tau = \\ &= k \int_{t_0-t}^{t_0} z(\tau' + t - t_0)s_{\text{ВХ}}(\tau')d\tau'. \end{aligned} \quad (3.2.17)$$

Если считать, что  $t_0 = T$ , а в качестве  $s_{\text{ВХ}}(t)$  фигурирует обнаруживаемый

(опорный) сигнал  $s(t)$ , то получаем

$$u_{\text{ВЫХ}}(T) = k \int_0^T z(t)s(t)dt,$$

т. е. с точностью до множителя искомый корреляционный интеграл.

Из представленных выше соотношений следует, что и коррелятор, и согласованный фильтр при реализации алгоритмов оптимальной обработки выполняют одну и ту же операцию — вычисление корреляционного интеграла; в этом смысле они эквивалентны. Тем не менее, считается, что согласованный фильтр является “более общим” понятием, поэтому в дальнейшем вопросы реализации алгоритмов будут обсуждаться в большей степени применительно к согласованному фильтру.

Когда на вход согласованного фильтра (3.2.16) воздействует аддитивная смесь полезного сигнала и шума

$$z(t) = s(t) + n(t),$$

напряжение на выходе  $u_{\text{ВЫХ}}(t)$  на основании (3.2.17) можно представить в виде суммы полезной  $u_s(t)$  и шумовой  $u_n(t)$  компонент:

$$u_{\text{ВЫХ}}(t) = u_s(t) + u_n(t),$$

где

$$u_s(t) = k \int_{t_0-t}^{t_0} s(\tau' + t - t_0)s(\tau')d\tau';$$

$$u_n(t) = k \int_{t_0-t}^{t_0} n(\tau' + t - t_0)s(\tau')d\tau'.$$

Найдём дисперсию шумовой компоненты.

Входной шум  $n(t)$  согласно модели АБГШ — это стационарный процесс с нулевым средним и спектральной плотностью  $G(\omega) = N_0/2$ . Тогда  $u_n(t)$  представляют собой реализации также стационарного процесса с нулевым средним, дисперсия которого определяется как

$$\sigma_{\text{ВЫХ}}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) |H(\omega)|^2 d\omega = \frac{N_0}{2} \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega.$$

Согласно соотношению Парсеваля, вычисление характеристик линейных устройств в частотной области может быть осуществлено во временной

области на основе операций с соответствующими временными аналогами. В данном случае

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega = \int_0^{\infty} h^2(t) dt,$$

и

$$\sigma_{\text{ВЫХ}}^2 = \frac{N_0}{2} k^2 \int_0^{\infty} s^2(T-t) dt = \frac{N_0}{2} k^2 \int_{-\infty}^T s^2(t) dt.$$

Из физических соображений следует, что  $s(t) = 0$  при  $t < 0$ , поэтому

$$\sigma_{\text{ВЫХ}}^2 = \frac{N_0}{2} k^2 \int_0^T s^2(t) dt = \frac{N_0 E}{2} k^2.$$

Полагая  $k = 1$ , получаем

$$\sigma_{\text{ВЫХ}}^2 = N_0 E / 2.$$

Такое же значение, согласно (3.2.12), имеет дисперсия гауссовской случайной величины

$$\xi = \int_0^T n(t) s(t) dt.$$

Теперь можно выразить отношение сигнал-шум на входе согласованного фильтра, осуществляющего оптимальное обнаружение, через значение дисперсии выходного шума:

$$h^2 = \frac{\mu^2 E}{N_0} = \frac{2\mu^2 E N_0}{N_0^2 2} = \frac{2\mu^2 \sigma_{\text{ВЫХ}}^2}{N_0^2}. \quad (3.2.18)$$

### 3.2.2. Различение сигналов

Задачи различения сигналов являются центральными для всей теории связи. Их особенность состоит в том, что априорные вероятности гипотез — появления различных сигналов на входе приёмник — как правило, одинаковы. Кроме того, в таких задачах представляется естественным и обоснованным использование стратегии идеального наблюдателя, что в итоге приводит к единичному порогу  $\Lambda_0$ . Таким образом, логарифм функционала отношения (3.2.8) сравнивается с нулевым значением.

Например, в задаче различения двоичных сигналов  $s_1(t)$  и  $s_2(t)$ , имеющих энергии  $E_1$  и  $E_2$  соответственно, алгоритм оптимального приёма имеет

следующий вид: *выносится решение в пользу гипотезы  $H_1$  о наличии на входе сигнала  $s_1(t)$ , если выполняется неравенство*

$$\int_0^T z(t)s_1(t)dt - \frac{\mu E_1}{2} > \int_0^T z(t)s_2(t)dt - \frac{\mu E_2}{2}; \quad (3.2.19)$$

*в противном случае выносится решение в пользу гипотезы  $H_2$  о наличии на входе сигнала  $s_2(t)$ .*

Для произвольной пары используемых сигналов алгоритм оптимального различения предполагает вычисление двух корреляционных интегралов с опорными сигналами  $s_1(t)$  и  $s_2(t)$ , а также оценку коэффициента передачи канала  $\mu$ , например, посредством предварительного анализа состояния канала (рис. 3.5).

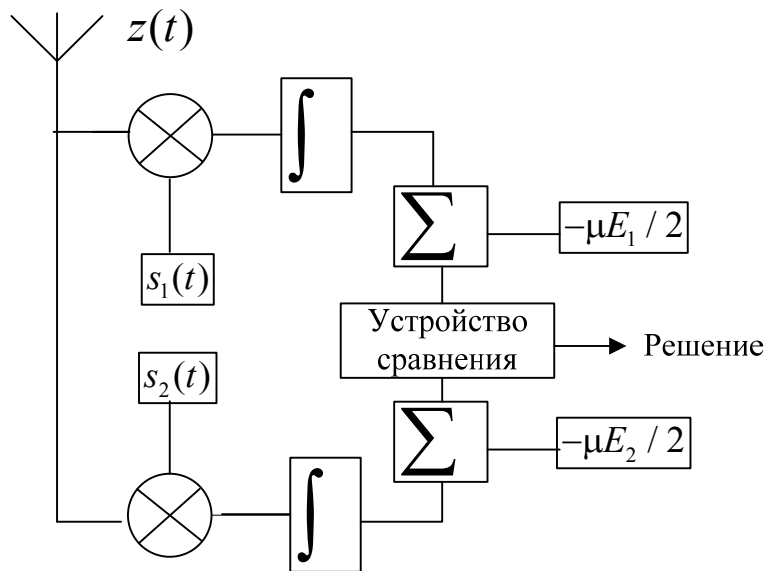


Рис. 3.5. Структурная схема оптимального устройства различения двоичных сигналов

Алгоритм (3.2.19) и, соответственно, структурная схема реализующего устройство упрощается при использовании равномошных сигналов, для которых  $E_1 = E_2 = E$ . В этом случае внеинтегральные слагаемые взаимно уничтожаются и алгоритм различения принимает следующий вид: *выносится решение в пользу гипотезы  $H_1$  о наличии на входе сигнала  $s_1(t)$ , если выполняется неравенство*

$$\int_0^T z(t)s_1(t)dt > \int_0^T z(t)s_2(t)dt; \quad (3.2.20)$$

в противном случае выносится решение в пользу гипотезы  $H_2$  о наличии на входе сигнала  $s_2(t)$ .

Структурная схема устройства, реализующего алгоритм (3.2.20), представлена на рис. 3.6. Подчеркнём, что для реализации алгоритма (3.2.20), в отличие от (3.2.19), не требуется априорного знания коэффициента передачи  $\mu$  и энергий принимаемых сигналов.

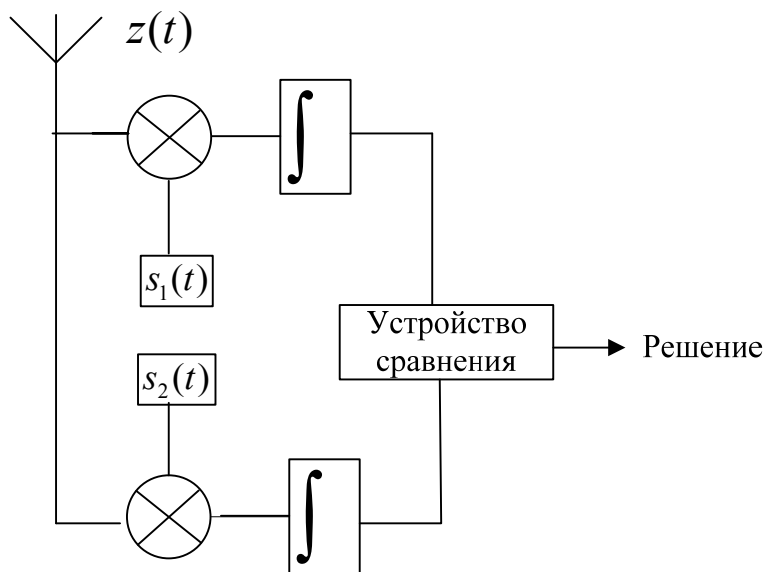


Рис. 3.6. Структурная схема оптимального устройства различения двоичных равномошных сигналов

Для получения алгоритмов оптимального различения сигналов многопозиционного ансамбля ( $m > 2$ ) обратимся к результатам решения многоальтернативной задачи проверки статистических гипотез (разд. 1.4). Согласно (1.4.10), среди всех конкурирующих гипотез выбирается та, для которой максимизируется отношение правдоподобия анализируемой выборки. Применительно к анализу непрерывных сигналов алгоритм оптимального различения можно сформулировать в следующем виде: *выносится решение в пользу гипотезы  $H_k$  о наличии на входе приёмного устройства сигнала  $s_k(t)$ ,  $k = 1, \dots, m$ , для которой*

$$\int_0^T z(t)s_k(t)dt - \frac{\mu E_k}{2} > \int_0^T z(t)s_i(t)dt - \frac{\mu E_i}{2}, \text{ для всех } i \neq k. \quad (3.2.21)$$

Другими словами, оптимальный приёмник выносит решение в пользу сигнала  $s_k(t)$ , максимизирующего величину

$$\int_0^T z(t)s_k(t)dt - \frac{\mu E_k}{2}, k = 1, \dots, m.$$

Для упрощения формулировки в литературе часто используется следующая запись:

$$s_k(t) = \arg \max_i \left[ \int_0^T z(t)s_i(t)dt - \frac{\mu E_i}{2} \right], k = 1, \dots, m.$$

Структурная схема устройства, реализующего алгоритм (3.2.21), представлена на рис. 3.7. Нетрудно видеть, что такая схема идентична схеме на рис. 3.5 с той лишь разницей, что на решающее устройство подаётся не две, а, в общем случае,  $m$  параллельных ветвей, в каждой из которых производится вычисление корреляционного интеграла.

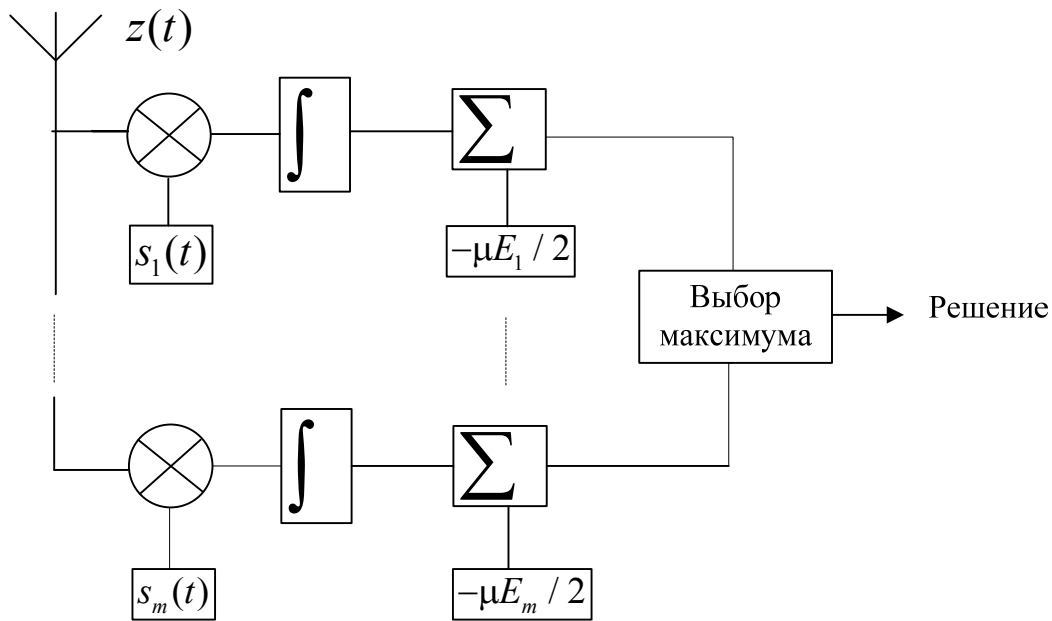


Рис. 3.7. Структурная схема оптимального устройства различения многопозиционных сигналов

Как и для двоичной гипотезы, алгоритм заметно упрощается (не требуется знания коэффициента передачи канала и энергий принимаемых сигналов) при использовании равномошных сигналов. В этом случае

$$s_k(t) = \arg \max_i \left[ \int_0^T z(t)s_i(t)dt \right], k = 1, \dots, m. \quad (3.2.22)$$

Структурная схема устройства, реализующего алгоритм (3.2.22), представлена на рис. 3.8.

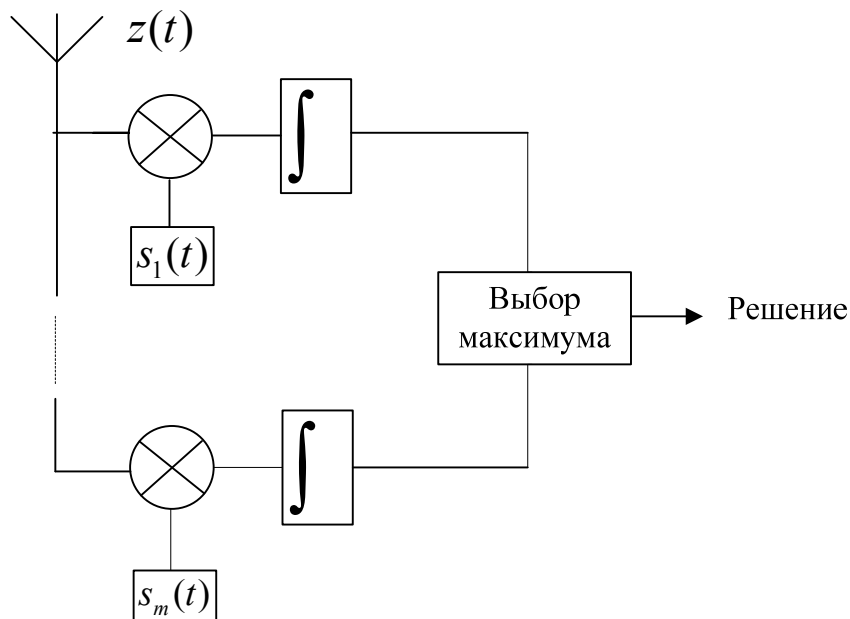


Рис. 3.8. Структурная схема оптимального устройства различения многопозиционных равномоощных сигналов

Рассмотрим помехоустойчивость оптимальных алгоритмов различения сигналов.

Основной вероятностной характеристикой в задачах различения сигналов является вероятность ошибочного различения, т. е. вероятность  $p_{\text{ош}k}$  принятия решения в пользу сигнала  $s_k(t)$ , когда в действительности на входе приёмного устройства присутствует сигнал  $s_i(t)$ ,  $i \neq k$ . Тогда при известных априорных вероятностях  $p(s_k)$  появления сигналов средняя по всему используемому ансамблю вероятность ошибочного различения  $p_{\text{ош}}$ , или, как её называют кратко, *вероятность ошибки* равна

$$p_{\text{ош}} = \sum_{k=1}^m p_{\text{ош}k} p(s_k).$$

Как уже неоднократно говорилось, в связанных задачах априорные вероятности появления сигналов, как правило, одинаковы и равны  $1/m$ , тогда

$$p_{\text{ош}} = \frac{1}{m} \sum_{k=1}^m p_{\text{ош}k}. \quad (3.2.23)$$

Более того, исходя из условий передачи сигналов по физической среде, определяемой выбранным каналом связи (электромагнитные или акусти-



ческие волны, ток или напряжение в кабеле и др.), вероятности ошибочного различения  $p_{\text{ошк}}$  обычно одинаковы и равны средней вероятности  $p_{\text{ош}}$ .

Начнём с двоичного случая.

Пусть в смеси с аддитивным шумом в действительности передавался сигнал  $s_1(t)$ :

$$z(t) = \mu s_1(t) + n(t),$$

но, например, в результате локального выброса шума в устройстве различения оказалось, что на рассматриваемом интервале справедливо неравенство

$$\int_0^T z(t) s_1(t) dt - \frac{\mu E_1}{2} \leq \int_0^T z(t) s_2(t) dt - \frac{\mu E_2}{2},$$

так что, следуя алгоритму, было принято ошибочное решение в пользу сигнала  $s_2(t)$ . Вероятность такого события и есть вероятность ошибочного различения. Для её нахождения подставим в обе части неравенства анализируемый процесс  $z(t)$ , содержащий полезный сигнал  $s_1(t)$ :

$$\int_0^T (\mu s_1(t) + n(t)) s_1(t) dt - \frac{\mu E_1}{2} \leq \int_0^T (\mu s_1(t) + n(t)) s_2(t) dt - \frac{\mu E_2}{2}$$

и представим это неравенство в виде, при котором в одной части неравенства находятся случайные слагаемые, а в другой — детерминированные. Имеем:

$$\int_0^T n(t)(s_1(t) - s_2(t)) dt \leq -\frac{\mu}{2} \int_0^T (s_1(t) - s_2(t))^2 dt. \quad (3.2.24)$$

Левая часть неравенства (3.2.24) представляет собой интеграл от стационарного гауссовского процесса, т. е. гауссовскую случайную величину

$$\zeta = \int_0^T n(t)(s_1(t) - s_2(t)) dt,$$

математическое ожидание которой

$$\mathbf{E}[\zeta] = \int_0^T \mathbf{E}[n(t)](s_1(t) - s_2(t)) dt = \int_0^T 0 \cdot (s_1(t) - s_2(t)) dt = 0,$$

а дисперсия, аналогично (3.2.12),

$$\begin{aligned}
\mathbf{D}[\zeta] &= \mathbf{E} \left[ \int_0^T n(t_1)(s_1(t_1) - s_2(t_1)) dt_1 \int_0^T n(t_2)(s_1(t_2) - s_2(t_2)) dt_2 \right] = \\
&= \int_0^T \int_0^T \mathbf{E}[n(t_1)n(t_2)](s_1(t_1) - s_2(t_1))(s_1(t_2) - s_2(t_2)) dt_1 dt_2 = \\
&= \frac{N_0}{2} \int_0^T (s_1(t) - s_2(t))^2 dt. \tag{3.2.25}
\end{aligned}$$

Таким образом, вероятность выполнения неравенства (3.2.24) — это вероятность того, что гауссовская случайная величина с нулевым средним и дисперсией (3.2.25) окажется не больше детерминированной величины, стоящей в правой части:

$$p_{\text{ош1}} = P\left(\zeta \leq -\frac{\mu}{2} \Delta_{12}\right), \tag{3.2.26}$$

где “энергия разностного сигнала”

$$\Delta_{12} = \int_0^T (s_1(t) - s_2(t))^2 dt.$$

С геометрической точки зрения  $\Delta_{12}$  представляет собой квадрат эвклидова расстояния в функциональном пространстве сигналов. В этой связи сами сигналы часто интерпретируют как точки в соответствующем функциональном пространстве. Как будет видно далее,  $\Delta_{12}$  играет главную роль при определении помехоустойчивости алгоритмов различения.

Хотя  $\Delta_{12}$  является, строго говоря, квадратом эвклидова расстояния, тем не менее, в дальнейшем для краткости часто будем называть эту величину просто эвклидовым расстоянием.

Выражая дисперсию  $\zeta$  через  $\Delta_{12}$ :

$$\mathbf{D}[\zeta] = \frac{N_0}{2} \Delta_{12},$$

из (3.2.26) получаем:

$$\begin{aligned}
p_{\text{ош1}} &= \int_{-\infty}^{-\mu\Delta_{12}/2} w_{\zeta}(z) dz = \frac{1}{\sqrt{2\pi\mathbf{D}[\zeta]}} \int_{-\infty}^{-\mu\Delta_{12}/2} \exp\left(-\frac{z^2}{2\mathbf{D}[\zeta]}\right) dz = \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\mu\Delta_{12}/(2\sqrt{\mathbf{D}[\zeta]})} \exp(-y^2/2) dy = 1 - F\left(\mu\sqrt{\frac{\Delta_{12}}{2N_0}}\right), \tag{3.2.27a}
\end{aligned}$$

где, по-прежнему,  $F$  — функция Лапласа.

Выражение вероятности ошибки через функцию ошибок имеет следующий вид:

$$P_{\text{ош1}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left( \frac{\mu}{2} \sqrt{\frac{\Delta_{12}}{N_0}} \right). \quad (3.2.27б)$$

Такое же значение имеет вероятность  $p_{\text{ош2}}$  ошибочного различение второго сигнала и, следовательно, средняя вероятность ошибки  $p_{\text{ош}}$ . В литературе зависимость вероятности ошибки от энергетических параметров полезного сигнала и канального шума (фактически — от отношения таких параметров) называется *кривая помехоустойчивости*.

Соотношение (3.2.27), определяющее помехоустойчивость произвольной пары сигналов, содержит три параметра, два из которых — коэффициент передачи канала  $\mu$  и спектральная плотность канального шума  $N_0$  считаются фиксированными на интервале анализа. Значение третьего параметра  $\Delta_{12}$  может изменяться в зависимости от того, какие конкретно выбраны сигналы. Следовательно, возможно существование такой пары сигналов, для которых обеспечивается минимальная вероятность ошибки — так называемая *потенциальная помехоустойчивость*.

Будем полагать, что потенциальная помехоустойчивость достигается в классе равномошных сигналов (использование двоичных сигналов с разными энергиями лишено практического смысла). Тогда, поскольку функция  $\operatorname{erf}$  монотонно возрастает, минимальное значение (3.2.27) обеспечивается выбором максимального значения аргумента, отсюда

$$s_2(t) = -s_1(t),$$

т. е. *потенциальная помехоустойчивость достигается при использовании противоположных сигналов*. При этом минимальная вероятность ошибки равна

$$P_{\text{ош min}} = 1 - F \left( \mu \sqrt{\frac{2E}{N_0}} \right) = 1 - F(h\sqrt{2}) \quad (3.2.28a)$$

или

$$P_{\text{ош min}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left( \mu \sqrt{\frac{E}{N_0}} \right) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}(h). \quad (3.2.28б)$$

Примерами противоположных сигналов, рассматриваемых на интер-

вале  $[0; T]$  являются:

- *разнополярные видеосигналы:*

$$s_1(t) = A_0;$$

$$s_2(t) = -A_0;$$

- *радиосигналы с фазовой манипуляцией на  $180^\circ$ :*

$$s_1(t) = A_0 \cos(\omega_0 t);$$

$$s_2(t) = A_0 \cos(\omega_0 t + \pi) = -A_0 \cos(\omega_0 t).$$

Отметим, что в обоих случаях алгоритм приёма предельно упрощается: выносится решение о наличии на входе приёмного устройства сигнала  $s_1(t)$ , если

$$\int_0^T z(t)s_1(t)dt > 0, \quad (3.2.29)$$

т. е. фактически алгоритм сводится к определению знака корреляционного интеграла. Структурная схема (рис. 3.9) содержит лишь одну ветвь и устройство определения знака.

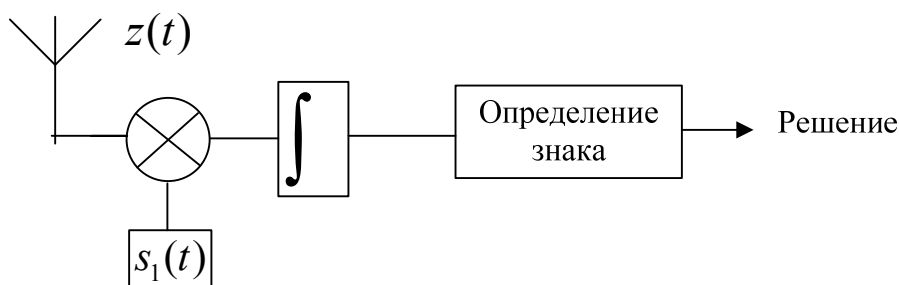


Рис. 3.9. Структурная схема оптимального устройства различения противоположных сигналов

Также, как и для задач обнаружения сигналов, помехоустойчивость алгоритмов различения зависит от энергии используемых сигналов и не зависит от их формы. Следовательно, одна и та же помехоустойчивость может быть обеспечена и простыми сигналами, например, отрезками гармонических колебаний, и сложными шумоподобными сигналами с большой базой<sup>1</sup>.

<sup>1</sup> Базой сигнала называется произведение длительности сигнала на эффективную полосу занимаемых частот. Сигналы с большой базой не имеют сколько-нибудь заметных преимуществ в задачах передачи данных по каналу с АБГШ, однако обладают ин-

Несмотря на то, что противоположные сигналы обеспечивают минимальную вероятность ошибки среди всех двоичных сигналов, практический интерес представляют и другие сигналы. Запишем квадрат эвклидова расстояния в виде

$$\Delta_{12} = \int_0^T [s_1(t) - s_2(t)]^2 dt = E_1 + E_2 - 2r_{12}\sqrt{E_1E_2},$$

где  $r_{12}$  — коэффициент корреляции<sup>1</sup> сигналов  $s_1(t)$  и  $s_2(t)$ , равный

$$r_{12} = \frac{1}{\sqrt{E_1E_2}} \int_0^T s_1(t)s_2(t)dt. \quad (1.2.30)$$

Тогда вероятность ошибки (3.2.27) для равномошных сигналов можно записать как

$$p_{\text{ош}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(h\sqrt{1-r_{12}} / \sqrt{2}\right), \quad (3.2.31)$$

Для противоположных сигналов коэффициент корреляции равен  $-1$ , и из (3.2.31) следует выражение (3.2.28) для потенциальной помехоустойчивости.

Сигналы, для которых коэффициент корреляции равен нулю, называются *ортогональными*. Для таких сигналов кривая помехоустойчивости принимает вид

$$p_{\text{ош орт}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(h / \sqrt{2}\right). \quad (3.2.32)$$

Приведём некоторые примеры сигналов, ортогональных на интервале  $[0; T]$ .

- Радиосигналы с фазовой манипуляцией на  $90^\circ$ :

$$\begin{aligned} s_1(t) &= A_0 \cos(\omega_0 t); \\ s_2(t) &= A_0 \cos(\omega_0 t + \pi / 2) = -A_0 \sin(\omega_0 t). \end{aligned}$$

В данном случае

$$E_1 = \int_0^T A_0^2 \cos^2(\omega_0 t) dt = \frac{A_0^2 T}{2} + \frac{A_0^2 T}{4\omega_0 T} \sin(2\omega_0 T).$$

тересными свойствами, позволяющими их использовать в многолучевых каналах, а также при организации механизмов множественного доступа.

<sup>1</sup> Данное понятие не следует путать с вероятностным (статистическим) коэффициентом корреляции.

Поскольку  $\omega_0$  — несущая частота, значение  $\omega_0 T$  много больше единицы. Это означает, что вторым слагаемым можно пренебречь, и

$$E_1 = E_2 \approx A_0^2 T / 2.$$

Тогда

$$r_{12} = -\frac{2}{A_0^2 T} A_0^2 \int_0^T \cos(\omega_0 t) \sin(\omega_0 t) dt = \frac{\cos(2\omega_0 T) - 1}{2\omega_0 T}.$$

Коэффициент корреляции при произвольном соотношении между  $\omega_0$  и  $T$ , вообще говоря, не равен нулю, однако ввиду того, что  $\omega_0 T \gg 1$ , абсолютное числовое значение  $r_{12}$  достаточно мало, и сигналы  $s_1(t)$  и  $s_2(t)$  практически можно считать ортогональными.

- Радиосигналы с частотной манипуляцией:

$$s_1(t) = A_0 \cos(\omega_1 t);$$

$$s_2(t) = A_0 \cos(\omega_2 t).$$

По-прежнему

$$E_1 = E_2 = A_0^2 T / 2.$$

Тогда

$$r_{12} = \frac{2}{A_0^2 T} A_0^2 \int_0^T \cos(\omega_1 t) \cos(\omega_2 t) dt = \frac{\sin(\omega_1 + \omega_2)T}{(\omega_1 + \omega_2)T} + \frac{\sin(\omega_1 - \omega_2)T}{(\omega_1 - \omega_2)T}.$$

Если  $\omega_1$  и  $\omega_2$  — радиочастоты, то первое слагаемое мало по абсолютному значению, однако это нельзя утверждать про второе слагаемое; например, при достаточно близких значениях  $\omega_1$  и  $\omega_2$  оно стремится к единице. Следовательно, равенство нулю коэффициента корреляции возможно лишь при условии

$$\omega_1 - \omega_2 = \pi q / T,$$

где  $q$  — отличное от нуля целое число.

- Ортогональными являются сигналы, которые на интервале  $[0; T]$  не перекрываются во времени, например,

$$s_1(t) = A_0, \quad t \in [0; T / 2];$$

$$s_2(t) = B_0, \quad t \in [T / 2; T];$$

- Ортогональными являются сигналы, которые в отведённой системе

полосе частот не перекрываются<sup>1</sup> по спектру.

Действительно, исходя из соотношения Парсеваля, для любой временной характеристики существует её эквивалент в спектральной области. В данном случае

$$\int_0^T s_1(t)s_2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_1(\omega)S_2^*(\omega)d\omega,$$

где  $S_1(\omega)$  и  $S_2(\omega)$  — спектры (преобразования Фурье) сигналов  $s_1(t)$  и  $s_2(t)$ . Поэтому при неперекрывающихся спектрах получаем нулевой коэффициент корреляции.

Ортогональными также являются *сигналы с пассивной паузой*, когда один из сигналов тождественно равен нулю на всём интервале  $[0; T]$ . Например, при *двоичной амплитудной манипуляции*

$$\begin{aligned} s_1(t) &= A_0 \cos(\omega_1 t); \\ s_2(t) &= 0. \end{aligned}$$

Такие сигналы не являются равномошными, для них  $\Delta_{12} = E_1$  и

$$P_{\text{ош}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf}(h_1 / 2), \quad (3.2.33)$$

где  $h_1$  — отношение сигнал-шум для ненулевого сигнала  $s_1(t)$ .

Сравнение двух различных пар используемых сигналов позволяет ввести понятие *энергетического выигрыша (проигрыша)*  $\gamma$ , показывающего, во сколько раз необходимо изменить энергию одной пары сигналов по сравнению с другой парой, чтобы добиться такого же значения вероятности ошибки. Обычно значения энергетического выигрыша представляются в децибелах:

$$\gamma_{[\text{дБ}]} = 10 \lg \gamma.$$

При фиксированной длительности интервала, на котором рассматриваются сигналы, сравнение энергии эквивалентно сравнению по средней мощности

$$P = \frac{1}{T} E = \frac{1}{T} \int_0^T s^2(t)dt,$$

поэтому часто энергетический выигрыш оценивают именно по изменению

---

<sup>1</sup> Теоретически по спектру перекрываются любые финитные сигналы; речь идёт о практически значимом влиянии спектров друг на друга.

средней мощности сигнала.

Найдём энергетический выигрыш использования противоположных сигналов по сравнению с ортогональными. Приравнявая (3.2.28) и (3.2.32), получаем

$$\sqrt{\mu^2 E_{\text{орт}} / (2N_0)} = \sqrt{\mu^2 E_{\text{прот}} / N_0},$$

где  $E_{\text{орт}}$  и  $E_{\text{прот}}$  — энергия (равномощных) ортогональных и противоположных сигналов соответственно. Отсюда

$$E_{\text{орт}} = 2E_{\text{прот}},$$

т. е. переход от ортогональных сигналов к противоположным позволяет обеспечить неизменное качество связи (сохранить такую вероятность ошибки) при понижении энергии сигнала в два раза или на 3 дБ.

На рис. 3.10 показаны кривые помехоустойчивости (с логарифмическим масштабом по оси ординат) оптимального различения двоичных противоположных и ортогональных сигналов.

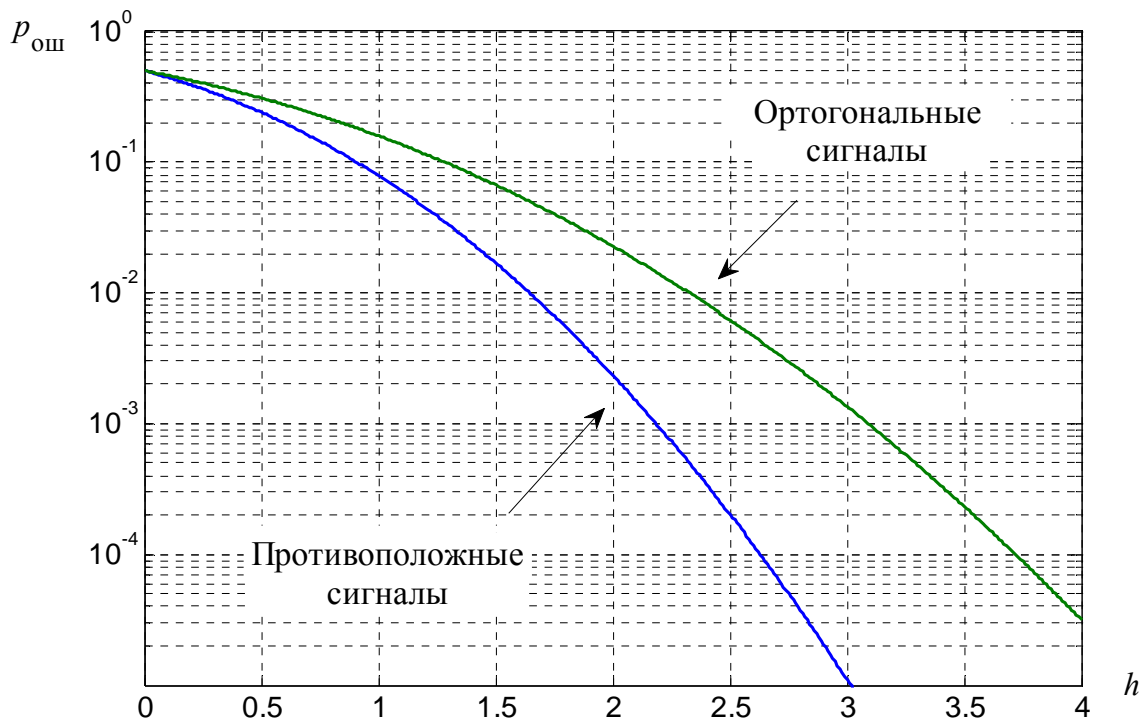


Рис. 3.10. Кривые помехоустойчивости оптимального различения двоичных сигналов



Заметим, что в литературе кривые помехоустойчивости часто приводятся в виде, когда по оси абсцисс отложены значения не  $h$ , а  $h^2$ . Более того, зачастую (и в дальнейшем это будет сделано) по оси абсцисс значения сигнал-шум представляются в децибелах, т. е. откладываются  $10\lg h$  или  $10\lg h^2$ . Для того, чтобы в подобных ситуациях не происходило путаницы, необходимо внимательно следить за тем, как вводятся и сравниваются значения отношения сигнал-шум.

Обратимся к вычислению помехоустойчивости алгоритма оптимального различения многопозиционного ансамбля сигналов. В общем случае, согласно структурной схеме, представленной на рис. 3.7, необходимо рассматривать совместные вероятности

$$\begin{aligned}\alpha_{ik} &\equiv \alpha_i(s_k(t)) = \int_0^T z(t)s_i(t)dt - \frac{\mu}{2}E_i = \int_0^T (\mu s_k(t) + n(t))s_i(t)dt - \frac{\mu}{2}E_i = \\ &= \int_0^T n(t)s_i(t)dt + \mu r_{ik}\sqrt{E_i E_k} - \frac{\mu}{2}E_i, \quad k, i = 1, \dots, m,\end{aligned}\quad (3.2.34)$$

формирующихся в каждой ветви корреляционной обработки. Каждая из  $\alpha_{ik}$  является гауссовской величиной с математическим ожиданием

$$\mathbf{E}[\alpha_{ik}] = \mu r_{ik}\sqrt{E_i E_k} - \mu E_i / 2, \quad (3.2.35)$$

дисперсией, аналогично (3.2.12),

$$\mathbf{D}[\alpha_{ik}] = \mathbf{D}\left[\int_0^T n(t)s_i(t)dt\right] = \frac{N_0 E_i}{2} \quad (3.2.36)$$

и ковариацией

$$\begin{aligned}\text{Cov}[\alpha_{ik}, \alpha_{jk}] &= \mathbf{E}[\alpha_{ik}\alpha_{jk}] - \mathbf{E}[\alpha_{ik}]\mathbf{E}[\alpha_{jk}] = \mathbf{E}\left[\int_0^T n(t_1)s_i(t_1)dt_1 \int_0^T n(t_2)s_j(t_2)dt_2\right] = \\ &= \frac{N_0}{2} \int_0^T s_i(t)s_j(t)dt = \frac{N_0}{2} r_{ij}\sqrt{E_i E_j}, \quad i, j = 1, \dots, m.\end{aligned}\quad (3.2.37)$$

Видно, что для произвольного класса используемых сигналов ковариация при  $i \neq j$  отлична от нуля, так что совместная (по всем ветвям, т. е. по индексам  $i$  и  $j$ ) вероятность случайных величин  $\alpha_{1k}, \dots, \alpha_{mk}$  — это многомерное гауссовское распределение с “полноценной” недиагональной ковариационной матрицей, что делает невозможной факторизацию многомерного

распределения и, следовательно, практическую невозможность расчёта вероятностей, описывающих помехоустойчивость алгоритма различения.

Так, условная вероятность  $p_{\text{прав}}(s_k)$  правильного различения при условии передачи сигнала  $s_k(t)$  равна вероятности того, что величин  $\alpha_{kk}$  в соответствующей  $k$ -й ветви структурной схемы окажется больше всех других величин  $\alpha_{ik}$  в остальных ветвях, т. е. вероятности одновременного выполнения  $m - 1$  неравенства:

$$p_{\text{прав}}(s_k) = P(\alpha_{1k} < \alpha_{kk}, \dots, \alpha_{k-1,k} < \alpha_{kk}, \alpha_{k+1,k} < \alpha_{kk}, \dots, \alpha_{mk} < \alpha_{kk}).$$

В этом выражении  $k$ -я ветвь фигурирует в середине; в действительности она может оказаться как в серединной, так и в начальной (сигнал  $s_1(t)$ ) или конечной (сигнал  $s_m(t)$ ) позициях.

Для нахождения полной (безусловной) вероятности  $p_{\text{прав}}$  правильного различения необходимо усреднить  $p_{\text{прав}}(s_k)$  по всем возможным значениям случайной величины  $\alpha_{kk}$ :

$$p_{\text{прав}} = \int_{-\infty}^{\infty} w_{\alpha_{kk}}(a) p_{\text{прав}}(s_k) da.$$

Тогда вероятность ошибочного различения равна

$$\begin{aligned} p_{\text{ош}} &= 1 - p_{\text{прав}} = 1 - \int_{-\infty}^{\infty} w_{\alpha_{kk}}(a) p_{\text{прав}}(s_k) da = \\ &= 1 - \int_{-\infty}^{\infty} w_{\alpha_{kk}}(a) P(\alpha_{1k} < \alpha_{kk}, \dots, \alpha_{k-1,k} < \alpha_{kk}, \alpha_{k+1,k} < \alpha_{kk}, \dots, \alpha_{mk} < \alpha_{kk}) da. \end{aligned} \quad (3.2.38)$$

К сожалению, при  $m > 2$  не удаётся получить универсальных и достаточно простых, подобно (3.2.27), выражений  $p_{\text{ош}}$  для произвольных ансамблей сигналов. Решение в “замкнутой форме”, хотя и требующее привлечения специальных вычислительных методов, удаётся получить лишь для некоторых классов, характеризующихся особыми свойствами составляющих их сигналов.

### **Эквидистантные сигналы**

Рассмотрим ансамбль *эквидистантных сигналов*, характеризующихся тем, что эвклидово расстояние для любой пары сигналов одинаково:

$$\Delta_{ki} = \int_0^T (s_k(t) - s_i(t))^2 dt \equiv \Delta = \text{const}, k, i = 1, \dots, m. \quad (3.2.39)$$

В геометрической интерпретации эквидистантные сигналы можно трактовать как точки, расположенные в вершинах правильных фигур (пространственных тел). В одномерном пространстве, т. е. на прямой, эквидистантную систему образует любая пара точек. В двумерном пространстве (на плоскости) таких точек три, и они располагаются в вершинах равностороннего треугольника. В трёхмерном пространстве возможны четыре эквидистантных точки — в вершинах правильного тетраэдра. Вообще, для  $K$ -мерного евклидова пространства максимально число эквидистантных точек равно  $m = K + 1$ , и они располагаются в вершинах  $K$ -мерного симплекса.

Приведём ряд примеров сигналов, образующих на интервале  $[0; T]$  эквидистантный ансамбль в трёхмерном пространстве.

- Трои́чная фазовая манипуляция на  $120^\circ$ :

$$\begin{aligned} s_1(t) &= A_0 \cos(\omega_1 t); \\ s_2(t) &= A_0 \cos(\omega_1 t + 2\pi / 3); \\ s_3(t) &= A_0 \cos(\omega_1 t + 4\pi / 3). \end{aligned}$$

Здесь

$$\begin{aligned} E_1 = E_2 = E_3 = E &= A_0^2 T / 2, \\ \Delta &= 3A_0^2 T / 2 = 3E. \end{aligned}$$

- Трои́чная частотная манипуляция

$$\begin{aligned} s_1(t) &= A_0 \cos(\omega_1 t); \\ s_2(t) &= A_0 \cos(\omega_2 t); \\ s_3(t) &= A_0 \cos(\omega_3 t); \end{aligned}$$

при условии, что

$$\omega_k - \omega_i = \pi q / T, k, i = 1, 2, 3,$$

где  $q$  — отличное от нуля целое число.

Здесь

$$E_1 = E_2 = E_3 = A_0^2 T / 2, \Delta = A_0^2 T.$$

- Сигналы вида

$$\begin{aligned}
s_1(t) &= A_0 \cos(\omega_1 t); \\
s_2(t) &= -(A_0 / 2) \cos(\omega_1 t) + (\sqrt{3} A_0 / 2) \cos(\omega_2 t); \\
s_3(t) &= -(A_0 / 2) \cos(\omega_1 t) - (\sqrt{3} A_0 / 2) \cos(\omega_2 t);
\end{aligned}$$

при условии, что

$$\omega_1 - \omega_2 = 2\pi q / T,$$

где  $q$  — отличное от нуля целое число.

Здесь

$$E_1 = E_2 = E_3 = A_0^2 T / 2, \quad \Delta = 3A_0^2 T / 2.$$

- Сигналы вида

$$\begin{aligned}
s_1(t) &= (3A_0 / 2) \cos(\omega_1 t); \\
s_2(t) &= (\sqrt{3} A_0 / 2) \sin(\omega_1 t); \\
s_3(t) &= -(\sqrt{3} A_0 / 2) \sin(\omega_1 t).
\end{aligned}$$

Здесь

$$E_1 = 9A_0^2 T / 8, \quad E_2 = E_3 = 3A_0^2 T / 8, \quad \Delta = 3A_0^2 T / 2.$$

Первые три примера показывают наборы равномошных сигналов; мощности сигналов представленных в четвёртом примере, различны.

Ансамбли эквидистантных сигналов обладают свойством *изоморфности*, заключающимся в том, что если два ансамбля имеют одинаковое значение (квадрата) эвклидова расстояния  $\Delta$  между точками, то такие ансамбли обладают одинаковой помехоустойчивостью. Это означает, что для расчёта помехоустойчивости заданного ансамбля можно перейти к другому, эквидистантному ансамблю, для которого нахождение вероятности ошибки заметно проще.

### Ортогональные сигналы

Обратимся к частному случаю, но, тем не менее, имеющему наибольшую практическую значимость среди эквидистантных сигналов, ансамблю *многопозиционных ортогональных сигналов*, характеризующихся одинаковой энергией и нулевым коэффициентом корреляции для несовпадающих сигналов:

$$E_k = \int_0^T s_k^2(t) dt = E;$$

$$r_{ik} = \frac{1}{\sqrt{E_i E_k}} \int_0^T s_i(t) s_k(t) dt = \delta_{ik}, \quad k, i = 1, \dots, m.$$

Пример троичного ансамбля ортогональных сигналов был представлен выше.

Для ортогональных сигналов фигурирующие в совместной вероятности (3.2.38) характеристики гауссовских случайных величин  $\alpha_{1k}, \dots, \alpha_{mk}$  равны:

$$\mathbf{E}[\alpha_{ik}] = \begin{cases} \mu E / 2, & i = k; \\ -\mu E / 2, & i \neq k; \end{cases}$$

$$\mathbf{D}[\alpha_{ik}] = N_0 E / 2;$$

$$\text{Cov}[\alpha_{ik}, \alpha_{jk}] = \begin{cases} N_0 E / 2, & i = j; \\ 0, & i \neq j, \end{cases}$$

так что для  $i \neq k$  плотность вероятности

$$w_{\alpha_{ik}}(a) = \frac{1}{\sqrt{\pi N_0 E}} \exp\left[-\frac{(a + \mu E / 2)^2}{N_0 E}\right].$$

Равенство нулю элементов, находящихся вне главной диагонали ковариационной матрицы, означает некоррелированность, а, следовательно, независимость величин  $\alpha_{1k}, \dots, \alpha_{mk}$ . Таким образом, совместная вероятность случайных величин  $\alpha_{1k}, \dots, \alpha_{mk}$  факторизуется, и

$$P(\alpha_{1k} < \alpha_{kk}, \dots, \alpha_{k-1,k} < \alpha_{kk}, \alpha_{k+1,k} < \alpha_{kk}, \dots, \alpha_{mk} < \alpha_{kk}) =$$

$$= \left( \int_{-\infty}^{\alpha_{kk}} w_{\alpha_{ik}}(a) da \right)^{m-1} = \left( \frac{1}{\sqrt{\pi N_0 E}} \int_{-\infty}^{\alpha_{kk}} \exp\left[-\frac{(a + \mu E / 2)^2}{N_0 E}\right] da \right)^{m-1}.$$

Делая замену переменных

$$\frac{a + \mu E / 2}{\sqrt{N_0 E / 2}} = u,$$

получаем

$$p_{\text{прав}}(s_k) = \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\alpha_{kk} + \mu E / 2) / \sqrt{N_0 E / 2}} \exp(-u^2 / 2) du \right)^{m-1} = F^{m-1} \left( \frac{\alpha_{kk} + \mu E / 2}{\sqrt{N_0 E / 2}} \right).$$

Тогда, согласно, (3.2.38), вероятность ошибки равна

$$p_{\text{ош}} = 1 - \frac{1}{\sqrt{\pi N_0 E}} \int_{-\infty}^{\infty} \exp \left[ -\frac{(a - \mu E / 2)^2}{N_0 E} \right] F^{m-1} \left( \frac{a + \mu E / 2}{\sqrt{N_0 E / 2}} \right) da.$$

Вновь совершая замену переменных

$$\frac{a + \mu E / 2}{\sqrt{N_0 E / 2}} = x,$$

окончательно получаем следующее выражение для вероятности ошибочного различения в ансамбле из  $m$  ортогональных сигналов:

$$\begin{aligned} p_{\text{ош}} &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[ -\frac{(x - \mu \sqrt{2E / N_0})^2}{2} \right] F^{m-1}(x) dx = \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[ -(x - h\sqrt{2})^2 / 2 \right] F^{m-1}(x) dx. \end{aligned} \quad (3.2.40a)$$

При использовании функции erf вероятность ошибки записывается в виде

$$p_{\text{ош}} = 1 - \frac{1}{2^{m-1} \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[ -(x - h\sqrt{2})^2 / 2 \right] \left[ 1 + \operatorname{erf} \left( x / \sqrt{2} \right) \right]^{m-1} dx. \quad (3.2.40б)$$

Теперь, на основе (3.2.40), можно записать выражения для вероятности ошибочного различения в любом ансамбле, изоморфном с  $m$ -позиционным ортогональным ансамблем, у которого эвклидово расстояние  $\Delta$  и энергия  $E$  связаны соотношением  $\Delta = 2E$ :

$$p_{\text{ош}} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[ -(x - \sqrt{\mu \Delta / (2N_0)})^2 / 2 \right] F^{m-1}(x) dx. \quad (3.2.41)$$

Нетрудно показать, что соотношения (3.2.40) и (3.2.41) переходят в формулы (3.2.32) и (3.2.27) для соответствующих двоичных случаев.

Для произвольного  $m$  интегралы, фигурирующие в выражениях для  $p_{\text{ош}}$ , не выражаются в элементарных или известных табулированных функциях, но могут быть вычислены с помощью методов численного интегрирования<sup>1</sup>.

Кривые помехоустойчивости ансамбля ортогональных сигналов при различных значениях  $m$  представлены на рис. 3.11.

---

<sup>1</sup> Дополнительную сложность создают бесконечные пределы интегрирования; погрешности, обусловленные их влиянием, могут сказываться при малых значениях  $p_{\text{ош}}$ .

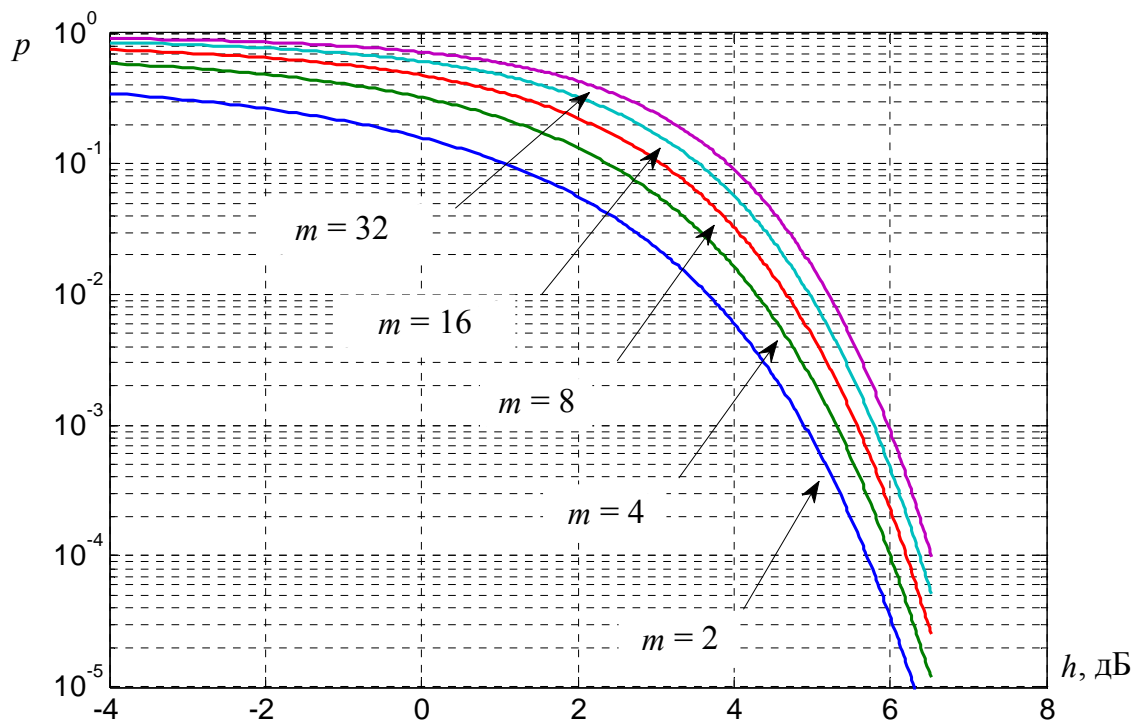


Рис. 3.11. Кривые помехоустойчивости для ортогональных сигналов

Полученные выражения показывают, что вероятность ошибки ортогональных сигналов, как и в двоичном случае, не зависит от формы сигналов и определяется отношением энергии сигнала к спектральной плотности средней мощности шума.

### Биортогональные сигналы

С ортогональными сигналами тесно связаны *биортогональные сигналы*. В таком равномоном ансамбле  $m$  всегда чётно, и он конструируется посредством дополнения  $m/2$  ортогональных сигналов им противоположными. Примером биортогональных сигналов на интервале  $[0; T]$  является следующий ансамбль:

$$\begin{aligned}
 s_1(t) &= A_0 \cos \omega_1 t; \\
 &\dots\dots\dots \\
 s_{m/2}(t) &= A_0 \cos \omega_{m/2} t; \\
 s_{m/2+1}(t) &= -A_0 \cos \omega_1 t; \\
 &\dots\dots\dots \\
 s_m(t) &= -A_0 \cos \omega_{m/2} t,
 \end{aligned}$$

где, как обычно, частоты находятся в “ортогональном” соотношении:

$$\omega_k - \omega_i = \pi q / T, \quad k, i = 1, m/2,$$

$q$  — отличное от нуля целое число.

Точки сигнального созвездия, соответствующие ансамблю биортогональных сигналов, располагаются в вершинах геометрического тела, называемого *кроссполитопом*: в двумерном пространстве это квадрат; в трёхмерном — октаэдр.

Биортогональные сигналы не являются эквидистантными, и формула (3.2.41) к ним не применима. Однако, используя рассмотренную выше методику, нетрудно получить (задача 3.4) для них следующее выражение для вероятности ошибки:

$$p_{\text{ош}} = 1 - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left[-(x - h\sqrt{2})^2 / 2\right] [2F(x) - 1]^{m/2-1} dx. \quad (3.2.42a)$$

$$p_{\text{ош}} = 1 - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left[-(x - h\sqrt{2})^2 / 2\right] [\text{erf}(x / \sqrt{2})]^{m/2-1} dx. \quad (3.2.42б)$$

Выражение (3.2.42) при одинаковых значениях  $h$  и  $m$  приводит к несколько большему, чем (3.2.40) значению  $p_{\text{ош}}$ , однако биортогональные сигналы позволяют получить заданное число  $m$  (число сигналов в ансамбле) при меньшей занимаемой полосе частот, что иногда является преимуществом.

При большом объёме ансамбля ( $m \geq 128$ ) помехоустойчивость ортогональных и биортогональных сигналов практически совпадает. Это понятно из того, что в ортогональном ансамбле каждому из  $m$  сигналов ортогональны  $m - 1$  остальных, а в биортогональном ансамбле  $m - 2$  сигналов ортогональны и лишь один из сигналов протовоположен.

### Сигналы с фазовой манипуляцией

Другим примером, в общем случае, неэквидистантной системы является многопозиционный ансамбль сигналов с *фазовой манипуляцией*:

$$s_k(t) = A_0 \cos\left(\omega_0 t + \frac{2\pi(k-1)}{m} + \varphi_0\right), \quad k = 1, \dots, m, \quad 0 \leq t \leq T. \quad (3.2.43)$$

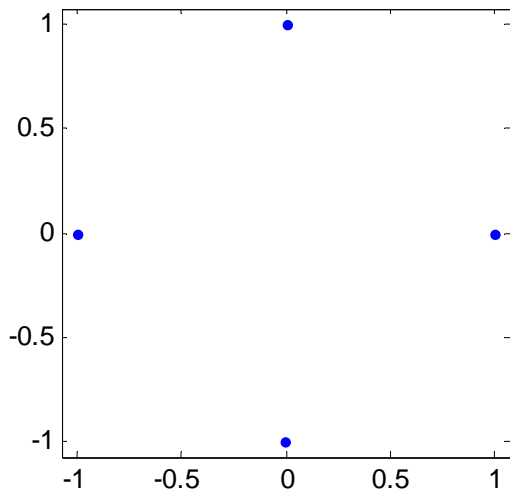
Здесь  $\varphi_0$  — случайная начальная фаза, которую обычно полагаю равномерно распределённой на интервале  $[0; 2\pi)$  или на интервале  $[-\pi; \pi)$ , а



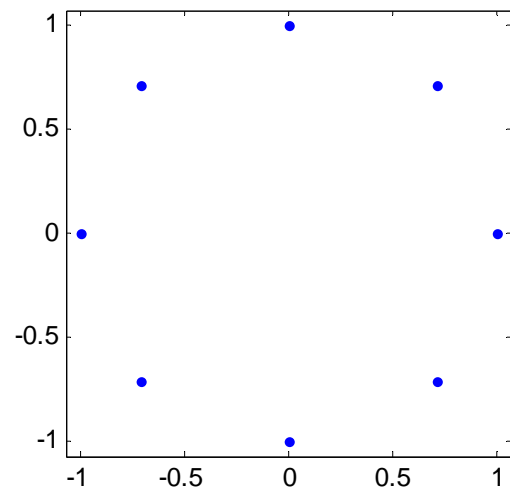
$$\varphi_k = 2\pi(k-1)/m$$

является модуляционным параметром, определяющим передаваемое сообщение.

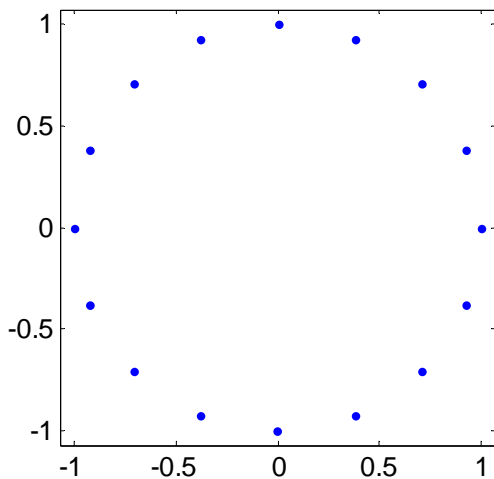
На рис. 3.12 показано схематичное расположение сигналов с фазовой манипуляцией на сигнальной плоскости (вращающейся с частотой  $\omega_0$ ) — так называемое *сигнальное созвездие*. Для простоты выбраны значения  $\varphi_0 = 0$  и  $A_0 = 1$ .



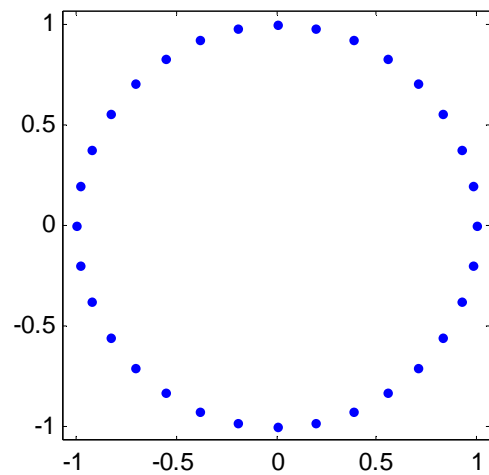
$m = 4$



$m = 8$



$m = 16$



$m = 32$

Рис. 3.12. Расположение на плоскости точек ансамбля сигналов с ФМ

При  $m = 2$  соотношение (3.2.43) описывает противоположные сигналы ФМ-2; при  $m = 3$  — эквидистантный ансамбль с троичной фазовой манипуляцией ФМ-3, а при  $m = 4$  — широко используемые в различных телекоммуникационных и информационных системах сигналы ФМ-4 (для упрощения записи выбрано  $\varphi_0 = 0$ ):

$$\begin{aligned} s_1(t) &= A_0 \cos(\omega_1 t + 0); \\ s_2(t) &= A_0 \cos(\omega_1 t + \pi / 2); \\ s_3(t) &= A_0 \cos(\omega_1 t + \pi); \\ s_4(t) &= A_0 \cos(\omega_1 t + 3\pi / 2), \end{aligned}$$

являющиеся простейшим биортогональным набором, в котором сигналы  $s_1$  и  $s_2$  — ортогональны, а  $s_3$  и  $s_4$  — им противоположны.

Для сигналов с ФМ-4 общее соотношение (3.2.42) заметно упрощается, и вероятность ошибки может быть вычислена через табулированные функции. Действительно, в выражении

$$\begin{aligned} &1 - \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left[-(x - h\sqrt{2})^2 / 2\right] \left[\operatorname{erf}\left(x / \sqrt{2}\right)\right] dx = \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left[-(x - h\sqrt{2})^2 / 2\right] \frac{2}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} \exp(-y^2) dy dx = \\ &= 1 - \frac{\sqrt{2}}{\pi} \int_0^\infty \int_0^{x/\sqrt{2}} \exp\left[-\frac{(x - h\sqrt{2})^2 + 2y^2}{2}\right] dy dx = \\ &= 1 - \frac{1}{\pi\sqrt{2}} \int_0^\infty \int_{-x/\sqrt{2}}^{x/\sqrt{2}} \exp\left[-\frac{(x - h\sqrt{2})^2 + 2y^2}{2}\right] dy dx \end{aligned}$$

произведём замену переменных:

$$\begin{aligned} \xi &= x / \sqrt{2} + y - h; \\ \eta &= x / \sqrt{2} - y - h. \end{aligned}$$

Обратное преобразование имеет вид

$$\begin{aligned} x &= (\xi + \eta + 2h) / \sqrt{2}; \\ y &= (\xi - \eta) / 2, \end{aligned}$$

и модуль якобиана преобразования равен

$$\left| \frac{D(x, y)}{D(\xi, \eta)} \right| = \frac{1}{\sqrt{2}}.$$

При этом клинообразная область интегрирования по переменным  $x, y$  переводится в прямоугольную область интегрирования по переменным  $\xi, \eta$  (рис. 3.13).

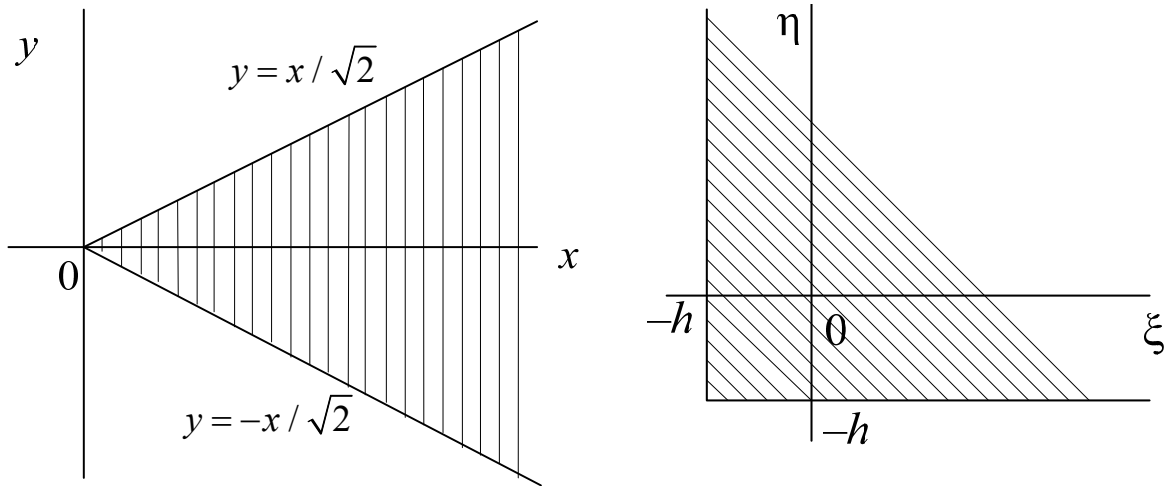


Рис. 3.13. Изменение области интегрирования при замене переменных

Нетрудно видеть, что в подынтегральном выражении

$$(x - h\sqrt{2})^2 + 2y^2 = \xi^2 + \eta^2,$$

тогда

$$p_{\text{ош}} = 1 - \frac{1}{2\pi} \int_{-h}^{\infty} \int_{-h}^{\infty} \exp\left[-(\xi^2 + \eta^2) / 2\right] d\xi d\eta = 1 - F^2(h) \quad (3.2.44a)$$

или

$$p_{\text{ош}} = 1 - \frac{1}{4} \left[ 1 + \operatorname{erf}\left(h / \sqrt{2}\right) \right]^2. \quad (3.2.44b)$$

Получение выражения для вероятности ошибки многопозиционных сигналов с фазовой манипуляцией основано на использовании распределения фазы  $\varphi(t)$  гармонического сигнала

$$u(t) = A(t) \cos[\omega_0 t + \varphi(t)] = u_s(t) \cos \omega_0 t - u_c(t) \sin \omega_0 t,$$

где

$$A(t) = \sqrt{u_s^2(t) + u_c^2(t)};$$

$$\operatorname{tg} \varphi(t) = u_s(t) / u_c(t),$$

наблюдаемого в аддитивной смеси с гауссовским шумом:

$$\zeta(t) = u(t) + n(t),$$

имеющего нулевое среднее и дисперсию  $\sigma^2$ .

Если процесс  $n(t)$  представить<sup>1</sup> в квадратурном виде

$$n(t) = n_c(t) \cos \omega_0 t - n_s(t) \sin \omega_0 t,$$

где квадратурные компоненты  $n_c(t)$  и  $n_s(t)$  являются стационарными и стационарно связанными процессами, то случайное колебание  $\zeta(t)$  также можно представить в квадратурном виде:

$$\zeta(t) = Z(t) \cos[\omega_0 t + \theta(t)].$$

Здесь

$$Z(t) = \sqrt{(u_c(t) + n_c(t))^2 + (u_s(t) + n_s(t))^2};$$

$$\operatorname{tg} \theta(t) = (u_s(t) + n_s(t)) / (u_c(t) + n_c(t)).$$

При квадратурном представлении результирующего процесса плотность распределения фазы  $\theta(t)$  имеет следующий вид [12]:

$$w(\theta, t) = \frac{1}{2\pi} \exp\left(-\frac{A^2(t)}{2\sigma^2}\right) + \frac{A(t) \cos(\theta(t) - \varphi(t))}{\sigma\sqrt{2\pi}} \times$$

$$\times F\left[\frac{A(t)}{\sigma} \cos(\theta(t) - \varphi(t))\right] \exp\left[-\frac{A^2(t)}{2\sigma^2} \sin^2(\theta(t) - \varphi(t))\right], |\theta(t) - \varphi(t)| \leq \pi.$$

(3.2.45)

При приёме сигналов с фазовой манипуляцией огибающая  $A(t)$  постоянна и равна  $\mu A_0$ , а дисперсия  $\sigma^2$  имеет физический смысл средней мощности шума. Если  $\Delta F$  — эффективная полоса частот, занимаемая радиосигналом, то

$$\frac{A^2(t)}{\sigma^2} = \frac{\mu^2 A_0^2}{(N_0 / 2) 2\Delta F} = \frac{\mu^2 A_0^2}{N_0 \Delta F}.$$

Будем считать, что полоса частот определяется по первому нулю спектра, т. е. значению  $1/T$ . Тогда эффективная полоса частот равна  $2/T$ , и

---

<sup>1</sup> Квадратурное представление стационарных случайных процессов основано на комплексном представлении процессов, когда в качестве мнимой компоненты выбирается преобразование Гильберта от исходного процесса. При этом фактически происходит переход от модели АБГШ к физически гораздо более адекватной модели узкополосного гауссовского процесса на выходе фильтра, реализующего корреляционную обработку.

$$\frac{A^2(t)}{\sigma^2} = \frac{\mu^2 A_0^2 T}{2N_0} = \frac{\mu^2 E}{N_0} = h^2.$$

Таким образом, можно сделать вывод о том, что распределение фазы (3.2.45) выражается через стандартный параметр  $h$ :

$$w(\theta, t) = \frac{1}{2\pi} \exp\left(-\frac{h^2}{2}\right) + \frac{h \cos \Delta\theta}{\sqrt{2\pi}} F[h \cos \Delta\theta] \exp\left[-\frac{h^2}{2} \sin^2 \Delta\theta\right], \quad (3.2.45a)$$

где

$$\Delta\theta = \theta(t) - \frac{2\pi(k-1)}{m}, \quad k = 1, \dots, m$$

при передаче сигнала  $s_k(t)$ .

Из алгоритма оптимального различения сигналов с ФМ следует, что решение в пользу сигнала  $s_k(t)$  принимается в том случае, когда начальная фаза  $\varphi_k$  находится в пределах от  $(2k-3)\pi/m$  до  $(2k-1)\pi/m$ ; в противном случае возникнет ошибка. Таким образом, можно записать:

$$p = 1 - \int_{(2k-3)\pi/m}^{(2k-1)\pi/m} w(\theta, t) d\theta = 1 - \frac{1}{m} \exp\left(-\frac{h^2}{2}\right) - \frac{1}{\sqrt{2\pi}} \int_{(2k-3)\pi/m}^{(2k-1)\pi/m} h \cos(\Delta\theta) F[h \cos \Delta\theta] \exp\left[-\sin^2(\Delta\theta)h^2/2\right] d\theta. \quad (3.2.46)$$

Вычисление (3.2.46) возможно либо численным образом, либо через известные табулированные функции. Так, при использовании табулированной функции *Никольсона*

$$V(a, b) = \frac{1}{2\pi} \int_0^a \int_0^{bx/a} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy$$

вероятность ошибки выражается следующим образом [13]:

$$p = 1 - \frac{1}{m} - \frac{1}{2} \operatorname{erf}\left(h \sin \frac{\pi}{m}\right) - 2V\left(h\sqrt{2} \sin \frac{\pi}{m}, h\sqrt{2} \cos \frac{\pi}{m}\right). \quad (3.2.47)$$

Полученные выражения (3.2.40), (3.2.42) или (3.2.47) для точной вероятности ошибочного различения различных многопозиционных ансамблей достаточно сложны и требуют либо численного (в том числе, многомерного) интегрирования, либо использования табулированных функций. Ещё более сложными оказываются подобные подходы при получении выражений, описывающих вероятность ошибки для многопозиционных сигналов с

квадратурной амплитудной (КАМ) или амплитудно-фазовой (АФМ) манипуляцией. В связи с этим представляет интерес поиск приближённых методов вычисления (оценки) вероятностей ошибок для многопозиционных ансамблей сигналов.

Одним из простых, но, тем не менее, достаточно точных и, поэтому, широко используемых методов оценки вероятностей ошибок является *метод аддитивной границы*.

Теоретико-вероятностная иллюстрация метода аддитивной границы представлена на рис. 3.14. Если  $A$ ,  $B$  и  $C$  — некоторые события (подмножества) пространства элементарных исходов  $\Omega$ , то вероятность объединения этих событий не превосходит сумму отдельных событий:

$$P(A \cup B \cup C) \leq P(A) + P(B) + P(C).$$

Графически это означает, что площадь результирующей фигуры, получившейся в результате наложения отдельных фигур, не превосходит суммы площадей отдельных фигур.

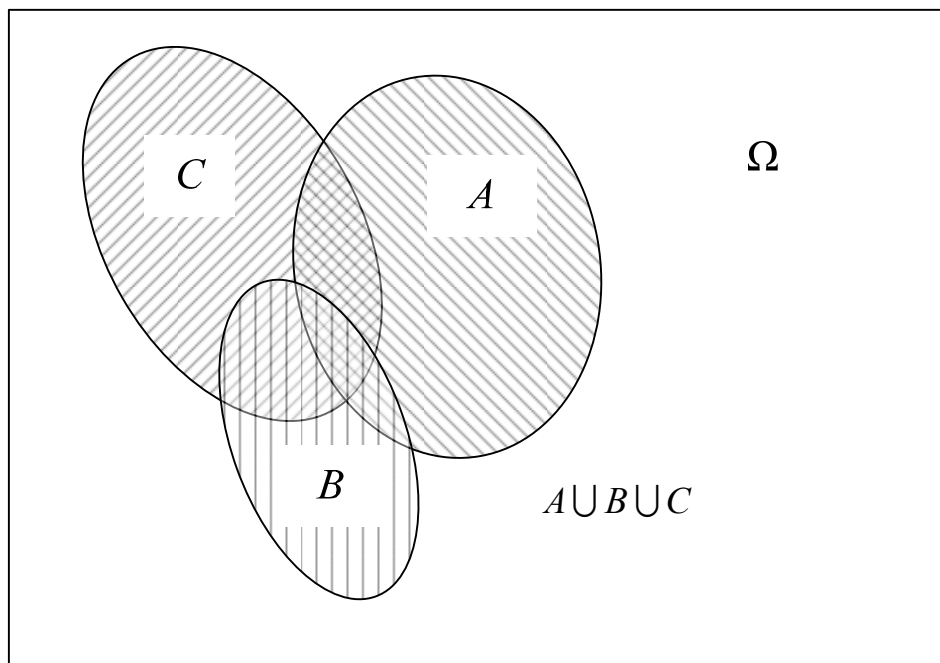


Рис. 3.14. К иллюстрации аддитивной границы

Применительно к оценке вероятности ошибки аддитивная граница означает, что совместная вероятность выполнения неравенств (3.2.34) не превосходит суммы вероятностей соответствующих неравенств в каждой

из ветвей корреляционной обработки. Тем самым, анализ помехоустойчивости одного многопозиционного ансамбля сигналов заменяется анализом помехоустойчивости некоторого гипотетического множества ансамблей двоичных сигналов, для которых вероятность ошибки в общем виде описывается выражением (3.2.27). Таким образом, для  $m$ -позиционного ансамбля справедливо неравенство

$$p_k \leq \sum_{\substack{i=1 \\ i \neq k}}^m p_{ki}, \quad (3.2.48)$$

где  $p_k$  — вероятность ошибки в многопозиционном ансамбле при передаче сигнала  $s_k(t)$ , а  $p_{ki}$  — вероятность ошибки в гипотетическом двоичном ансамбле, в котором используются сигналы  $s_k(t)$  и  $s_i(t)$ . Тогда средняя по всему ансамблю вероятность ошибки при равенстве априорных вероятностей равна

$$p = \frac{1}{m} \sum_{k=1}^m p_k \leq \frac{1}{m} \sum_{k=1}^m \sum_{\substack{i=1 \\ i \neq k}}^m p_{ki}. \quad (3.2.49)$$

Аддитивная граница (3.2.49) содержит весь “спектр” помехоустойчивости — набор вероятностей  $p_{ki}$ , среди которых есть и сравнительно малые, и сравнительно большие, что определяется эвклидовым расстоянием  $\Delta_{ki}$ . Очевидно, что наибольший вклад в вероятность ошибки будут вносить именно сигналы с минимальным эвклидовым расстоянием:

$$p_{\text{ош max}} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left( \frac{\mu}{2} \sqrt{\frac{\min \Delta_{ki}}{N_0}} \right).$$

Следовательно, можно использовать (например, для “быстрого” анализа помехоустойчивости) чуть менее точную, но заметно более простую оценку вероятности ошибки:

$$p \leq (m-1) p_{\text{ош max}}. \quad (3.2.50)$$

Отметим, что аддитивная граница является более точной для больших и менее точной — для малых значений сигнал-шум. Более того, может оказаться так, что при малых значениях  $h$  вероятность ошибки превысит единицу — это свидетельствует лишь о том, что предложенный теоретико-вероятностный подход перестаёт быть адекватным, и аддитивная граница “не работает”.

Расчёт помехоустойчивости согласно (3.2.49) предполагает вычисление суммы из  $m(m - 1)$  слагаемого, однако в силу симметричности реально используемых ансамблей достаточно рассмотреть половину или, даже, четверть ансамбля, а затем пропорционально увеличить итоговый результат.

Рассмотрим применение аддитивной границы для оценки помехоустойчивости некоторых многопозиционных ансамблей сигналов.

При использовании ортогональных сигналов вероятность ошибки для любой пары  $s_k(t)$  и  $s_i(t)$  определяется выражением (3.2.32). Следовательно, все вероятности  $p_{ki}$ , фигурирующие в (3.2.49), одинаковы, и “точная” аддитивная граница совпадает с “грубой” границей (3.2.50):

$$p_{\text{ош}} \leq p_{\text{ош орт, } m} = \frac{m-1}{2} \left[ 1 - \text{erf} \left( h / \sqrt{2} \right) \right]. \quad (3.2.51)$$

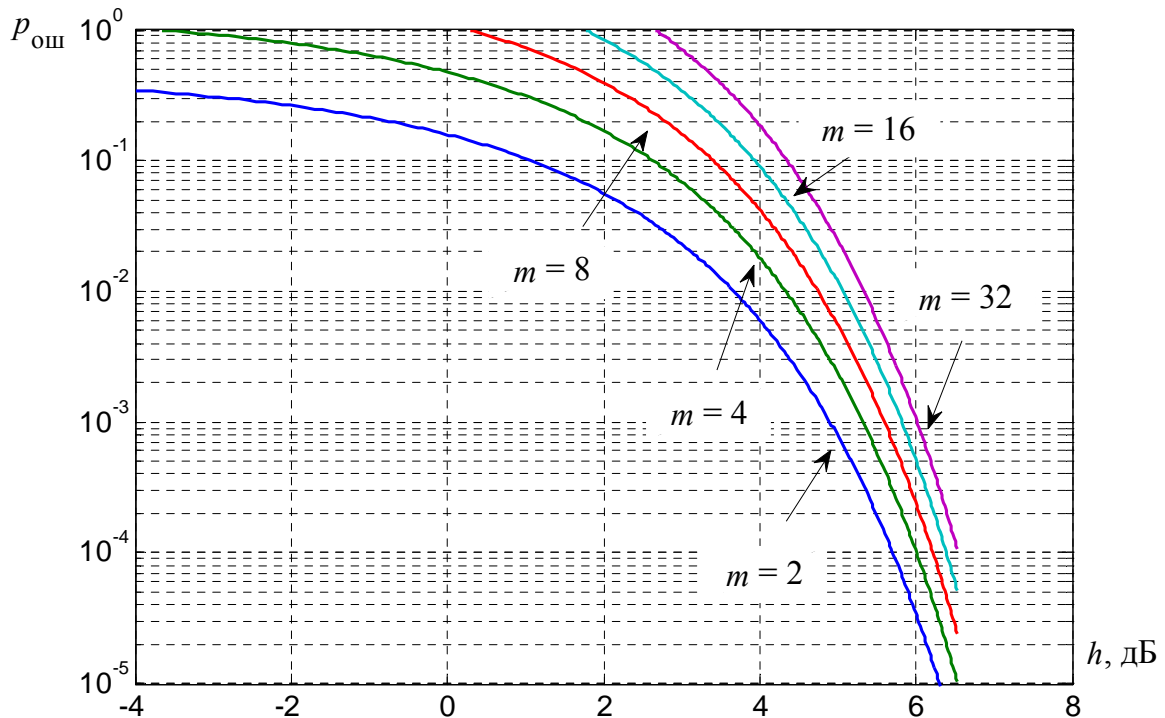


Рис. 3.15. Аддитивная граница для ортогональных сигналов

На рис. 3.15 представлена аддитивная граница (3.2.51) для различных значений  $m$ . Сравнение зависимостей на рис. 3.15 с аналогичными зависимостями, представленными на рис. 3.11, которые вычислены по точной формуле (3.2.40), показывает, что для малых значений  $m$  ( $m = 2, 4$ ) аддитивная граница даёт практически такие же значения, как и точная формула.



Однако, начиная с  $m = 8$ , проявляются заметные расхождения, достигающие к  $m = 32$  около  $+0,5$  дБ.

Отметим также, что для  $m > 2$  начиная с некоторого значения  $h$  кривые помехоустойчивости превышают единичное значение. Это, как уже было сказано, свидетельствует лишь о том, что для таких значений сигнал-шум модель аддитивной границы перестает быть адекватной.

Для многопозиционных сигналов с фазовой манипуляцией (3.2.43), имеющих одинаковую энергию  $E = A_0^2 T / 2$ , коэффициент корреляции

$$\begin{aligned} r_{ki} &= \frac{A_0^2 T}{E} \int_0^T \cos[\omega_0 t + 2\pi(k-1)/M] \cos[\omega_0 t + 2\pi(i-1)/M] dt \approx \\ &\approx \frac{A_0^2 T}{2E} \cos(2\pi(k-i)/m) = \cos(2\pi(k-i)/m). \end{aligned}$$

Тогда

$$\Delta_{ki} = 2E - 2r_{ki}E = 2E[1 - \cos(2\pi(k-i)/m)].$$

В частности, минимальное евклидово расстояние равно

$$\Delta_{\min} = 2E[1 - \cos(2\pi/M)]. \quad (3.2.52)$$

Тогда аддитивная граница имеет вид

$$\begin{aligned} P_{\text{ош ФМ, } m} \leq p &= \frac{1}{2m} \sum_{k=1}^m \sum_{\substack{i=1 \\ i \neq k}}^m \left(1 - \operatorname{erf}\left(h\sqrt{1-r_{ki}}/\sqrt{2}\right)\right) = \\ &= \frac{1}{2m} \sum_{k=1}^m \sum_{\substack{i=1 \\ i \neq k}}^m \left(1 - \operatorname{erf}\left(h\sqrt{1-\cos(2\pi(k-i)/m)}/\sqrt{2}\right)\right), \end{aligned} \quad (3.2.53)$$

а её грубая оценка

$$\begin{aligned} p &\leq \frac{m-1}{2} \left[1 - \operatorname{erf}\left(h\sqrt{1-\cos(2\pi/m)}/\sqrt{2}\right)\right] = \\ &= \frac{m-1}{2} \left[1 - \operatorname{erf}\left(h\sqrt{2\sin^2(\pi/m)}/2\right)\right] = \frac{m-1}{2} \left[1 - \operatorname{erf}(h\sin(\pi/m))\right]. \end{aligned} \quad (3.2.54)$$

На рис. 3.16 показаны кривые помехоустойчивости ансамбля сигналов с фазовой манипуляцией для различных значений  $m$ . Сплошными линиями изображены зависимости, полученные на основе аддитивной границы (3.2.53), а штрихпунктирными — на основе грубой оценки (3.2.54). Кроме

того, пунктирной линией показана кривая помехоустойчивости приёма сигналов с ФМ-4, рассчитанная по точной формуле (3.2.44).

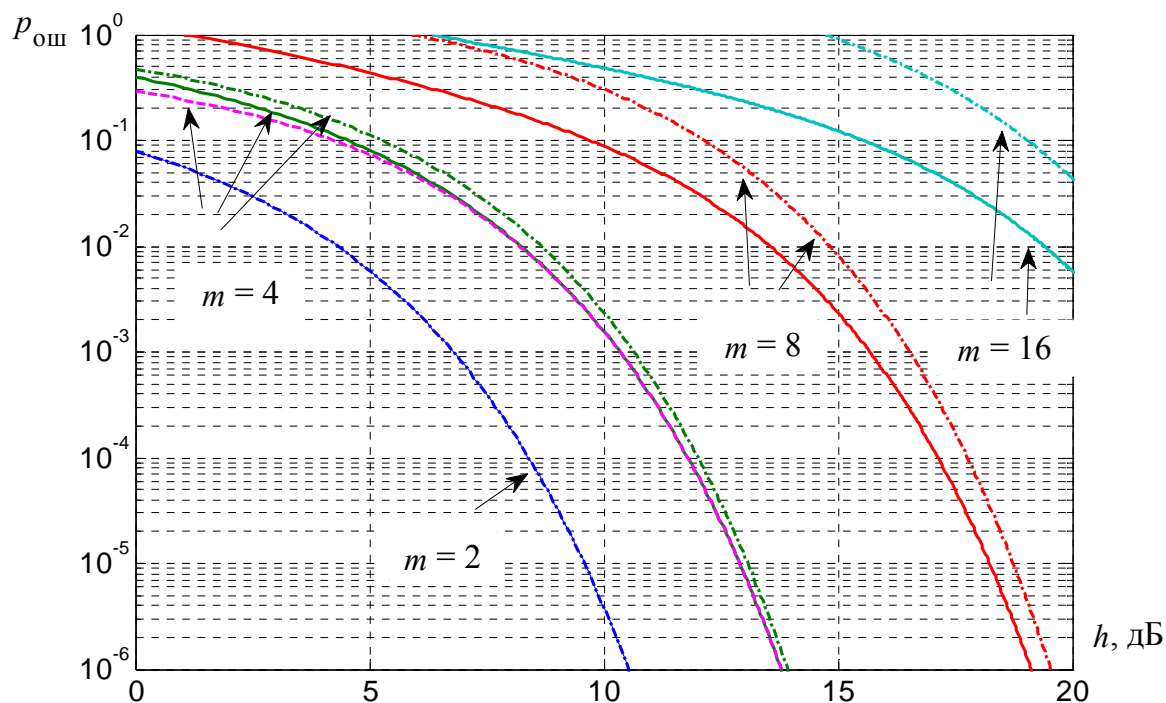


Рис. 3.16. Аддитивная граница для сигналов с ФМ

Кривая для  $m = 2$  соответствует приёму противоположных сигналов и совпадает с аналогичной зависимостью на рис. 3.10; для неё вычисления по формулам (3.2.53) и (3.2.54) совпадают.

Для четырёхпозиционного ансамбля зависимости, полученные на основе точной формулы, аддитивной границы и её грубой оценки весьма близки: отклонения составляют около 0,1 дБ.

При  $m \geq 8$  становятся ощутимыми различия между кривыми помехоустойчивости, рассчитанными по точной и грубой границам; в области значений вероятности ошибки  $10^{-2} \dots 10^{-3}$  это различие составляет около 1 дБ для  $m = 8$  и около 1,5 дБ для  $m = 16$ .

Из анализа кривых, представленных на рис. 3.16, следует, что увеличение объёма ансамбля сигналов с фазовой манипуляцией приводит к существенному энергетическому проигрышу, тем большему, чем больше  $m$ . Так, если при переходе от  $m = 2$  к  $m = 4$  проигрыш составляет 1,6 дБ в области значений вероятности ошибки  $10^{-2} \dots 10^{-3}$ , то при переходе от  $m = 4$  к

$m = 8$  проигрыш составляет уже 2,5 дБ и ещё более возрастает при переходе от  $m = 8$  к  $m = 16$ .

Тот факт, что кривые помехоустойчивости для  $m = 4$ , построенные по точной формуле и на основе аддитивной границы, практически совпадают, свидетельствует о том, что метод расчёта помехоустойчивости, основанный на вычислении аддитивной границы, является достаточно точным. Единственным его недостатком является, как уже было сказано, то, что он перестаёт работать при малых значениях сигнал-шум, причём область значений некорректной работы смещается вправо при увеличении числа  $m$  сигналов в ансамбле.

### Сигналы с фазово-амплитудной манипуляцией

Рассмотрим ансамбль сигналов с *фазово-амплитудной манипуляцией* (ФАМ), характеризуемых тем, что в передаваемом колебании возможны как изменение амплитуды, так и скачки начальной фазы на  $180^\circ$ :

$$s_k(t) = A_0 d_k \cos(\omega_0 t + \varphi_0), \quad k = 1, \dots, m, \quad 0 \leq t \leq T, \quad (3.2.55)$$

где информационным параметром является

$$d_k = \frac{m - 2k + 1}{m - 1}, \quad k = 1, \dots, m, \quad (3.2.56)$$

а начальная фаза  $\varphi_0$  (неинформационный параметр) равномерно распределена на интервале  $[0; 2\pi]$ . Абсолютное значение параметра  $d_k$  определяет амплитудную составляющую модуляции ( $m/2$  значений амплитуды), а его знак — фазовую составляющую (скачки фазы на  $180^\circ$ ).

На рис. 3.17 показаны сигнальные точки сигналов с ФАМ для различных значений  $m$ .

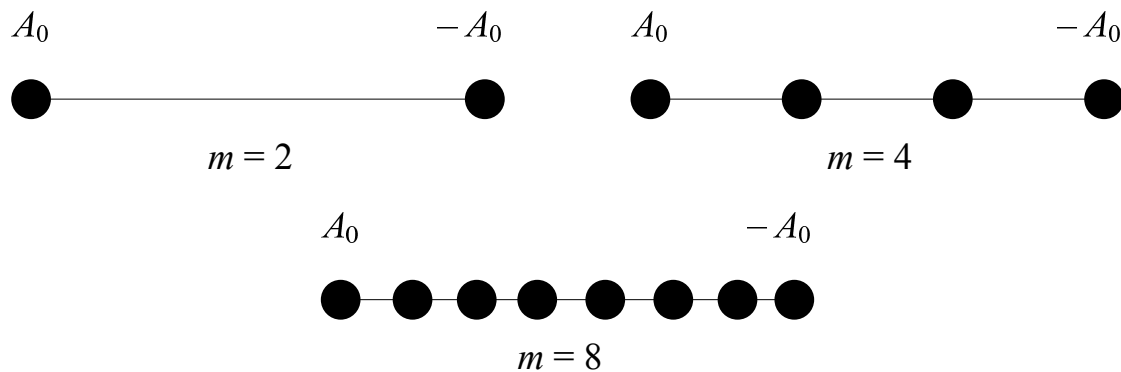


Рис. 3.17. Расположение на плоскости точек ансамбля сигналов с ФАМ

Коэффициент корреляции рассматриваемых сигналов равен

$$r_{ki} = \frac{A_0^2 d_k d_i}{\sqrt{E_k E_i}} \int_0^T \cos^2(\omega_0 t + \varphi_0) dt \approx \frac{A_0^2 T d_k d_i / 2}{\sqrt{A_0^2 T d_k^2 / 2} \sqrt{A_0^2 T d_i^2 / 2}} =$$

$$= \frac{d_k d_i}{|d_k| |d_i|} = \text{sign}(d_k) \text{sign}(d_i). \quad (3.2.57)$$

Здесь  $\text{sign}$  — функция знака:

$$\text{sign}(x) = \begin{cases} +1, & x > 0; \\ 0, & x = 0; \\ -1, & x < 0. \end{cases}$$

Следовательно, эвклидово расстояние составляет

$$\Delta_{ki} = E_k + E_i - 2r_{ki} \sqrt{E_k E_i} = \frac{A_0^2 T}{2} (d_k^2 - 2d_k d_i + d_i^2) =$$

$$= \frac{A_0^2 T}{2} (d_k - d_i)^2 = \frac{2A_0^2 T (k-i)^2}{(m-1)^2},$$

а его минимальное значение равно

$$\Delta_{\min} = \frac{2A_0^2 T}{(m-1)^2}.$$

Сигналы с ФАМ не являются равномошными — их энергии различны. Пусть  $E = A_0^2 T / 2$  — максимальная энергия (энергия сигналов, для которых  $|d_k| = 1$ ). Тогда

$$\Delta_{ki} = \frac{4E(k-i)^2}{(m-1)^2} \quad (3.2.58)$$

и

$$\Delta_{\min} = \frac{4E}{(m-1)^2}. \quad (3.2.59)$$

На основе полученных соотношений аддитивная граница для сигналов с ФАМ имеет вид

$$p \leq p_{\text{ош ФАМ}, m} = \frac{1}{2m} \sum_{k=1}^m \sum_{\substack{i=1 \\ i \neq k}}^m \left( 1 - \text{erf} \left( \frac{h|k-i|}{m-1} \right) \right) \quad (3.2.60)$$

а её грубая оценка

$$p \leq \frac{m-1}{2} \left[ 1 - \operatorname{erf} \left( \frac{h}{m-1} \right) \right]. \quad (3.2.61)$$

На рис. 3.18 показаны кривые помехоустойчивости сигналов с ФАМ полученные на основе аддитивной границы и её грубой оценки.

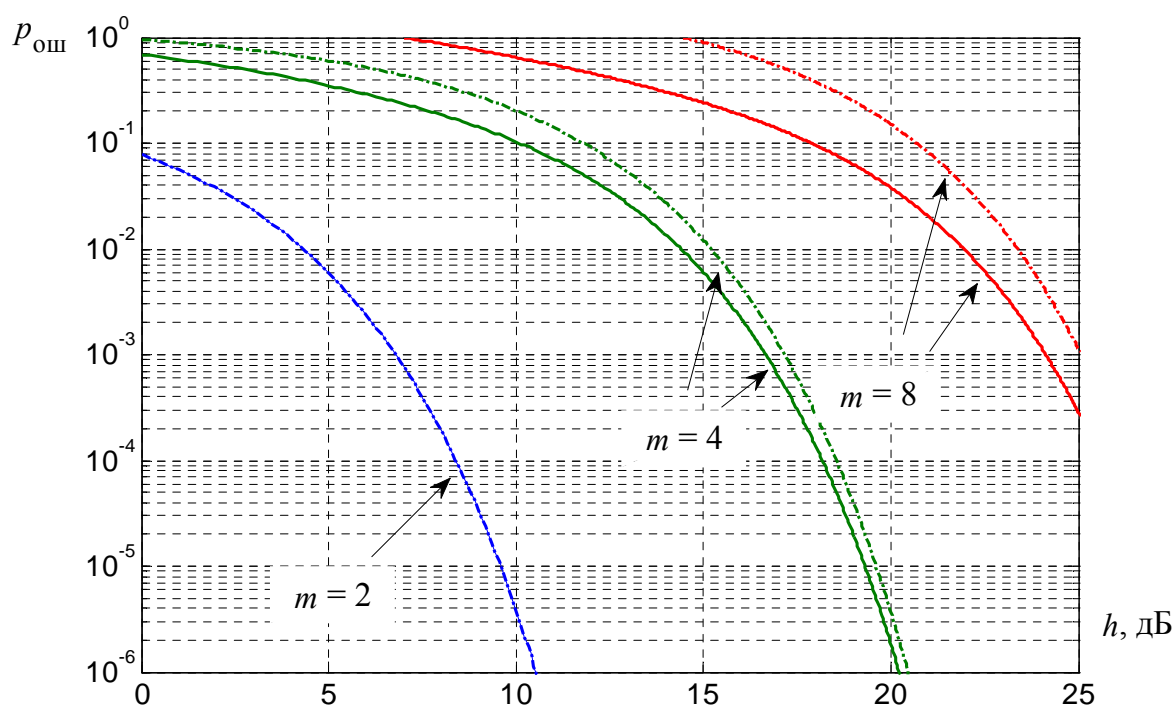


Рис. 3.18. Аддитивная граница для сигналов с ФАМ

Кривая для  $m = 2$  соответствует приёму противоположных сигналов и, как и для сигналов с ФМ, совпадает с аналогичной зависимостью, представленной на рис. 3.10

Как следует из анализа кривых, представленных на рис. 3.18, сигналы с ФАМ существенно проигрывают в помехоустойчивости сигналам с ФМ, имеющим такое же значение  $m$ . Так, в области значений вероятности ошибок  $10^{-2} \dots 10^{-3}$  энергетический проигрыш для  $m = 4$  составляет около 5 дБ, а для  $m = 8$  — около 8 дБ. Поэтому сами по себе такие сигналы не имеют самостоятельного практического применения, однако они являются основой для построения многих многопозиционных ансамблей с амплитудно-фазовой и частотной модуляцией.

## Сигналы с квадратурной амплитудной манипуляцией

Одним из наиболее часто используемых на практике многопозиционных ансамблей, построенных на основе сигналов с ФАМ, является ансамбль сигналов с *квадратурной амплитудной манипуляцией* (КАМ). Традиционный  $m$ -элементный ансамбль сигналов с КАМ формируется посредством использования двух  $\sqrt{m}$ -элементных ансамблей сигналов с ФАМ в сдвинутых на  $\pi/2$  квадратурах:

$$s_q(t) = A_0 d_k \cos(\omega_0 t + \varphi_0) - A_0 d_i \sin(\omega_0 t + \varphi_0), \quad 0 \leq t \leq T,$$

$$k, i = 1, \dots, \sqrt{m}, \quad q = 1, \dots, m. \quad (3.2.62)$$

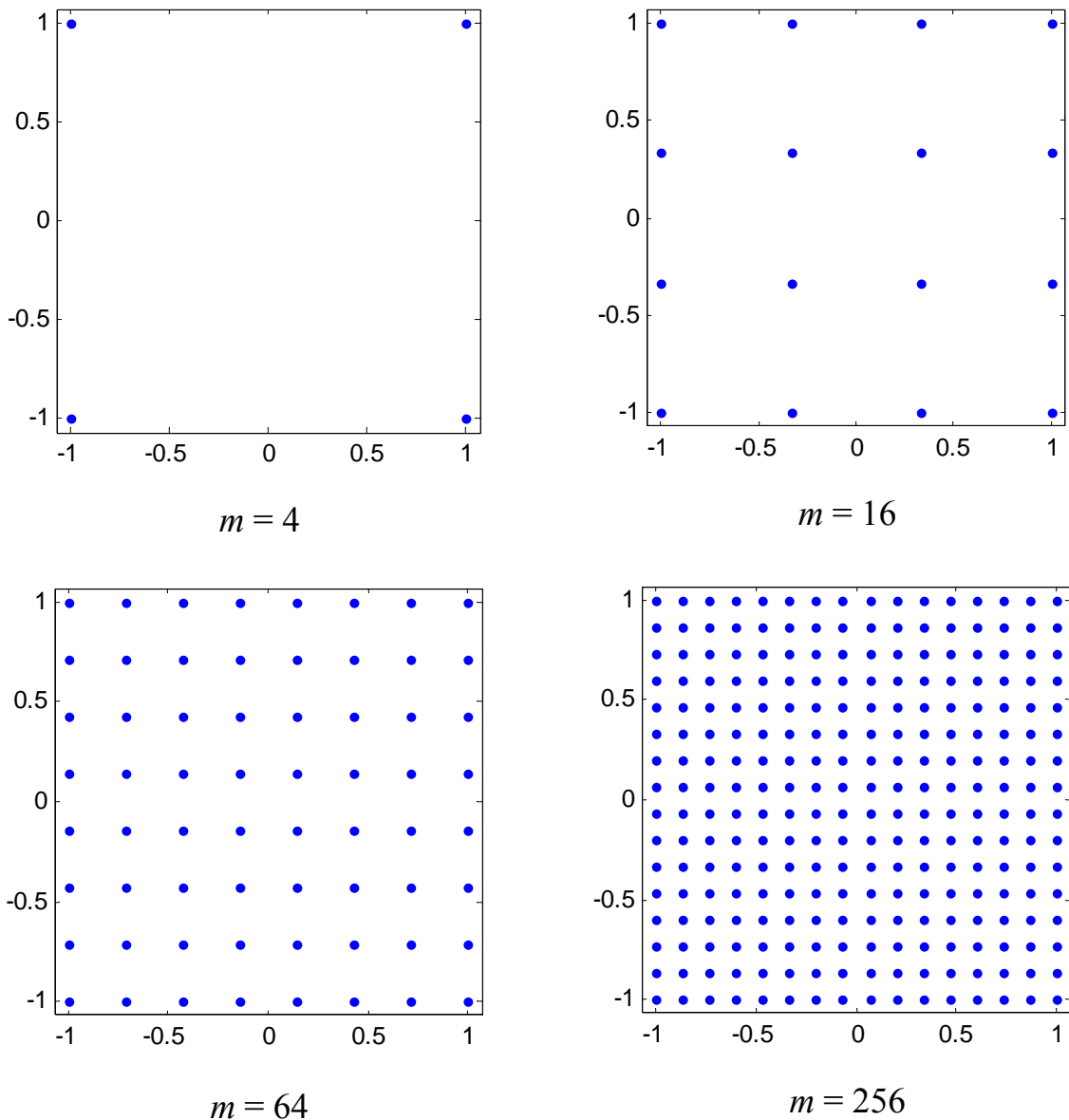


Рис. 3.18. Расположение на плоскости точек ансамбля сигналов с КАМ

Соотношение (3.2.62) можно записать в квадратурном виде:

$$s_q(t) = A_0 V_{ki} \cos(\omega_0 t + \varphi_{ki} + \varphi_0), \quad (3.2.62a)$$

где

$$V_{ki} = \sqrt{d_k^2 + d_i^2},$$

$$\operatorname{tg} \varphi_{ki} = d_i / d_k,$$

а

$$d_k = \frac{\sqrt{m} - 2k + 1}{\sqrt{m} - 1}.$$

На рис. 3.18 показаны сигнальные точки сигналов с КАМ для различных значений  $m$ ; для простоты выбраны значения  $\varphi_0 = 0$  и  $A_0 = 1$ .

Как видно, сигналы с КАМ в общем случае характеризуются множественными значениями амплитуды и фазы. Например, для  $m = 16$  сигнальные точки содержат 3 значения амплитуды и 12 значений фазы.

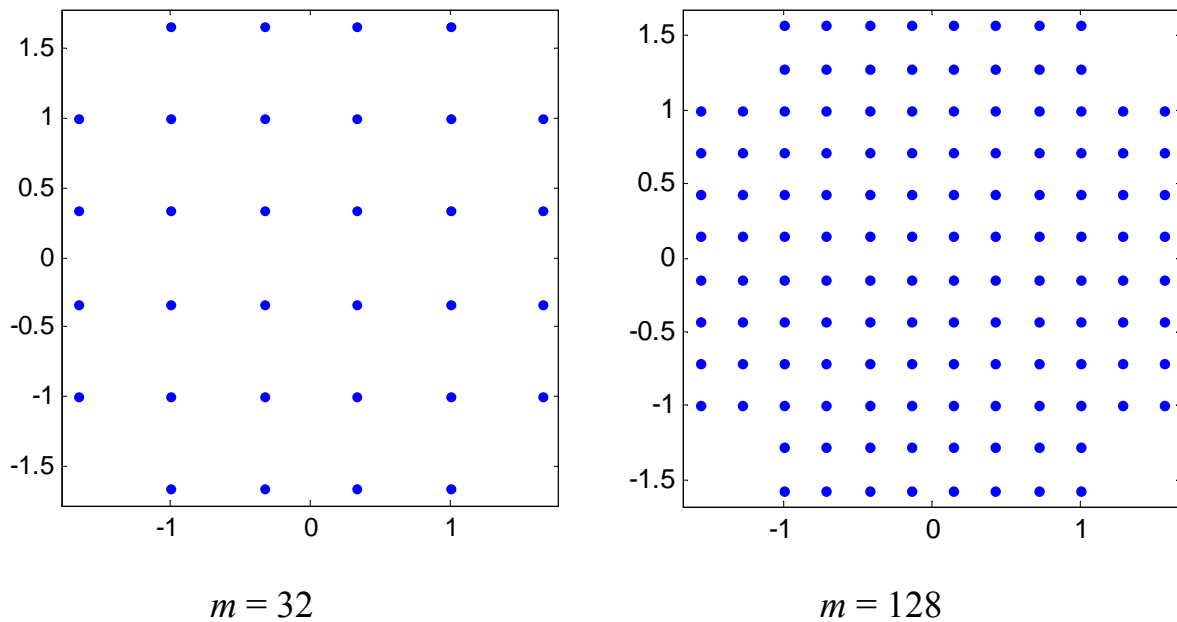


Рис. 3.19. Сигнальные созвездия крестообразных ансамблей сигналов с КАМ

Соответствие между нумерацией сигналов с ФАМ (индексы  $k$  и  $i$ ) и нумерацией сигналов с КАМ (индекс  $q$ ) произвольно. Обычно значения  $r$  устанавливают при последовательном возрастании индексов  $k$  и  $i$ :

$$s_1(t) \div d_1, d_1;$$

$$s_2(t) \div d_1, d_2;$$

$$s_3(t) \div d_2, d_1;$$

$$s_4(t) \div d_2, d_2;$$

и т. д.

В (3.2.62) предполагается, что  $\sqrt{m}$  — целое число, т. е.  $m$  является чётной степенью целого основания (обычно двойки):  $4 = 2^{2^1}$ ,  $16 = 2^{2^2}$ ,  $64 = 2^{2^3}$ ,  $64 = 2^{2^4}$  и т. д. Однако заметим, что в настоящее время используются “нетрадиционные” ансамбли с КАМ, в которых это соотношение не выполняется, например,  $m = 32$  или  $m = 128$ . В таких ансамблях сигнальные точки не располагаются равномерно внутри квадрата; сигнальное созвездие имеет крестообразный вид (рис. 3.19).

Найдём эвклидово расстояние для произвольной пары сигнальных точек традиционных сигналов с КАМ. Энергия определяется амплитудной составляющей:

$$E_q = A_0^2 V_{ki}^2 \int_0^T \cos^2(\omega_0 t + \varphi_{ki} + \varphi_0) dt \approx \frac{A_0^2 T V_{ki}^2}{2}.$$

Тогда коэффициент корреляции между парой сигналов  $s_{q(k,i)}(t)$  и  $s_{p(l,j)}(t)$  равен

$$r_{qp} = \frac{A_0^2}{\sqrt{(A_0^2 T V_{ki}^2)/2} \sqrt{(A_0^2 T V_{lj}^2)/2}} \int_0^T [d_k \cos(\omega_0 t + \varphi_0) - d_i \sin(\omega_0 t + \varphi_0)] \times$$

$$\times [d_l \cos(\omega_0 t + \varphi_0) - d_j \sin(\omega_0 t + \varphi_0)] dt \approx$$

$$\approx \frac{d_k d_l + d_i d_j}{\sqrt{(d_k^2 + d_i^2)(d_l^2 + d_j^2)}}. \quad (3.2.63)$$

Отсюда получаем эвклидово расстояние для сигналов с КАМ:

$$\Delta_{qp} = \frac{A_0^2 T}{2} (d_k^2 + d_i^2) + \frac{A_0^2 T}{2} (d_l^2 + d_j^2) -$$

$$- \frac{A_0^2 T}{2} \frac{2(d_k d_l + d_i d_j)}{\sqrt{(d_k^2 + d_i^2)(d_l^2 + d_j^2)}} \sqrt{(d_k^2 + d_i^2)(d_l^2 + d_j^2)} =$$

$$= \frac{A_0^2 T}{2} [(d_k - d_l)^2 + (d_i - d_j)^2] = E[(d_k - d_l)^2 + (d_i - d_j)^2] =$$



$$= \frac{4E}{(\sqrt{m}-1)^2} [(k-l)^2 + (i-j)^2], \quad (3.2.64)$$

минимальное значение которого

$$\Delta_{\min} = \frac{4E}{(\sqrt{m}-1)^2}. \quad (3.2.65)$$

совпадает с минимальным евклидовым расстоянием сигналов с ФАМ (с заменой  $m$  на  $\sqrt{m}$ ), лежащих в основе ансамбля сигналов с КАМ. В формулах (3.2.64) и (3.2.65), по прежнему,  $E = A_0^2 T / 2$  — энергия сигнала с максимальной амплитудой.

На рис. 3.20 показаны кривые помехоустойчивости сигналов с КАМ полученные на основе аддитивной границы и её грубой оценки, вычисленные на основе соотношений (3.2.59) и (3.2.50).

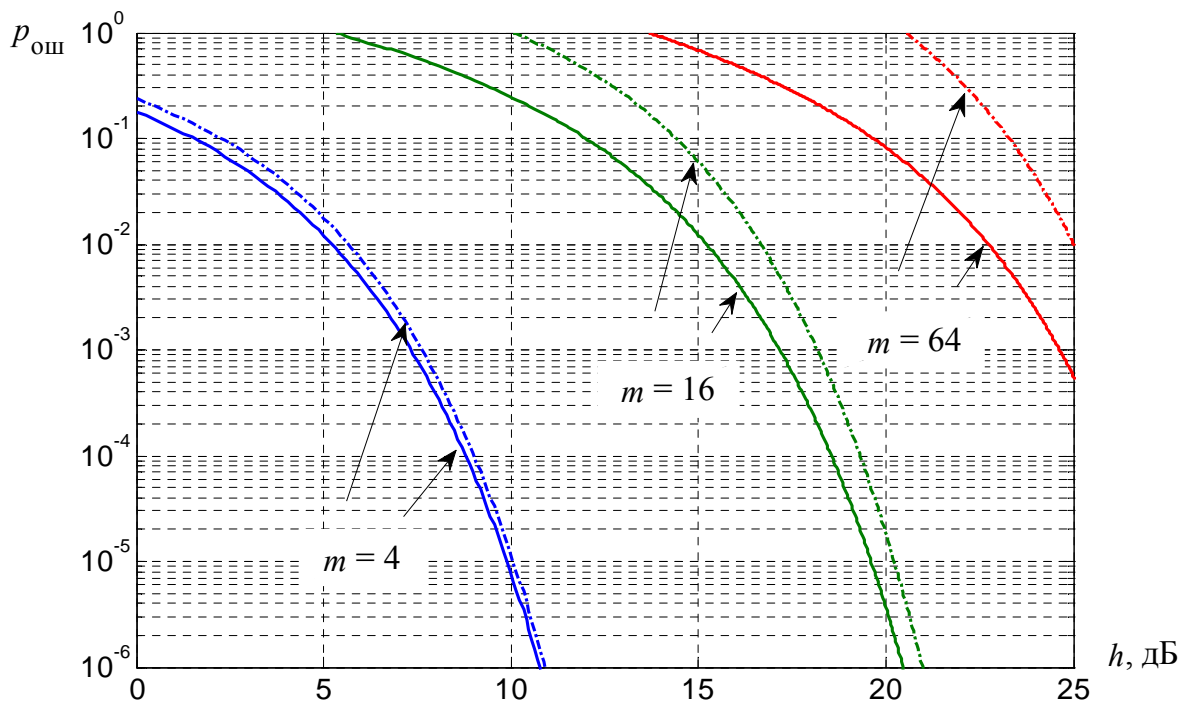


Рис. 3.20. Аддитивная граница для сигналов с КАМ

Как следует из сравнения рис. 3.20 с рис. 3.16 и 3.18, сигналы с КАМ при одинаковом объёме ансамбля обладают существенно более высокой помехоустойчивостью по сравнению с сигналами с ФМ и, тем более, сиг-

налами с ФАМ. Так, энергетический выигрыш сигналов КАМ по сравнению с сигналами с ФМ в области значений вероятности ошибок  $10^{-2} \dots 10^{-3}$  для  $m = 4$  составляет 3...3,5 дБ, а для  $m = 4$  — 7...8 дБ. Причина этому — большее эвклидово расстояние. Однако, несмотря на очевидный проигрыш в помехоустойчивости сигналы с ФМ обладают, как минимум, одним существенным преимуществом: все сигналы ансамбля имеют одинаковую амплитуду, что предопределяет более высокую эффективность их использования в условиях энергонапряжённой работы радиопередающих средств.

В завершении данного раздела сделаем одно важное замечание, касающееся сравнения помехоустойчивости различных ансамблей сигналов.

Прежде всего, отметим, что следует с большой аккуратностью проводить сравнение кривых помехоустойчивости, приведённых в научной и справочной литературе, поскольку зачастую под одними и теми же обозначениями используются разные величины. В частности, многие авторы под отношением сигнал-шум понимают (и представляют на графиках) не величину  $h$ , а её квадрат, т. е. в выражении (3.2.13) определяют

$$h = \mu^2 E / N_0.$$

Это означает, что соответствующие кривые помехоустойчивости оказываются смещёнными вправо в 2 раза (на 3 дБ), и некорректное сопоставление таких кривых с аналогичными зависимостями, в которых под  $h$  понимается

$$h = \sqrt{\mu^2 E / N_0},$$

может привести к ошибочным выводам.

Далее, обратим внимание, что анализ помехоустойчивости различных ансамблей сигналов проводился на основе “абсолютных” значений вероятностей ошибки без учёта каких-либо соображений, касающихся того, каким образом с помощью выбранных сигналов осуществляется передача информации. Например, из сравнения кривых помехоустойчивости следует, что использование сигналов с ФМ-4 приводит к заметному энергетическому проигрышу около 3,5 дБ по отношению к сигналам с ФМ-2. Однако при этом необходимо учесть, что при ФМ-4 передаётся в 2 раза больше количества информации [1], чем при ФМ-2, и, возможно, этот факт окажет корректирующее влияние на окончательный вывод о преимуществах использования одного ансамбля над другим. В этой связи желательно иметь более общую характеристику помехоустойчивости, позволяющую сравни-

вать между собой различные, в том числе, разнопозиционные ансамбли сигналов.

В качестве такой характеристики обычно используют предложенную Л. М. Финком *эквивалентную вероятность ошибки*  $p_3$  [13], суть которой состоит в том, что для анализа помехоустойчивости ансамбля, содержащего  $m$  сигналов, рассматривают зависимость  $p_3 = p_{\text{ош}} / \log_2 m$  от  $h / \sqrt{\log_2 m}$  или, при вычислении в децибелах, от  $10 \lg(h / \sqrt{\log_2 m})$ .

### 3.3. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ АЛГОРИТМОВ ОПТИМАЛЬНОЙ ОБРАБОТКИ СИГНАЛОВ

В предыдущем разделе были получены алгоритмы оптимального приёма сигналов для модели аддитивного белого гауссовского шума и предложены структурные схемы устройств, реализующих такие алгоритмы для наиболее часто встречающихся задач обнаружения и различения сигналов.

Основу алгоритмов оптимальной обработки составляет вычисление корреляционного интеграла, и, как было указано, его практическая реализация возможна либо на основе корреляторов, либо на основе согласованных фильтров [14]. Несмотря на то, что современная элементная база позволяет реализовывать различные схемотехнические решения весьма эффективно (высокое быстродействие, малые энергозатраты, низкая стоимость и т. п.), тем не менее, такие “прямые” подходы в настоящее время представляются устаревшими. Практическая реализация алгоритмов оптимального приёма основана на применении методов цифровой обработки сигналов либо с использованием программируемой логики (цифровых сигнальных процессоров, программируемых логических интегральных схем), либо с непосредственным включением в систему универсального или специализированного вычислителя.

Рассмотрим обобщённую структурную схему типичного приёмного устройства, входящего в состав программно-аппаратного комплекса, решающего различные задачи приёма, обработки и анализа сообщений (рис. 3.14).

Принятый из эфира сигнал  $z(t)$  преобразуется на нулевую частоту (на рис. 3.14 показан только один преобразователь частоты; реально таких преобразователей может быть несколько), фильтрами нижних частот ФНЧ

выделяются низкочастотные синфазная  $z_c(t)$  и квадратурная  $z_s(t)$  компоненты. Далее в аналого-цифровых преобразователях АЦП происходит их дискретизация с частотой  $F_d$  и квантование на заданное количество уровней, например  $2^{16}$ . Оцифрованные отсчеты  $i_l$  и  $q_l$ , где  $l = 0, 1, 2, \dots$ , подаются на блок цифровой обработки (БЦО), где они будут рассматриваться как единый комплексный отсчёт  $c_l = i_l + jq_l$ . Результатом такой обработки, как правило, являются битовые информационные последовательности, поступающие непосредственно на устройства воспроизведения информации: ЖК-дисплей, звуковая система и т. д. Если устройство является активным, то полученные последовательности также могут быть использованы в качестве управляющей информации для других блоков устройства.

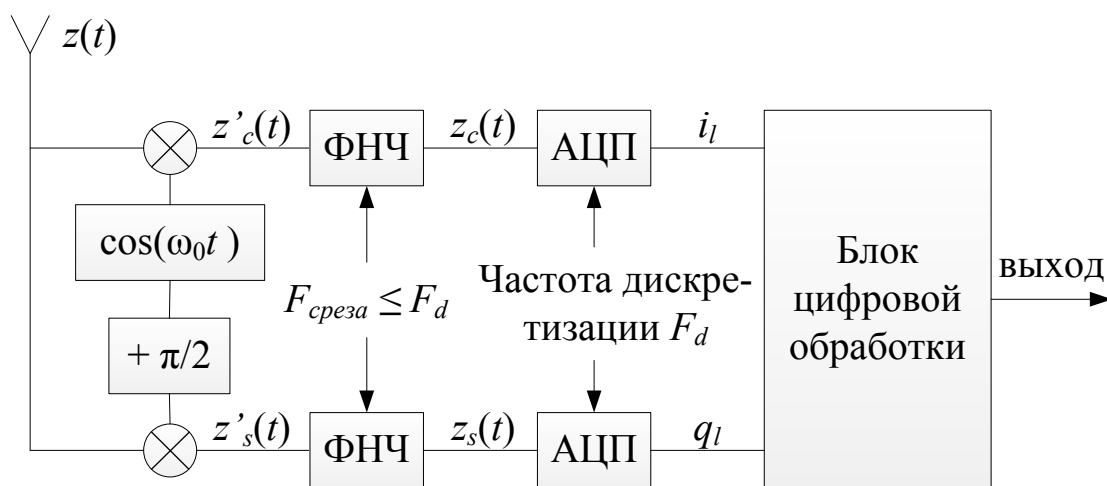


Рис. 3.14. Обобщённая структурная схема приёмного устройства

Видно, что в схеме на рис. 3.14 присутствует минимальная аналоговая обработка сигнала. Это является общей тенденцией в построении всех современных телекоммуникационных устройств и обусловлено быстрым развитием цифровой схемотехники и повышением уровня универсализации при цифровой обработке сигналов.

Рассмотрим подробнее операции переноса на нулевую частоту и дискретизации сигнала. Согласно теореме Котельникова, для восстановления сигнала с финитным спектром по его дискретным отсчётам достаточно выбрать частоту дискретизации, вдвое превосходящую максимальную частоту спектра сигнала. Понятно, что при попытке дискретизации сигнала на несущей частоте, во-первых, понадобился бы сверхбыстродействующий

АЦП; во-вторых, необходимая скорость АЦП зависела бы от несущей частоты сигнала; наконец, в-третьих, в получаемой последовательности отсчётов была бы большая избыточность, т. е. сильная зависимость значений соседних отсчётов. Для избавления от указанных недостатков прямой дискретизации используется перенос сигнала на нулевую частоту.

Следует отметить, что, конечно, все реальные сигналы имеют нефинитный спектр, так как сами сигналы финитны во времени, тем не менее, спектры всех реальных сигналов убывают при отстройке от некоторой центральной частоты и, следовательно, большая часть информации содержится в конечной полосе. Именно на эту полосу (определяемую по тому или иному критерию) настраиваются ФНЧ, которые необходимы для преобразования спектра сигнала к “почти” финитному виду перед процедурой аналого-цифрового преобразования. Преобразование на нулевую частоту в двух квадратурах позволяет в каждом АЦП производить дискретизацию с частотой, равной верхней компоненте спектра сигнала, а не удвоенной, как это требовалось бы при рассмотрении одной квадратуры.

Кроме дискретизации по времени в АЦП производится квантование по уровню входного сигнала, что приводит к дополнительной потере информации, тем не менее, для современных АЦП с количеством уровней  $2^{16} \dots 2^{32}$  потери практически незаметны. Выбор степеней 16 и 32 обусловлен байтовым представлением данных в цифровых устройствах<sup>1</sup>. Как правило, осуществляется равномерное симметричное квантование со знаком. Отметим, что степень потери информации при квантовании, конечно, зависит от того как соотносятся динамический диапазон АЦП и входного сигнала. Например, если максимальный уровень сигнала перед АЦП превосходит максимальный уровень АЦП, то будет происходить усечение сигнала и, следовательно, потеря части информации; если, наоборот, максимальный уровень сигнала меньше максимального уровня АЦП, то часть уровней квантования не будут задействованы и, следовательно, сигнал будет квантован менее подробно, чем позволяет АЦП — потенциально часть информации также будет потеряна. Таким образом, лучшим вариантом с точки зрения минимума потери информации является совпадение динами-

---

<sup>1</sup> В большинстве языков программирования, в том числе в MATLAB, соответствующие типы данных называются `int16` и `int32`.

ческих диапазонов сигнала и АЦП. Для выравнивания динамических диапазонов, очевидно, должна быть предусмотрена система автоматической регулировки усиления, которая на рис. 3.14 не показана.

Итак, на входе БЦО имеется последовательность комплексных чисел, вещественная и мнимая части которой являются целыми числами со знаком, не превосходящими по модулю заданное число ( $2^{15}$  либо  $2^{31}$ ). Последовательность либо записывается в файл и затем обрабатывается в режиме отложенного времени, либо обрабатывается в режиме реального времени непрерывно (как правило, определёнными “порциями”-сегментами). Обработка в режиме реального времени принципиально отличается только тем, что необходимо согласовывать скорость и количество операций, выполняемых для получения полезной информации, со скоростью поступления цифровой информации (т. е. последовательности  $c_l$ ) в БЦО.

Прежде чем обратиться к решению задачи приёма и обработки сигналов в БЦО, проведём ряд преобразований полученного выше алгоритма оптимального различения сигналов (3.2.21)<sup>1</sup>.

Во-первых, умножим обе части всех неравенств на  $-2\mu$ , во-вторых, добавим энергию принятого сигнала:

$$\int_0^T z^2(t)dt - 2\mu \int_0^T z(t)s_r(t)dt + \mu^2 E_r < \int_0^T z^2(t)dt - 2\mu \int_0^T z(t)s_q(t)dt + \mu^2 E_q, \quad \forall r \neq q,$$

и, в-третьих, внесём слагаемые каждой части всех неравенств под общий знак интеграла:

$$\int_0^T \left( z^2(t) - 2\mu z(t)s_r(t) + \mu^2 s_r^2(t) \right) dt < \int_0^T \left( z^2(t) - 2\mu z(t)s_q(t) + \mu^2 s_q^2(t) \right) dt$$

или

$$\int_0^T \left( z(t) - \mu s_r(t) \right)^2 dt < \int_0^T \left( z(t) - \mu s_q(t) \right)^2 dt, \quad \text{для всех } r \neq q. \quad (3.3.1)$$

---

<sup>1</sup> В данном разделе переменная  $i$  использована под обозначение синфазной цифровой последовательности на входе БЦО, а переменная  $k$  будет задействована под обозначение часто употребляемого номера тактового интервала. Поэтому для обозначения индексов гипотез в оптимальном алгоритме различения сигналов будут использоваться переменные  $r$  и  $q$ .

В обеих частях всех неравенств (3.3.1) записаны квадраты евклидовых расстояний между принятым и опорными сигналами. Таким образом, можно сформулировать алгоритм оптимального различения сигналов следующим образом: *выносится решение в пользу той гипотезы  $H_r$  о наличии на входе приёмного устройства сигнала  $s_r(t)$ ,  $r = 1, \dots, m$ , для которой евклидово расстояние между принятым сигналом  $z(t)$  и сигналом  $s_r(t)$  окажется минимальным.* По-видимому, такое определение является интуитивно более очевидным, чем данное выше.

Перейдём к решению задачи приёма сигналов с амплитудно-фазовой модуляцией. Рассмотрим случайную последовательность  $y_N(t)$ , состоящую из  $N$  независимых сигналов с произвольным  $m$ -элементным сигнальным созвездием, сформированных на несущей частоте  $f_0$  и следующих с тактовым интервалом  $T$ :

$$y_N(t) = \sum_{k=0}^{N-1} A_r^{(k)} a(t - kT) \cos(2\pi f_0 t + \varphi_r^{(k)}), \quad 0 \leq t \leq LT + (N-1)T, \quad (3.3.2)$$

где  $a(t)$  — ограниченная на интервале  $[0; LT]$  огибающая,  $A_r^{(k)}$ ,  $\varphi_r^{(k)}$  — амплитуда и фаза, соответствующие точке сигнального созвездия, причём нижний индекс  $r = 1, 2, \dots, m$  соответствует номеру в канальном алфавите, а верхний индекс  $k = 0, 1, \dots, N-1$  (здесь и в других обозначениях ниже) отражает принадлежность к соответствующему тактовому интервалу; при этом  $A_r^{(k)} \in [0; 1]$  и  $\varphi_r^{(k)} \in [-\pi; \pi]$ . Так, для созвездия ФМ-4 имеем:  $A_r^{(k)} = 1$ ,  $\varphi_r^{(k)} = -3\pi/4 + r\pi/2$ ,  $r = 1, 2, 3, 4$ .

Форма огибающей  $a(t)$  длительности  $LT$  выбирается, исходя из требований к компактности энергетического спектра случайной последовательности таких сигналов<sup>1</sup>. Заметим, что при  $L > 1$  выражение (3.3.2) соответствует сигналам с межсимвольной интерференцией (МСИ), для которых применение поэлементного алгоритма приёма в общем случае неэффективно.

---

<sup>1</sup> Как известно [12], спектр случайной последовательности сигналов (3.3.2) полностью определяется видом огибающей  $a(t)$  и не зависит от вида сигнального созвездия. Например, чем больше производных  $a(t)$  на краях тактового интервала обращаются в нуль, тем больше скорость спада спектра случайной последовательности сигналов (3.3.2) в области внеполосных излучений.

Рассмотрим случай передачи сигналов без МСИ, когда  $L = 1$ . В качестве примера соответствующих огибающих можно представить следующее семейство кривых:

$$a(t) = \sin^d(\pi t/T), \quad d = 0, 1, \dots, \quad 0 \leq t \leq T. \quad (3.3.3)$$

Для  $d = 0$  получаем традиционную прямоугольную огибающую. Замечательным свойством семейства (3.3.3) является то, что каждая его огибающая обеспечивает убывание частотного спектра последовательности сигналов (3.3.2) не медленнее, чем  $C/\omega^{(d+1)}$ , где  $C$  — некоторая константа.

В соответствии с алгоритмом (3.3.1) для каждого тактового интервала в БЦО необходимо находить минимум следующего выражения:

$$\int_{kT}^{T+kT} \left( z^{(k)}(t) - \mu s_q^{(k)}(t) \right)^2 dt, \quad q = 1, \dots, m, \quad (3.3.4)$$

где

$$z^{(k)}(t) = \mu A_r^{(k)} a(t - kT) \cos(2\pi f_0 t + \varphi_r^{(k)}) + n^{(k)}(t),$$

$$s_q^{(k)}(t) = A_q^{(k)} a(t - kT) \cos(2\pi f_0 t + \varphi_q^{(k)}).$$

Здесь  $z^{(k)}(t)$  — принятый на  $k$ -м тактовом интервале входной сигнал;  $n^{(k)}(t)$  — реализация белого гауссовского шума, имеющего нулевое среднее и одностороннюю спектральную плотность средней мощности  $N_0/2$ ;  $\mu$ , как и прежде, — коэффициент ослабления сигнала.

Для преобразования (3.3.4) разложим принятый и опорный сигналы по низкочастотным квадратурам:

$$z^{(k)}(t) = \left[ \mu A_r^{(k)} a(t - kT) \cos \varphi_r^{(k)} + n_c^{(k)}(t) \right] \cos(2\pi f_0 t) - \left[ \mu A_r^{(k)} a(t - kT) \sin \varphi_r^{(k)} + n_s^{(k)}(t) \right] \sin(2\pi f_0 t), \quad (3.3.5)$$

$$s_q^{(k)}(t) = A_q^{(k)} a(t - kT) \cos \varphi_q^{(k)} \cos(2\pi f_0 t) - A_q^{(k)} a(t - kT) \sin \varphi_q^{(k)} \sin(2\pi f_0 t). \quad (3.3.6)$$

где<sup>1</sup>

$$n_c^{(k)}(t) = n^{(k)}(t) \cos(2\pi f_0 t),$$

$$n_s^{(k)}(t) = -n^{(k)}(t) \sin(2\pi f_0 t).$$

---

<sup>1</sup>Величины  $n_c^{(k)}$  и  $n_s^{(k)}$  не являются низкочастотными квадратурными компонентами при представлении шума  $n(t)$  в квадратурном виде, обозначаемыми обычно такими же символами.



Вернёмся к преобразованию (3.3.4), используя (3.3.5) и (3.3.6):

$$\begin{aligned} & \int_{kT}^{T+kT} \left( z^{(k)}(t) - \mu s_q^{(k)}(t) \right)^2 dt = \\ & = \int_{kT}^{T+kT} \left[ \left( \mu A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} + n_c^{(k)} - \mu A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)} \right) \cos(2\pi f_0 t) - \right. \\ & \quad \left. - \left( \mu A_r^{(k)} a(t-kT) \sin \varphi_r^{(k)} + n_s^{(k)} - \mu A_q^{(k)} a(t-kT) \sin \varphi_q^{(k)} \right) \sin(2\pi f_0 t) \right]^2 dt. \end{aligned}$$

Раскрыв скобки и отбросив быстроосциллирующие слагаемые, получим

$$\begin{aligned} & \int_{kT}^{T+kT} \left( z^{(k)}(t) - \mu s_q^{(k)}(t) \right)^2 dt = \\ & = \frac{1}{2} \int_{kT}^{T+kT} \left[ \left( \mu A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} + n_c^{(k)} - \mu A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)} \right)^2 + \right. \\ & \quad \left. + \left( \mu A_r^{(k)} a(t-kT) \sin \varphi_r^{(k)} + n_s^{(k)} - \mu A_q^{(k)} a(t-kT) \sin \varphi_q^{(k)} \right)^2 \right] dt = \\ & = \frac{1}{2} \int_{kT}^{T+kT} \left| Z^{(k)}(t) - \mu S_q^{(k)}(t) \right|^2 dt, \end{aligned}$$

где

$$\begin{aligned} Z^{(k)}(t) &= \left( \mu A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} + n_c^{(k)} \right) + j \left( \mu A_r^{(k)} a(t-kT) \sin \varphi_r^{(k)} + n_s^{(k)} \right), \\ S_q^{(k)}(t) &= A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)} + j A_q^{(k)} a(t-kT) \sin \varphi_q^{(k)}. \end{aligned}$$

Множитель 1/2 в полученном выражении можно опустить, так как он не влияет на результат поиска минимума.

Покажем, как связан комплексный сигнал  $Z^{(k)}(t)$  с низкочастотными квадратурными составляющими  $z_c(t)$  и  $z_s(t)$ . Для этого запишем последовательные преобразования сигнала  $z(t)$  в соответствии с рис. 3.14. Первым этапом происходит преобразование на нулевую частоту в квадратурах:

$$\begin{aligned} z'_c(t) &= z(t) \cos(2\pi f_0 t); \\ z'_s(t) &= -z(t) \sin(2\pi f_0 t), \end{aligned}$$

а затем фильтрация в ФНЧ. Подставляя (3.3.5) и отбрасывая высокочастотные компоненты, получим:

$$z_c^{(k)}(t) = \left( \mu A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} + n_c^{(k)} \right) / 2, \quad (3.3.7a)$$

$$z_s^{(k)}(t) = (\mu A_r^{(k)} a(t - kT) \sin \varphi_r^{(k)} + n_s^{(k)}) / 2. \quad (3.3.7б)$$

Видно, что

$$Z^{(k)}(t) = 2(z_c^{(k)}(t) + jz_s^{(k)}(t)).$$

Запишем преобразованный алгоритм различения сигналов без МСИ: на  $k$ -м тактовом интервале выносится решение в пользу той гипотезы  $H_q$  о наличии на входе приёмного устройства сигнала  $s_q(t)$ ,  $q = 1, \dots, m$ , соответствующего передаче  $q$ -го символа канального алфавита, для которой эвклидово расстояние между удвоенным принятым комплексным сигналом  $z_c^{(k)}(t) + jz_s^{(k)}(t)$  и опорным комплексным сигналом  $S_q^{(k)}(t)$ , взятым с учётом коэффициента ослабления  $\mu$ , окажется минимальным:

$$\int_{kT}^{T+kT} \left( 2(z_c^{(k)}(t) + jz_s^{(k)}(t)) - \mu S_q^{(k)}(t) \right)^2 dt, \quad q = 1, \dots, m. \quad (3.3.8)$$

Последним шагом в преобразовании алгоритма различения сигналов для БЦО является переход к дискретным сигналам. Такой переход означает замену вычисления интеграла на вычисление дискретной суммы (например, суммы площадей прямоугольников). Для пояснения предположим, что входной сигнал оцифрован с частотой  $F_d = K/T$ , где  $K$  — натуральное число. Это значит, что на каждый тактовый интервал сигнала приходится ровно  $K$  отсчётов. Для удобства будем считать, что первый отсчёт комплексной последовательности  $c_l$ , присутствующей на входе БЦО (рис. 3.14), является первым отсчётом первого тактового интервала принятого сигнала. Тогда для нахождения оценки номера символа  $\hat{r}$  сигнального созвездия, передаваемого на  $k$ -м тактовом интервале, необходимо найти аргумент минимума следующего выражения:

$$\hat{r} = \arg \min_{q=1, \dots, m} \left\{ \sum_{l=1+(k-1)K}^{l=K+(k-1)K} (2c_l - \mu S_q(l'T))^2 \right\},$$

$$l'T = (l - (k-1)K - 1/2)T / K, \quad (3.3.9)$$

где  $l'T$  — равномерно дискретизированная на длительности тактового интервала ось времени.

Обратимся теперь к приёму сигналов с МСИ, т. е.  $L > 1$ . Использование таких сигналов, с одной стороны, позволяет достигать большей степени компактности спектра, чем для сигналов без МСИ; с другой стороны,

наличие интерференции существенно ухудшает эффективность алгоритма поэлементного приёма. В настоящее время подавляющее большинство одностотных систем передачи информации используют сигналы с МСИ и так называемой *найквистовской огибающей*. Форма огибающей, используемая при построении таких сигналов, описывается, вообще говоря, нефи- нитной функцией

$$a(t) = \frac{\sin(\pi t(1 - \beta) / T) + 4\beta \frac{t}{T} \cos(\pi t(1 + \beta) / T)}{\pi t \left[ 1 - (4\beta t / T)^2 \right] / T}, \quad (3.3.10)$$

где  $\beta$  — коэффициент скругления, выбираемый из диапазона  $[0, 1]$  и определяющий степень компактности спектра. Чем меньше  $\beta$ , тем компактнее спектр и наоборот. При  $\beta = 0$  получаем огибающую вида  $\sin(x)/x$ , т. е. сигналы с предельно компактным прямоугольным спектром; при  $\beta = 1$  спектр будет иметь вид корня из приподнятого косинуса. Заметим, что при значениях  $t = 0$  и  $t = \pm T/4\beta$  в (3.3.10) необходимо вычислять пределы

$$a(0) = \lim_{t \rightarrow 0} a(t) = 1 - \beta + 4\beta / \pi,$$

и

$$a\left(\pm \frac{T}{4\beta}\right) = \lim_{t \rightarrow \pm T/4\beta} a(t) = \frac{\beta}{\sqrt{2}} \left[ \left(1 + \frac{2}{\pi}\right) \sin\left(\frac{\pi}{4\beta}\right) + \left(1 - \frac{2}{\pi}\right) \cos\left(\frac{\pi}{4\beta}\right) \right].$$

Вид найквистовской огибающей с различными значениями коэффициента скругления показан на рис. 3.15.

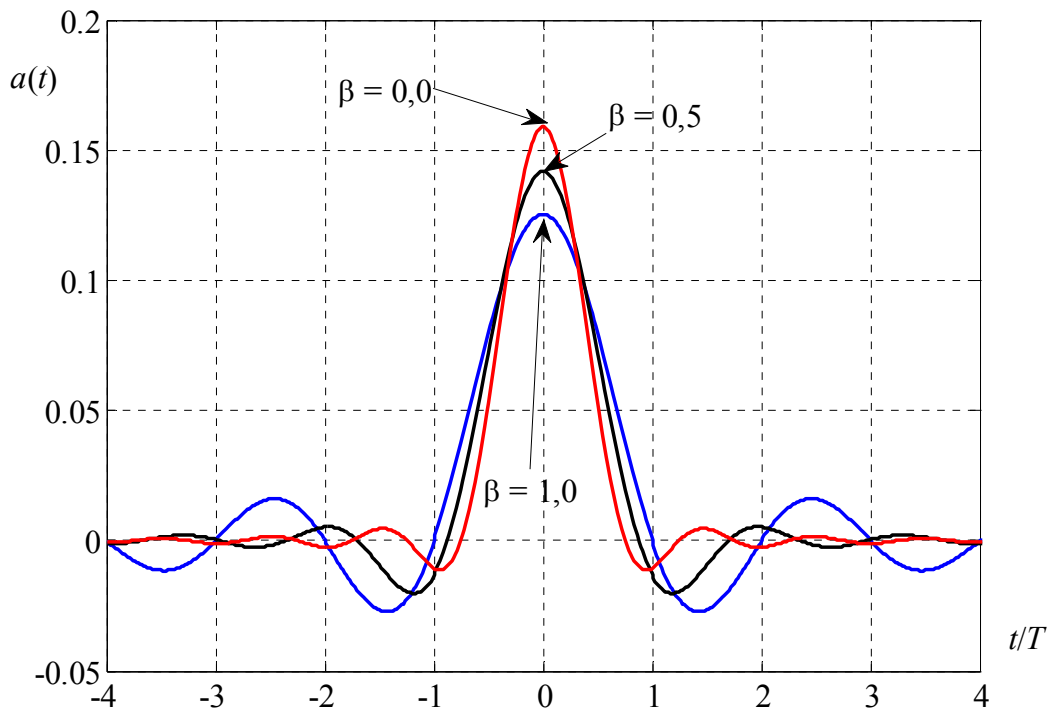


Рис. 3.15. Вид Найквистовской огибающей для различных значений коэффициента скругления

Важным свойством семейства огибающих (3.3.10) является равенство нулю их автокорреляционных функций в моменты времени кратные длительности тактового интервала  $T$ :

$$\int_{-\infty}^{+\infty} a(t)a(t - kT)dt = 0, \quad k = \pm 1, \pm 2, \pm 3, \dots \quad (3.3.11)$$

Как будет показано ниже, свойство (3.3.11) позволяет использовать для сигналов с МСИ и найквистовской огибающей алгоритм поэлементного принятия решений.

Хотя огибающая (3.3.10) является нефинитной, на практике обычно выбирают  $0,2 \leq \beta \leq 0,3$ , что обеспечивает достаточную степень компактности спектра и быстрый спад значений огибающей при удалении от  $t = 0$ , и отбрасывают “хвосты” огибающей, т. е., например, выбирают только<sup>1</sup>  $-10T \leq t \leq 10T$ . При этом все свойства найквистовской огибающей, описанные выше, остаются верными с высокой степенью точности.

<sup>1</sup> Заметим, что в этом случае  $L = 12$  и в качестве огибающей в (3.3.2) должна быть использована  $a(t - 10T)$ .

Обратимся к приёму сигналов с найквистовской огибающей. Как было сказано выше, из-за наличия МСИ прямое применение алгоритма поэлементного приёма нецелесообразно. Однако можно рассматривать всю последовательность (3.3.2) как один сигнал. В таком случае для реализации алгоритма различения придётся сравнивать принятый сигнал с  $m^N$  опорными сигналами, что для реальных значений  $N$  практически невозможно. Тем не менее, покажем, что благодаря свойству (3.3.11) можно реализовать поэлементный приём для сигналов с найквистовской огибающей и произвольным значением  $N$ . Для этого воспользуемся преобразованиями алгоритма различения, проведёнными для сигналов без МСИ.

Итак, для различения сигналов с МСИ в соответствии с алгоритмом (3.3.1) в БЦО необходимо находить минимум следующего выражения:

$$\int_0^{LT+(N-1)T} (z(t) - \mu s_Q(t))^2 dt, \quad Q=1, \dots, m^N, \quad (3.3.12)$$

где

$$z(t) = \mu \sum_{k=0}^{N-1} A_r^{(k)} a(t - kT) \cos(2\pi f_0 t + \varphi_r^{(k)}) + n(t),$$

$$s_Q(t) = \sum_{k=0}^{N-1} A_q^{(k)} a(t - kT) \cos(2\pi f_0 t + \varphi_q^{(k)}).$$

Здесь, как и прежде,  $z(t)$  — принятый сигнал;  $n(t)$  — реализация белого гауссовского шума, имеющего нулевое среднее и спектральную плотность средней мощности  $N_0/2$ ;  $\mu$  — коэффициент ослабления сигнала.

Для преобразования (3.3.12) разложим принятый и опорный сигналы по низкочастотным квадратурам:

$$z(t) = \left\{ \mu \sum_{k=0}^{N-1} A_r^{(k)} a(t - kT) \cos \varphi_r^{(k)} + n_c(t) \right\} \cos(2\pi f_0 t) -$$

$$- \left\{ \mu \sum_{k=0}^{N-1} A_r^{(k)} a(t - kT) \sin \varphi_r^{(k)} + n_s(t) \right\} \sin(2\pi f_0 t); \quad (3.3.13)$$

$$s_Q(t) = \left\{ \sum_{k=0}^{N-1} A_q^{(k)} a(t - kT) \cos \varphi_q^{(k)} \right\} \cos(2\pi f_0 t) -$$

$$-\left\{ \sum_{k=0}^{N-1} A_q^{(k)} a(t-kT) \sin \varphi_q^{(k)} \right\} \sin(2\pi f_0 t), \quad (3.3.14)$$

где

$$\begin{aligned} n_c(t) &= n(t) \cos(2\pi f_0 t); \\ n_s(t) &= -n(t) \sin(2\pi f_0 t). \end{aligned}$$

Вернёмся к преобразованию (3.3.12), используя (3.3.13) и (3.3.14):

$$\begin{aligned} & \int_0^{LT+(N-1)T} (z(t) - \mu s_Q(t))^2 dt = \\ &= \int_0^{LT+(N-1)T} \left[ \left( n_c(t) + \mu \sum_{k=0}^{N-1} (A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} - A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)}) \right) \times \right. \\ & \quad \times \cos(2\pi f_0 t) - \\ & \quad \left. - \left( n_s(t) + \mu \sum_{k=0}^{N-1} (A_r^{(k)} a(t-kT) \sin \varphi_r^{(k)} - A_q^{(k)} a(t-kT) \sin \varphi_q^{(k)}) \right) \times \right. \\ & \quad \left. \times \sin(2\pi f_0 t) \right]^2 dt. \end{aligned}$$

Раскрыв скобки и отбросив быстро осциллирующие слагаемые, получим

$$\begin{aligned} & \int_0^{LT+(N-1)T} (z(t) - \mu s_Q(t))^2 dt = \\ &= \frac{1}{2} \int_0^{LT+(N-1)T} \left[ \left( n_c(t) + \mu \sum_{k=0}^{N-1} (A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} - A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)}) \right)^2 + \right. \\ & \quad \left. + \left( n_s(t) + \mu \sum_{k=0}^{N-1} (A_r^{(k)} a(t-kT) \sin \varphi_r^{(k)} - A_q^{(k)} a(t-kT) \sin \varphi_q^{(k)}) \right)^2 \right] dt. \end{aligned} \quad (3.3.15)$$

Далее будем рассматривать только первое слагаемое (3.3.15), так как полученные результаты легко будут применимы и ко второму, а также опустим коэффициент 1/2, поскольку он не влияет на результат поиска минимума выражения.

Раскроем скобки в первом слагаемом и с учётом свойства (3.3.11) получим:

$$\begin{aligned}
& \int_0^{LT+(N-1)T} \left( n_c(t) + \mu \sum_{k=0}^{N-1} \left( A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} - A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)} \right) \right)^2 dt = \\
& = \int_0^{LT+(N-1)T} n_c^2(t) dt + \mu^2 E_a \sum_{k=0}^{N-1} \left[ A_r^{(k)} \cos \varphi_r^{(k)} \right]^2 + \mu^2 E_a \sum_{k=0}^{N-1} \left[ A_q^{(k)} \cos \varphi_q^{(k)} \right]^2 + \\
& \quad + 2\mu \sum_{k=0}^{N-1} A_r^{(k)} \cos \varphi_r^{(k)} \int_{kT}^{kT+LT} a(t-kT) n_c(t) dt - \\
& \quad - 2\mu \sum_{k=0}^{N-1} A_q^{(k)} \cos \varphi_q^{(k)} \int_{kT}^{kT+LT} a(t-kT) n_c(t) dt - \\
& \quad - 2\mu^2 E_a \sum_{k=0}^{N-1} A_r^{(k)} \cos \varphi_r^{(k)} A_q^{(k)} \cos \varphi_q^{(k)},
\end{aligned}$$

где  $E_a = \int_0^{LT} a^2(t) dt$  — энергия огибающей. Тогда

$$\begin{aligned}
& \int_0^{LT+(N-1)T} \left( n_c(t) + \mu \sum_{k=0}^{N-1} \left( A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} - A_q^{(k)} a(t-kT) \cos \varphi_q^{(k)} \right) \right)^2 dt = \\
& = \frac{1}{E_a} \sum_{k=0}^{N-1} \left( \mu E_a A_r^{(k)} \cos \varphi_r^{(k)} + \int_{kT}^{kT+LT} a(t-kT) n_c(t) dt - \mu E_a A_q^{(k)} \cos \varphi_q^{(k)} \right)^2 + \Delta n_c,
\end{aligned} \tag{3.3.16}$$

где

$$\Delta n_c = \int_0^{LT+(N-1)T} n_c^2(t) dt - \frac{1}{E_a} \sum_{k=0}^{N-1} \left( \int_{kT}^{kT+LT} a(t-kT) n_c(t) dt \right)^2.$$

В (3.3.15) оба слагаемых неотрицательны, тогда, учитывая, что в (3.3.16) все слагаемые, кроме  $\Delta n_c$ , неотрицательны, получаем, что  $\Delta n_c$  также должно быть неотрицательно. Отсюда имеем выражение, в котором все слагаемые неотрицательны, и для которого необходимо найти минимум. Очевидно, что процедуру минимизации можно проводить независимо для каждого слагаемого. Аналогичные результаты можно получить и для второго слагаемого (3.3.15).

В итоге для различения сигналов с найквистовской огибающей необходимо искать минимум следующего выражения:

$$\begin{aligned} & \frac{1}{E_a} \sum_{k=0}^{N-1} \left( \mu E_a A_r^{(k)} \cos \varphi_r^{(k)} + \int_{kT}^{kT+LT} a(t-kT) n_c(t) dt - \mu E_a A_q^{(k)} \cos \varphi_q^{(k)} \right)^2 + \Delta n_c + \\ & + \frac{1}{E_a} \sum_{k=0}^{N-1} \left( \mu E_a A_r^{(k)} \sin \varphi_r^{(k)} + \int_{kT}^{kT+LT} a(t-kT) n_s(t) dt - \mu E_a A_q^{(k)} \sin \varphi_q^{(k)} \right)^2 + \Delta n_s, \end{aligned} \quad (3.3.17)$$

где

$$\Delta n_s = \int_0^{LT+(N-1)T} n_s^2(t) dt - \frac{1}{E_a} \sum_{k=0}^{N-1} \left( \int_{kT}^{kT+LT} a(t-kT) n_s(t) dt \right)^2 \geq 0.$$

Покажем, что минимизацию слагаемых (3.3.17) можно проводить на основе значений сигналов с выхода согласованного фильтра, взятых в отсчётные моменты времени. В соответствии с (3.2.17a), а также с учётом симметричности найквистовской огибающей относительно начала координат получаем, что импульсный отклик фильтра, согласованного с рассматриваемыми сигналами, совпадает с найквистовской огибающей. Учтём, что на основании (3.3.7) квадратурные составляющие принятого сигнала  $z(t)$  перед АЦП имеют вид

$$\begin{aligned} z_s(t) &= \left( \mu \sum_{k=0}^{N-1} A_r^{(k)} a(t-kT) \cos \varphi_r^{(k)} + n_s \right) / 2, \\ z_s(t) &= \left( \mu \sum_{k=0}^{N-1} A_r^{(k)} a(t-kT) \sin \varphi_r^{(k)} + n_s \right) / 2. \end{aligned}$$

Выше было введено обозначение комплексного сигнала перед АЦП:

$$Z(t) = 2(z_c(t) + jz_s(t)). \quad (3.3.18)$$

Тогда сигнал  $Z_f(t)$  с выхода согласованного фильтра для входного комплексного сигнала (3.3.18) можно записать следующим образом:

$$Z_f(t) = \int_{-\infty}^{\infty} a(\tau) Z(t-\tau) d\tau = \int_0^{LT} a(\tau) Z(t-\tau) d\tau. \quad (3.3.19)$$

Подставляя (3.3.18) в (3.3.19) и учитывая свойство (3.3.11), несложно видеть, что

$$Z_f(kT) = \left( \mu E_a A_r^{(k)} \cos \varphi_r^{(k)} + \int_{kT}^{kT+LT} a(t-kT) n_c(t) dt \right) +$$



$$+j \left( \mu E_a A_r^{(k)} \sin \varphi_r^{(k)} + \int_{kT}^{kT+LT} a(t-kT) n_s(t) dt \right), \quad k = 0, 1, \dots, N-1. \quad (3.3.20)$$

Подставив теперь (3.3.20) в (3.3.17), получим выражение, поиск минимума которого соответствует алгоритму оптимального различения сигналов с МСИ и найквистовской огибающей:

$$\frac{1}{E_a} \sum_{k=0}^{N-1} |Z_f(kT) - \mu E_a S(q, k)|^2 + \Delta n_c + \Delta n_s, \quad (3.3.21)$$

где

$$S(q, k) = A_q^{(k)} \cos \varphi_q^{(k)} + j A_q^{(k)} \sin \varphi_q^{(k)}$$

представляет собой точку сигнального созвездия.

Обратим внимание, что в (3.3.20) присутствуют только числа, т. е. нет интегралов. Все интегральные вычисления проводятся один раз для каждого тактового интервала при определении сигнала с выхода согласованного фильтра.

Слагаемые  $\Delta n_c$  и  $\Delta n_s$ , стоящие в (3.3.21) вне суммы, являются фиксированными и, поэтому, не влияют на процедуру нахождения минимума.

Окончательно сформулируем правило различения сигналов с МСИ и Найквистовской огибающей: *выносится решение в пользу той гипотезы  $H_q$  о передаче на  $k$ -м тактовом интервале символа канального алфавита  $q$ ,  $q = 1, \dots, t$ , для которой эвклидово расстояние между значением сигнала с выхода согласованного фильтра для удвоенного принятого комплексного сигнала  $(z_c(t) + jz_s(t))$ , взятого в момент времени  $kT$ , и комплексной точкой сигнального созвездия  $S(q, k)$ , взятой с учётом коэффициентов  $\mu$  и  $E$ , окажется минимальным.*

Заметим, что алгоритм (3.3.21) уже является дискретизированным, поэтому для БЦО достаточно лишь отметить, что значения  $Z_f(kT)$  могут быть получены при цифровой фильтрации последовательности  $c_l$  КИХ-фильтром (т. е. фильтром с конечной импульсной характеристикой), характеристика которого получается путём дискретизации (3.3.10).

Выражения (3.3.8) и (3.3.21) справедливы только при следующих существенных ограничениях: во-первых, несущая частота и начальная фаза принимаемого сигнала и генератора приёмника точно совпадают; во-вторых, коэффициент ослабления сигнала при распространении до приём-

ника постоянен (для (3.3.8) достаточно постоянства хотя бы в пределах каждого тактового интервала). Таким образом, в реальном приёмном устройстве в БЦО должны быть предусмотрены следующие процедуры:

- подстройка частоты и фазы принимаемого сигнала;
- оценки и исправления неравномерной АЧХ канала передачи сигнала — так называемый *эквайзинг*;
- тактовая синхронизация.

В случае приёма сигналов известной системы указанные задачи, как правило, решаются посредством использования специальных *пилотных сигналов*, все параметры которых заранее полностью известны. В остальных случаях используются различные “слепые” алгоритмы.

Рассмотрим реализацию процедур частотной и фазовой подстройки, а также оценки коэффициента ослабления сигнала на простом примере. Будем считать, что на входе БЦО имеется последовательность сигналов с найквистовской огибающей и коэффициентом скругления  $\beta = 0,3$  (что является типичным значением); при этом коэффициент ослабления сигнала постоянен и равен  $\mu = 0,1$ . Длительность огибающей  $L = 20T$ , а несущая частота сигнала отличается от частоты генератора в приёмнике на  $\Delta f = 0,1/T$ , т. е. частота приёмного генератора  $f_0$ , а несущая частота сигнала  $f_0 + \Delta f$ . Сигнал передаётся кадрами по 1024 информационных символов, где первые 128 символа являются пилотными. Для передачи полезных данных используется созвездие КАМ-16, для пилотных — ФМ-2, причём максимальные амплитуды созвездий совпадают (рис. 3.16). Отношение сигнал-шум на бит для информационного созвездия равно  $h^2 = 30$  дБ. Значения символов информационной последовательности распределены равномерно, а значения символов пилотной последовательности определяются двоичными элементами  $\pm 1$  псевдослучайной последовательности<sup>1</sup>  $\text{pilot}_k$ ,  $k = 1, \dots, 128$ .

---

<sup>1</sup> Конкретный вид псевдослучайной последовательности в данном примере не будет важен и поэтому не приводится.

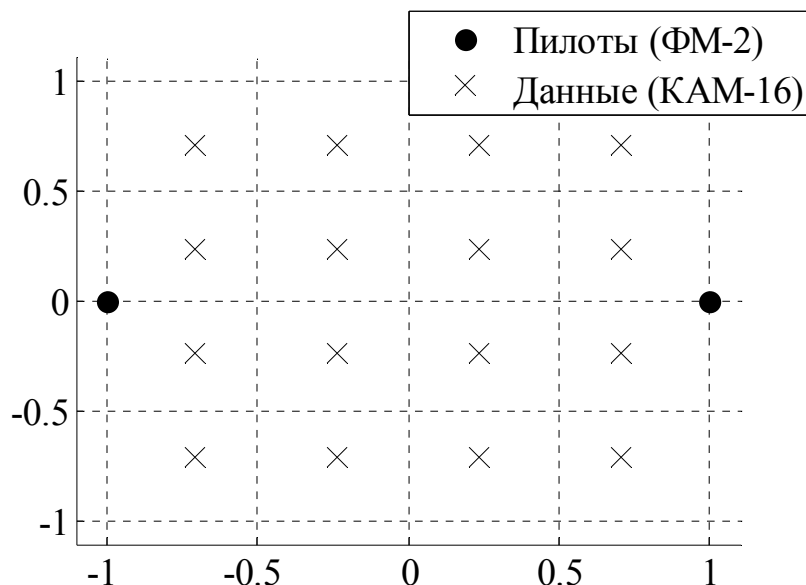


Рис. 3.16. Сигнальные созвездия пилотов и данных

Наличие частотной отстройки между несущей частотой принятого сигнала и частотой опорного генератора в приёмнике может быть представлено как изменение фазы полезной составляющей принятого сигнала от значения  $\varphi_r^{(k)}$  к значению  $(\varphi_r^{(k)} + 2\pi\Delta f t)$ . Если  $\Delta f \ll 1/LT$ , то значение  $\cos(\varphi_r^{(k)} + 2\pi\Delta f t)$  не успеваает заметно измениться на периоде  $LT$ , и сигнал с выхода согласованного фильтра  $Z_f(t)$  в отсчётные моменты времени будет преобразован следующим образом:

$$Z_f(kT) = \left( \mu EA_r^{(k)} \cos(\varphi_r^{(k)} + 2\pi\Delta f (kT + LT/2)) + \int_{kT}^{kT+LT} a(t - kT) n_c(t) dt \right) + j \left( \mu EA_r^{(k)} \sin(\varphi_r^{(k)} + 2\pi\Delta f (kT + LT/2)) + \int_{kT}^{kT+LT} a(t - kT) n_s(t) dt \right),$$

$$k = 0, 1, \dots, N - 1. \quad (3.3.22)$$

Видно, что наличие частотной отстройки приводит к повороту значений комплексного сигнала с выхода согласованного фильтра на угол, зависящий от значения частотной отстройки и номера тактового интервала. Следовательно, разница фаз соседних пилотных сигналов на выходе согласованного фильтра будет отличаться от заданной на  $2\pi\Delta f T$ .

Воспользовавшись сказанным, можно предложить следующую процедуру оценки значения  $\Delta f$ : для каждой пары пилотов вычисляется набег фаз относительно ожидаемой разницы фаз пилотов, далее проводится усредне-

ние набегов, и путём деления полученного среднего значения на  $2\pi T$  получается оценка частотной отстройки:

$$\widehat{\Delta f} = \frac{1}{2\pi T} \frac{1}{127} \sum_{k=1}^{127} \text{angle} \left( \overline{Z}_f(kT) Z_f((k+1)T) \text{pilot}_k \text{pilot}_{k+1} \right), \quad (3.3.23)$$

где  $\text{angle}$  — операция вычисления фазы комплексного числа,  $-\pi \leq \text{angle}(z) < \pi$ , и сигнал с выхода согласованного фильтра синхронизирован с началом кадра. Отметим, что алгоритм (3.3.23) имеет теоретическую полосу захвата:

$$-\frac{\varphi}{2\pi T} \leq \Delta f \leq \frac{\varphi}{2\pi T}, \quad (3.3.24)$$

где  $\varphi$  — минимальная разница фаз между соседними пилотами (в нашем примере  $\varphi = \pi$ ). Такая полоса захвата обусловлена периодичностью функции  $\text{angle}$ . Фактическая полоса захвата оказывается ещё меньше и зависит от отношения сигнал-шум: чем меньше отношение сигнал-шум, тем больше вероятность “переполнения” функции  $\text{angle}$  и, следовательно, получения неверной оценки, тем меньше фактическая полоса захвата. Часто в качестве практической полосы захвата выбирают в 2 раза более узкую полосу, чем (3.3.24).

Вернёмся к изучению влияния частотной отстройки на сигнал  $Z_f(t)$ . Следует отметить, что даже в рассматриваемом примере, когда  $\Delta f = 0,1/T$ , условие  $\Delta f \ll 1/LT$  не выполняется, и выражение (3.3.22) не будет верным. В этих условиях точный анализ изменения сигнала с выхода согласованного фильтра затруднителен. Поэтому для качественной оценки на рис. 3.17 представлены значения сигнала с выхода согласованного фильтра в отсчётные моменты времени для следующих параметров: на входе присутствует только полезный сигнал с ФМ-2, содержащий 1024 случайных информационных символа;  $\beta = 0,3$ ,  $L = 20$  и  $\Delta f = 0,1/T$ .

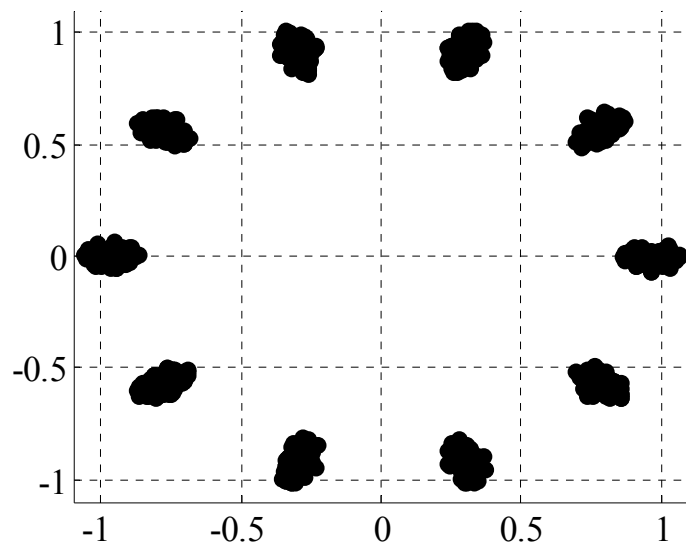


Рис. 3.17. Значения сигнала с выхода согласованного фильтра при подаче на вход сигнала с ФМ-2 и  $\Delta f = 0,1/T$

Из анализа рис. 3.17 следует, что помимо ожидаемого набега фазы на  $2\pi/10$  между соседними значениями сигнала, к самим значениям добавлена некоторая случайная составляющая. Напомним, что на вход согласованного фильтра поступал сигнал без шума. Наличие дополнительной шумовой составляющей на выходе согласованного фильтра дополнительно ухудшает качество оценки частотной отстройки, но не изменяет принципиально алгоритм (3.3.23).

На рис. 3.18 представлены значения сигнала в отсчётные моменты времени с выхода согласованного фильтра, соответствующие пилотным и информационным отсчётам. Из анализа представленных данных видно, что в информационных отсчётах явно просматриваются три уровня амплитуды, что соответствует созвездию КАМ16. Также заметно ослабление амплитуды сигнала примерно в 10 раз по отношению к исходным созвездиям, что соответствует выбранному значению коэффициента ослабления  $\mu = 0,1$ .

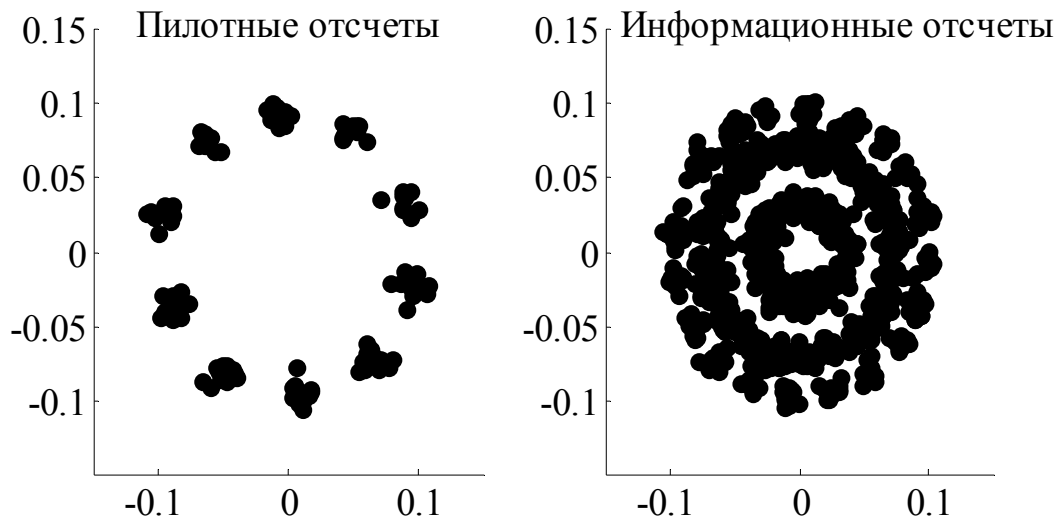


Рис. 3.18. Значения сигнала с выхода согласованного фильтра до процедур подстройки частоты, фазы и коэффициента ослабления

Для данных представленных на рис. 3.18 была проведена процедура оценки частотной отстройки. По полученному значению  $\widehat{\Delta f}$  была проведена подстройка частоты сигнала перед входом согласованного фильтра и полученный сигнал вновь пропущен через согласованный фильтр. На рис. 3.19 представлены отсчеты  $Z_f^{(1)}(kT)$  подстроенного по частоте сигнала после прохождения согласованного фильтра.

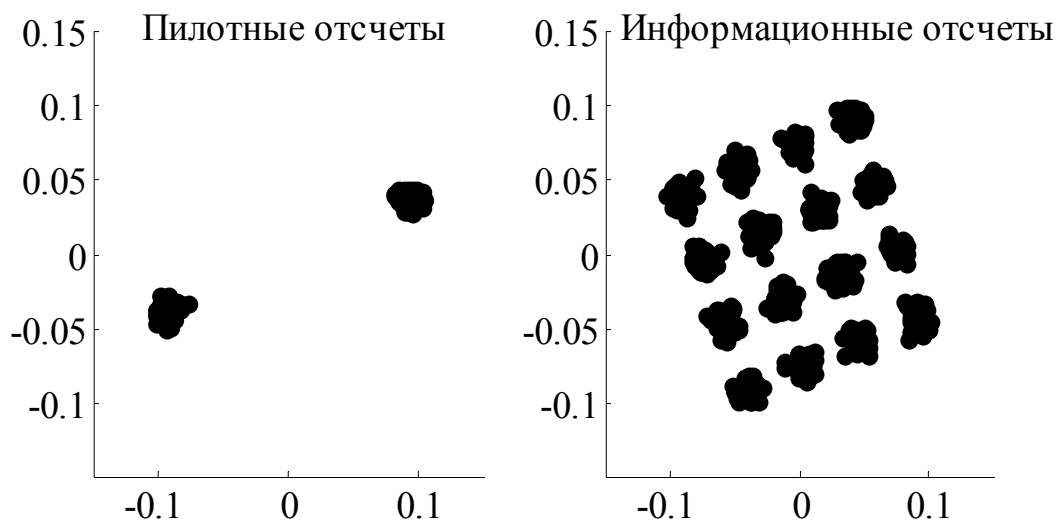


Рис. 3.19. Значения сигнала с выхода согласованного фильтра после процедуры подстройки частоты

Из анализа данных, представленных на рис. 3.19, видно, что оценка и исправление частотной отстройки проведены достаточно качественно, чтобы можно было различить созвездия пилотов и данных. Однако осталась неисправленной фаза поворота созвездий, и не оценен коэффициент ослабления сигнала. Как и прежде, оценку фазы и коэффициента ослабления будем проводить по пилотным отсчётам.

Для оценки фазы предлагается следующий алгоритм: пилотные отсчеты, образующие два “облака” на рис. 3.19, преобразуются к одному “облаку”, соответствующему точке сигнального созвездия (0, 1), путём поворота на 0 либо  $\pi$  в соответствии с пилотной ПСП. В качестве оценки фазы выбирается значение фазы среднего пилотного вектора из общего “облака”:

$$\widehat{\Delta\varphi} = \frac{1}{128} \text{angle} \sum_{k=1}^{128} Z_f^{(1)}(kT) \text{pilot}_k.$$

Для оценки коэффициента ослабления предлагается рассматривать среднее значения модуля пилотных векторов:

$$\widehat{\mu} = \frac{1}{128} \sum_{k=1}^{128} |Z_f^{(1)}(kT)|.$$

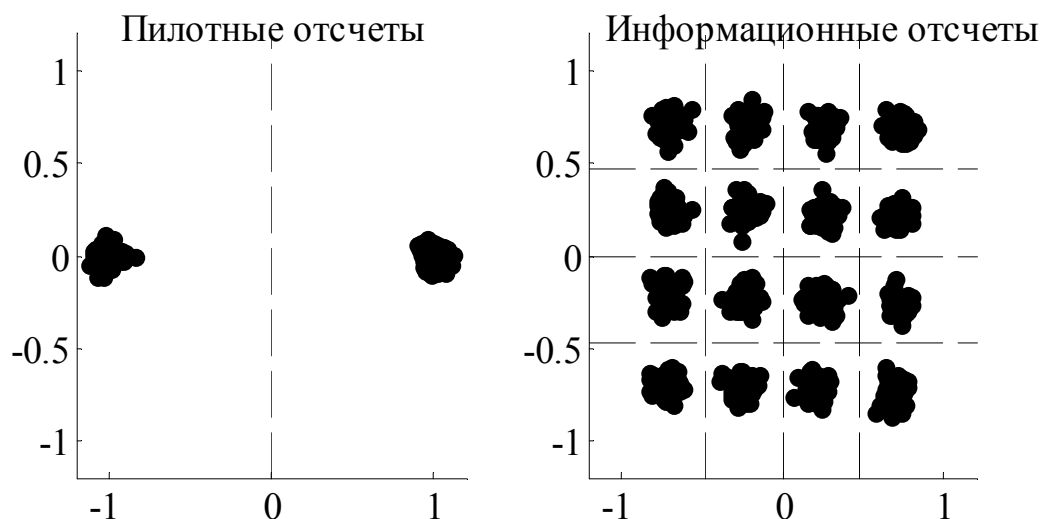


Рис. 3.20. Значения сигнала с выхода согласованного фильтра после процедур подстройки частоты, фазы и коэффициента ослабления

На рис. 3.20 представлены пилотные и информационные отсчеты после процедур оценки и восстановления фазы и оценки и исправления ко-

эфициента ослабления сигнала:

$$Z_f^{(2)}(kT) = Z_f^{(1)}(kT) \exp(-j\widehat{\Delta\phi}) / \widehat{\mu}, \quad k = 1, \dots, 1024.$$

Также на рис. 3.20 представлены пунктирные линии, соответствующие границам областей принятия решения в пользу того или иного символа канального алфавита для ФМ-2 и КАМ-16.

Из анализа представленных на рис. 3.20 данных следует, что процедуры приёма были проведены успешно, так как все полученные “облака” вписываются в области принятия решений.

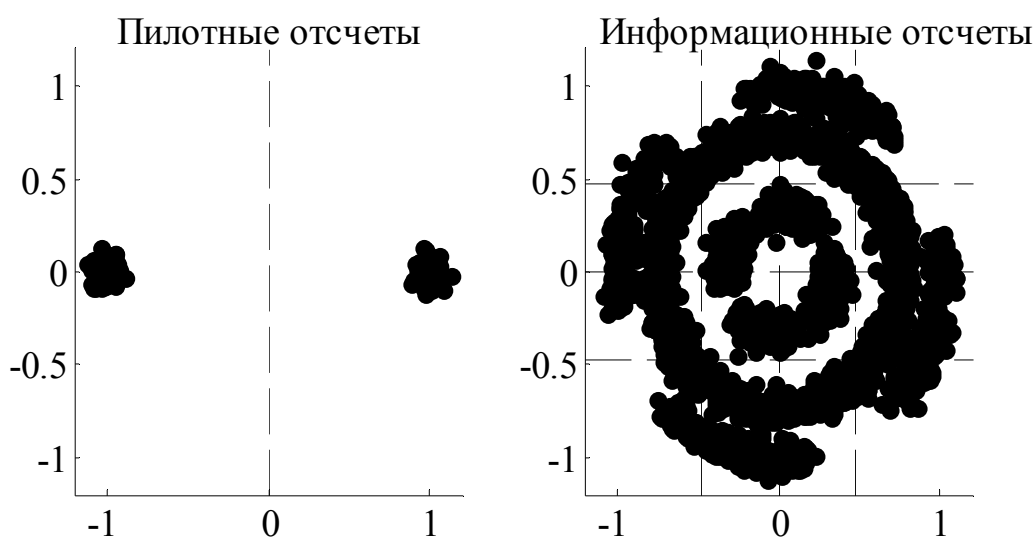


Рис. 3.21. Значения сигнала с выхода согласованного фильтра после процедур подстройки частоты, фазы и коэффициента ослабления для сигнала с длиной кадра 4096 информационных символов

В виду того, что оценка частотной отстройки не может быть произведена точно, с увеличением длины кадра при сохранении числа пилотов “облака” информационных отсчётов могут начать вытягиваться в дугу, что сделает невозможным правильное различение сигналов. Так, на рис. 3.21 представлены результаты аналогичные рис. 3.20, но для кадра длиной 4096 информационных символов. Видно, что информационные отсчёты за длительность кадра вытянулись в дугу, поворачиваясь вокруг начала координат против часовой стрелки. “Облака” внешней и внутренней окружностей растянулись примерно на  $\pi/3$  и, учитывая, что в каждой из этих окружностей присутствует по 4 точки, полученные дуги различимы. В средней окружности присутствует 8 точек, поэтому их дуги не различимы и обра-



зуют единое кольцо. Исходя из сказанного, а также из того, что коэффициент ослабления сигнала  $\mu$  оказывается зависимым от времени, все процедуры подстроек параметров и синхронизации должны повторяться с определённой периодичностью. И именно поэтому сигналы, как правило, передаются кадрами с повторением пилотных последовательностей.

## ВОПРОСЫ И ЗАДАНИЯ К ГЛАВЕ 3

1. В чём состоит цель представления случайного процесса в виде разложения Карунена–Лоэва?

2. Каковы значения собственных чисел интегрального уравнения Фредгольма, возникающего в разложении Карунена–Лоэва, если в качестве модели канального шума рассматривается АБГШ?

3. Что такое функционал отношения правдоподобия?

4. С какими параметрами сигнала связана импульсная характеристика фильтра, согласованного с этим сигналом?

5. Чем определяются характеристики различения двух детерминированных сигналов? Какие сигналы имеют наилучшую помехоустойчивость?

6. Определить помехоустойчивость приёма для случая различения в белом гауссовском шуме двух равновероятных противоположных сигналов, имеющих постоянное значение  $a$  на интервале  $[0; T]$ , если приёмник, оптимальный по критерию идеального наблюдателя принимает решение: а) по одному отсчёту; б) по  $N$  независимым отсчётам.

7. Найти энергетический выигрыш (проигрыш) при оптимальном когерентном различении равновероятных двоичных сигналов при переходе от сигналов с АМ к сигналам с ФМ и с ортогональной ЧМ, а также при переходе от сигналов с ЧМ к сигналам ФМ, если у сигналов: а) фиксирована средняя мощность; б) фиксирована пиковая мощность.

8. Производится максимально правдоподобное различение двух равновероятных сигналов постоянного уровня  $s(t) = \pm a$ ,  $t \in [0; T]$  в смеси сигнала и аддитивного шума, имеющего распределение  $w(n) = (1/c)(1 - |n|/c)$ ,  $|n| \leq c$ , посредством взятия  $N$  независимых отсчётов. Какое решение будет принято в случае трёхэлементной выборки  $\{0,4 \text{ мВ } 3,2 \text{ мВ } - 1,9 \text{ мВ}\}$ , если  $a = 1 \text{ мВ}$ , а дисперсия шума равна  $1,5 \text{ мкВт}$ ?

9. Производится различение двух сигналов  $s_r(t) = (-1)^{r-1} a$ ,  $t \in [0; T]$ ,  $r = 1, 2$ , по однократному отсчёту принимаемого колебания  $x(t) = s_r(t) + n(t)$ , где значения стационарного шума  $n(t)$  имеют распределение  $w(n) = (\alpha/2) \exp(-\alpha |n|)$ . Найти среднюю вероятность ошибки, если: а) сигналы равновероятны,  $a = 1 \text{ мВ}$ , а дисперсия шума  $\sigma^2 = 1 \text{ мкВт}$ ; б) априорные вероятности появления сигналов есть  $P(s_1) = 1/3$ ,

$P(s_2) = 2/3$ ,  $a = 1$  мВ, а дисперсия шума  $\sigma^2 = 9$  мкВт.

10. Производится оптимальное различение двух сигналов с ЧМ:  $s_i(t) = A_0 \cos(\omega_i t + \varphi_0)$ ,  $i = 1, 2$ ;  $\omega_1 \neq \omega_2$ ;  $t \in [0; T]$ . Найти частотный разнос  $\Delta\omega = |\omega_1 - \omega_2|$ , при котором достигается максимальная помехоустойчивость. Каков при этом энергетический проигрыш по сравнению с приёмом противоположных сигналов?

11. В присутствии аддитивного белого гауссовского шума со спектральной плотностью  $N_0/2$  производится оптимальное различение противоположных сигналов  $s_r(t) = (-1)^{r-1} A(t) \cos \omega_0 t$ ,  $r = 1, 2$ ,  $t \in [0; T]$ , где  $A(t)$  — вещественная огибающая, являющаяся медленной по сравнению с  $\omega_0 t$  функцией, имеющая вид: а)  $A(t) = A_0$ ; б)  $A(t) = A_0 \sin(\pi t / T)$ ; в)  $A(t) = A_0 \sin^2(\pi t / T)$ . Каковы значения взаимного энергетического проигрыша в условиях фиксированной пиковой мощности сигналов?

12. В чём разница структурных схем устройств, реализующих алгоритмы оптимального различения двух и  $m > 2$  сигналов?

13. Получить формулу (3.2.42) для вероятности ошибки в ансамбле биортогональных сигналов.

14. В чём заключается свойство изоморфности ансамбля сигналов?

15. В чём проявляется неточности подстройки частоты и начальной фазы опорного сигналы при реализации алгоритмов оптимального различения?

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Гельгор А.Л.* Основы теории информации: учеб. пособие / А. Л. Гельгор, Е. А. Попов. — СПб. : Изд-во Политехн. ун-та, 2008. — 158 с.
2. *Вентцель Е. С.* Теория вероятностей. — М. : Кнорус, 2010. — 658 с.
3. *Ивченко Г. И.* Введение в математическую статистику / Г. И. Ивченко, Ю. И. Медведев. — М. : Издательство ЛКИ, 2010. — 600 с.
4. *Ильин В. А.* Линейная алгебра / В. А. Ильин, Э. Г. Позняк. — М. : Физматлит, 2002. — 320 с.
5. *Фихтенгольц Г. М.* Курс дифференциального и интегрального исчисления. Т. 1. — М. : Физматлит, 2001. — 616 с.
6. *Гоноровский И. С.* Радиотехнические цепи и сигналы. — М. : Дрофа, 2006. — 719 с.
7. *Тихонов А. Н.* Уравнения математической физики / А. Н. Тихонов, А. А. Самарский. — М. : Изд-во МГУ, 1999. — 799 с.
8. *Янке Е.* Специальные функции / Е. Янке, Ф. Эмде, В. Лёш. — М. : Наука, 1977. — 342 с.
9. *Тихонов В. И.* Оптимальный приём сигналов. — М. : Радио и связь, 1983. — 320 с.
10. *Тихонов В. И.* Статистический анализ и синтез радиотехнических устройств и систем / В. И. Тихонов, В. Н. Харисов. — М. : Горячая линия — Телеком, 2004. — 608 с.
11. *Биккенин Р. Р.* Теория электрической связи / Р. Р. Биккенин, М. Н. Чесноков. — М. : Академия, 2010. — 329 с.
12. *Левин Б. Р.* Теоретические основы статистической радиотехники. — М. : Радио и связь, 1989. — 656 с.
13. *Финк Л. М.* Теория передачи дискретных сообщений. — М. : Советское радио, 1970. — 728 с.
14. *Бунтов В.Д.* Цифровые и микропроцессорные радиотехнические устройства : учеб. пособие / В. Д. Бунтов, С. Б. Макаров. — СПб. : Изд-во Политехн. ун-та, 2005, — 399 с.
15. *Прокис Дж.* Цифровая связь. — М. : Радио и связь, 2000. — 800 с.
16. *Скляр Б.* Цифровая связь. — М. : Вильямс, 2003. — 1004 с.
17. *Горяинов В. Т.* Примеры и задачи по статистической радиотехнике / В. Т. Горяинов, А. Г. Журавлёв, В. И. Тихонов. — М. : Сов. радио, 1980. — 544 с.

Гельгор Александр Леонидович  
Горлов Антон Игоревич  
Попов Евгений Александрович

**ОБЩАЯ ТЕОРИЯ СВЯЗИ**  
**ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ**  
**ОЦЕНИВАНИЕ ПАРАМЕТРОВ**  
**ОПТИМАЛЬНЫЙ ПРИЁМ СИГНАЛОВ**  
Учебное пособие

Лицензия ЛР № 020593 от 07.08.97

Налоговая льгота – общероссийский классификатор продукции ОК 005-93, т. 2;  
95 3004 – научная и производственная литература

---

Подписано в печать .2013 г. Формат 60×84/16. Печать цифровая.

Усл. печ. л. 14,19. Уч.-изд. л. 8,4. Тираж . Заказ

---

Отпечатано с готового оригинал-макета, предоставленного Корпоративным центром качества СПбГПУ,  
в Цифровом типографском центре Издательства Политехнического университета.

195251, Санкт-Петербург, Политехническая ул., 29.

Тел.: (812) 550-40-14

Тел./факс: (812) 297-57-76