

1



Г. ДЖЕНКИНС, Д. ВАТТС
*
СПЕКТРАЛЬНЫЙ
АНАЛИЗ
И ЕГО ПРИЛОЖЕНИЯ



SPECTRAL ANALYSIS
AND ITS APPLICATIONS

GWILYM M. JENKINS
University of Lancaster, U.K.
and
DONALD G. WATTS
University of Wisconsin, U.S.A.

HOLDEN-DAY
San Francisco, Cambridge, London, Amsterdam
1969

Г. ДЖЕНКИНС, Д. ВАТТС

СПЕКТРАЛЬНЫЙ
АНАЛИЗ
И ЕГО ПРИЛОЖЕНИЯ

ВЫПУСК 1

Перевод с английского
В. Ф. ПИСАРЕНКО
С предисловием
А. М. ЯГЛОМА

ИЗДАТЕЛЬСТВО „МИР“
МОСКВА 1971



06

Спектральный анализ — новая и очень важная отрасль прикладной математики, посвященная выделению из наблюдаемых явлений или процессов периодических компонент, т. е. правильно меняющихся со временем составляющих. Подобные процессы очень часто встречаются в инженерном деле, различных отделах физики и геофизики, а также в экономике.

Задача данной книги — дать инженеру или физику руководство, позволяющее овладеть приемами и методами спектрального анализа и применить их в своей практической работе. Для удобства читателей русское издание разделено на два выпуска. Выпуск 1 выйдет в 1971 г., выпуск 2 — в начале 1972 г.

В данный выпуск вошли общие принципы спектрального анализа, анализ Фурье, основы теории вероятностей и математической статистики, оценки корреляционных функций и спектров стационарных процессов.

Книга будет полезна инженерно-техническим работникам, физикам, геофизикам, математикам-прикладникам и экономистам, а также студентам старших курсов, для которых она послужит ценным учебным пособием.

Редакция космических исследований, астрономии и геофизики

2-6-2; 2-3-2; 2-9-2

1971

Подписное издание

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Предлагаемая читателю монография известного английского специалиста в области математической статистики Г. Дженкинса и американского ученого Д. Ваттса посвящена прикладным аспектам теории временных рядов, т. е. рядов наблюдений $x(t)$, зависящих от дискретного или непрерывно меняющегося аргумента t (обычно времени наблюдения). При этом авторы рассматривают лишь ряды, подверженные нерегулярным флуктуациям, создаваемым или ошибками наблюдений, или какими-то иными неустранимыми помехами («шумами»), искажающими эти наблюдения, или, наконец, помехами, заложенными в самой природе величины x . Ряды такого рода встречаются буквально на каждом шагу в геофизике (метеорологии, океанологии, сейсмологии, учении о земном магнетизме и аэрономии) и астрономии, экономике, технических дисциплинах (особенно радиотехнике, электронике и автоматике) и даже в биологии и медицине, причём их роль с течением времени все возрастает. Поэтому неудивительно, что и литература по вопросам, касающимся таких рядов, также очень быстро растёт; так, например, одной только статистической радиотехнике (т. е. фактически изучению комплекса проблем, связанных с временными рядами радиотехнического происхождения) на русском языке посвящено по крайней мере полтора десятка монографий и несколько сотен научных работ. Однако до сих пор на русском языке не было ни одной книги, предназначенной сразу для читателей-прикладников всех специальностей, имеющих дело с временными рядами, и излагающей с единой точки зрения и на современном уровне общие математические приемы их изучения и обработки. Именно такую цель и преследует настоящая книга.

Естественно, что временные ряды, подверженные нерегулярным флуктуациям, можно изучать только статистически — на основе широкого использования аппарата теории вероятностей и математической статистики. При таком подходе ряд $x(t)$ рассматривается как одна реализация, выбранная из статистического ансамбля функций, описываемого определенным распределением вероятностей в функциональном пространстве, т. е. как выборочная функция случайного процесса $X(t)$, зависящего от непрерывного или дискретного аргумента. Тем самым, анализ временных рядов оказывается частью

теории случайных процессов, являющейся одним из наиболее глубоких и сложных разделов современной теории вероятностей. Очень большое место в этом анализе составляют вопросы, относящиеся к статистике случайных процессов — области, лежащей на пересечении теории случайных процессов и математической статистики, все развитие которой относится к последнему двадцатилетию. Заметим также, что до сих пор большинство результатов теории случайных процессов относится не к совершенно произвольным процессам $X(t)$, а лишь к процессам того или иного частного вида; анализ временных рядов по традиции имеет дело только со стационарными случайными процессами, являющимися удобной моделью широкого класса (но, разумеется, все же не всех) рядов реальных наблюдений. Этого ограничения придерживаются и авторы настоящей книги.

Центральной задачей статистического анализа временных рядов бесспорно является чрезвычайно важная задача об определении спектра процесса по одной его реализации; именно ей Дженкинс и Ваттс посвящают больше всего внимания (с этим обстоятельством связан и выбор названия их книги). Эта задача имеет длинную и интересную историю (достаточно сказать, что основное для всего прикладного спектрального анализа понятие периодограммы временного ряда впервые было введено — правда, для других целей — известным физиком А. Шустером еще в конце прошлого века). Однако математическое ее исследование началось лишь после того, как около 1950 г. было строго доказано, что при широких условиях, накладываемых на процесс $X(t)$, периодограмма не стремится ни к какому пределу при стремлении к бесконечности интервала наблюдения. На русском языке кое-какие сведения относительно методов оценки спектра по данным наблюдений могут быть найдены в книгах М. С. Бартлетта («Введение в теорию случайных процессов», ИЛ, М., 1958) и Э. Хеннана («Анализ временных рядов», изд-во «Наука», М., 1964), но помимо явной неполноты и отрывочности содержащегося здесь материала, частично уже заметно устаревшего (последнее особенно относится к содержанию первой из указанных книг), надо также отметить, что обе они предназначены не для прикладников, а для математиков, и ввиду крайней сжатости изложения даже и для профессионалов-математиков являются довольно трудными. В чисто же прикладной литературе этому вопросу особенно не повезло — достаточно сказать, что в ряде книг по статистической радиотехнике или автоматике фактически утверждается, что спектр процесса можно определить как предел соответствующей периодограммы (правда, термин «периодограмма» при этом обычно не упоминается). В мировой литературе вопросу об оценке спектра специально посвящена книга Блэкмана и Тьюки (R. V. Blackman, J. W. Tukey, *The measurement of power spectra from the point of view of communications engineering*, Dover, New York, 1959), сыгравшая очень большую роль в развитии соответствующей статистиче-

ской теории и в пропаганде правильных представлений по этому вопросу, но написанная нарочито усложненным («замысловатым») языком с большим числом методических и терминологических ухищрений, интересных для специалиста, хорошо знакомого с предметом, но очень затрудняющих изучение этой книги для новичка (такой стиль вообще характерен для Тьюки — старшего автора указанной книги). Тем не менее ввиду отсутствия другого подходящего изложения ссылки на книгу Блэкмана и Тьюки до самого последнего времени очень часто встречались в прикладной литературе; известно также, что в прикладных научно-исследовательских учреждениях нашей страны было изготовлено даже несколько разных переводов этой книги на русский язык, часто, к сожалению, мало-квалифицированных и содержащих ошибки, способные вконец запутать неопытного читателя.

Книга Дженкинса и Ваттса рассчитана примерно на тот же круг читателей, что и книга Блэкмана и Тьюки: обе они не содержат строгих доказательств используемых математических предложений и основной упор делают на рецептурную сторону дела, т. е. на формулировку конкретных рекомендаций, предназначенных для практика. Однако настоящая книга имеет то большое преимущество, что написана она относительно просто и ясно, хотя и достаточно строго и с учетом всех основных достижений математической теории; кое в чем она оказывается также заметно более современной, чем ее предшественница, со времени появления которой прошло уже более десяти лет (так, например, стоит отметить краткое изложение в приложении П7.3 очень важной для приложений на современных вычислительных машинах техники «быстрых преобразований Фурье», созданной при активном участии Тьюки, но заметно позже опубликования совместной с Блэкманом книги, в которой, естественно, эта техника никак не отражена). Следует также отметить, что содержание книги Дженкинса и Ваттса (опять же в отличие от книги Блэкмана и Тьюки) не ограничивается одним лишь вопросом о вычислении спектров; в частности, весьма полезными являются также разделы этой книги, посвященные оценке корреляционной функции или каких-то параметров процесса по материалам наблюдений в течение конечного промежутка времени. Надо надеяться, что появление этой книги в русском переводе будет приветствоваться широкими кругами читателей-прикладников различных специальностей, имеющих дело с рядами наблюдений, и даст им, наконец, в руки доступный источник сведений о том, как следует математически грамотно обрабатывать такие ряды для извлечения из них основной информации о статистических характеристиках исследуемого процесса.

Для удобства наших читателей русское издание разделено на два выпуска. В первый выпуск включены гл. 1—6, содержащие общие принципы спектрального анализа, анализ Фурье, основы теории

вероятностей и математической статистики, оценки корреляционных функций и спектров стационарных процессов. Во втором выпуске (гл. 7—11) спектральный анализ иллюстрируется искусственными и практическими примерами и обобщается на случай многомерных стационарных процессов.

В конце данного выпуска дается список дополнительной литературы, составленный переводчиком книги. Ссылки в тексте на дополнительную литературу даются цифрой со звездочкой.

А. М. Яглом

ПРЕДИСЛОВИЕ

Анализ временных рядов в настоящее время широко используется во многих отраслях техники, в физических науках и экономике.

Одним из важных видов анализа временных рядов является спектральный анализ, имеющий дело с разделением временных рядов на различные частотные составляющие.

Применения спектрального анализа покрывают широкий диапазон задач, например влияние морского волнения на вибрацию кораблей и влияние возмущений, или шума, на работу систем электрического управления и химических реакторов.

Эта книга предназначена в первую очередь для инженеров, повышающих квалификацию после окончания учебного заведения, поскольку большинство применений спектрального анализа фактически осуществляется инженерами и физиками. Одна из трудностей, встречающихся при использовании спектрального анализа, состоит в том, что большая часть теории спектрального анализа была развита статистиками за последние пятнадцать лет. К сожалению, многое из литературы, посвященной этому вопросу, представляет собой трудный для чтения материал. Поэтому потребность в книге, рассчитанной в основном на инженеров, ощущалась уже давно. Мы надеемся, однако, что настоящая книга привлечет внимание гораздо более широкой аудитории, в том числе математиков, статистиков, экономистов, физиков и биологов.

Одна из трудностей при написании этой книги состояла в том, что спектральный анализ использует довольно сложные статистические методы, в то время как многие инженеры испытывают недостаток знаний по элементарной статистике. Это справедливо даже для некоторых инженеров-электриков, имеющих солидные сведения по теории вероятностей.

Так, например, винеровская теория предсказания и управления показывает, что оптимальный фильтр или система контроля могут быть рассчитаны при условии, что известны различные спектры, характеризующие сигнал и шум в системе.

Однако в книгах по теории управления уделяется мало внимания очень важному практическому вопросу о том, как *оценить* эти спектры по *записи конечной длины*. Именно с такими вопросами мы будем иметь дело в этой книге.

Чтобы постепенно подойти к вопросам оценивания временных рядов, мы были вынуждены заниматься в первых главах элементарными статистическими задачами. Это может отвлечь математика или статистика, но нам кажется (на основании нашего опыта изложения этих идей инженерам), что введение, не использующее других источников и включающее большинство статистических понятий, которые понадобятся впоследствии в книге, необходимо.

Читатели, знакомые с материалом гл. 2, 3 и 4, могут, конечно, начать чтение с гл. 5.

Гл. 1 посвящена краткому описанию затрагиваемых вопросов и классу задач, которые могут быть решены с помощью спектрального анализа. В гл. 2 рассматриваются важные понятия анализа Фурье; эта глава является основной для всего последующего материала.

Содержание большей ее части известно инженерам, но весь материал собран здесь в том виде, в каком он нужен для спектрального анализа. В гл. 3 мы вводим некоторые основные понятия теории вероятностей, являющиеся фундаментальными для последующих глав. В гл. 4 вводятся многие важные понятия теории статистических выводов и обсуждается использование выборочных распределений в теории оценивания и теории наименьших квадратов, а также дается краткое изложение способов получения статистических выводов с помощью функции правдоподобия. Не весь этот материал необходим для понимания спектральных методов, обсуждаемых ниже, и читатели-инженеры могут при желании пропустить последнюю часть этой главы при первом чтении. Для спектрального анализа наиболее существенными из этой главы являются разделы о применении выборочных распределений в теории оценивания и теории наименьших квадратов. Последняя является важнейшим оружием в арсенале статистики и, как показывает наш опыт, часто неправильно понимается инженерами.

Гл. 5 содержит некоторые элементарные понятия теории случайных процессов, такие, например, как стационарность, автокорреляционная функция и понятие о процессе скользящего среднего — авторегрессии. Изложены и проиллюстрированы примерами методы оценки автокорреляционных функций и параметров линейных процессов. В гл. 6 понятия анализа Фурье и теории случайных процессов объединяются для получения способа описания стационарного случайного процесса с помощью его спектра. Показано, как должны быть модифицированы методы анализа Фурье для того, чтобы оценить спектр процесса по реализации конечной длины. Затем выводятся выборочные свойства спектральных оценок и вво-

дится важное понятие сглаживания этих оценок. Гл. 7 содержит много искусственных и практических примеров спектрального оценивания и дает стандартный способ, названный «стягиванием окна» (window closing), предназначенный для определения требуемой степени сглаживания.

В гл. 8 понятия, введенные в гл. 5—7, распространяются на случай пары временных рядов, что приводит к определению взаимной корреляционной функции, взаимного спектра и спектра квадрата коэффициента когерентности.

Гл. 9 посвящена оцениванию взаимного спектра и понятию выравнивания двух временных рядов. Анализ взаимных спектров применяется в гл. 10 для оценивания частотной характеристики линейной системы. Наконец, в гл. 11 мы рассматриваем спектральный анализ векторного временного ряда и оценивание матрицы частотных характеристик линейной системы.

Настоящая книга написана в то время, когда в этой области еще ведется активная работа и когда еще очень не хватает опыта применения спектральных методов. Тем не менее многое, по-видимому, уже достигнуто, чтобы такую попытку можно было считать оправданной. Мы надеемся, что эта книга послужит ученым-прикладникам и инженерам всесторонним и полезным справочником по применению спектрального анализа к практическим задачам с временными рядами, а также окажется полезным пособием для аспирантов и лиц, повышающих свою квалификацию.

Мы приносим глубокую благодарность проф. Стантону из Технической школы университета Пэрдью за предоставленные в наше распоряжение данные о работе электростанции, использованные в последующих главах, и проф. Уэрцу из Висконсинского университета за полезные советы относительно программ для вычислительных машин. Мы очень благодарны Маккормику из Отдела статистики Висконсинского университета, а также Макклелану из Математического исследовательского центра армии США (Висконсинский университет) за составление некоторых программ для вычислительных машин и расчеты по ним. Мы также благодарны Маккормику и Алави из Ланкастерского университета за проверку всей рукописи.

Ланкастер, Великобритания
Мадисон, Висконсин, США

Гвиллим Дженкинс
Дональд Ваттс

ОБОЗНАЧЕНИЯ

Система обозначений была выбрана с возможной тщательностью, чтобы избежать путаницы и в то же время сохранить четкую разницу между выборочными оценками (estimates)*), оценками (estimators)**) и теоретическими***) значениями, а также между функциями, зависящими от частоты, и функциями, зависящими от времени. Читатель не будет иметь затруднений, если он запомнит следующие соглашения, которых мы придерживались всюду за небольшими исключениями.

Символы, обозначения и условные соглашения

Предмет	Шрифт	Примеры
Наблюдения, временные ряды	Латинский, малые буквы	$x_1, x_2; z_t, z(t)$
Случайные величины, случайные процессы	Латинский, большие буквы	$X_1, X_2; Z_t, Z(t)$
Выборочные оценки	Латинский и греческий с крышкой, малые буквы	$\bar{x}, s^2; \hat{\theta}, \hat{\alpha}_1, \hat{\alpha}_2$
Оценки	Латинский и греческий с крышкой, большие буквы	$\bar{X}, S^2, \hat{\Theta}$
Параметры	Греческий, малые буквы	$\theta, \alpha_1, \alpha_2$

Ковариации и корреляции (показаны символы и нижние индексы)

	Выборочные оценки	Оценки	Теоретические
Автоковариации****)	$c_{xx}(u), c_{22}(k)$	$c_{XX}(u), c_{22}(k)$	$\gamma_{XX}(u), \gamma_{22}(k)$
Взаимные ковариации	$c_{xy}(u), c_{12}(k)$	$c_{XY}(u), c_{12}(k)$	$\gamma_{XY}(u), \gamma_{12}(k)$

*) Под термином «estimates» авторы имеют в виду конкретные, наблюдаемые значения оценок.— *Прим. перев.*

**) Термином «estimators» авторы называют оценки, рассматриваемые как случайные величины.— *Прим. перев.*

***) Словами «теоретические значения» авторы называют истинные значения неизвестных параметров или функций.— *Прим. перев.*

****) В нашей литературе чаще используются названия ковариации.— *Прим. перев.*

Четная часть	$I_{xy}(u), I_{12}(k)$	$I_{XY}(u), I_{12}(k)$	$\lambda_{XY}(u), \lambda_{12}(k)$
Нечетная часть	$q_{xy}(u), q_{12}(k)$	$q_{XY}(u), q_{12}(k)$	$\psi_{XY}(u), \psi_{12}(k)$
Автокорреляции *)	$r_{xx}(u), r_{22}(k)$	$r_{XX}(u), r_{22}(k)$	$\rho_{XX}(u), \rho_{22}(k)$
Взаимные корреляции	$r_{xy}(u), r_{12}(k)$	$r_{XY}(u), r_{12}(k)$	$\rho_{XY}(u), \rho_{12}(k)$

Спектры (черта сверху обозначает сглаживание, звездочка — комплексное сопряжение)

	Выборочные оценки	Оценки	Теоретические
Автоспектры **)	$C_{xx}(f), C_{11}(f)$	$C_{XX}(f), C_{11}(f)$	$\Gamma_{XX}(f), \Gamma_{11}(f)$
Взаимные спектры	$C_{xy}(f), C_{12}(f)$	$C_{XY}(f), C_{12}(f)$	$\Gamma_{XY}(f), \Gamma_{12}(f)$
Коспектры	$L_{12}(f)$	$L_{12}(f)$	$\Delta_{12}(f)$
Квадратурные спектры	$Q_{12}(f)$	$Q_{12}(f)$	$\Psi_{12}(f)$
Амплитудные спектры	$A_{12}(f)$	$A_{12}(f)$	$\alpha_{12}(f)$
Фазовые спектры	$F_{12}(f)$	$F_{12}(f)$	$\varphi_{12}(f)$
Квадрат коэффициента когерентности	$K_{12}^2(f)$	$K_{12}^2(f)$	$\kappa_{12}^2(f)$

*) В нашей литературе чаще используются названия корреляции. — Прим. перев.
 **) В нашей литературе чаще используется название спектры. — Прим. перев.

ЦЕЛИ И СРЕДСТВА АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Чтобы помочь читателю разобраться в спектральном анализе, в этой главе дается краткий обзор важнейших понятий и основных целей анализа временных рядов.

1.1. ВРЕМЕННЫЕ РЯДЫ И СЛУЧАЙНЫЕ ПРОЦЕССЫ

1.1.1. Детерминированные и недетерминированные функции

Трудно найти какую-либо отрасль науки, которая не приводила бы к изучению данных, представляемых в виде *временных рядов*. Временной ряд — это *случайная*, или *недетерминированная*, функция x независимой переменной t . В большинстве ситуаций функция

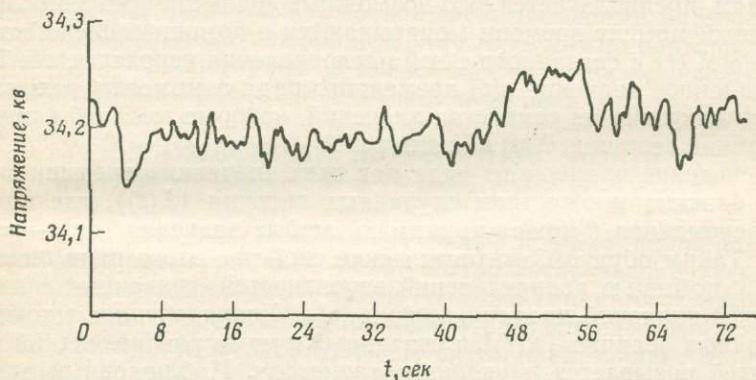


Рис. 1.1. Отклонения напряжения на клеммах статора турбогенератора на 50 Мвт.

$x(t)$ будет функцией времени, хотя в некоторых случаях она может быть функцией некоторого другого физического параметра t , например координаты.

Характерное свойство временного ряда состоит в том, что его будущее поведение не может быть предсказано точно, что можно было бы сделать в случае *детерминированной* функции времени. Во

многих отраслях прикладной математики удобно считать, что некоторые физические процессы могут быть описаны с помощью детерминированных, или математических, функций времени. Например, в большинстве электротехнических расчетов напряжение в сети представляют косинусоидальной функцией

$$x(t) = a \cos(2\pi f_0 t + \varphi), \quad (1.1.1)$$

где f_0 — частота сети и a — амплитуда напряжения. Однако более тщательное исследование амплитуды показывает, что она не является константой, а флуктуирует во времени. Так, например, на рис. 1.1 показаны отклонения напряжения на клеммах турбогенератора в зависимости от времени. Если сравнить две записи зависимости напряжения от времени, можно не обнаружить их сходства. Однако, если сравнить их *статистическое* поведение, или поведение *в среднем*, сходство начнет выявляться. Это замечание приводит к понятию случайного процесса.

1.1.2. Случайные процессы

Поскольку различные участки временного ряда обнаруживают сходство только в их осредненных свойствах, необходимо описывать эти ряды с помощью вероятностных законов, или моделей. Таким образом, предполагается, что возможные значения временного ряда в данный момент времени t описываются с помощью *случайной величины* $X(t)$ и связанного с ней распределения вероятностей. Тогда наблюдаемое значение $x(t)$ временного ряда в момент t рассматривается как одно из семейства значений, которые могла бы принять случайная величина $X(t)$ в момент t .

Поведение временного ряда при всех значениях времени может быть описано множеством случайных величин $\{X(t)\}$, где временная переменная t может принимать любые значения от $-\infty$ до $+\infty$. Таким образом, статистические свойства этого ряда описываются с помощью распределений вероятностей, связанных с любым набором значений времени t_1, t_2, \dots, t_N . Упорядоченное множество случайных величин $\{X(t)\}$ и связанных с ними распределений вероятностей называется *случайным процессом*. Наблюдаемый временной ряд $x(t)$, таким образом, рассматривается как одна из дважды бесконечного множества функций, которые могли бы быть порождены этим случайным процессом. Это множество дважды бесконечно, так как возможно бесконечное множество значений в любой заданный момент времени и имеется бесконечно много моментов времени.

Временные ряды, которые встречаются на практике, являются *дискретными* или *непрерывными*. Примерами дискретных временных рядов являются месячные показатели импорта и экспорта или выход продукции в последовательных партиях химического про-

цесса при переработке материала порциями (см. рис. 5.2). Примерами непрерывных временных рядов являются данные турбогенератора на рис. 1.1 или отраженный радиолокационный сигнал на рис. 5.1.

1.1.3. Данные, получаемые из эксперимента, и неэкспериментальные данные

Более существенная особенность временных рядов определяется тем, являются ли данные *неэкспериментальными* или же они получаются из некоторого запланированного эксперимента. Так, временные ряды в экономике и социальных науках являются примерами неэкспериментальных данных. Экономист обычно в состоянии лишь *наблюдать* экономическую систему и редко может проводить планируемые эксперименты.

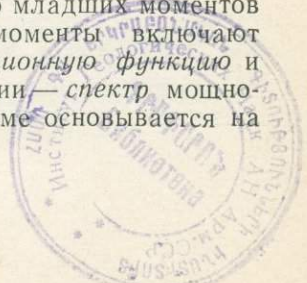
Дальнейшая трудность, связанная с анализом экономических временных рядов, состоит в том, что обычно они содержат мало наблюдений. Из-за этого крайне трудно проверить, хорошо ли согласуется предлагаемая случайная модель с данными. Тем не менее методы анализа временных рядов играют существенную роль в анализе экономических данных [1].

С другой стороны, в технике и физических науках масштаб времени, в течение которого нужно собирать полезные данные, обычно намного меньше, так что можно получить временные ряды, содержащие гораздо больше значений. Кроме того, можно повторить эксперименты при аналогичных условиях, так что справедливость анализа и различных моделей может быть проверена.

1.2. ОПИСАНИЕ ВРЕМЕННЫХ РЯДОВ ВО ВРЕМЕННОЙ И ЧАСТОТНОЙ ОБЛАСТЯХ

В разд. 1.1 утверждалось, что случайный процесс, порождающий наблюдаемый временной ряд, может быть описан распределениями вероятностей, связанными со всеми возможными множествами моментов времени. Определение природы этих распределений вероятности по одному или по малому числу рядов представляет собой невозможное или даже бессмысленное занятие. В этом разделе мы обсудим некоторые из наиболее важных упрощений, которые приняты в анализе временных рядов для того, чтобы сделать этот анализ выполнимым и в то же время плодотворным.

Важнейшие предположения о временных рядах заключаются в том, что соответствующий случайный процесс является *стационарным* и может быть адекватно описан с помощью младших моментов его распределения вероятностей. Младшие моменты включают в себя *среднее значение, дисперсию, ковариационную функцию* и преобразование Фурье ковариационной функции — *спектр мощности*. Другой подход к вышеизложенной проблеме основывается на



предположении, что случайный процесс может быть адекватно описан с помощью некоторой модели, содержащей небольшое число параметров, которые могут быть оценены по данным. Обсудим теперь кратко эти упрощающие предположения.

1.2.1. Стационарность

Проверка выходного сигнала генератора шума в течение ограниченного промежутка времени показывает, что различные участки выходного сигнала похожи. Напротив, характерная черта экономического ряда, такого, как валовой национальный продукт индустриальной страны, состоит в том, что его уровень стремится увеличиться с течением времени, и поэтому различные участки этого ряда не будут сравнимы. Говорят, что выходной сигнал генератора шума является *стационарным* временным рядом, в то время как про временной ряд валового национального продукта говорят, что он *нестационарный*.

Качественно стационарный ряд — это такой ряд, который находится в статистическом равновесии, в том смысле, что он не содержит никаких трендов, тогда как нестационарный ряд таков, что его свойства изменяются со временем. Ряды, встречающиеся на практике, принадлежат обычно к одному из трех видов: ряды, проявляющие свойства стационарности в течение долгих периодов времени, например выходные сигналы генераторов шумов; ряды, достаточно стационарные в течение коротких периодов времени, например измерения атмосферной турбулентности, и ряды, которые являются явно нестационарными в том смысле, что их видимые свойства непрерывно изменяются со временем.

Большинство методов, имеющих дело с нестационарными временными рядами, основано на способах устранения или отфильтровывания нестационарной части, так что остается ряд, с которым можно обращаться как со стационарным. В одной из недавних работ [2] приведены модели, которые могут описывать нестационарные ряды.

Так как статистические свойства стационарных рядов не изменяются со временем, то эти свойства можно накопить и выявить с помощью вычисления некоторых функций от данных. Функция, которую впервые использовали для этой цели, является автоковариационной функцией.

1.2.2. Автоковариационная функция

При классической статистической обработке измерения x_t ($t = 1, 2, \dots, N$) некоторого физического параметра можно считать независимыми, поскольку эксперименты, порождающие эти наблюдения, физически независимы. Если связанное с этими измерениями рас-

пределение вероятностей $f_X(x)$ является *нормальным*, или *гауссовским*, то его можно полностью задать своим средним значением *)

$$\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (1.2.1)$$

и дисперсией

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx. \quad (1.2.2)$$

Среднее значение определяет расположение, или центр тяжести распределения, а дисперсия характеризует его изменчивость, или разброс.

Если наблюдения x_t образуют часть временного ряда, то только для чисто случайного ряда соседние величины будут независимы, т. е. на значение величины x_t не влияют значения величин x_{t-1} , x_{t-2} , ... В общем случае соседние величины временного ряда будут коррелированы. Поэтому в случае стационарного нормального ряда, кроме среднего значения μ и дисперсии σ^2 , необходимо задать его автоковариационную функцию

$$\gamma(u) = E[(X(t) - \mu)(X(t+u) - \mu)]. \quad (1.2.3)$$

На практике $\gamma(u)$ может быть оценена с помощью

$$c(u) = \frac{1}{N} \sum_{t=1}^{N-u} (x_t - \bar{x})(x_{t+u} - \bar{x}), \quad (1.2.4)$$

где

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$$

является *средним значением* наблюдаемого временного ряда. Функция $c(u)$ называется *выборочной автоковариационной функцией* временного ряда. Иногда удобно для сравнения рядов с разными масштабами измерений нормировать (1.2.4) с помощью деления на дисперсию $c(0)$. Таким образом, определяется *выборочная автокорреляционная функция*

$$r(u) = \frac{c(u)}{c(0)}. \quad (1.2.5)$$

Выборочная автокорреляционная функция для данных турбогенератора, изображенных на рис. 1.1, приведена на рис. 1.2. Видно, что напряжение имеет высокую положительную корреляцию при

*) Символом $E[...]$ всюду в этой книге обозначается математическое ожидание случайной величины. — Прим. перев.

сдвиге на одну точку, что соответствует $1/2$ сек, сохраняет некоторую положительную корреляцию после 1 сек, но в интервале от $1 1/2$ до $2 1/2$ сек проявляет явную отрицательную корреляцию. Это означает, что если имеет место большое напряжение, превышающее среднее значение, то весьма вероятно, что примерно через 2 сек напряжение спадет ниже среднего значения, и наоборот. Выборочные оценки $r(u)$ для сдвигов от $2 1/2$ до 10 сек очень малы, но устойчиво отрицательны; это означает, что в среднем положительное отклонение от

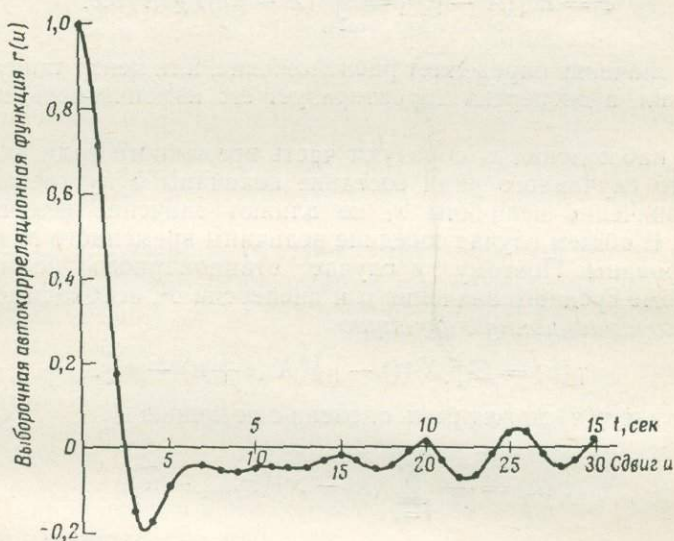


Рис. 1.2. Выборочная автокорреляционная функция для данных, изображенных на рис. 1.1, $N=1000$.

среднего значения имеет тенденцию к последующему отрицательному отклонению с задержкой от 2 до 10 сек. Однако значения $r(u)$ в этой области крайне малы, и поэтому выводы, основанные на них, могут быть ненадежны. При больших значениях аргумента выборочная корреляционная функция обнаруживает периодичность формы напряжения с периодом примерно 3 сек. Эта периодичность, возможно, также может давать некоторый вклад в отрицательную корреляцию для сдвига около 2 сек.

Автокорреляционная функция полезна в некоторых случаях, поскольку она дает наглядную картину того, как зависимость в ряде затухает с увеличением задержки или разделяющего промежутка u между точками ряда. Однако иногда автокорреляционная функция с трудом поддается интерпретации, так как соседние значения могут быть сильно коррелированы. Это означает, что выборочная ав-

токорреляционная функция может иметь видимые искажения. Более детальное описание свойств и применений автокорреляционной функции дается в гл. 5. В этой книге мы будем использовать ее главным образом как промежуточную ступень при оценке спектра.

1.2.3. Спектр

Предположим, что временной ряд x_t состоит из значений косинусоидальной функции (1.1.1), отсчитываемых в дискретные моменты. Тогда можно проверить, что для частот f_0 , кратных основной частоте $1/N$, дисперсия, подсчитанная по формуле (1.2.4), равна $a^2/2$. Если x_t измеряется в вольтах, то это означает, что средняя мощность переменного тока, или дисперсия ряда, равна $a^2/2$ вт. В более общем случае, когда x_t состоит из смеси нескольких косинусоидальных волн с частотами f_i и амплитудами a_i , дисперсия равна

$$c(0) = \sum_i \frac{1}{2} a_i^2. \quad (1.2.6)$$

Результат (1.2.6) показывает, что если x_t можно считать состоящим из смеси косинусоидальных волн, то его дисперсию можно разложить на компоненты со средней мощностью, или дисперсией, $a_i^2/2$, соответствующие различным частотам f_i . В гл. 6 будет показано, что если x_t является стационарным временным рядом, то дисперсию соответствующего случайного процесса можно разложить на компоненты, интегрируемые по *непрерывной области частот*, согласно формуле

$$\sigma^2 = \gamma(0) = \int_{-\infty}^{\infty} \Gamma(f) df,$$

где $\Gamma(f)$ называется *спектром мощности* этого случайного процесса. Таким образом, $\Gamma(f) df$ есть приближенная мера средней мощности, или дисперсии, в полосе частот от f до $f + df$.

Выборочная оценка спектра данных турбогенератора, приведенных на рис. 1.1, показана на рис. 1.3. Отличительная черта этого спектра состоит в том, что высокая мощность сосредоточена на низких частотах, а на высоких частотах мощность невелика. Это происходит главным образом из-за больших положительных значений выборочной автокорреляционной функции при сдвигах, равных 1 и 2. Заметим также, что мощность не спадает равномерно от низких к высоким частотам. Вместо этого имеется плоская область в районе $0-1/2$ гц. Имеется также хорошо выраженный небольшой пик на частоте 0,39 гц, или периоде 2,54 сек, который, возможно, объясняет небольшую периодичность выборочной корреляционной функции на рис. 1.2 при больших значениях аргумента.

В гл. 6 будет также показано, что спектр и автоковариационная функция связаны соотношением преобразования Фурье

$$\Gamma(f) = \int_{-\infty}^{\infty} \gamma(u) \cos 2\pi fu \, du, \quad (1.2.7)$$

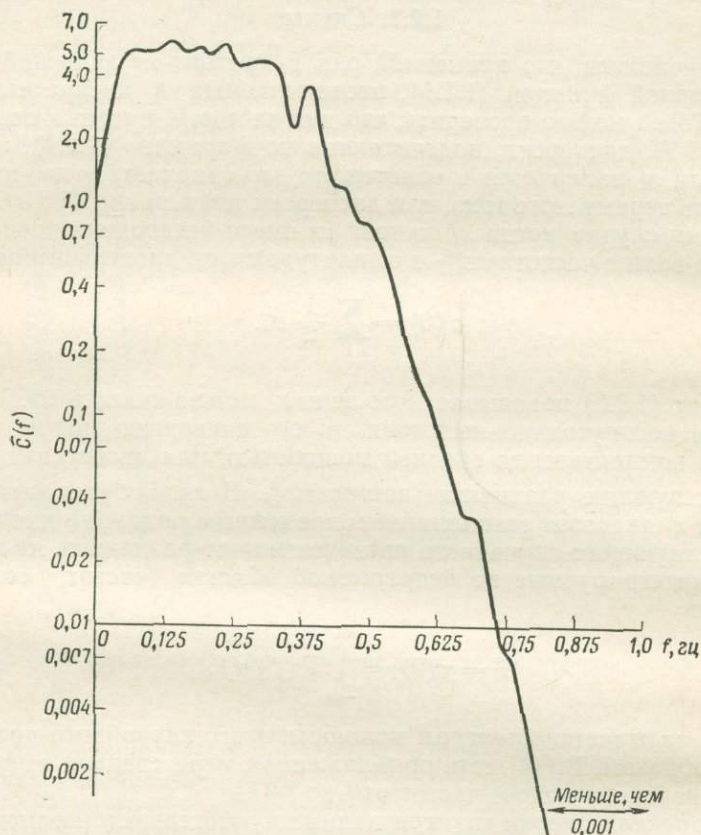


Рис. 1.3. Выборочная оценка спектра для изображенных на рис. 1.1 данных, $N=1000$.

и поэтому знание автоковариационной функции процесса эквивалентно знанию спектра процесса.

Однако при анализе записей конечной длины спектр часто предпочтительней, чем автоковариационная функция. Во-первых, оценки спектра на соседних частотах приблизительно независимы, и поэтому выборочный спектр обычно легче интерпретировать, чем выборочную автоковариационную функцию. И во-вторых, что важнее, во многих физических задачах спектр представляет непосредственный

физический интерес. Примеры использования спектрального анализа будут даны в разд. 1.3.

Цифровые фильтры. Хотя для описания случайного процесса с помощью его спектра и необходимо предполагать стационарность, на практике *предположение стационарности не представляет серьезной проблемы*. Это происходит из-за того, что спектр отделяет

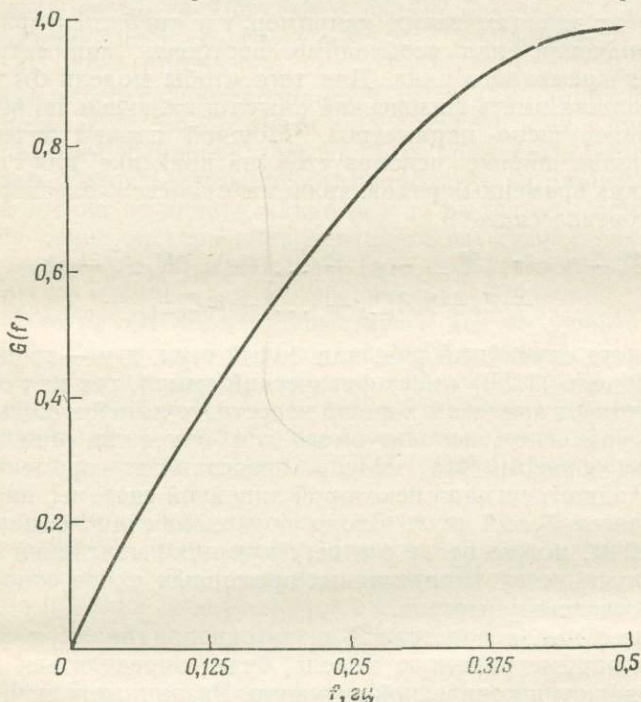


Рис. 1.4. Функция усиления для фильтра первых разностей.

вклады во временной ряд, которые можно приписывать различным частотным полосам. Нестационарный ряд обычно характеризуется присутствием большой мощности на низких частотах. Однако во многих практических приложениях представляющая интерес информация может быть сосредоточена на высоких частотах. В таких случаях все, что нужно сделать, — это отфильтровать нестационарные низкочастотные компоненты и использовать оставшийся ряд для спектрального анализа.

Особенно простой вид цифрового фильтра для устранения низкочастотных компонент представляет собой фильтр первых разностей

$$y_t = (x_t - x_{t-1}). \quad (1.2.8)$$

Коэффициент усиления $G(f)$ этого фильтра показан на рис. 1.4. Он характеризует степень пропускания фильтром косинусоидальной волны частоты f . Видно, что низкие частоты значительно ослаблены и, следовательно, будут менее заметны на выходе фильтра.

1.2.4. Параметрические модели временных рядов

Во многих задачах, таких, например, где требуется *предсказать* будущие значения ряда, необходимо построить параметрическую модель для временного ряда. Для того чтобы модель была полезной, она должна иметь физический смысл и включать по возможности небольшое число параметров. Мощной параметрической моделью, которая широко используется на практике для описания эмпирических временных рядов, является *процесс скользящего среднего* — авторегрессии:

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \dots + \alpha_m (X_{t-m} - \mu) + Z_t + \beta_1 Z_{t-1} + \dots + \beta_l Z_{t-l}, \quad (1.2.9)$$

где Z_t — чисто случайный ряд, или белый шум, и μ — средний уровень X_t . Модель (1.2.9) имеет физический смысл, так как она является дискретным аналогом хорошо известного линейного дифференциального уравнения, используемого для описания линейных систем. Таким образом, эта модель представляет временной ряд в виде выходного сигнала некоторой линейной системы, на вход которой подается белый шум. Вводя подходящее число параметров α и β в (1.2.9), можно после соответствующих вычислений [2] сопоставить большinstву эмпирических временных рядов относительно наибольшее число параметров.

Решение о том, использовать ли автоковариационную функцию, спектр или параметрическую модель, будет определяться на практике требованиями конкретной ситуации. Различные условия потребуют различных методов подхода, и поэтому важно уяснить, что *не существует единого подхода, который нужно было бы применять к анализу временных рядов во всех ситуациях.*

1.3. ЦЕЛИ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

В этой книге будет описано несколько различных применений спектрального анализа. Поскольку спектральный анализ является почти единственным оружием, имеющимся в распоряжении для анализа временных рядов, полезно обсудить природу задач, касающихся временных рядов, в несколько более общей постановке.

Задачи о временных рядах можно классифицировать, рискуя упасть в слишком сильное упрощение, на те, которые требуют в той или иной форме *построения моделей*, и те, которые приводят к изу-

чению частотных характеристик. Эти категории неизбежно должны несколько перекрываться.

1.3.1. Построение моделей

Легко можно отличить друг от друга несколько различных видов моделей, например *пробную* и *усовершенствованную*, *эмпирическую* и *физическую*, *параметрическую* и *непараметрическую*.

Пробная и усовершенствованная модели. На ранних стадиях работы исследователь может знать очень мало о каком-нибудь конкретном явлении. Основная цель анализа временных рядов на этом этапе может состоять в том, чтобы посмотреть на данные с различных точек зрения и увидеть, какие можно выдвинуть гипотезы. Например, изучение спектра поля вертикальных скоростей атмосферной турбулентности [3] показывало, что пик этого спектра сдвигается в сторону низких частот с увеличением солнечной радиации. Это наводило на мысль, что существуют две различные причины флуктуаций атмосферной турбулентности: высокочастотная компонента, обусловленная силами трения, и низкочастотная компонента, обусловленная тепловой конвекцией, вызванной солнечной радиацией. В результате после этого пробного анализа оказалось возможным начать построение более реальной модели атмосферной турбулентности.

Часто случается, как это было в только что изложенном примере, что первоначально предполагаемая модель временного ряда впоследствии может оказаться несовершенной. Природа этого несовершенства пробной модели может быть использована тогда для ее модификации и построения более совершенной модели.

Эмпирические и физические модели. В некоторых случаях можно построить детальные модели временного ряда, основанные на физике, лежащей в основе явления. Например, большие усилия были затрачены на построение моделей атмосферной турбулентности [3] и гидродинамической турбулентности [4]. В других ситуациях об исследуемом явлении известно так мало, что нужно прибегать к подгонке эмпирических моделей, таких, как модель скользящего среднего — авторегрессии (1.2.9). Большое преимущество физических моделей состоит в том, что они обычно требуют меньшего количества параметров, чем эмпирические модели. Чтобы принять решение о том, тратить ли время и усилия для нахождения физической модели или же прибегнуть к помощи эмпирической модели, требуется рассудительность и интуиция. Вообще необходимо идти на компромисс и использовать любые доступные физические сведения, чтобы иметь основу в начале построения.

Параметрические и непараметрические модели. Модель скользящего среднего — авторегрессии (1.2.9) является *параметрической* моделью. Чтобы подобрать такую модель, нужно оценить по наблюдаемым данным небольшой набор параметров. С другой стороны, описание временного ряда, даваемое автокорреляционной функцией или спектром, является *непараметрическим* (или многопараметрическим, так как для того, чтобы задать весь процесс, требуется действительно бесконечное число параметров).

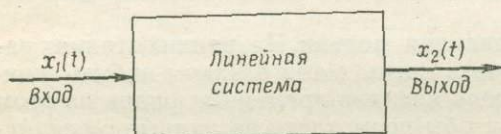


Рис. 1.5. Схематическое изображение линейной системы.

Оценка динамической модели линейной системы, показанной на рис. 1.5, представляет собой задачу, к которой можно применить оба эти метода. В простейшем случае, когда имеется один вход x_1 и один выход x_2 , динамическую модель можно оценить по записям $x_1(t)$, $x_2(t)$ входного и выходного сигналов. Например, может быть известно, что некоторая простая параметрическая модель, такая, как

$$T \frac{dx_2(t)}{dt} + x_2(t) = x_1(t),$$

является подходящей. Эта модель содержит один параметр T , называемый *постоянной времени* этой системы. В других случаях более подходящим является описание системы с помощью некоторой непараметрической модели, требующей задания функции усиления $G(f)$ и фазовой характеристики $\varphi(f)$ этой системы. Они являются функциями, описывающими отклик системы на косинусоидальные волны с различными частотами f . Так, входная косинусоидальная волна $x_1(t) = a \cos 2\pi ft$ появляется на выходе как косинусоидальная волна $x_2(t) = aG(f) \cos(2\pi ft + \varphi(f))$, т. е. ее амплитуда изменилась от a до $aG(f)$, а фаза сдвинулась на $\varphi(f)$. В гл. 10 будет показано, что анализ взаимных спектров можно использовать для оценки коэффициента усиления и сдвига фазы линейной системы.

В некоторых ситуациях амплитудно-фазовое описание предпочтительнее, так как модель может быть нужна лишь в ограниченной полосе частот. В других может оказаться лучше полное описание, даваемое параметрической моделью.

Поскольку спектральный анализ является непараметрическим методом, его применение в области построения моделей ограничено. Однако он полезен иногда, как это было в упомянутом выше примере с турбулентностью, для выдвижения пробных моделей, которые можно затем подобрать параметрически.

1.3.2. Использование моделей временных рядов

Модели временных рядов используются для различных целей. Наиболее распространенными из них являются: а) прогнозирование, б) оценивание передаточных функций, в) фильтрация и управление, г) имитация и оптимизация и д) создание новых физических теорий.

Прогнозирование. Под прогнозированием понимается оценивание будущих значений $x(t+T)$ временного ряда из некоторого интервала будущих значений $0 \leq T \leq l$ по значениям временного ряда до момента t включительно. Прогнозирование экономических и промышленных временных рядов представляет очень важное применение временных рядов и обсуждается в работе [2].

Оценивание передаточных функций. Это применение обсуждается выше.

Фильтрация и управление. Более общей задачей, чем описанное выше прогнозирование, является задача линейной фильтрации в том виде, как она была сформулирована Винером [5, 1*]. Линейный фильтр представляет собой устройство, которое, действуя на входной сигнал $x_1(t)$, дает выходной сигнал $x_2(t)$ согласно формуле

$$x_2(t) = \int_0^{\infty} h(u) x_1(t-u) du, \quad (1.3.1)$$

где $h(u)$ — *весовая функция*, или *функция отклика* фильтра на *единичный импульс*. Предположим, что на вход подается процесс $x_1(t) = s(t) + z(t)$, где $s(t)$ — *сигнал*, или *полезная информация*, а $z(t)$ — *шум*, или *нежелательная информация*. Требуется найти, как показано на рис. 1.6, такой фильтр, который давал бы на выходе некоторую функцию от сигнала в будущий момент времени T . Например, желаемый выход мог бы быть следующим:

$$g[s(t+T)] = \frac{d}{dt} [s(t+T)].$$

В этом случае оптимальный фильтр определяется как весовая функция, которая минимизирует средний квадрат сигнала ошибки

$$\varepsilon(t) = x_2(t) - g[s(t+T)]$$

между действительным и желаемым выходными процессами. Если в распоряжении имеются модели для случайных процессов, описывающих сигнал $s(t)$ и шум $z(t)$, то можно использовать разработанные методы [6] вычисления коэффициента усиления и фазовой характеристики оптимального фильтра. Фактически вычисление

оптимального фильтра осуществляется легче по спектрам сигнала $s(t)$ и шума $z(t)$.

В теории управления требуется следить за некоторым заданным сигналом $s(t)$ с возможно меньшей ошибкой. Можно показать [6], что эта задача сводится к частному случаю задачи фильтрации, описанной выше.

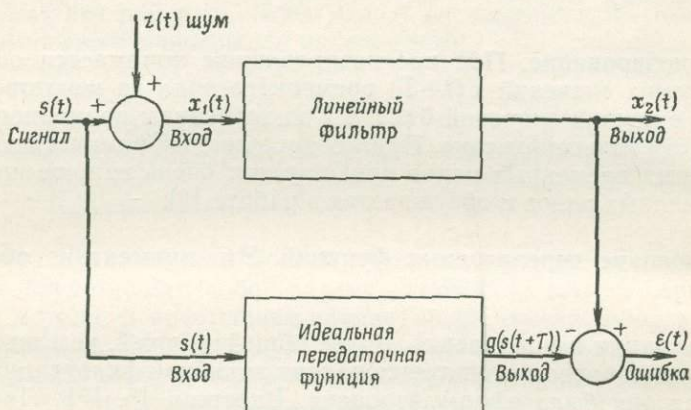


Рис. 1.6. Схематическое изображение общей задачи фильтрации.

Имитация и оптимизация. Многие системы, например системы электронного управления, являются слишком сложными, для того чтобы их изучать или оптимизировать аналитически. В таком случае система может быть имитирована с помощью аналоговых, числовых и гибридных вычислительных машин. Возмущения, попадающие в реальную систему в различных местах, можно измерить, и, если таких данных достаточно, они могут быть поданы непосредственно в имитатор. Однако объем данных, требуемых для изучения имитации, обычно так велик, что возникает необходимость подобрать модели для возмущений. Эти модели затем можно использовать для генерирования неограниченного количества искусственных данных, которые затем можно использовать в имитации.

Создание новых физических теорий. Изложенные выше применения моделей временных рядов относятся к инженерным задачам. Применение анализа временных рядов в физике отличается несколько иным подходом. Физик интересуется созданием теорий физических явлений, которые можно использовать для предсказания в возможно более широком диапазоне ситуаций. Поэтому изобретаемые им модели являются более детальными, чем те, которые используют инженеры, и должны постоянно модифицироваться и расширяться по мере все большего понимания физической сущности

явлений. Использование анализа временных рядов при построении физических моделей атмосферной турбулентности хорошо проиллюстрировано в [3].

1.3.3. Изучение частотных характеристик

Хотя спектральный анализ и играет большую роль в построении моделей временных рядов, однако наиболее подходящим он является в задачах изучения частотных характеристик.

Выше упоминалось, что линейная система (1.3.1) преобразует косинусоидальную входную волну $x_1(t) = a \cos 2\pi ft$ в косинусоидальную выходную волну $x_2(t) = aG(f) \cos(2\pi ft + \varphi(f))$. В гл. 6 будет показано, что если $x_1(t)$ — стационарный временной ряд, то изменение спектра входного процесса состоит в умножении его на квадрат коэффициента усиления фильтра. Значит, спектр $x_2(t)$ равен

$$\Gamma_2(f) = G^2(f) \Gamma_1(f). \quad (1.3.2)$$

Рис 1.7 показывает результат прохождения входного сигнала со спектром, обозначенным буквой *a* этого рисунка, через три системы, квадраты коэффициентов усиления которых даны у буквы *b*. В примерах, проиллюстрированных на рис. 1.7, входной сигнал характеризует неровности взлетно-посадочной полосы, система представляет собой шасси самолета, а выходным сигналом является типичная реакция самолета, такая, как ускорение его центра тяжести. Используя результат (1.3.2), нетрудно увидеть, что сочетание входного спектра с частотной характеристикой шасси, отмеченной цифрой 3, дает выходной спектр с очень острым пиком, как показано на рис. 1.7, в. Это показывает, что на данной резонансной частоте будут возникать большие ускорения, создающие неприятные ощущения у пассажиров и приводящие к большим напряжениям в шасси. Зная графики коэффициентов усиления для шасси типичных самолетов при типичных посадочных скоростях, можно составить нормы для неровностей взлетно-посадочных полос.

В только что разобранным примере можно было изменять входной спектр, конструируя соответствующие взлетно-посадочные полосы, но характеристики шасси определяются обычно из других соображений и поэтому должны рассматриваться как фиксированные. Наоборот, при конструировании узлов подвески для мотоциклов и автомобилей картина обратная. В последние годы конструирование систем подвески в некоторых компаниях основывается на измерениях неровностей дорог в конкретной стране. Характеристики дорог сильно изменяются от страны к стране, и поэтому системы подвески должны проектироваться соответственным образом.

Изучение частотных характеристик можно применять также при конструировании самолетов и для планирования экспериментов

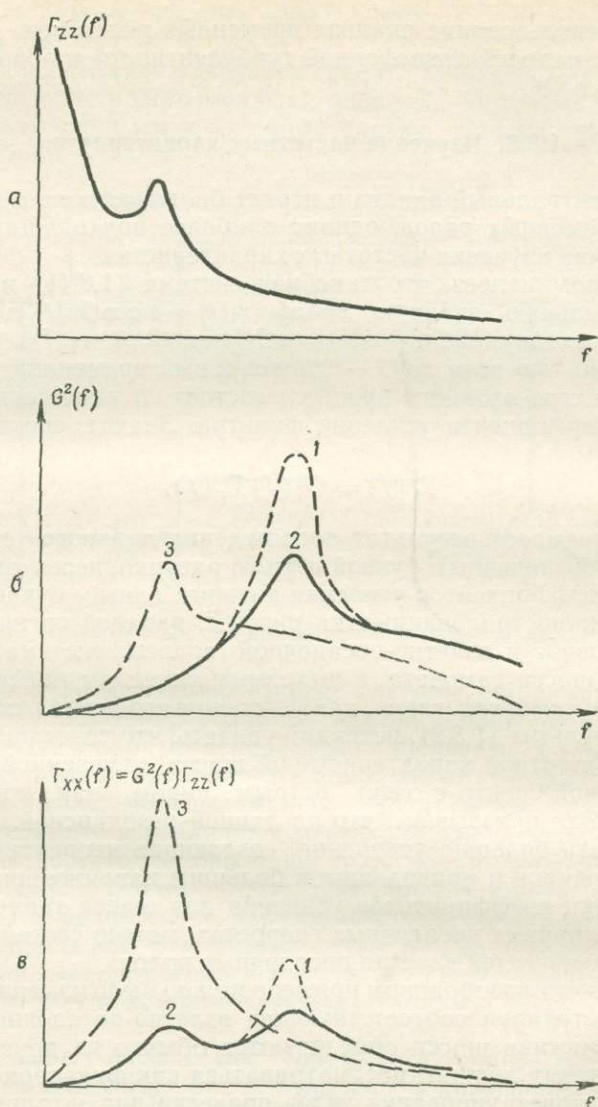


Рис. 1.7. Результат воздействия пиков входного спектра, амортизации и скорости на отклик самолета.

с целью оптимизации промышленных процессов. Эти применения обсуждаются подробнее в гл. 7.

1.4. КРУГ ВОПРОСОВ, РАССМОТРЕННЫХ В ДАННОЙ КНИГЕ

Нет никакого сомнения, что спектральные методы призваны сыграть важную роль в анализе временных рядов. Однако важно уяснить, что они действительно имеют ограничения и должны применяться разумно. Первые результаты по оцениванию спектров, основанному на записях конечной длины, можно найти в книгах Барлента [7] и Блэкмана и Тьюки [8].

В этих книгах рассматриваются главным образом вопросы оценивания спектров одиночных временных рядов. В настоящей книге эти понятия распространяются на случай оценивания спектров и взаимных спектров нескольких временных рядов и их последующего использования для оценки коэффициентов усиления и фазовых характеристик линейных систем.

Несколько разделов спектрального анализа не включены в эту книгу. Среди них важным является спектральный анализ случайных процессов, зависящих от нескольких переменных^{*)}, например высоты океанских волн как функции земных координат. Другой опущенный раздел — спектры высших порядков, например биспектр. Спектры высших порядков полезны при анализе негауссовских процессов и нелинейных систем. Случайные поля были опущены из-за того, что книга и так уже очень велика. Что касается нелинейных спектров, то они были опущены главным образом потому, что, по нашему мнению, дополнительные усложнения, вносимые этими спектрами, затрудняют их широкое использование. По данным, имеющимся к настоящему времени, чувствуется, что параметрические методы больше подходят в этих ситуациях.

ЛИТЕРАТУРА

1. Granger C. W. J., Spectral Analysis of Economic Time Series, Princeton Univ. Press, Princeton, 1964.
2. Box G. E. P., Jenkins G. M., Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, 1970.
3. Lumley J. L., Panofsky H. A., The Structure of Atmospheric Turbulence, John Wiley, New York, 1964. (Русский перевод: Ламли Дж., Пановский Г., «Структура атмосферной турбулентности», М., изд-во «Мир», 1966.)
4. Batchelor G. K., The Theory of Homogeneous Turbulence, Cambridge Univ. Press, Cambridge, 1953.
5. Wiener N., The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, John Wiley, New York, 1949.

^{*)} В советской литературе такие случайные процессы называются случайными полями.— Прим. перев.

6. Laning J. H., Battin R. H., Random Processes in Automatic Control, McGraw-Hill, New York, 1956. (Русский перевод: Лэнинг Дж. Х., Баттин Р. Г., Случайные процессы в задачах автоматического управления, М., ИЛ, 1958.)
7. Bartlett M. S., An Introduction to Stochastic Processes with Special Reference to Methods and Applications, Cambridge Univ. Press, Cambridge, 1953. (Русский перевод: Бартлетт М. С., Введение в теорию случайных процессов, М., ИЛ, 1958).
8. Blackman R. B., Tukey J. W., The Measurement of Power Spectra from the Point of View of Communications Engineering, Dover, New York, 1958.

АНАЛИЗ ФУРЬЕ

Спектральный анализ объединяет два важных теоретических подхода: статистический анализ временных рядов и методы анализа Фурье. Последние не нуждаются в подробном изложении для инженеров, так как значительная часть инженерной подготовки базируется на этих методах. Однако ради полноты изложения и для удобства других читателей в этой главе будут описаны те понятия анализа Фурье, которые необходимы для анализа временных рядов. В последующих главах будет показано, как должны быть модифицированы методы Фурье для обработки функций времени, которые являются скорее статистическими, чем детерминированными.

2.1. ВВЕДЕНИЕ

2.1.1. Роль анализа Фурье в прикладной математике и в технических науках

Аналитические методы, развитые Жаном Батистом Жозефом Фурье (1768—1830), сыграли важную роль в развитии прикладной математики. Особенно важны они для трех приложений: а) для изучения периодических решений физических задач, описываемых дифференциальными уравнениями, особенно уравнениями в частных производных, например, для изучения волновых колебаний струн, возбужденных щипком, или для передачи электромагнитных волн по волноводам или кабелям; б) как операционный способ решения дифференциальных уравнений; например, обыкновенные дифференциальные уравнения с постоянными коэффициентами можно перевести с помощью преобразования Фурье в алгебраические уравнения; в) для приближения непериодических функций.

В этой книге мы будем иметь дело в основном с последним случаем и лишь эпизодически с решением дифференциальных уравнений. Периодические решения физических задач рассматриваться не будут. В качестве примера приближения непериодической функции рассмотрим детерминированную функцию $s(t)$ времени t , которую будем называть *сигналом* и которую нужно аппроксимировать с помощью выбранных подходящим образом периодических функций. Детерминированный сигнал является функцией, которая

известна точно для всех моментов времени, и поэтому представляет математическую идеализацию. Примерами детерминированных сигналов являются

$$s(t) = e^{-|t|}, \quad -\infty \leq t \leq \infty,$$

или

$$s(t) = a \cos 2\pi f_1 t, \quad -\infty \leq t \leq \infty.$$

Многие сигналы, встречающиеся на практике, полезно рассматривать как детерминированные, например следующие: напряжение в сети как функцию времени; выход генератора прямоугольных волн; перемещение предмета, подверженного внезапному воздействию постоянной силы; ток, протекающий через сопротивление, когда оно внезапно замыкается на заряженный конденсатор. Размерность первых двух сигналов выражается в вольтах, третьего — в метрах и четвертого — в амперах. Однако размерность сигнала могла бы быть и метром в секунду, если бы сигнал был скоростью, и единицей температуры, давления и т. д. Для того чтобы не возникало противоречий, всегда будет предполагаться, что t измеряется в секундах, а $s(t)$ — в вольтах, поскольку в большинстве практических приложений изучаемая физическая величина перед регистрацией переводится в напряжение.

Детерминированная функция, упомянутая в первом случае, является *непериодической*, в то время как во втором случае функция — *периодическая*. Слово «периодическая» означает, что существует число T , называемое *периодом* функции, такое, что

$$s(t) = s(t + T) \quad (2.1.1)$$

для всех t .

Между моментами времени t и $t+T$ функция может иметь совершенно произвольную форму. Особенно простой формой обладает косинусоидальная функция в упомянутом выше примере, которая имеет период $T = 1/f_1$, так как

$$a \cos 2\pi f_1 \left[t + \frac{1}{f_1} \right] = a \cos 2\pi f_1 t.$$

Непериодическую функцию можно представить, используя любой класс периодических функций. В анализе Фурье такими функциями являются синусоидальная и косинусоидальная. Они обладают важным свойством ортогональности, так что коэффициенты можно находить независимо друг от друга.

2.1.2. Конечные ряды Фурье

Рассмотрим сигнал, заданный лишь в дискретные моменты времени, и предположим, что нужно разложить его по периодическим функциям. Дискретный сигнал можно рассматривать как полученный из непрерывного сигнала $s(t)$ длительности T при отсчете зна-

чений сигнала через интервалы времени Δ , как показано на рис. 2.1а. Это дает $N = T/\Delta$ выбранных значений s_r , где

$$s_r = s(t = r\Delta). \quad (2.1.2)$$

Для удобства будем считать, что N четное и равно $2n$, так что r может изменяться по целым числам $-n, \dots, 0, 1, \dots, n-1$.

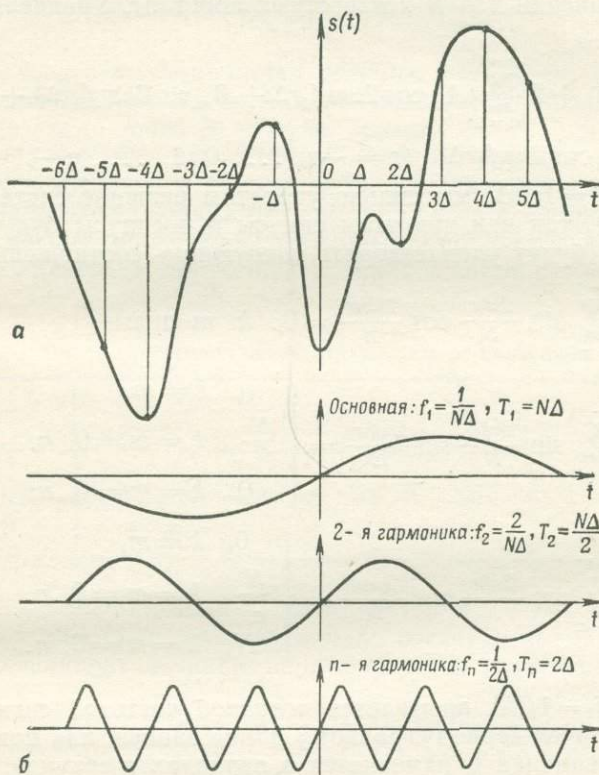


Рис. 2.1. а — дискретный сигнал, полученный выбором из непрерывного сигнала; б — основная синусоида и гармоники.

Заметим, что периодические функции, проходящие через значения сигнала в указанные N моментов времени, могут быть выбраны бесконечным множеством способов. Например, конечный ряд Фурье

$$\tilde{s}(t) = A_0 + 2 \sum_{m=1}^{n-1} (A_m \cos 2\pi m f_1 t + B_m \sin 2\pi m f_1 t) + A_n \cos 2\pi n f_1 t \quad (2.1.3)$$

содержит N констант A_m и B_m , которые можно определить так, чтобы дискретные и непрерывные значения совпадали в точках $t = r\Delta$, т. е. $\tilde{s}_r(t) = s_r$. Следовательно, функция $\tilde{s}(t)$ дает *приближение* к исходной непрерывной функции $s(t)$ в интервале $-T/2 \leq t < T/2$. Заменяя t на $r\Delta$ в (2.1.3) и полагая $\tilde{s}(r\Delta) = s_r$, получаем систему N уравнений для N неизвестных констант. Уравнения имеют вид

$$s_r = A_0 + 2 \sum_{m=1}^{n-1} (A_m \cos 2\pi m f_1 r\Delta + B_m \sin 2\pi m f_1 r\Delta) + A_n \cos 2\pi n f_1 r\Delta \quad (r = -n, \dots, 0, 1, \dots, n-1). \quad (2.1.4)$$

Выбрав $f_1 = 1/N\Delta$, мы сильно упростим решение системы уравнений (2.1.4), так как при этом синусы и косинусы будут *ортogonalны*, т. е. будут удовлетворять следующим соотношениям:

$$\begin{aligned} \sum_{r=-n}^{n-1} \sin \frac{2\pi kr}{N} \cos \frac{2\pi mr}{N} &= 0, \quad k, m \text{ целые;} \\ \sum_{r=-n}^{n-1} \sin \frac{2\pi kr}{N} \sin \frac{2\pi mr}{N} &= \begin{cases} 0, & k \neq m, \\ \frac{N}{2}, & k = m \neq 0, n, \\ 0, & k = m = 0, n; \end{cases} \\ \sum_{r=-n}^{n-1} \cos \frac{2\pi kr}{N} \cos \frac{2\pi mr}{N} &= \begin{cases} 0, & k \neq m, \\ \frac{N}{2}, & k = m \neq 0, n, \\ N, & k = m = 0, n. \end{cases} \end{aligned} \quad (2.1.5)$$

Частота $f_1 = 1/N\Delta$ называется *основной частотой* сигнала $\tilde{s}(t)$; она соответствует периоду, равному длине записи, как показано на рис. 2.1, б. Величина f_1 измеряется в периодах в секунду, или герцах (гц), если t измеряется в секундах (сек).

Таким образом, функция $\tilde{s}(t)$ в (2.1.3) составлена из суммы синусоидальных и косинусоидальных функций, частоты которых кратны основной частоте f_1 , т. е. являются *гармониками* основной частоты, как показано на рис. 2.1, б. Наивысшей из присутствующих частот является $n/N\Delta = 1/2\Delta$ гц, что соответствует периоду, равному двум интервалам отсчета.

Коэффициенты A_m или B_m в случае $f_1 = 1/N\Delta$ можно найти, умножая обе части (2.1.4) на $\cos(2\pi mr/N)$ или $\sin(2\pi mr/N)$ и суммируя по r , а затем воспользовавшись соотношениями ортогональности (2.1.5).

Окончательные выражения для коэффициентов следующие:

$$A_m = \frac{1}{N} \sum_{r=-n}^{n-1} s_r \cos \frac{2\pi mr}{N}, \quad (2.1.6)$$

$$B_m = \frac{1}{N} \sum_{r=-n}^{n-1} s_r \sin \frac{2\pi mr}{N}, \quad (2.1.7)$$

где $m=0, 1, \dots, n$. A_0 является *средним значением*, или средним арифметическим, величин s_r . Аналогичные выражения можно получить, когда число точек N нечетно, скажем $2n-1$, причем единственное отличие будет лишь в том, что член A_n исчезает.

Пример. Рассмотрим данные табл. 2.1., которая дает интенсивность сигналов, отраженных от одного из слоев E в ионосфере. Приведенные цифры являются осредненными по нескольким месяцам значениями интенсивности в фиксированное время суток.

Таблица 2.1

Интенсивности сигналов, отраженных от ионосферы

Время	0	1	2	3	4	5	6	7	8	9	10	11
Средняя интенсивность	-6	-20	-28	-8	-1	7	-20	-6	-7	14	19	12

Табл. 2.2 дает значения коэффициентов A_m и B_m , вычисленные по (2.1.6) и (2.1.7), причем за начало отсчета времени бралось 6 час. Коэффициент A_2 , например, получается следующим образом:

$$A_2 = \frac{1}{12} \left\{ (-6) \cos(-2\pi) + (-20) \cos\left(-\frac{5\pi}{3}\right) + \dots + \right. \\ \left. + (12) \cos\left(\frac{5\pi}{3}\right) \right\} = \frac{1}{12} \{-6 - 10 + \dots + 6\} = -2,25.$$

Амплитудное и фазовое представление. Иногда удобнее записывать (2.1.3) в виде

$$\tilde{s}(t) = R_0 + 2 \sum_{m=1}^{n-1} R_m \cos(2\pi m f_1 t + \varphi_m) + R_n \cos 2\pi n f_1 t, \quad (2.1.8)$$

где $R_m = \sqrt{A_m^2 + B_m^2}$, $\varphi_m = \arctg\left(-\frac{B_m}{A_m}\right)$ (2.1.9)

и $A_m = R_m \cos \varphi_m$, $B_m = -R_m \sin \varphi_m$. (2.1.10)

R_m называется *амплитудой* и φ_m — *фазой* m -й гармоники относительно некоторого произвольного начала отсчета времени. В приведенных выше формулах начало отсчета времени бралось в точке, расположенной примерно посередине между первым и последним значениями s_r . Если бы мы изменили это начало отсчета, то амплитуда осталась бы прежней, а фаза изменилась соответствующим образом. Амплитуды и фазы для ионосферных данных приведены в табл. 2.2.

Таблица 2.2

Разложение Фурье среднеквадратичного значения сигнала, представляющего ионосферные данные

Источник	m	A_m	B_m	R_m	φ_m	Вклад в среднеквадратичное значение
Среднее значение	0	-3,667	0	3,667	180°	13,44
Основная гармоника	1	-0,475	5,584	5,604	85	62,81
2-я гармоника	2	-2,250	-7,073	7,422	-72	110,17
3-я гармоника	3	-1,250	-0,250	1,275	-11	3,25
4-я гармоника	4	-0,667	0,577	0,882	41	1,56
5-я гармоника	5	-1,775	-0,334	1,806	-11	6,52
6-я гармоника	6	-3,500	0	3,500	0	12,25
Полное количество						210,00

Теорема Парсеваля. Среднеквадратичная величина, или средняя мощность, сигнала s_r равна

$$\frac{1}{N} \sum_{r=-n}^{n-1} s_r^2.$$

Используя (2.1.3) и свойства ортогональности (2.1.5), можно убедиться в том, что эта величина записывается в виде

$$\frac{1}{N} \sum_{r=-n}^{n-1} s_r^2 = R_0^2 + 2 \sum_{m=1}^{n-1} R_m^2 + R_n^2, \quad (2.1.11)$$

что является частным случаем *теоремы Парсеваля*. Другими словами, эта теорема утверждает, что среднеквадратичное значение сигнала s_r , или *средняя мощность*, рассеиваемая сигналом s_r , может быть разложена на составляющие, даваемые каждой гармони-

кой. Для нулевой и n -й гармоник вклад равен R_m^2 ($m=0$, $m=n$), а для m -й гармоники ($m \neq 0$, $m \neq n$) средняя мощность равна $2R_m^2$.

Более удобной мерой является среднеквадратичное значение сигнала s_r относительно среднего R_0 . Оно просто равно дисперсии

$$\sigma^2 = \frac{1}{N} \sum_{r=-n}^{n-1} (s_r - R_0)^2 = 2 \sum_{m=1}^{n-1} R_m^2 + R_n^2, \quad (2.1.12)$$

или, в терминах электротехники, *средней мощности переменного тока*.

Разложение среднеквадратичного значения s_r для ионосферных данных приведено в табл. 2.2. Мы видим, что среднее значение, основная и вторая гармоники составляют около 89% всей среднеквадратичной суммы, что указывает на то, что данные очень хорошо приближаются с помощью модели

$$s_r = -3,67 + 11,2 \cos\left(\frac{\pi r}{6} + 85^\circ\right) + 14,8 \cos\left(\frac{\pi r}{3} - 72^\circ\right).$$

Разложение среднеквадратичной суммы можно представить, нанеся на график среднюю мощность гармоники против частоты этой гар-

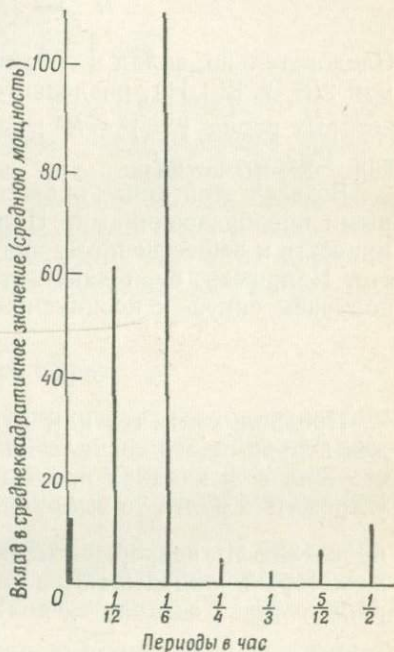


Рис. 2.2. Линейчатый спектр Фурье (периодограмма).

моники. Такой график называется *линейчатым спектром Фурье*; для ионосферных данных он показан на рис. 2.2.

Комплексные ряды Фурье. Приведенные выше формулы громоздки в обращении, поэтому для удобства в работе с ними лучше выразить сигнал s_r через *комплексные* амплитуды S_m , где

$$S_m = R_m e^{+j\varphi_m} = A_m - jB_m, \quad j^2 = -1. \quad (2.1.13)$$

Таким образом, (2.1.3) можно записать в виде

$$\tilde{s}(t) = \sum_{m=-n}^{n-1} S_m e^{j(2\pi m t / N \Delta)}, \quad (2.1.14)$$

где $S_{-m} = S_m^*$, причем звездочка означает комплексное сопряжение. Аналогично формулы (2.1.6) и (2.1.7) переходят в

$$S_m = \frac{1}{N} \sum_{r=-n}^{n-1} s_r e^{-j(2\pi mr/N)}, \quad -n \leq m \leq n-1, \quad (2.1.15)$$

и теорема Парсеваля (2.1.11) записывается как

$$\frac{1}{N} \sum_{r=-n}^{n-1} s_r^2 = \sum_{m=-n}^{n-1} |S_m|^2. \quad (2.1.16)$$

Следовательно, вклад в среднеквадратичную сумму, вносимый членом $2R_m^2$ в (2.1.11), разделяется в (2.1.16) на две части, каждая из которых равна $|S_m|^2 = R_m^2$; одна соответствует частоте mf_1 , а другая — частоте $-mf_1$.

Во всей этой книге окажется удобнее оперировать с комплексными преобразованиями. Получаемые при этом формулы можно привести к вещественному виду, взяв действительную и мнимую части. Например, беря действительную и мнимую части от (2.1.15), получаем синус- и косинус-преобразования (2.1.6) и (2.1.7).

2.1.3. Ряды Фурье

Предположим, что нам нужно получить представление Фурье для непрерывного сигнала на интервале от $-T/2$ до $T/2$. Заметим, что если в выкладках предыдущего раздела интервал отсчета Δ устремить к нулю, то выбранные точки сигнала s_r будут все полнее

прослеживать непрерывный сигнал $s(t)$. Непрерывный сигнал $\tilde{s}(t)$, на который накладываются условия, чтобы он проходил через выбранные точки сигнала s_r , должен при этом совпадать с $s(t)$, и поэтому в этом предельном случае представление Фурье $\tilde{s}(t)$ будет *точным* представлением сигнала $s(t)$ на интервале от $-T/2$ до $T/2$.

Коэффициенты Фурье S_m , определяемые в (2.1.15), можно переписать в виде

$$S_m = \frac{1}{N\Delta} \sum_{r=-n}^{n-1} s_r \Delta e^{-j(2\pi mr\Delta/N\Delta)}, \quad (2.1.17)$$

и если $\Delta \rightarrow 0$ и $N \rightarrow \infty$, так что $N \cdot \Delta = T$, то $r\Delta \rightarrow t$, $s_r \Delta \rightarrow s(t)dt$ и сумма (2.1.17) стремится к интегралу

$$S_m = \frac{1}{T} \int_{-T/2}^{T/2} s(t) e^{-j(2\pi mt/T)} dt. \quad (2.1.18)$$

Аналогично (2.1.14) стремится к

$$s(t) = \sum_{m=-\infty}^{\infty} S_m e^{j(2\pi mt/T)}. \quad (2.1.19)$$

Теорема Парсевала (2.1.16) теперь переходит в

$$\frac{1}{T} \int_{-T/2}^{T/2} s^2(t) dt = \sum_{m=-\infty}^{\infty} |S_m|^2, \quad (2.1.20)$$

поскольку (2.1.16) можно записать в виде

$$\frac{1}{N\Delta} \sum_{r=n}^{n-1} s_r^2 \Delta = \sum_{m=-n}^{n-1} |S_m|^2$$

и $s_r^2 \Delta \rightarrow s^2(t) dt$, когда $\Delta \rightarrow 0$ и $N \rightarrow \infty$. Уравнение (2.1.20) утверждает, что средний квадрат непрерывного периодического сигнала $s(t)$ можно разложить на бесконечное число вкладов от гармоник $f_m = m/T$ ($-\infty < m < +\infty$) основной частоты $1/T$ гц. Уравнение (2.1.19) называется представлением функции $s(t)$ в виде *ряда Фурье* на интервале $-T/2 \leq t < T/2$. Заметим, что хотя приведенные выше рассуждения являются эвристическими, они могут быть строго обоснованы.

2.1.4. Интегралы Фурье

До сих пор было показано, что с помощью тригонометрических рядов можно представить два типа сигналов. Сигналы первого типа s_r состояли из конечного числа N ординат, отстоящих на Δ сек друг от друга. Сигналы этого типа можно было бы представить на дан-

ном интервале с помощью непрерывного сигнала $\tilde{s}(t)$, образованного N гармониками основной частоты $1/N\Delta$ гц. Максимальной из присутствующих частот является $1/2\Delta$ гц, и поэтому про сигнал

$\tilde{s}(t)$ говорят, что он имеет *ограниченную полосу частот*. Сигналы второго типа $s(t)$ были непрерывными сигналами, заданными на интервале $-T/2 \leq t \leq T/2$. Мы видели, что сигналы такого типа можно представить на этом интервале с помощью некоторого сигнала, состоящего из бесконечного числа гармоник основной частоты $1/T$ гц.

В более общем случае нужно рассматривать сигналы $s(t)$ третьего типа, определенные на *бесконечном* интервале $-\infty \leq t \leq \infty$. Соответствующий подход является предельным случаем анализа Фурье, изложенного в разд. 2.1.3, в котором рассматриваются неограниченно увеличивающиеся отрезки бесконечной записи. По мере того как T стремится к бесконечности, частотный интервал $1/T$ между соседними гармониками становится бесконечно малым, что приводит к непрерывному распределению амплитуд по частоте.

Чтобы продемонстрировать эти предельные рассуждения, можно переписать (2.1.19) в виде

$$s(t) = \sum_{m=-\infty}^{\infty} (TS_m) e^{j(2\pi mt/T)} \frac{1}{T}. \quad (2.1.21)$$

В пределе, когда $T \rightarrow \infty$, $m/T \rightarrow f$, $1/T \rightarrow df$ и $TS_m \rightarrow S(f)$. Поэтому (2.1.21) стремится к интегралу

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} df. \quad (2.1.22)$$

Аналогично (2.1.18) можно переписать в виде

$$TS_m = \int_{-T/2}^{T/2} s(t) e^{-j2\pi (m/T)t} dt, \quad (2.1.23)$$

что стремится к

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt, \quad (2.1.24)$$

когда $T \rightarrow \infty$. Функция $S(f)$ называется *преобразованием Фурье* функции $s(t)$.

Соотношение Парсеваля (2.1.20) для случая бесконечного интервала можно записать в виде

$$\int_{-T/2}^{T/2} s^2(t) dt = \sum_{m=-\infty}^{\infty} |TS_m|^2 \frac{1}{T}, \quad (2.1.25)$$

что стремится к

$$\int_{-\infty}^{\infty} s^2(t) dt = \int_{-\infty}^{\infty} |S(f)|^2 df. \quad (2.1.26)$$

Предельные операции в (2.1.25) можно представить себе следующим образом: сначала считаем, что мощность, или дисперсия, $|S_m|^2$ на частоте m/T распределяется на полосе частот шириной $1/T$, что дает среднюю мощность $T|S_m|^2$ в этой полосе; затем эта средняя мощность стремится к непрерывному распределению мощности по частоте по мере того, как ширина полосы становится бесконечно малой.

Физически преобразование Фурье $S(f)$ представляет собой *распределение интенсивности сигнала по частоте*, т. е. является функцией плотности. Если s измеряется в вольтах и t — в секундах, то размерность $S(f)$ есть «вольт-секунда», или «вольт на единицу частоты», так как f имеет размерность частоты, т. е. сек^{-1} .

В математических руководствах по анализу Фурье приводится множество достаточных условий для существования интегралов

Краткая сводка формул преобразований Фурье

Описание	Функция	Преобразование	Обратное преобразование
Конечный дискретный ряд	$s_r,$ $r = -n, \dots, 0, 1, \dots, n-1$	$S_m = \frac{1}{N} \sum_{r=-n}^{n-1} s_r e^{-j(2\pi mr/N)},$ $m = -n, \dots, 0, 1, \dots, n-1$	$\tilde{s}(t) = \sum_{m=-n}^{n-1} S_m e^{j(2\pi mt/N\Delta)},$ $-\infty \leq t \leq \infty$ $\tilde{s}(t = r\Delta) = s_r,$ $\tilde{s}(t) = s(t + N\Delta)$
Непрерывная периодическая функция	$s(t) = s(t + T),$ $-\infty \leq t \leq \infty$	$S_m = \frac{1}{T} \int_{-T/2}^{T/2} s(t) e^{-j(2\pi mt/T)} dt,$ $m = 0, \pm 1, \pm 2, \dots$	$s(t) = \sum_{m=-\infty}^{\infty} S_m e^{j(2\pi mt/T)},$ $-\infty \leq t \leq \infty$ $s(t) = s(t + T)$
Непрерывная апериодическая функция	$s(t),$ $-\infty \leq t \leq \infty$	$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt,$ $-\infty \leq f \leq \infty$	$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} df,$ $-\infty \leq t \leq \infty$

(2.1.22) и (2.1.24). В этой книге мы обходим эти условия, используя теорию обобщенных функций, начало которой было положено Дираком и которая впоследствии была строго обоснована Шварцем. Превосходное описание этой теории дано в [1, 4*]; можно рекомендовать также [2]. Согласно этой теории, каждая обобщенная функция имеет преобразование Фурье, которое само является обобщенной функцией. Одно из следствий этой теории заключается в том, что ряд Фурье можно рассматривать как частный случай интеграла Фурье, как мы увидим впоследствии. Результаты разд. 2.1 резюмированы в табл. 2.3 на стр. 43.

2.2. ПРЕОБРАЗОВАНИЯ ФУРЬЕ И ИХ СВОЙСТВА

2.2.1. Функции с хорошим поведением

В качестве примера применения (2.1.24) рассмотрим преобразование Фурье простой функции $s(t) = e^{-|t|}$. Тогда

$$\begin{aligned} S(f) &= \int_{-\infty}^{\infty} e^{-|t|} e^{-j2\pi ft} dt = \\ &= \frac{1}{1-j2\pi f} + \frac{1}{1+j2\pi f} = \frac{2}{1+(2\pi f)^2}. \end{aligned}$$

В табл. 2.4 приведены преобразования Фурье некоторых сигналов $s(t)$, которые нам понадобятся позднее.

Таблица 2.4

Некоторые простые функции и их преобразования Фурье

$s(t)$	$S(f)$
$e^{- t } (-\infty \leq t \leq \infty)$	$2/[1+(2\pi f)^2]$
$\begin{cases} 0, & t > b \\ a & t \leq b \end{cases}$	$2ab \frac{\sin 2\pi fb}{2\pi fb}$
$e^{-a t } \cos 2\pi f_0 t$	$\frac{a}{a^2 + [2\pi(f+f_0)]^2} + \frac{a}{a^2 + [2\pi(f-f_0)]^2}$

Эти сигналы и их преобразования изображены на рис. 2.3. Вспомня, что $S(f)$ дает распределение интенсивности сигнала по частоте, отметим, что сигнал на рис. 2.3, a является вполне плавным, и поэтому в его преобразовании доминируют низкие частоты. Заме-

тим также, что острые углы в $s(t)$, как на рис. 2.3, б, создают волнистую рябь, или боковые лепестки, в преобразовании, а периодич-

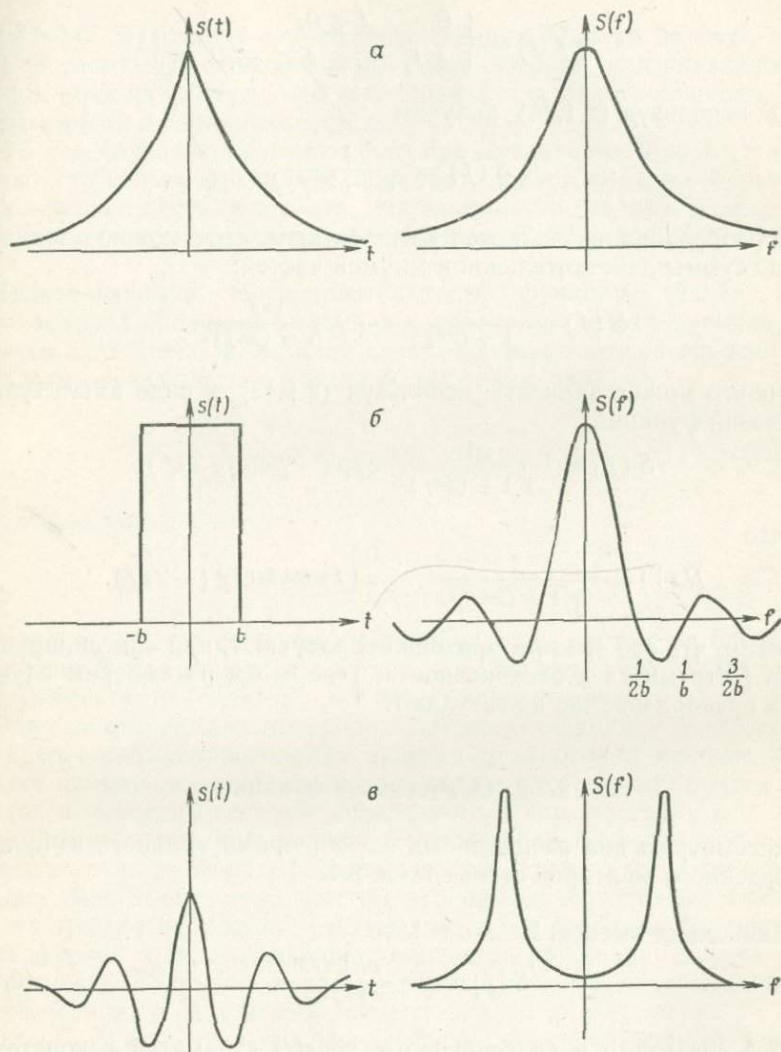


Рис. 2.3. Некоторые простые сигналы и их преобразования Фурье.

ности в $s(t)$ появляются в преобразовании в виде пиков, что видно на рис. 2.3, в.

Все сигналы в табл. 2.4 являются четными функциями t , и поэтому их преобразования Фурье являются действительными и

четными функциями. В общем случае это не так. Например, предположим, что $s(t)$ не является четной функцией

$$s(t) = \begin{cases} 0, & t < 0, \\ e^{-t}, & 0 \leq t \leq \infty. \end{cases}$$

Тогда, используя (2.1.24), получим

$$S(f) = \frac{1}{1 + j2\pi f}.$$

Это преобразование является комплексным, его можно записать в виде суммы действительной и мнимой частей:

$$S(f) = \frac{1}{1 + (2\pi f)^2} - j \frac{2\pi f}{1 + (2\pi f)^2}.$$

Иначе его можно записать, используя (2.1.13), в виде амплитудной и фазовой функций

$$S(f) = \frac{1}{\sqrt{1 + (2\pi f)^2}} \exp(-j \arctg 2\pi f),$$

так что

$$R(f) = \frac{1}{\sqrt{1 + (2\pi f)^2}}, \quad \varphi(f) = \arctg(-2\pi f).$$

Отметим, что все эти преобразования затухают, или «диссипируют», когда f стремится к бесконечности. Теперь мы рассмотрим случаи, когда преобразования не затухают.

2.2.2. Обобщенные функции

Рассмотрим два специальных случая прямоугольного импульса, приведенного во второй строке табл. 2.4.

Единичная высота. Если $a = 1$, то

$$S(f) = 2b \frac{\sin 2\pi fb}{2\pi fb}. \quad (2.2.1)$$

Если b стремится к бесконечности, то $s(t)$ стремится к константе, равной 1 всюду. Поведение $S(f)$ при увеличении b проиллюстрировано на рис. 2.4, где можно видеть, что $S(f)$ стремится стать острым пиком бесконечной высоты при $f=0$ и ограничена во всех остальных точках. Такая функция понимается как дельта-функция Дирака, или импульсная функция. Поэтому *преобразование Фурье от константы есть дельта-функция.*

Единичная площадь. Если $2ab = 1$, то

$$S(f) = \frac{\sin 2\pi fb}{2\pi fb}. \quad (2.2.2)$$

Когда $b \rightarrow 0$, $S(f)$ всюду стремится к единице. Однако по мере того как b убывает, $s(t)$ становится все более высокой, как показано на рис. 2.5. Отсюда следует, что $s(t)$ стремится к дельта-функции, сосредоточенной в начале координат.

Эти два случая показывают, что преобразование Фурье от константы есть дельта-функция и, наоборот, преобразование Фурье от дельта-функции есть константа. Эту взаимность следовало ожидать из-за симметрии равенств преобразования (2.1.22) и (2.1.24).

Дельта-функции. Последовательность функций (2.2.1) при $b \rightarrow \infty$, которая послужила нам для определения дельта-функции, не является единственной. Вообще дельта-функцию можно определить как последовательность функций $\delta_n(t)$, таких, что

$$\int_{-\infty}^{\infty} \delta_n(t) dt = 1 \quad \text{для каждого } n, \quad (2.2.3)$$

и в пределе, когда $n \rightarrow \infty$,

$$\delta(t) = \begin{cases} 0, & t \neq 0, \\ \infty, & t = 0. \end{cases} \quad (2.2.4)$$

Примеры таких последовательностей функций вместе с их преобразованиями Фурье приведены в табл. 2.5. Заметим, что $S_n(f)$ стремится к константе (единице) для всех f , когда $n \rightarrow \infty$.

Одну из физических интерпретаций дельта-функции дает описание процессов преобразования энергии в некоторой системе. Используя пример из механики, предположим, что твердый брусок находится в покое на плоской поверхности. Если выстрелить в этот брусок очень маленькой пулей, летящей с большой скоростью, то при ударе пули произойдет обмен энергии. Предполагая, что столкновение происходит столь быстро, что брусок не успевает сдвинуться за это время, можно считать, что пуля передала бруску импульс энергии в виде изменения количества движения. Другую интерпретацию, взятую из теории электромагнетизма, дает единичный точечный заряд в начале координат.

Дельта-функцию можно использовать как операторный прием для выбора значения сигнала в данный момент времени. Следующая выкладка поясняет это:

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \delta_n(t - t_0) s(t) dt = \int_{-\infty}^{\infty} \delta(t - t_0) s(t) dt = s(t_0). \quad (2.2.5)$$

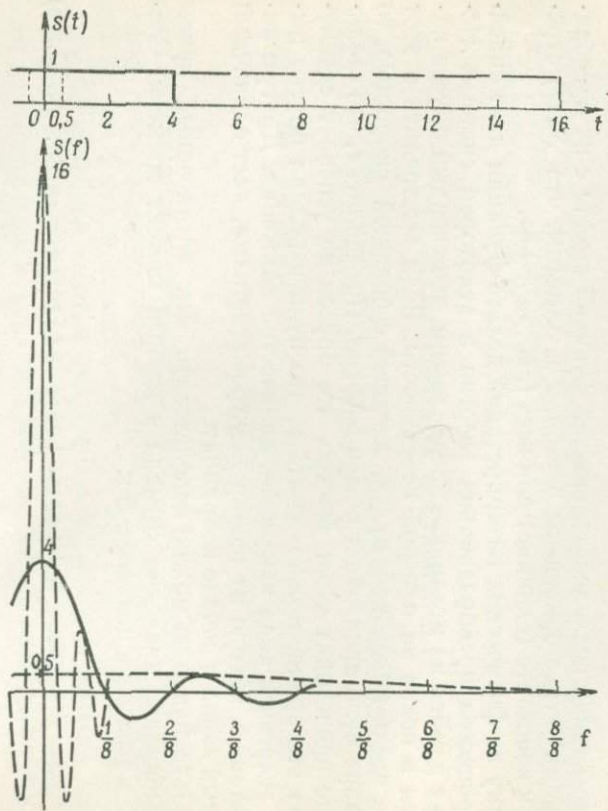


Рис. 2.4. Прямоугольные импульсы единичной высоты и их преобразования Фурье.

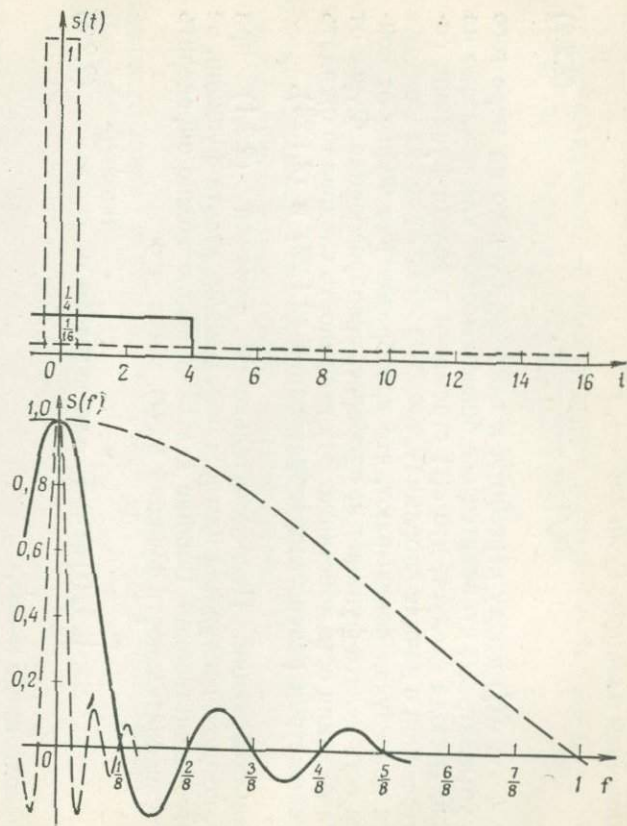


Рис. 2.5. Прямоугольные импульсы единичной площади и их преобразования Фурье.

Таблица 2.5

Последовательности, определяющие дельта-функции

$\delta_n(t)$	$S_n(f)$
1) $2n \frac{\sin 2\pi nt}{2\pi nt}$	1, $ f \leq n$ 0, $ f > n$
2) $\sqrt{n} e^{-\pi n t^2}$	$e^{-\pi f^2/n}$
3) $\frac{n}{2} e^{-n t }$	$\frac{n^2}{n^2 + (2\pi f)^2}$
4) $\frac{1}{\pi} \frac{n}{n^2 t^2 + 1}$	$e^{- 2\pi f/n }$
5) $\frac{n \sin^2(\pi n t)}{(\pi n t)^2}$	$1 - \frac{ f }{n}, \quad f \leq n$ 0, $ f > n$

Рассматривая аналогичным способом предел последовательности m -х производных $\delta_n(t)$ [1], можно определить m -ю производную дельта-функции, а именно $\delta^{(m)}(t)$. Ее можно использовать для выбора m -й производной некоторой функции в данной точке. Это приводит к обобщению (2.2.5), а именно

$$\int_{-\infty}^{\infty} \delta^{(m)}(t - t_0) s(t) dt = (-1)^m s^{(m)}(t_0). \quad (2.2.6)$$

Возвращаясь к интерпретации дельта-функции как единичного заряда в начале координат, можно сказать, что $\delta'(t)$ соответствует математической идеализации единичного диполя. Это обусловлено тем, что первый момент $\delta'(t)$ равен

$$\int_{-\infty}^{\infty} t \delta'(t) dt = -1,$$

где мы воспользовались (2.2.6). Поэтому абсолютный момент $\delta'(t)$ равен единице, что является стандартным определением единичного диполя.

Функция единичного скачка ^{*}). С $\delta(t)$ тесно связана функция единичного скачка. Физически она соответствует приложению единичной силы, которая затем остается постоянной, или переключению крана, которое меняет поток в трубе. Математически она является сигналом, задаваемым равенствами

$$U(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{2}, & t = 0, \\ 1, & t > 0. \end{cases} \quad (2.2.7)$$

Функцию $U(t)$ можно рассматривать как предел последовательности функций $U_n(t)$ при $n \rightarrow \infty$, например

$$U_n(t) = \begin{cases} \frac{1}{2} e^{nt}, & t < 0, \\ 1 - \frac{1}{2} e^{-nt}, & t \geq 0. \end{cases} \quad (2.2.8)$$

Когда $n \rightarrow \infty$, то $U_n(t) \rightarrow 0$ для отрицательных t и к единице для положительных t . Дифференцирование $U_n(t)$ дает

$$\frac{d}{dt} U_n(t) = \frac{n}{2} e^{-n|t|} = \delta_n(t), \quad (2.2.9)$$

что иллюстрирует важный результат: производная функции скачка есть дельта-функция.

Преобразованием Фурье функции единичного скачка (2.2.7) является

$$S(f) = \frac{1}{2} \delta(f) + \frac{1}{j2\pi f}.$$

2.2.3. Ряды Фурье как преобразования Фурье

Рассмотрим преобразование Фурье следующего сигнала:

$$s_T(t) = \begin{cases} a \cos \frac{2\pi t}{\Delta}, & -\frac{T}{2} \leq t \leq \frac{T}{2}, \\ 0, & |t| > \frac{T}{2}, \end{cases} \quad (2.2.10)$$

который является «периодическим» сигналом в интервале $(-T/2, +T/2)$. Непосредственно используя (2.1.24), получаем, что его преобразование Фурье равно

$$S_T(f) = \frac{a}{2} \left\{ T \frac{\sin \pi T [f - (1/\Delta)]}{\pi T [f - (1/\Delta)]} + T \frac{\sin \pi T [f + (1/\Delta)]}{\pi T [f + (1/\Delta)]} \right\}. \quad (2.2.11)$$

^{*}) Unit step function. В операционном исчислении применяется также название «единичная функция (Хевисайда)». — Прим. перев.

Когда T стремится к бесконечности, сигнал $s_T(t)$ становится действительно периодическим сигналом $s(t)$ (периодическим для всех моментов времени), в то время как преобразование $S_T(f)$ стремится к

$$S(f) = \frac{a}{2} \left\{ \delta \left(f - \frac{1}{\Delta} \right) + \delta \left(f + \frac{1}{\Delta} \right) \right\}, \quad (2.2.12)$$

поскольку каждый из членов внутри фигурных скобок в (2.2.11) является последовательностью, сходящейся к дельта-функции. Поэтому преобразование Фурье действительно периодической косинусоидальной волны (бесконечного протяжения) состоит из дельта-функции амплитуды $a/2$, сосредоточенной в $f = +1/\Delta$, и дельта-функции амплитуды $a/2$, сосредоточенной в $f = -1/\Delta$.

Аналогично комплексный сигнал

$$s_T(t) = e^{j(2\pi mt/\Delta)}, \quad -\frac{T}{2} \leq t \leq \frac{T}{2},$$

имеет преобразование Фурье

$$S_T(f) = T \frac{\sin \pi T [f - (m/\Delta)]}{\pi T [f - (m/\Delta)]}.$$

Поэтому, когда $T \rightarrow \infty$, $S_T(f)$ стремится к $S(f) = \delta [f - (m/\Delta)]$. Отсюда следует, что периодический сигнал с периодом Δ , представляемый рядом Фурье

$$s(t) = \sum_{m=-\infty}^{\infty} S_m e^{j(2\pi mt/\Delta)}, \quad (2.2.13)$$

имеет преобразование Фурье

$$S(f) = \sum_{m=-\infty}^{\infty} S_m \delta \left(f - \frac{m}{\Delta} \right), \quad (2.2.14)$$

которое представляет собой ряд, состоящий из дельта-функций. Таким образом, допуская обобщенные функции, ряды Фурье можно рассматривать как частный случай преобразований Фурье.

Для того чтобы найти коэффициенты Фурье S_m , соответствующие некоторой обобщенной функции, уже нельзя применять классическую формулу (2.1.18), так как обобщенная функция может оказаться неинтегрируемой в конечных пределах. Соответствующая формула, которую нужно использовать в таких случаях, приводится в [1].

В частности, можно показать, что преобразованием Фурье ряда, состоящего из дельта-функций

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - n\Delta), \quad (2.2.15)$$

является

$$S(f) = \frac{1}{\Delta} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\Delta}\right). \quad (2.2.16)$$

Таким образом, ряд из дельта-функций переходит в ряд также из дельта-функций. Отметим, что этот результат симметричен по отношению к частотной и временной областям.

Ряд, состоящий из дельта-функций, не является единственной функцией, симметричной относительно преобразования Фурье. Более простая функция, обладающая этим свойством, дается примером 2 в табл. 2.5 при $n=1$. Таким образом, $s(t) = \exp(-\pi t^2)$ преобразуется в $S(f) = \exp(-\pi f^2)$.

В этом месте читатель должен убедиться, что он хорошо знаком с различными операторными свойствами преобразований Фурье, которые резюмированы в приложении П2.1.

2.3. ЛИНЕЙНЫЕ СИСТЕМЫ И СВЕРТКИ

2.3.1. Линейные дифференциальные уравнения

Важность практического применения анализа Фурье и спектрального анализа определяется тем, что они упрощают анализ *инвариантных во времени линейных систем*, т. е. систем, поведение которых можно описать с помощью линейных интегро-дифференциальных уравнений с постоянными коэффициентами. Можно показать вообще [3], что решение такого уравнения может быть записано в виде интеграла свертки

$$y(t) = \int_{-\infty}^{\infty} h(u) x(t-u) du, \quad (2.3.1)$$

где $y(t)$ — решение и $x(t)$ — вынуждающая функция. В разд. 2.3.4 будет показано, что это решение упрощается, если перейти к преобразованиям Фурье. Преобразование решения дает

$$Y(f) = H(f) X(f),$$

где $Y(f)$, $H(f)$ и $X(f)$ — преобразования Фурье от $y(t)$, $h(t)$ и $x(t)$ соответственно. Таким образом, свертка во временной области преобразуется в произведение в частотной области.

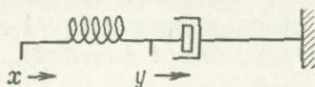
Иллюстрация свертки. Чтобы проиллюстрировать интеграл свертки, рассмотрим простую линейную систему, состоящую из пружины и буфера, показанную схематически на рис. 2.6. Одно из назначений такого устройства состоит в том, чтобы двери не хлопали. Сила, приложенная к пружине, производит *входное* смещение $x(t)$,

которое вызывает выходное смещение $y(t)$ буферного хомута. Дифференциальное уравнение, полученное приравнованием сил, имеет вид

$$K [x(t) - y(t)] = D \frac{dy}{dt},$$

где K — постоянная пружины, измеряемая в кг/м, и D — постоянная скорости буфера, измеряемая в кг · м/сек. Перегруппировав члены

Рис. 2.6. Механическая система первого порядка.



этого уравнения, получим

$$T \frac{dy}{dt} + y(t) = x(t), \quad (2.3.2)$$

где $T = D/K$ — постоянная времени этой системы (в секундах).

Уравнение (2.3.2) можно использовать для описания поведения многих других физических систем, например температуры у выпускного отверстия химического реактора, когда температура у впускного отверстия равна $x(t)$. В этом случае уравнение (2.3.2) показывает, что скорость изменения температуры у выпускного отверстия прямо пропорциональна температурному градиенту в реакторе.

Решение уравнения (2.3.2) можно записать в виде интеграла свертки, вводя интегрирующий множитель $e^{t/T}$. Таким образом, получим

$$y(t) = \int_{-\infty}^t x(u) \frac{e^{-(t-u)/T}}{T} du = \int_{-\infty}^{\infty} x(u) h(t-u) du, \quad (2.3.3)$$

где

$$h(u) = \begin{cases} \frac{1}{T} e^{-u/T}, & u \geq 0, \\ 0, & u < 0. \end{cases}$$

Следовательно, выход $y(t)$ можно записать в виде взвешенной суммы прошлых значений входа $x(t)$, т. е. выходной сигнал является сверткой входного сигнала с весовой функцией $h(u)$.

Вообще можно показать [3], что решение любого линейного инвариантного во времени дифференциального уравнения можно записать так же, как и в (2.3.3), или же, сделав замену переменных, в виде

$$y(t) = \int_{-\infty}^{\infty} h(u) x(t-u) du. \quad (2.3.4)$$

Весовая функция полностью характеризует поведение системы, точно так же, как это делает дифференциальное уравнение.

Инвариантные во времени линейные системы. Уравнения (2.3.3) и (2.3.4) изображают в общем виде то, что известно под именем *инвариантных во времени линейных систем*, или фильтров. Они характеризуются следующими свойствами.

а) *Свойство линейности*: если $x_1(t)$ и $x_2(t)$ — два входных сигнала, а $y_1(t)$, $y_2(t)$ — соответствующие им выходные сигналы, то линейная комбинация $\mu_1 x_1(t) + \mu_2 x_2(t)$ входных сигналов дает на выходе *ту же самую* линейную комбинацию выходных сигналов $\mu_1 y_1(t) + \mu_2 y_2(t)$.

б) *Свойство неизменности во времени*: если входной сигнал $x(t)$ задержать на время τ , так что получится $x(t - \tau)$, то выходной сигнал задержится на то же самое время и будет равен $y(t - \tau)$.

Именно свойство (б) обеспечивает то, что весовая функция $h(u)$ не зависит от времени. Линейная система без свойства инвариантности во времени имела бы весовую функцию, зависящую от времени t . Можно показать, что системы, которые могут быть описаны с помощью линейных дифференциальных уравнений с постоянными коэффициентами, имеют инвариантное во времени представление (2.3.3). Впрочем, многие нелинейные системы можно *линеаризовать* так, что для малых возмущений на входе можно использовать (2.3.3) как приближенное изображение системы.

2.3.2. Функции скачка и импульсные функции

Для любой физической системы весовая функция $h(u)$ должна быть равна нулю для отрицательных значений u ; это означает, что система не может давать отклик на входные сигналы, которые она еще не приняла. Это условие называется условием физической реализуемости. Для физически реализуемых систем уравнения (2.3.3) и (2.3.4) можно записать в виде

$$y(t) = \int_0^{\infty} h(u) x(t - u) du, \quad (2.3.5)$$

или же

$$y(t) = \int_{-\infty}^t x(u) h(t - u) du. \quad (2.3.6)$$

Функции отклика на единичный импульс *). Предположим, что на систему воздействует резкий импульс в момент времени $t = 0$,

*) В нашей литературе используются также следующие названия: весовая функция, импульсная передаточная функция, импульсная переходная функция, функция импульсной реакции, импульсная характеристика. — Прим. перев.

так что $x(t) = \delta(t)$. Тогда

$$y(t) = \int_{-\infty}^{\infty} h(u) \delta(t-u) du, \quad (2.3.7)$$

и, используя (2.2.5), получаем, что последний интеграл равен $h(t)$. Весовая функция $h(t)$ называется функцией отклика этой системы на единичный импульс [4], так как она дает выходной сигнал в момент t для системы, подверженной действию импульса при $t=0$.

Отклики на единичный импульс для некоторых простых систем приведены в первом столбце табл. 2.6. На рис. 2.7 приведены отклики на единичный импульс для трех из этих систем. В первом примере (а) система представляет собой простую задержку, для которой выходной сигнал, или отклик на единичный импульс, является таким же импульсом, задержанным на время τ . Во втором примере (б) система описывается одной постоянной времени и изображается дифференциальным уравнением (2.3.2); для этой системы отклик на единичный импульс является экспоненциальной кривой, изображенной на рис. 2.7, б. Третий пример (в) представляет собой систему второго порядка, изображаемую дифференциальным уравнением

$$\frac{1}{\omega_n^2} \frac{d^2 y}{dt^2} + \frac{2\zeta}{\omega_n} \frac{dy}{dt} + y = x(t). \quad (2.3.8)$$

Для этой системы откликом на единичный импульс является затухающая синусоида, показанная на рис. 2.7, в.

Функции отклика на единичный скачок *). Линейную систему можно также охарактеризовать с помощью ее отклика на функцию единичного скачка (2.2.7). Предположим, например, что входным сигналом является скорость притока холодной воды в теплообменник, а выходной сигнал — температура воды у выпускного отверстия. Тогда откликом на единичный скачок будут изменения температуры со временем у выпускного отверстия, после того как сделано единичное изменение входной скорости потока. Из (2.3.5) получаем, что отклик в момент времени t на единичный скачок при $t=0$ равен

$$y(t) = \int_0^t h(u) du, \quad (2.3.9)$$

так что отклик на единичный скачок равен интегралу от отклика на единичный импульс.

Из рис. 2.7 можно видеть, что отклик на единичный скачок для системы, являющейся чистой задержкой τ , есть также единичный

*) Иногда называется также «переходная реакция на скачок». — Прим. перев.

Система	Частотная характеристика $H(f)$	Усиление $G(f)$	Фаза $\varphi(f)$
1) Усиление	g	g	0
2) Задержка	$e^{-j2\pi f\tau}$	1	$-2\pi f\tau$
3) Интегрирование	$\frac{1}{j2\pi f}$	$\frac{1}{2\pi f}$	$-\frac{\pi}{2}$
4) Одиночная экспонента	$\frac{1}{1 + j2\pi fT}$	$\frac{1}{\{1 + (2\pi fT)^2\}^{1/2}}$	$-\arctg 2\pi fT$
5) Одиночная экспонента с задержкой	$\frac{e^{-j2\pi f\tau}}{(1 + j2\pi fT)}$	$\frac{1}{\{1 + (2\pi fT)^2\}^{1/2}}$	$-2\pi f\tau - \arctg 2\pi fT$
6) Две последовательные экспоненты	$\frac{1}{(1 + j2\pi fT_1)(1 + j2\pi fT_2)}$	$\frac{1}{\{ [1 + (2\pi fT_1)^2] [1 + (2\pi fT_2)^2] \}^{1/2}}$	$-\arctg 2\pi fT_1$ $-\arctg 2\pi fT_2$
7) Задержка по квадратичному закону (квадратичная задержка)	$\frac{1}{1 - (f/f_n)^2 + j2\zeta(f/f_n)}$, $\omega_n = 2\pi f_n$	$\frac{1}{\{ [1 - (f/f_n)^2]^2 + [2(\zeta/f_n)]^2 \}^{1/2}}$	$-\arctg \frac{2\zeta(f/f_n)}{[1 - (f/f_n)^2]}$
8) Дифференцирование	$j2\pi f$	$2\pi f$	$\frac{\pi}{2}$
9) Линейное упреждение	$1 + j2\pi fT$	$\{1 + (2\pi fT)^2\}^{1/2}$	$\arctg 2\pi fT$

Функции отклика на единичный импульс, на единичный скачок
и частотные характеристики для некоторых простых систем

Система	Импульсный отклик $h(t)$	Отклик на скачок $y(t)$
1) Усиление	$g\delta(t)$	$0, \quad t < 0$ $g, \quad t \geq 0$
2) Задержка	$\delta(t - \tau)$	$0, \quad t < \tau$ $1, \quad t \geq \tau$
3) Интегрирование	1	t
4) Одиночная экспонента	$\frac{1}{T} e^{-t/T}$	$(1 - e^{-t/T})$
5) Одиночная экспонента с задержкой	$0, \quad t < \tau$ $\frac{1}{T} e^{-(t-\tau)/T}, \quad t \geq \tau$	$0, \quad t < \tau$ $(1 - e^{-(t-\tau)/T}), \quad t \geq \tau$
6) Две последовательные экспоненты	$\frac{e^{-t/T_1} - e^{-t/T_2}}{T_1 - T_2}$	$\left(1 - \frac{T_1 e^{-t/T_1} - T_2 e^{-t/T_2}}{T_1 - T_2}\right)$
7) Задержка по квадратичному закону (квадратичная задержка)	$\frac{\omega_n e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1-\zeta^2} t)}{\sqrt{1-\zeta^2}}$	$1 - \frac{e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1-\zeta^2} t + \varphi)}{\sqrt{1-\zeta^2}},$ $\sin \varphi = \sqrt{1-\zeta^2}$
8) Дифференцирование	$\delta'(t)$	$-\delta(t)$
9) Линейное упреждение	$\delta(t) + T\delta'(t)$	$1 - T\delta(t)$

скачок, начинающийся на τ сек позднее, как показано на рис. 2.7, а. Для экспоненциального отклика на единичный импульс отклик на

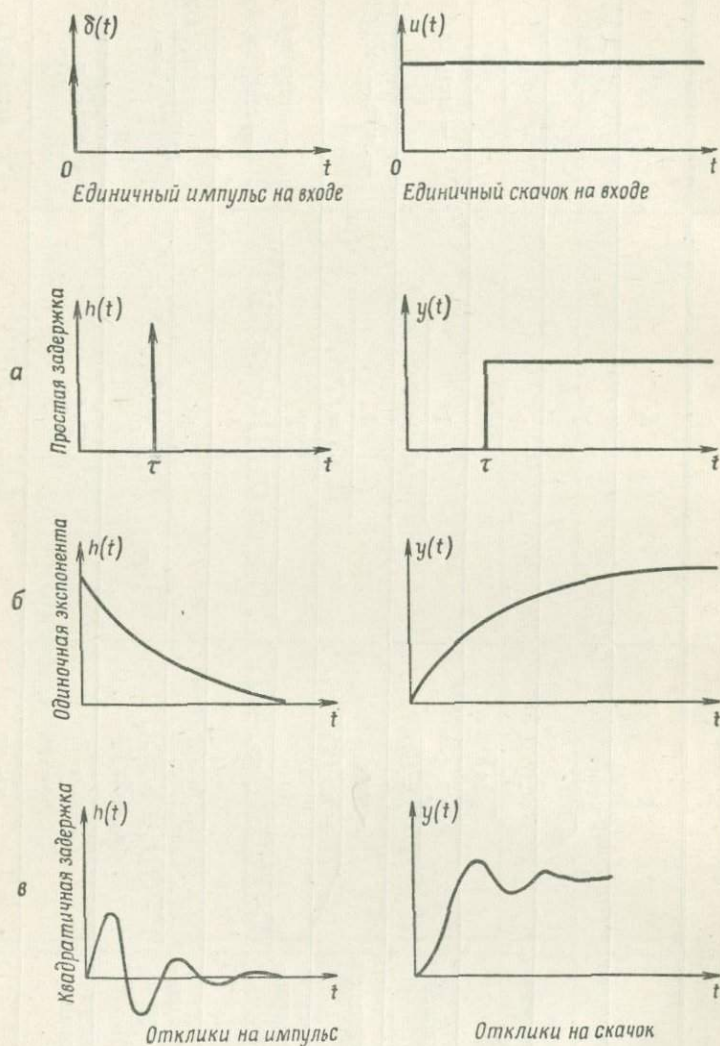


Рис. 2.7. Отклики на единичный импульс и единичный скачок для некоторых простых систем.

единичный скачок экспоненциально возрастает, стремясь к своему предельному значению, как показано на рис. 2.7, б. Для системы второго порядка (рис. 2.7, в) отклик на единичный скачок перехо-

дит свое предельное значение и затем колеблется около него с уменьшающейся амплитудой.

Когда $t \rightarrow \infty$, отклик на единичный скачок (2.3.9) стремится к значению

$$g = \int_0^{\infty} h(u) du, \quad (2.3.10)$$

которое называется *установившимся усилением* системы, так как оно измеряет предельное значение усиления после того, как система возмущена единичным скачком и ей дана возможность прийти до нового установившегося значения.

Устойчивость. Система называется *устойчивой* [4], если ограниченные входные сигналы создают ограниченные сигналы на выходе. Ясно, что такое свойство желательно, так как в противном случае выходной сигнал неограниченно возрастал бы. Предположим, что $|x(t)| \leq K_1$ в (2.3.5), где K_1 — некоторая конечная константа. Тогда

$$\begin{aligned} |y(t)| &= \left| \int_{-\infty}^{\infty} h(u) x(t-u) du \right| \leq \int_{-\infty}^{\infty} |h(u)| |x(t-u)| du \leq \\ &\leq K_1 \int_{-\infty}^{\infty} |h(u)| du, \end{aligned}$$

так что достаточным условием для того, чтобы система была устойчивой, является

$$\int_{-\infty}^{\infty} |h(u)| du < K_2, \quad (2.3.11)$$

где K_2 — также некоторая конечная константа. Другая форма условия устойчивости будет дана в следующем разделе.

2.3.3. Частотные характеристики

Для входных сигналов, более сложных, чем импульс или скачок, вычисление выходного сигнала с помощью интеграла свертки (2.3.5) становится утомительным. Эта задача значительно упрощается при использовании анализа Фурье. Метод состоит в следующем: сигнал $s(t)$ разлагают на его компоненты Фурье $S(f)$ по формуле (2.1.24), затем находят отклик системы на периодический сигнал $s_1(t) = e^{j2\pi f t}$ и, наконец, суммируют все отклики по формуле (2.1.22), что и дает окончательный выходной сигнал. Сначала нужно узнать

отклик системы на входной сигнал $x(t) = \cos 2\pi ft$. Подстановка этого сигнала в (2.3.5) дает

$$y(t) = \int_0^{\infty} h(u) \cos 2\pi f(t-u) du = \int_0^{\infty} h(u) [\cos 2\pi ft \cos 2\pi fu + \sin 2\pi ft \sin 2\pi fu] du = A(f) \cos 2\pi ft + B(f) \sin 2\pi ft, \quad (2.3.12)$$

где

$$A(f) = \int_0^{\infty} h(u) \cos 2\pi fu du \quad (2.3.13)$$

и

$$B(f) = \int_0^{\infty} h(u) \sin 2\pi fu du. \quad (2.3.14)$$

Иначе (2.3.12) можно переписать в виде

$$y(t) = G(f) \cos [2\pi ft + \varphi(f)], \quad (2.3.15)$$

где

$$G(f) = \sqrt{A^2(f) + B^2(f)}$$

и

$$\varphi(f) = \operatorname{arctg} \left(-\frac{B(f)}{A(f)} \right).$$

Отсюда отклик на косинусоидальную волну частоты f является косинусоидальной волной той же частоты, но с амплитудой, умноженной на величину $G(f)$, называемую *коэффициентом усиления*, и с фазой, сдвинутой на величину $\varphi(f)$, называемую *фазовым углом*.

Как и прежде, для удобства оперирования с формулами рассмотрим отклик на комплексный входной сигнал

$$e^{j2\pi ft} = \cos 2\pi ft + j \sin 2\pi ft$$

частоты f . В этом случае выходным сигналом будет

$$y(t) = H(f) e^{j2\pi ft} = G(f) e^{j[2\pi ft + \varphi(f)]}, \quad (2.3.16)$$

где функция

$$H(f) = G(f) e^{j\varphi(f)} = \int_0^{\infty} h(u) e^{-j2\pi fu} du \quad (2.3.17)$$

называется *частотной характеристикой системы*. Следовательно, частотная характеристика является преобразованием Фурье от функции отклика на единичный импульс.

Графики Бодэ. Частотные характеристики, коэффициенты усиления и фазы для некоторых простых систем приведены

в табл. 2.6, а коэффициенты усиления и фазы изображены на рис. 2.8. Обычно на график наносят логарифм коэффициента усиления в зависимости от логарифма частоты и фазу в зависимости от логарифма частоты. Эти графики называют *графиками Бодэ* [5]. Графики на рис. 2.8 распадаются естественным образом на четыре категории.

1. Номера 1 и 2 имеют постоянный коэффициент усиления для всех частот и называются *широкополосными* системами *) (пропускающими все частоты).

2. Номера 3, 4, 5 и 6 таковы, что высокие частоты отфильтровываются или ослабляются системой, а низкие частоты пропускаются с различными коэффициентами усиления. Поэтому эти системы ведут себя как фильтры низких частот и соответствуют некоторой форме интегрирования или сглаживания входного сигнала.

3. Номер 7 соответствует колебательной системе, описываемой уравнением (2.3.8). Здесь график коэффициента усиления имеет *резонанс*, или пик, на частоте $f = f_n(1 - 2\xi^2)^{1/2}$, где f_n — естественная резонансная частота системы.

4. Номера 8 и 9 имеют графики коэффициентов усиления, такие, что более низкие частоты ослабляются, а более высокие частоты проходят. Эти системы действуют как *фильтры высоких частот* и включают в себя дифференцирование входного сигнала. Дальнейшее различие между категориями (2) и (4) состоит в том, что в (2) интегрирование входного сигнала приводит к отрицательным фазам $\varphi(f)$, т. е. выходной сигнал *запаздывает* по отношению к входному. С другой стороны, в (4) дифференцирование входного сигнала дает положительные фазы, так что выходной сигнал *опережает* входной, как это имеет место на графике номер 9.

Ширина полосы частот. Удобный способ описания функции усиления линейной системы можно получить, используя ее *ширину полосы частот* [5]. Были предложены различные определения ширины полосы частот; в простейшем из них для определения используется такая полоса, в которой мощность уменьшается до половины максимального значения. Для системы, имеющей максимальное усиление на частоте f_0 , ширина полосы частот определяется как разность $f_2 - f_1$, где f_1 и f_2 выбраны так, что

$$G^2(f_1) = G^2(f_2) = \frac{1}{2} G^2(f_0).$$

Например, для одиночной экспоненциальной системы максимальное усиление достигается при $f_0 = 0$, а усиление, равное половине максимального, — при $f_1 = 1/(2\pi T)$. Следовательно, если T велико,

*) В оригинале all-pass systems. Иногда их называют «фазовыми системами», поскольку они воздействуют лишь на фазу.— *Прим. перев.*

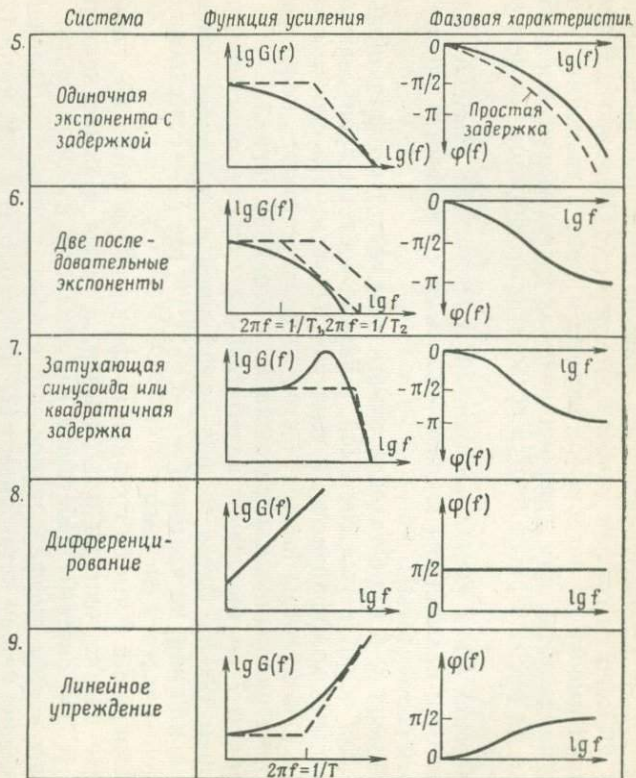
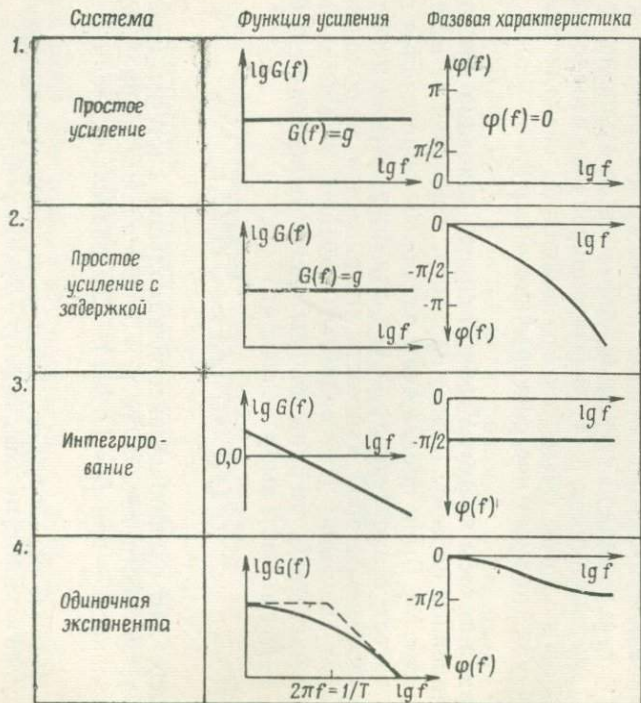


Рис. 2.8. Графики Бодэ для некоторых простых систем.

то ширина полосы частот очень мала, как можно увидеть на рис. 2.8. Таким образом, отклик на единичный импульс будет очень широким и небольшим по амплитуде. С другой стороны, для малых T ширина полосы частот велика и отклик на единичный импульс очень высокий и узкий. В пределе, когда $T \rightarrow 0$, ширина полосы частот становится бесконечной, как для простого усиления на рис. 2.8, и отклик на единичный импульс стремится к дельта-функции. Следовательно, широкие полосы частот соответствуют узким функциям отклика на единичный импульс, и наоборот, узкие полосы частот соответствуют широким функциям отклика на единичный импульс.

Устойчивость. Системы, приведенные в табл. 2.6, могут быть представлены дифференциальным уравнением следующего общего вида:

$$a_m \frac{d^m y(t)}{dt^m} + \dots + a_1 \frac{dy(t)}{dt} + a_0 y(t) = b_n \frac{d^n x(t-\tau)}{dt^n} + \dots + b_1 \frac{dx(t-\tau)}{dt} + b_0 x(t-\tau). \quad (2.3.18)$$

Подставляя в (2.3.18) $x(t) = e^{j2\pi ft}$, $y(t) = H(f)e^{j2\pi ft}$, получим, что частотная характеристика равна

$$H(f) = \frac{b_n (j2\pi f)^n + \dots + b_1 (j2\pi f) + b_0}{a_m (j2\pi f)^m + \dots + a_1 (j2\pi f) + a_0} e^{-j2\pi f\tau}. \quad (2.3.19)$$

Подставив в (2.3.19) $p = j2\pi f$ и приравняв знаменатель нулю, получим *характеристическое уравнение* системы, а именно

$$a_m p^m + \dots + a_1 p + a_0 = 0. \quad (2.3.20)$$

Можно показать [4], что условие устойчивости системы (2.3.11) эквивалентно условию, что все корни $\pi_1, \pi_2, \dots, \pi_m$ характеристического уравнения (2.3.20) имеют отрицательные действительные части.

2.3.4. Отклик на произвольный входной сигнал

Если известно, что отклик системы на входной сигнал $x(t) = e^{j2\pi ft}$ равен $y(t) = H(f)e^{j2\pi ft}$, то можно найти отклик на произвольный входной сигнал. Сначала надо взять преобразование Фурье от этого входного сигнала:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt. \quad (2.3.21)$$

Составляющая Фурье выходного сигнала на частоте f равна

$$Y(f) = \int_{-\infty}^{\infty} e^{-j2\pi ft} \left[\int_{-\infty}^{\infty} h(u)x(t-u) du \right] dt = \\ = \int_{-\infty}^{\infty} h(u) e^{-j2\pi fu} du \int_{-\infty}^{\infty} x(v) e^{-j2\pi fv} dv, \quad (2.3.22)$$

где $v = t - u$, т. е.

$$Y(f) = H(f)X(f). \quad (2.3.23)$$

Уравнение (2.3.23) показывает, что составляющая выходного сигнала на частоте f получается из составляющей входного сигнала на той же частоте с помощью умножения на $H(f)$ — значение частотной характеристики на этой же частоте. Наконец, чтобы возвратиться к $y(t)$, нужно синтезировать, или просуммировать, составляющие от всех частот при одном и том же значении t , что дает

$$y(t) = \int_{-\infty}^{\infty} Y(f) e^{j2\pi ft} df = \int_{-\infty}^{\infty} X(f) H(f) e^{j2\pi ft} df. \quad (2.3.24)$$

Равенства (2.3.22) — (2.3.24) показывают, что *свертка* во временной области эквивалентна *перемножению* в частотной области. Следовательно, если между двумя переменными существует соотношение в виде дифференциального уравнения (2.3.18), то решение равно (2.3.24), где частотная характеристика дается выражением (2.3.19). Следовательно, преобразование Фурье дает очень полезный операционный метод решения линейных дифференциальных уравнений.

Нахождение решения можно ускорить с помощью таблиц преобразований. Таблица преобразований обобщенных функций приведена в [1, 4*]; преобразования Фурье обычных функций имеются в [6, 5*].

Несколько линейных систем, соединенных последовательно. Рассмотрим k не влияющих друг на друга линейных систем, соединен-

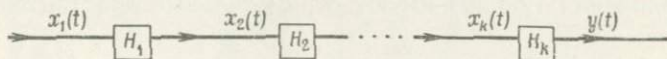


Рис. 2.9. Несколько линейных систем, соединенных последовательно.

ных последовательно, как показано на рис. 2.9. Повторное использование (2.3.23) дает

$$Y(f) = H_k(f)H_{k-1}(f) \dots H_1(f)X(f), \quad (2.3.25)$$

откуда видно, что для последовательно соединенных линейных систем полная частотная характеристика равна произведению частотных характеристик отдельных систем. Используя (2.3.17), полу-

чаем, что полный коэффициент усиления равен *произведению* отдельных коэффициентов усиления:

$$G(f) = G_1(f) G_2(f) \dots G_k(f), \quad (2.3.26)$$

а полный сдвиг фазы равен *сумме* отдельных фазовых сдвигов:

$$\varphi(f) = \varphi_1(f) + \varphi_2(f) + \dots + \varphi_k(f). \quad (2.3.27)$$

Выходной сигнал этой системы можно теперь вычислить, суммируя вклады от всех частот в виде

$$y(t) = \int_{-\infty}^{\infty} H_1(f) H_2(f) \dots H_k(f) X(f) e^{j2\pi ft} df. \quad (2.3.28)$$

Заметим, что при этом интегрирование проводится только один раз, в то время как выкладки во временной области потребовали бы вычисления k интегралов свертки.

2.3.5. Линейные уравнения в конечных разностях

В предыдущих разделах было показано, что систему, описываемую линейным дифференциальным уравнением, можно также описать с помощью функции отклика на единичный импульс $h(u)$ или же частотной характеристики $H(f)$, причем $h(u)$ и $H(f)$ образуют пару преобразований Фурье. Функции $h(u)$ и $H(f)$ легко получить из дифференциального уравнения, описывающего систему. В этом разделе показано, как можно использовать отклик на единичный импульс и частотную характеристику для описания системы, заданной с помощью линейного *разностного* уравнения.

Линейное разностное уравнение — это уравнение вида

$$y_r = \alpha_1 y_{r-1} + \alpha_2 y_{r-2} + \dots + \alpha_m y_{r-m} + \beta_0 x_r + \dots + \beta_n x_{r-n}. \quad (2.3.29)$$

Его общее решение имеет вид

$$y_r = \sum_{k=0}^{\infty} h_k x_{r-k}. \quad (2.3.30)$$

Величины $y_r, y_{r-1}, \dots, y_{r-m}$ и $x_r, x_{r-1}, \dots, x_{r-n}$ могли бы быть значениями непрерывных сигналов $y(t)$ и $x(t)$ в моменты времени $t = r\Delta, (r-1)\Delta, \dots, (r-m)\Delta, (r-n)\Delta$ соответственно, т. е.

$$y(t) = \alpha_1 y(t - \Delta) + \alpha_2 y(t - 2\Delta) + \dots + \alpha_m y(t - m\Delta) + \beta_0 x(t) + \dots + \beta_n x(t - n\Delta). \quad (2.3.31)$$

Преобразование Фурье от (2.3.31) можно записать в виде

$$Y(f) = \frac{\beta_0 + \beta_1 e^{-j2\pi f \Delta} + \dots + \beta_n e^{-j2\pi f n \Delta}}{1 - \alpha_1 e^{-j2\pi f \Delta} - \dots - \alpha_m e^{-j2\pi f m \Delta}},$$

так что частотная характеристика системы $H(f)$ равна, согласно (2.3.23),

$$H(f) = \frac{\beta_0 + \beta_1 e^{-j2\pi f \Delta} + \dots + \beta_n e^{-j2\pi f n \Delta}}{1 - \alpha_1 e^{-j2\pi f \Delta} - \dots - \alpha_m e^{-j2\pi f m \Delta}}. \quad (2.3.32)$$

Частотная характеристика $H(f)$ и дискретная функция отклика на единичный импульс h_k связаны соотношениями

$$H(f) = \sum_{k=0}^{\infty} h_k e^{-j2\pi f k \Delta}, \quad (2.3.33)$$

и

$$h_k = \Delta \int_{-1/(2\Delta)}^{1/(2\Delta)} H(f) e^{j2\pi f k \Delta} df. \quad (2.3.34)$$

z -преобразования. С частотной характеристикой (2.3.32) лучше всего обращаться, если сделать замену вида $z = e^{j2\pi f \Delta}$, что приводит к выражению

$$H(z) = \frac{\beta_0 + \beta_1 z^{-1} + \dots + \beta_n z^{-n}}{1 - \alpha_1 z^{-1} - \dots - \alpha_m z^{-m}} = \sum_{k=0}^{\infty} h_k z^{-k}. \quad (2.3.35)$$

Это выражение называется z -преобразованием [7] функции отклика на единичный импульс h_k .

С операционной точки зрения переменную z в (2.3.35) можно рассматривать как оператор сдвига, обладающий свойством

$$z^{-k} x_r = x_{r-k}. \quad (2.3.36)$$

Следовательно, разностное уравнение (2.3.39) можно записать в виде

$$\begin{aligned} (1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} - \dots - \alpha_m z^{-m}) y_r = \\ = (\beta_0 + \beta_1 z^{-1} + \dots + \beta_n z^{-n}) x_r, \end{aligned} \quad (2.3.37)$$

т. е.

$$y_r = \frac{(\beta_0 + \beta_1 z^{-1} + \dots + \beta_n z^{-n})}{(1 - \alpha_1 z^{-1} - \dots - \alpha_m z^{-m})} x_r = H(z) x_r,$$

где $H(z)$ является передаточной функцией дискретной системы. Разложение $H(z)$ по степеням z^{-1} дает

$$y_r = \sum_{k=0}^{\infty} h_k z^{-k} x_r = \sum_{k=0}^{\infty} h_k x_{r-k},$$

что является общим решением (2.3.30).

Устойчивость. Вынося множитель z^{-m} за скобки в знаменателе (2.3.35), заменяя z на p и приравнивая этот знаменатель нулю, по-

лучаем характеристическое уравнение дискретной системы

$$p^m - \alpha_1 p^{m-1} - \dots - \alpha_m = 0. \quad (2.3.38)$$

Условие устойчивости, соответствующее (2.3.11), будет иметь вид

$$\sum_{k=0}^{\infty} |h_k| < K_2. \quad (2.3.39)$$

Аналогично условие устойчивости, соответствующее (2.3.20), состоит в том, что корни π_1, \dots, π_m характеристического уравнения (2.3.38) должны лежать внутри единичного круга.

Пример. Рассмотрим разностное уравнение второго порядка:

$$y_r = \alpha_1 y_{r-1} + \alpha_2 y_{r-2} + x_r. \quad (2.3.40)$$

Оно имеет z -преобразование

$$(1 - \alpha_1 z^{-1} - \alpha_2 z^{-2}) y_r = x_r$$

и, следовательно, передаточную функцию

$$H(z) = \frac{1}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2}}. \quad (2.3.41)$$

Характеристическое уравнение имеет вид

$$p^2 - \alpha_1 p - \alpha_2 = 0,$$

а его корни равны

$$\pi_1 = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}, \quad \pi_2 = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2}. \quad (2.3.42)$$

Функция отклика на единичный импульс для этой системы имеет вид

$$h_k = \frac{1}{\pi_1 - \pi_2} \{ \pi_1^{k+1} - \pi_2^{k+1} \} \quad (2.3.43)$$

для действительных корней, т. е. когда $\alpha_1^2 \geq -4\alpha_2$. Когда корни комплексные, т. е. при $\alpha_1^2 < -4\alpha_2$,

$$h_k = R^k \frac{\sin 2\pi f_0 (k+1)}{\sin 2\pi f_0}, \quad (2.3.44)$$

где

$$\begin{aligned} \pi_1 &= R e^{j2\pi f_0}, \\ \pi_2 &= R e^{-j2\pi f_0}. \end{aligned}$$

Система устойчива при условии $|\pi_1| < 1$, $|\pi_2| < 1$, т. е. при условии, что α_1 и α_2 лежат внутри треугольной области:

$$\begin{aligned} \alpha_1 + \alpha_2 &< 1, \\ \alpha_1 - \alpha_2 &> -1, \\ -1 &< \alpha_2 < 1. \end{aligned} \quad (2.3.45)$$

2.4. ПРИМЕНЕНИЯ В АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

2.4.1. Записи конечной длины

На практике можно получать только записи конечной длины. Статистические вопросы, которые будут обсуждаться ниже, возникают из-за необходимости оценивать точность различных функций, получаемых из конечного объема данных. Даже, если $s(t)$ является детерминированной функцией, возникает смещение, или ошибка усечения, если $s(t)$ известна лишь на конечном интервале $-T/2 \leq t \leq T/2$. Чтобы понять влияние этого усечения, рассмотрим *временное окно*, определяемое с помощью соотношений

$$w(t) = \begin{cases} 1, & |t| \leq \frac{T}{2}, \\ 0, & |t| > \frac{T}{2}. \end{cases} \quad (2.4.1)$$

Если $s(t)$ является детерминированным сигналом в интервале $-\infty \leq t \leq \infty$, то сигнал, действительно измеренный на конечном интервале, можно записать в виде

$$s_T(t) = s(t) w(t). \quad (2.4.2)$$

Таким образом, операция взятия конечного участка записи равносильна умножению подлинного сигнала $s(t)$ на временное окно $w(t)$. Отсюда, используя (П2.1.8), получаем, что преобразование Фурье $S_T(f)$ сигнала $s_T(t)$ на конечном интервале является сверткой преобразований Фурье от $s(t)$ и $w(t)$:

$$S_T(f) = \int_{-\infty}^{\infty} S(g) W(f-g) dg, \quad (2.4.3)$$

где *спектральное окно* $W(f)$ является преобразованием Фурье от временного окна $w(t)$ и равно в этом случае

$$W(f) = T \frac{\sin \pi f T}{\pi f T}. \quad (2.4.4)$$

Совсем необязательно, чтобы временное окно имело в точности форму (2.4.1). Любое разумно выбранное окно $w(t)$ даст спектральное окно $W(f)$, сосредоточенное около нулевой частоты $f=0$, но с боковыми лепестками, или малыми всплесками, которые затухают при удалении f от нуля. Для небольших T преобразование $S_T(f)$ может дать очень искаженное изображение $S(f)$, так как окно $W(f-g)$ будет широким, а, следовательно, значения $S(g)$, отстоящие далеко от $g=f$, будут давать вклад в $S_T(f)$ согласно формуле (2.4.3). По мере того как T увеличивается, эти искажения будут уменьшаться. Наконец, когда T устремляется к бесконечности, со-

ставляющая преобразования Фурье на частоте f может быть определена полностью, так как временное окно будет стремиться к константе ($= 1$). Следовательно, если $T \rightarrow \infty$, то $W(f-g)$ стремится к дельта-функции, сосредоточенной в точке $g=f$, так что $S_T(f)$ стремится к $S(f)$.

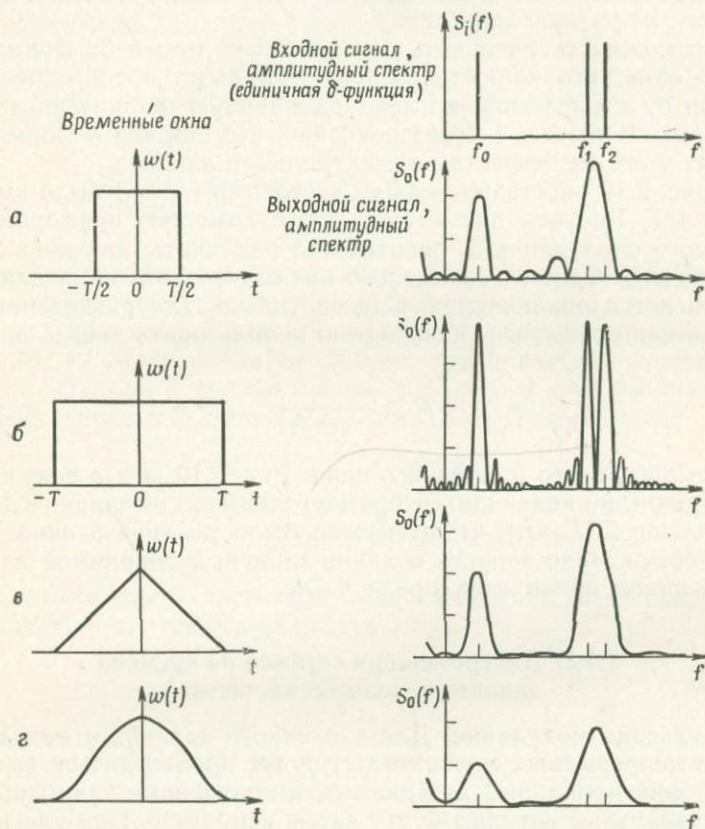


Рис. 2.10. Влияние формы и длительности временного окна на спектр сигнала.

Влияние формы и ширины окна на измеряемое преобразование Фурье проиллюстрировано на рис. 2.10 для одного частного входного сигнала $s(t)$, преобразование Фурье которого состоит из трех дельта-функций, сосредоточенных в f_0 , f_1 и f_2 . Отметим следующее.

1. Только два главных пика появляются в выходном преобразовании для окон *а*, *б*, и *г*, так как два входных пика на частотах f_1 и f_2 сливаются в один. Это происходит из-за использования слишком узкого временного окна.

2. Выходные преобразования для окон a и b имеют несколько ложных пиков между настоящими главными пиками. Эти ложные пики вызваны резкими углами временного окна.

3. Возможность различать пики (разрешающая способность) зависит от *ширины* временного окна, что иллюстрируется выходными преобразованиями для окон a и b , которые имеют одну и ту же форму, но разную ширину.

4. Возможность различать пики зависит также от *формы* временного окна, что иллюстрируется выходными преобразованиями для окон b , v и g , которые имеют одинаковую ширину, но различную форму. В гл. 6 и 7 будет показано, что ширина и форма окна приводят к тем же эффектам в спектральном анализе.

На рис. 2.10 расстояние между частотами ($f_2 - f_1$) было выбрано равным $1/T$. Рисунок показывает, что с помощью прямоугольного временного окна длины T невозможно различить два пика на частотах f_1 и f_2 . Однако, с помощью прямоугольного окна длины $2T$ эти пики легко различаются. Следовательно, для разделения двух пиков на частотах f_1 и f_2 необходимо использовать запись длины T порядка

$$T \geq \frac{1}{f_2 - f_1} \quad (2.4.5)$$

для прямоугольного временного окна. Рис. 2.10, v и g показывают, что для окон, не являющихся прямоугольными, их ширина должна быть больше $2/(f_2 - f_1)$, чтобы можно было различать пики. Дальнейшее обсуждение вопроса о длине записи, необходимой для различения пиков, приводится в разд. 6.4.4.

2.4.2. Дискретизация сигнала по времени и явление наложения частот

Импульсная модуляция. Для численного анализа отсчеты большинства непрерывных сигналов $s(t)$ будут производиться через некоторый фиксированный интервал Δ , и полученные таким образом *дискретизованные* сигналы будут затем использоваться для цифровых вычислений. Дискретизованный сигнал можно рассматривать как результат умножения первоначального непрерывного сигнала на сигнал $i(t)$, состоящий из *бесконечного ряда единичных импульсов*, или *дельта-функций*:

$$i(t) = \sum_{n=-\infty}^{\infty} \delta(t - n\Delta). \quad (2.4.6)$$

Это дает дискретизованный, или импульсно-модулированный, сигнал

$$s_i(t) = s(t) i(t). \quad (2.4.7)$$

Следовательно, воспользовавшись теоремой о свертке (П2.1.8), находим

$$S_i(f) = \int_{-\infty}^{\infty} S(f-g)I(g)dg, \quad (2.4.8)$$

где $I(g)$ является преобразованием Фурье от $i(t)$. Используя для $I(g)$ выражение (2.2.16), преобразуем (2.4.8):

$$\begin{aligned} S_i(f) &= \int_{-\infty}^{\infty} S(f-g) \frac{1}{\Delta} \sum_{n=-\infty}^{\infty} \delta\left(g - \frac{n}{\Delta}\right) dg = \\ &= \frac{1}{\Delta} \sum_{n=-\infty}^{\infty} S\left(f - \frac{n}{\Delta}\right). \end{aligned} \quad (2.4.9)$$

Равенство (2.4.9) показывает, что дискретизованный, или импульсно-модулированный, сигнал $s_i(t)$ имеет периодическое преобразование Фурье с периодом $1/\Delta$, и если $S(f)$ обращается в нуль при $|f| \geq 1/2\Delta$, то $S_i(f)$ является просто периодически повторяемой функцией $S(f)$, как показано на рис. 2.11, б и в. Это означает, что можно восстановить $S(f)$ по $S_i(f)$, умножив $S_i(f)$ на $H(f)$, где

$$H(f) = \begin{cases} \Delta, & |f| \leq \frac{1}{2\Delta}, \\ 0, & |f| > \frac{1}{2\Delta}. \end{cases} \quad (2.4.10)$$

Так как умножение в частотной области соответствует свертке во временной области, то отсюда следует, что

$$s(t) = \int_{-\infty}^{\infty} \frac{\sin(\pi u/\Delta)}{\pi u/\Delta} s_i(t-u) du. \quad (2.4.11)$$

Функция $\sin(\pi u/\Delta)/(\pi u/\Delta)$ является идеальным фильтром для восстановления непрерывного сигнала $s(t)$ из дискретизованного сигнала $s_i(t)$. Иначе говоря, функция $\sin(\pi u/\Delta)/(\pi u/\Delta)$ является идеальной интерполирующей функцией для равноотстоящих ординат, и формулу (2.4.11) иногда называют *интерполяционной формулой Уиттекера*.

Наложение частот. Если интервал отсчета таков, что $S(f)$ убывает до нуля, не доходя до $|f| = 1/2\Delta$, как в случаях б или в на рис. 2.11, то можно восстановить $s(t)$ по $s_i(t)$. С другой стороны, если $S(f)$ не равна нулю за частотой $f_N = 1/2\Delta$, то частотные компоненты от частот выше $1/2\Delta$ присутствуют в $S_i(f)$ в диапазоне частот $-1/2\Delta \leq f \leq 1/2\Delta$, как, например, в случае (г) на рис. 2.11. Частота $f_N = 1/2\Delta$ называется *частотой Найквиста* и является

наивысшей частотой, которую можно обнаружить на данных, полученных с интервалом отсчета Δ .

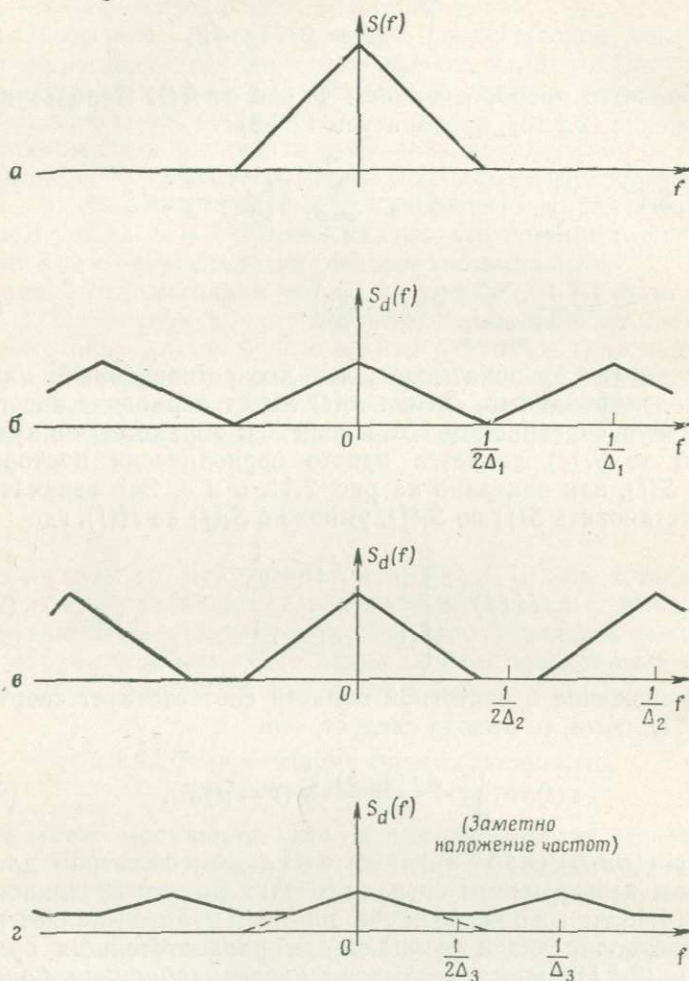


Рис. 2.11. Преобразования Фурье входного сигнала и дискретизованных сигналов для различных интервалов отсчета.

Если, например, $\Delta = 0,1$ сек, то частота Найквиста равна 5 гц. Преобразование Фурье $S_i(f)$ дискретизованного сигнала на 4 гц будет состоять из вкладов преобразования $S(f)$ на 4 гц, на $10+4=14$ гц, на $-10+4=-6$ гц, на $20+4=24$ гц, на $-20+4=-16$ гц и т. д. Все эти частоты, кроме первой, называются обычно *двойниками* (aliases) частоты 4 гц, а их влияние на преобразование Фу-

рпе — явлением наложения частот (aliasing). Следовательно, при дискретизации по времени непрерывных временных рядов нужно надлежащим образом позаботиться о выборе достаточно высокой частоты отсчетов $f_N = 1/2\Delta$, чтобы избежать искажающего влияния наложения частот на $S_i(f)$.

Явление наложения частот возникает в ряде практических ситуаций, например при использовании стробоскопа или в кинофильмах. Так, если в фильме колеса телеги приходят в движение, то вначале видно, что они вращаются в направлении движения, затем при возрастании скорости кажется, что направление вращения меняется на обратное и скорость колес уменьшается до полной остановки, затем они начинают вращаться с возрастающей скоростью в направлении движения и т. д.

Пример. Чтобы проиллюстрировать обсуждаемые в этом разделе вопросы, предположим, что желательны вычислить длину записи T и интервал отсчета Δ , необходимые для достижения некоторых целей. Предположим, известно, что изучаемый сигнал содержит две синусоидальные компоненты на частотах 100 и 99 гц. Тогда, если мы хотим различить эти пики в преобразовании Фурье, взятом от конечной записи, нам нужно, как показывает (2.4.5), взять $1/T$ порядка $100 - 99 = 1$ гц, т. е. T должно быть порядка 1 сек. Чтобы оценить частоты порядка 100 гц, величина $1/2\Delta$ должна быть по меньшей мере 100 гц и, следовательно, $\Delta \leq 5$ мсек. Таким образом, нужно взять по крайней мере 200 точек.

Если бы нам захотелось различить две частотные компоненты на 999 и 1000 гц, необходимая длительность записи была бы все еще 1 сек, однако интервал отсчета в этом случае нужно было бы взять 0,5 мсек, так что потребовалось бы 2000 точек.

Следовательно, длина записи T определяет степень различимости пиков в преобразовании Фурье, а интервал отсчета Δ определяет максимальную частоту, которую можно различить.

ЛИТЕРАТУРА

1. Lighthill M. J., An Introduction to Fourier Analysis Generalized Functions, Cambridge Univ. Press, Cambridge, 1959.
2. Papoulis A., The Fourier Integral and its Applications, McGraw-Hill, New York, 1962.
3. Courant R., Differential and Integral Calculus, Vol. II, Blackie and Son, London, 1952. (Русский перевод: Курант Р., Курс дифференциального и интегрального исчисления, М.—Л., ОНТИ, 1934.)
4. James H. M., Nichols N. B., Phillips R. S., Theory of Servomechanisms, McGraw-Hill, New York, 1947. (Русский перевод: Джеймс Х., Николс Н., Филлипс Р., Теория следящих систем, М., ИЛ, 1953.)
5. Bode H. W., Network Analysis and Feedback Amplifier Design, Van Nostrand, New York, 1945.
6. Campbell G. A., Foster R. M., Fourier Integrals for Practical Applications, Van Nostrand, New York, 1962.
7. Jury E. I., Theory and Applications of the z-Transform Method, John Wiley, New York, 1964.

ОПЕРАТОРНЫЕ СВОЙСТВА ПРЕОБРАЗОВАНИЙ ФУРЬЕ

На протяжении всей этой книги нам потребуется выполнять различные операции с преобразованиями Фурье. Ниже приводится их сводка.

Изменение масштаба времени и сдвиг начала координат. Если $s(t)$ имеет преобразование Фурье $S(f)$, то преобразование Фурье от $s(\alpha t + \beta)$ равно

$$\frac{1}{|\alpha|} e^{j(2\pi f \beta / \alpha)} S\left(\frac{f}{\alpha}\right). \quad (\text{П2.1.1})$$

Пример. Из табл. 2.5 мы видим, что преобразование Фурье от $e^{-\pi t^2}$ равно

$$S(f) = e^{-\pi f^2}.$$

Следовательно, преобразование Фурье от

$$\exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] = s\left(\frac{t-\mu}{\sqrt{2\pi}\sigma}\right)$$

равно

$$|\sqrt{2\pi}\sigma| e^{j2\pi f(-\mu)} S(\sqrt{2\pi}\sigma f) = \sqrt{2\pi}\sigma e^{-j2\pi f\mu} e^{-2\pi^2\sigma^2 f^2},$$

где

$$\alpha = \frac{1}{\sqrt{2\pi}\sigma}, \quad \beta = -\frac{\mu}{\sqrt{2\pi}\sigma}, \quad \frac{\beta}{\alpha} = -\mu.$$

Дифференцирование. Если $s(t)$ имеет преобразование Фурье $S(f)$, то m -я производная $s^{(m)}(t)$ имеет преобразование Фурье

$$(j2\pi f)^m S(f) \quad (\text{П2.1.2})$$

при условии, что эта производная существует.

Пример. Как и в предыдущем примере, используем пару преобразований из табл. 2.5:

$$s(t) = e^{-\pi t^2}, \quad S(f) = e^{-\pi f^2}.$$

Получаем, что преобразование Фурье от

$$\frac{d}{dt}s(t) = \frac{d}{dt}(e^{-\pi t^2}) = -2\pi t e^{-\pi t^2}$$

равно

$$j2\pi f S(f) = j2\pi f e^{-\pi f^2}.$$

Интегрирование. Если $s(t)$ имеет преобразование Фурье $S(f)$, то преобразование Фурье от $I^m s(t)$, где

$$I s(t) = \int_{-\infty}^t s(u) du,$$

равно

$$\left(\frac{1}{j2\pi f}\right)^m S(f) + K_1 \delta(f) + K_2 \delta'(f) + \dots + K_m \delta^{m-1}(f). \quad (\text{П2.1.3})$$

Константы K_1, K_2, \dots, K_m в (П2.1.3) можно определить, используя значения функций $s(t), ds/dt, \dots, d^m s/dt^m$ в нуле, например

$$s(0) = \int_{-\infty}^{\infty} S(f) df.$$

Пример. Функция предыдущего примера

$$s(t) = -2\pi t e^{-\pi t^2}$$

имеет преобразование Фурье

$$S(f) = j2\pi f e^{-\pi f^2}.$$

Следовательно, преобразование Фурье от

$$s_1(t) = \int_{-\infty}^t s(u) du = e^{-\pi t^2}$$

равно

$$S_1(f) = \frac{1}{j2\pi f} S(f) + K_1 \delta(f) = e^{-\pi f^2} + K_1 \delta(f).$$

Интегрирование обеих частей по f дает

$$s_1(0) = \int_{-\infty}^{\infty} e^{-\pi f^2} df + K_1 = 1 + K_1.$$

Но $s_1(0) = 1$ и, следовательно, $K_1 = 0$.

Симметрия. Если $S(f)$ есть преобразование Фурье от $s(t)$, то $s(f)$ есть преобразование Фурье от $S(-t)$.

Пример. Преобразование Фурье от функции

$$s(t) = \begin{cases} e^{-t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

равно $S(f) = 1/(1 + j2\pi f)$. Следовательно, преобразование Фурье от $s(t) = 1/(1 - j2\pi t)$ равно

$$S(f) = \begin{cases} e^{-f}, & f \geq 0, \\ 0, & f < 0. \end{cases}$$

Аналогично преобразование Фурье от $s(t) = 1/(1 + j2\pi t)$ равно

$$S(f) = \begin{cases} e^f, & f \leq 0, \\ 0, & f > 0. \end{cases}$$

Следовательно, преобразование Фурье от

$$\frac{2}{1 + (2\pi t)^2} = \frac{1}{1 - j2\pi t} + \frac{1}{1 + j2\pi t}$$

равно

$$e^{-|f|}, \quad -\infty \leq f \leq \infty.$$

Свертки и теорема Парсеваля. Мы приведем эту теорему в более общем виде, чем результаты (2.1.16), (2.1.20), (2.1.26), выведенные в разд. 2.1. Обобщение утверждает, что если $s_1(t)$ и $s_2(t)$ — два комплексных сигнала с преобразованиями Фурье $S_1(f)$ и $S_2(f)$ соответственно, то

$$\int_{-\infty}^{\infty} s_1(t) s_2^*(t) dt = \int_{-\infty}^{\infty} S_1(f) S_2^*(f) df, \quad (\text{П2.1.4})$$

где звездочка означает комплексное сопряжение.

Иногда бывают полезны три специальных случая формулы (П2.1.4):

а) Если $s_2^*(t) = h(u - t)$, то (П2.1.4) сводится к

$$\int_{-\infty}^{\infty} s_1(t) h(u - t) dt = \int_{-\infty}^{\infty} S_1(f) H(f) e^{j2\pi fu} df. \quad (\text{П2.1.5})$$

б) Если $s_1(t)$ и $s_2(t)$ действительны, то (П2.1.4) сводится к

$$\int_{-\infty}^{\infty} s_1(t) s_2(t) dt = \int_{-\infty}^{\infty} S_1(f) S_2(-f) df. \quad (\text{П2.1.6})$$

в) Если $s_1(t) = s_2(t) = s(t)$, то (П2.1.4) сводится к

$$\int_{-\infty}^{\infty} |s(t)|^2 dt = \int_{-\infty}^{\infty} |S(f)|^2 df. \quad (\text{П2.1.7})$$

Теорема Парсеваля в форме (П2.1.7) включает в себя эту же теорему в форме (2.1.26), выведенную в разд. 2.1.

Заметим, что из-за симметрии преобразования Фурье сигнал и его преобразование можно поменять ролями. Например,

$$\text{а) } \int_{-\infty}^{\infty} S_1(f) S_2(g-f) df = \int_{-\infty}^{\infty} s_1(t) s_2(t) e^{j2\pi gt} dt, \quad (\text{П2.1.8})$$

$$\text{б) } \int_{-\infty}^{\infty} S_1(f) S_2(f) df = \int_{-\infty}^{\infty} s_1(t) s_2(-t) dt, \quad (\text{П2.1.9})$$

а симметрия соотношения (в) видна непосредственно. Следует отметить, что упомянутые выше операторные свойства применимы точно так же к конечным и бесконечным рядам Фурье. Три формы теоремы Парсеваля, выведенные в разд. 2.1, служат примером этому.

Эта глава содержит краткое описание тех понятий теории вероятностей, которые необходимы для понимания задач с временными рядами. Разд. 3.1 иллюстрирует подход, с помощью которого статистик описывает физические явления, пользуясь выборочным пространством, случайной величиной и распределением вероятностей. В разд. 3.2 рассматриваются способы приближения распределения вероятностей с помощью его первых моментов. Наконец, в разд. 3.3 обсуждаются выборочные распределения некоторых полезных функций от случайных величин, таких как среднее значение и дисперсия.

3.1. ЧАСТОТНЫЕ РАСПРЕДЕЛЕНИЯ И РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

В гл. 1 было показано, что детерминистические модели не всегда могут адекватно описывать физические системы. Поэтому, когда системе свойственна *неопределенность* или она подвержена *случайному изменению*, необходимо использовать недетерминистические или случайные модели. Математическая теория, лежащая в основе таких случайных моделей, называется теорией вероятностей.

3.1.1. Дискретные случайные величины и распределения

В качестве примера физического процесса, которому свойственна неопределенность или случайная изменчивость, рассмотрим данные, приведенные на рис. 3.1. Они показывают флуктуации числа дефектных транзисторов в последовательных выборках объема 100, взятых случайным образом с выхода поточной линии. Такой выборочный контроль необходим для поддержания качества продукции, а график числа дефектных изделий x в зависимости от номера выборки называется диаграммой контроля качества.

Диаграммы контроля дают *наглядную* картину изменения данных и используются для получения заблаговременных предупреждений о том, что произойдут изменения качества. *Количественное* утверждение об изменчивости можно получить, построив *частотное*

распределение, как показано в табл. 3.1 и на рис. 3.2. Эти иллюстрации изображают n_x — число выборок с x дефектными изделиями — как функцию от x для пятидесяти выборок, приведенных

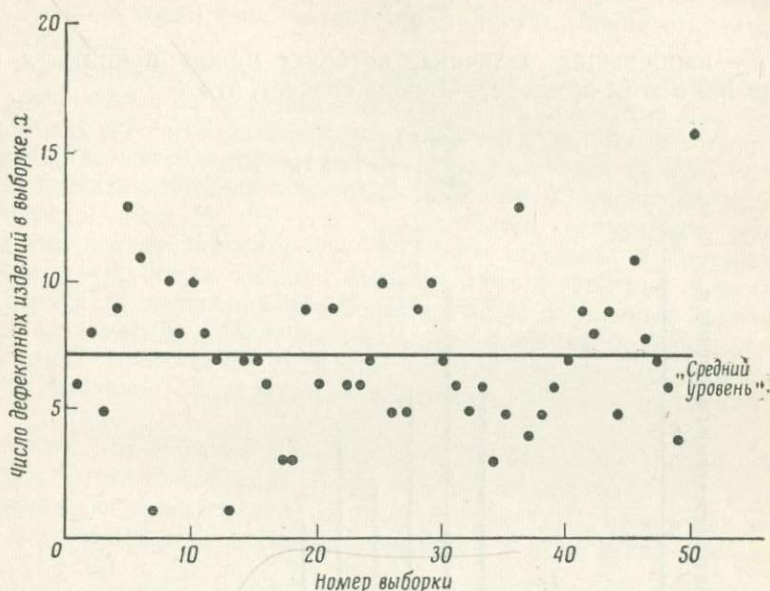


Рис. 3.1 Число дефектных транзисторов в 50 выборках объема 100.

Таблица 3.1

Частотное распределение числа дефектных транзисторов в выборке (по пятидесяти выборкам объема 100)

x	0	1	2	3	4	5	6	7	8
n_x	0	2	0	3	2	7	9	7	5
x	9	10	11	12	13	14	15	16	Полное количество выборок
n_x	6	4	2	0	2	0	0	1	50

на рис. 3.1. Частотное распределение показывает, что в то время, как число дефектных образцов в выборке изменяется от 1 до 16, большинство выборок (90%) имеет от 3 до 11 дефектных изделий.

Итак, полное число проверенных выборок равно

$$\sum_{x=0}^k n_x = N, \quad (3.1.1)$$

где k — наибольшая величина, которую может принять x (она равна 100 в этом примере). Отсюда следует, что

$$\sum_{x=0}^k \left(\frac{n_x}{N} \right) = 1, \quad (3.1.2)$$

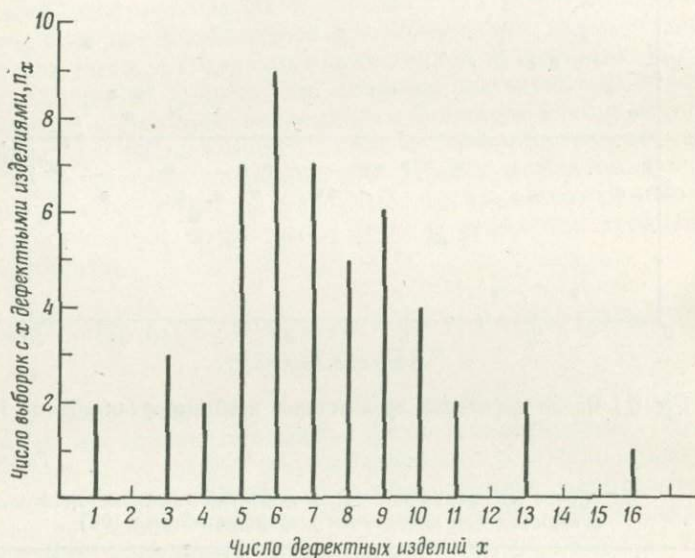


Рис. 3.2. Частотное распределение для данных рис. 3.1.

где n_x/N определяет долю выборок с x дефектными изделиями. Например, из рис. 3.2 видно, что 5 из 50, или одна десятая часть выборок, имеют равно 8 дефектных изделий.

Выборочные пространства, события, случайные величины и распределения вероятностей. Данные контроля качества можно описать, введя четыре основных понятия. Первым из них является *выборочное пространство*, которое представляет собой множество точек, соответствующих всем возможным исходам эксперимента. Например, при проверке 100 транзисторов выборочное пространство состоит из 101 точки P_0, P_1, \dots, P_{100} , которые соответствуют 0, 1, 2, ..., 100 дефектным изделиям.

Некоторая совокупность или подмножество точек выборочного пространства называется *событием*. Например, выборочные точки P_0, P_1 соответствуют событию «число дефектных изделий меньше двух». Каждая точка выборочного пространства соответствует *простому* событию.

Для того чтобы обращаться к различным событиям в выборочном пространстве, необходимо ввести понятие *случайной величины*. Например, точки выборочного пространства для данных о транзисторах можно обозначить по-другому, так, что точки P_0 и P_1 будут соответствовать событию «случайная величина Y принимает значение $y=0$ », а точки P_2, P_3, \dots, P_{100} — событию «случайная величина Y принимает значение $y=1$ ». Таким образом, Y принимает значение $y=0$, когда имеется меньше двух дефектных изделий, и $y=1$, когда имеется два или большее число дефектных изделий. Случайная величина обозначается обычно большой буквой, например X или Y , а численное значение, которое она принимает в конкретной выборке, обозначается маленькой буквой, например x или y .

Заметим, что события в выборочном пространстве можно обозначать многими способами. Например, некоторая случайная величина могла бы быть связана с числом дефектных изделий в выборке. В этом примере случайная величина X принимает значения $x=0, 1, \dots, 100$.

В общем случайная величина является функцией, которую можно использовать для обозначения множеств или событий в выборочном пространстве.

Основными понятиями, необходимыми для описания примера с контролем качества, являются вероятность и *распределение вероятностей*. Вероятность равна отношению числа событий, в которых случайная величина X принимает значение x , к общему числу событий; она записывается $p_X(x)$. Множество чисел $p_X(x)$, $x=0, 1, 2, \dots, 100$, является распределением вероятностей. Каждая из вероятностей является неотрицательной величиной, и их сумма равна единице. Оценку $p_X(x)$ можно получить из наблюдаемых отношений n_x/N , определенных в (3.1.2). При увеличении полного числа проверяемых транзисторов N отношения n_x/N дают все лучшие и лучшие оценки вероятностей $p_X(x)$.

Иногда можно вывести математическую формулу для $p_X(x)$, сделав разумные физические предположения. Например, подходящим распределением вероятностей для описания задачи с транзисторами является *биномиальное* распределение

$$p_X(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n, \quad (3.1.3)$$

где n — объем выборки и θ — вероятность того, что транзистор является дефектным.

Параметр θ можно оценить по наблюдаемым данным с помощью следующего соотношения:

$$\hat{\theta} = \frac{\text{Число дефектных транзисторов}}{\text{Полное число проверенных транзисторов}} = \frac{355}{5000} = 0,071.$$

Используя $\hat{\theta}$ вместо истинной величины θ , можно оценить вероятность того, что случайная величина X примет значение x , по формуле

$$\hat{p}_x(x) = \binom{100}{x} (0,071)^x (0,929)^{100-x}, \quad x=0, 1, \dots, 100. \quad (3.1.4)$$

Следовательно, в группе из 50 выборок, каждая из которых имеет объем 100, предсказываемое число выборок с x дефектными изделиями равно $\tilde{n}_x = 50\hat{p}_x(x)$.

Таблица 3.2

Сравнение наблюдаемых частот с ожидаемыми частотами, вычисленными по биномиальному распределению, подобранному к данным о транзисторах

x	0	1	2	3	4	5	6	7	8
n_x	0	2	0	3	2	7	9	7	5
\tilde{n}_x	<0,1	0,3	0,9	2,3	4,2	6,2	7,5	7,6	6,8
x		9	10	11	12	13	14	15	16
n_x		6	4	2	0	2	0	0	1
\tilde{n}_x		5,3	3,7	2,3	1,3	0,6	0,2	0,1	<0,1

В табл. 3.2 наблюдаемые частоты n_x сравниваются с ожидаемыми частотами \tilde{n}_x в предположении, что модель (3.1.4) верна. Мы видим, что наблюдается хорошее согласие и, следовательно, (3.1.4) является адекватной вероятностной моделью для этой ситуации.

Вопрос о том, какую из вероятностных моделей использовать в конкретной задаче, является важным, и для получения ответа на него нужно использовать все имеющиеся в распоряжении данные

и относящуюся к сути явления информацию. Ответ не может быть продиктован математикой, но должен быть получен в результате тщательного анализа физической ситуации.

3.1.2. Непрерывные случайные величины и распределения

Во многих случаях нужно описывать ситуацию с помощью *непрерывной* случайной величины, т. е. случайной величины, определенной на выборочном пространстве, которое является непрерывным. Например, рис. 3.3 показывает частотное распределение тока

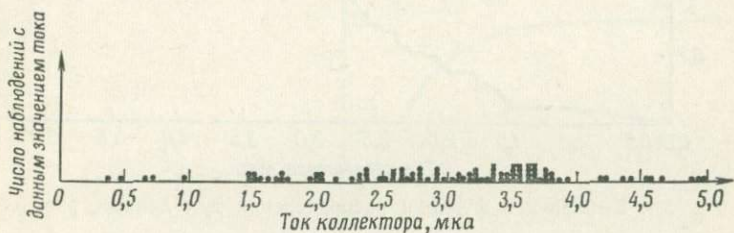


Рис. 3.3. Точечная диаграмма для токов коллектора.

коллектора для выборки, состоящей из $N = 100$ транзисторов. Поскольку величина тока может принимать любые неотрицательные значения, нужно ввести случайную величину X , принимающую значение x из непрерывного выборочного пространства $0 \leq x < \infty$.

Рис. 3.3 показывает, что иногда сразу четыре транзистора имеют одну и ту же величину тока. Однако если воспользоваться более чувствительным амперметром, то может случиться так, что никакие две точки на оси силы тока не совпадут и, таким образом, бессмысленно строить распределение частот. Следовательно, нельзя говорить о вероятности осуществления конкретного значения непрерывной случайной величины X , скажем $x = 2,000$ мкА.

Функция распределения. Хотя и бессмысленно рассматривать вероятность того, что некоторая непрерывная случайная величина X принимает конкретное значение x , тем не менее можно определить вероятность того, что X будет меньше некоторой величины x , т. е. $\text{Pr}\{X \leq x\}$. Эта вероятность записывается $F_X(x)$ и называется *функцией распределения*. Типичная форма такой функции показана на рис. 3.5, где видно, что она стремится к значению 1, поскольку $F_X(\infty) = 1$.

Функцию распределения можно оценить с помощью доли значений выборки, не превосходящих данной величины x . *Выборочная функция распределения* для данных, приведенных на рис. 3.3,

показана на рис. 3.4. Она состоит из ряда скачков высоты n_x/N , расположенных над значениями, из которых состоит выборка.

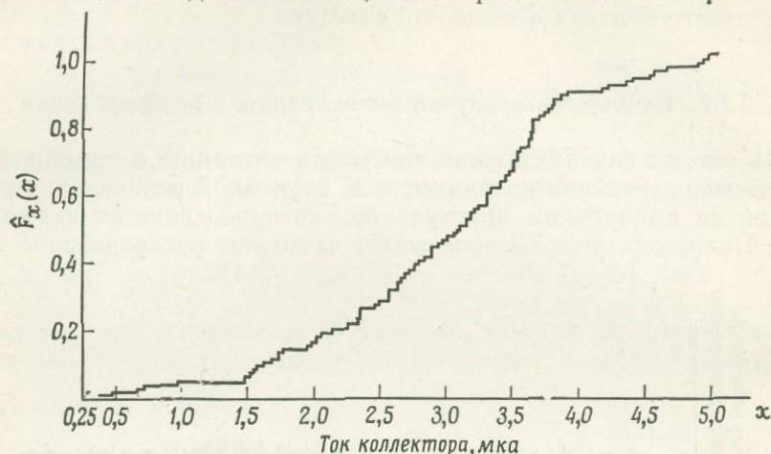


Рис. 3.4. Выборочная функция распределения для данных рис. 3.3.

Плотность вероятности. С функцией распределения $F_X(x)$ связана плотность вероятности $f_X(x)$. Она задается соотношением

$$f_X(x) = \frac{dF_X(x)}{dx}, \quad (3.1.5)$$

при условии, что функция распределения достаточно гладкая, так что ее производная существует. Этого не будет, если случайная величина является дискретной, так как функция распределения в этом случае имеет скачки, или разрывы, в точках, соответствующих дискретным значениям X .

Плотность вероятности *не* является *распределением* вероятностей, но ее можно использовать для вычисления вероятностей. Так, интегрируя (3.1.5), получаем вероятность того, что случайная величина X меньше x :

$$\Pr \{ X \leq x_1 \} = F_X(x_1) = \int_{-\infty}^{x_1} f_X(x) dx, \quad (3.1.6)$$

а вероятность, что X лежит в интервале от x_1 до x_2 , равна

$$\Pr \{ x_1 < X \leq x_2 \} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x) dx. \quad (3.1.7)$$

По определению $f_X(x)$ имеет следующие свойства:

$$f_X(x) \geq 0 \text{ для всех } x; \quad \int_{-\infty}^{\infty} f_X(x) dx = 1. \quad (3.1.8)$$

Нормальная плотность вероятности. Одной из наиболее важных плотностей вероятности в статистике является *нормальная*, или *гауссовская*, плотность вероятности

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad -\infty \leq x \leq \infty, \quad (3.1.9)$$

показанная на рис. 3.5 вместе с ее функцией распределения. Нормальная плотность вероятности полностью задается двумя *параметрами* μ и σ^2 и будет обозначаться $N(\mu, \sigma^2)$. Она может быть использована для описания многих практических ситуаций, например

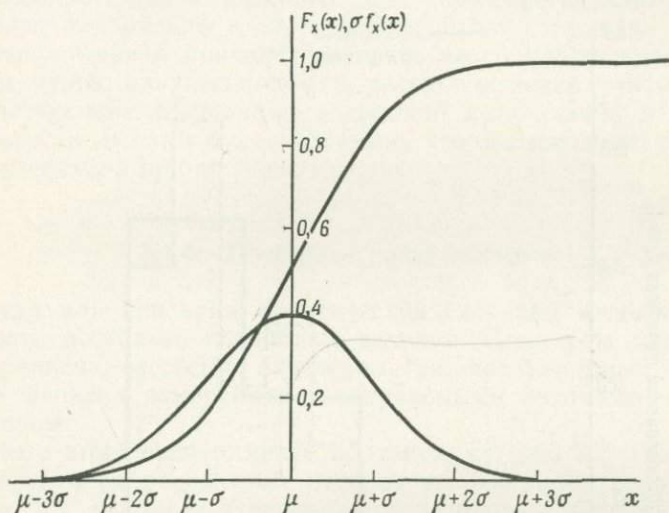


Рис. 3.5. Нормальная плотность вероятности и функция распределения.

для характеристики диаметра обрабатываемых на станке деталей или срока службы электрических ламп. Этот факт можно объяснить с помощью *центральной предельной теоремы*, которая утверждает, что плотность вероятности суммы n случайных величин $X = X_1 + X_2 + \dots + X_n$ сходится *очень быстро* к нормальной при увеличении n независимо от того, каковы плотности вероятности отдельных X_i . Таким образом, если окончательное измерение x является результатом многих мелких эффектов, действующих аддитивно, то следует ожидать, что нормальная плотность вероятности будет хорошо описывать ситуацию. Во многих других ситуациях может существовать некоторая подходящая функция $g(X)$ от случайной величины X , имеющая приближенно нормальное распределение. Например, плотность вероятности логарифма емкости конденсаторов на некоторой поточной линии хорошо описывается

нормальной плотностью. Как и в дискретном случае, вопрос о применимости конкретной плотности вероятности может быть решен только после тщательного анализа данных и относящейся к сути явления информации.

3.1.3. Оценка плотностей вероятности

Один из способов оценки плотности вероятности состоит в построении *гистограммы*. Она показывает долю $p(n)$ наблюдений,

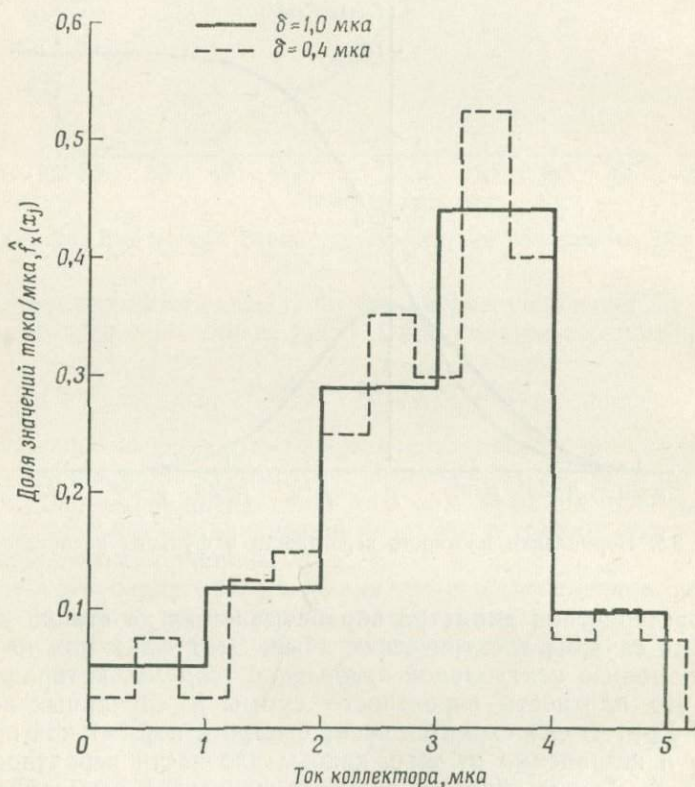


Рис. 3.6. Гистограммы для данных о коллекторном токе.

лежащих в интервале от $(n - \frac{1}{2})\delta$ до $(n + \frac{1}{2})\delta$. Так как

$$\text{Pr} \left\{ \left(n - \frac{1}{2} \right) \delta < X \leq \left(n + \frac{1}{2} \right) \delta \right\} \approx \delta f_X(n\delta), \quad n = 0, \pm 1, \pm 2, \dots,$$

то выборочной оценкой плотности вероятности является функция

$$\hat{f}_x(n\delta) = \frac{p(n)}{\delta}, \quad n=0, \pm 1, \dots$$

состоящая из прямоугольников ширины δ . Рис. 3.6 показывает гистограмму данных о токе, приведенных на рис. 3.3, для двух значений ширины интервала δ , а именно 0,4 мка и 1,0 мка. Выборочная оценка, использующая широкий интервал, является сравнительно плавной и скрывает большую часть тонкой структуры данных. Наоборот, узкий интервал дает более детальную картину, но выборочная оценка в этом случае более изменчива, так как в каждый интервал попадает меньшее число наблюдений. Таким образом, нужно принимать компромиссное решение, учитывая противодействующие требования подробной детализации и большой изменчивости. В гл. 6 будет показано, что аналогичные рассуждения применимы и при оценке спектров.

3.1.4. Двумерные распределения

Иногда для описания практической ситуации необходимо использовать несколько случайных величин. Примером может служить сравнение отсчетов акселерометра, производимых пилотом, с более точными измерениями, получаемыми автоматическим регистратором.

Данные этого эксперимента показаны на рис. 3.7, где на график нанесены одновременные отсчеты пилота (x_1) и регистратора (x_2). Рис. 3.7. называется *диаграммой разброса*; она может быть использована для построения двумерной гистограммы с помощью подсчета числа точек в прямоугольниках на плоскости (x_1, x_2).

Данные, приведенные на рис. 3.7, можно описать с помощью двух случайных величин X_1 и X_2 , где X_1 относится к отсчетам пилота, а X_2 — регистратора. Выборочное пространство для этого примера представляет собой область $x_1 \geq 0, x_2 \geq 0$, но в общем случае оно может быть и целой плоскостью (x_1, x_2). С этим общим выборочным пространством можно связать *двумерную функцию распределения*

$$F_{12}(x_1, x_2) = \text{Pr} \{ X_1 \leq x_1, X_2 \leq x_2 \}. \quad (3.1.10)$$

Как и в одномерном случае, если функция распределения является достаточно гладкой, ее можно продифференцировать, в результате чего получится двумерная плотность вероятности

$$f_{12}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{12}(x_1, x_2). \quad (3.1.11)$$

Следовательно, функцию распределения можно выразить через плотность вероятности с помощью

$$F_{12}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{12}(t_1, t_2) dt_1 dt_2. \quad (3.1.12)$$

Плотность вероятности (3.1.11) можно оценить по двумерной гистограмме точно так же, как была оценена плотность вероятности $f_X(x)$ по одномерной гистограмме.

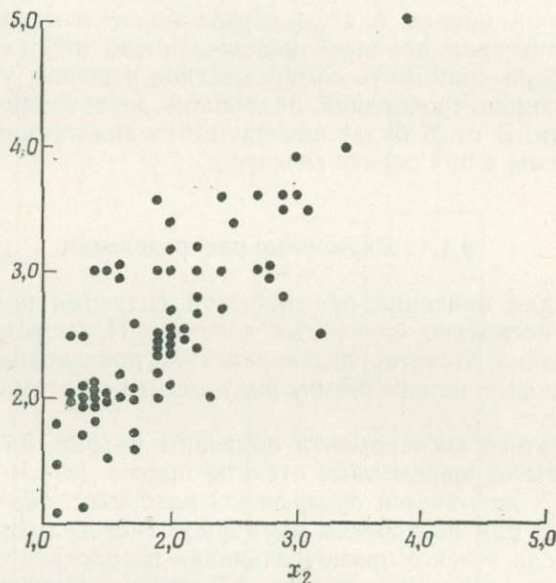


Рис. 3.7. Диаграмма разброса для измерений ускорения (в единицах g).

Для дискретных случайных величин совместная плотность вероятности записывается $p_{12}(x_1, x_2)$ и представляет собой вероятность того, что случайная величина X_1 примет значение x_1 , а X_2 — значение x_2 .

Условные распределения и независимость. Рассмотрим для двух дискретных случайных величин функцию, определяемую как долю случаев, в которых X_1 принимает значение x_1 , при условии, что X_2 зафиксировано на некотором значении x_2 . Эта функция называется *условным распределением вероятностей* X_1 при заданном X_2 и обозначается $p_{1|2}(x_1, x_2)$. Аналогично $p_{2|1}(x_1, x_2)$ обозначает условное распределение вероятностей X_2 при заданном X_1 . Сов-

местное распределение вероятностей X_1 и X_2 можно при этом записать в виде

$$p_{12}(x_1, x_2) = p_1(x_1) p_{2|1}(x_1, x_2) = p_2(x_2) p_{1|2}(x_1, x_2), \quad (3.1.13)$$

где, например, $p_1(x_1)$ — безусловное (маргинальное) распределение X_1 . Безусловное распределение вероятностей $p_1(x_1)$ можно получить из совместного распределения вероятностей с помощью

$$p_1(x_1) = \sum_{x_2=0}^{\infty} p_{12}(x_1, x_2).$$

Оно дает долю случаев, в которых X_1 равно x_1 вне зависимости от того, каково значение X_2 .

Если вероятность того, что случайная величина X_1 принимает значение x_1 не зависит от того, что случайная величина X_2 принимает значение x_2 , то условное распределение вероятностей $p_{1|2}(x_1, x_2) = p_1(x_1)$ и условное распределение вероятностей $p_{2|1}(x_1, x_2) = p_2(x_2)$. В этом случае говорят, что случайные величины X_1 и X_2 независимы, а выражение (3.1.13) для совместного распределения вероятностей разлагается на множители в виде

$$p_{12}(x_1, x_2) = p_1(x_1) p_2(x_2). \quad (3.1.14)$$

Аналогично для непрерывных случайных величин совместная плотность вероятности разлагается на множители вида

$$f_{12}(x_1, x_2) = f_1(x_1) f_{2|1}(x_1, x_2) = f_2(x_2) f_{1|2}(x_1, x_2) \quad (3.1.15)$$

в случае, если случайные величины зависимы, и на множители вида

$$f_{12}(x_1, x_2) = f_1(x_1) f_2(x_2), \quad (3.1.16)$$

если случайные величины независимы.

Двумерная нормальная плотность вероятности. Так же как нормальная плотность вероятности играет главную роль при описании одиночных случайных величин, двумерная нормальная плотность вероятности

$$f_{12}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho_{12}^2)^{1/2}} \exp \left\{ \frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right] \right\},$$

$$-\infty \leq x_1, x_2 \leq \infty, \quad (3.1.17)$$

играет столь же важную роль среди двумерных плотностей вероятности. Двумерная нормальная плотность вероятности зависит от

пяти параметров: $\mu_1, \mu_2, \sigma_1, \sigma_2$ и ρ_{12} . Если $\rho_{12} = 0$, то (3.1.17) распадается на произведение двух нормальных плотностей вероятности; это говорит о том, что в случае $\rho_{12} = 0$ случайные величины X_1 и X_2 независимы. Параметр ρ_{12} называется *коэффициентом корреляции*; он измеряет степень линейной зависимости между двумя случайными величинами.

3.1.5. Многомерные распределения

Когда измеряются одновременно n количеств, ситуацию можно описать с помощью n случайных величин с заданной n -мерной совместной функцией распределения

$$F_{12 \dots n}(x_1, x_2, \dots, x_n)$$

и плотностью вероятности

$$f_{12 \dots n}(x_1, x_2, \dots, x_n).$$

Если случайные величины взаимно независимы, то совместная плотность вероятности распадается на множители

$$f_{12 \dots n}(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n). \quad (3.1.18)$$

Важным частным случаем многомерной плотности вероятности является *многомерная нормальная плотность вероятности*, которую можно записать сжато, используя матричные обозначения, в виде

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (3.1.19)$$

где $\mathbf{x}' = (x_1, x_2, \dots, x_n)$, $\boldsymbol{\mu}' = (\mu_1, \mu_2, \dots, \mu_n)$ — векторы-строки и \mathbf{V}^{-1} — матрица, обратная матрице ковариаций \mathbf{V} , где

$$\mathbf{V} = \begin{vmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \sigma_1\sigma_3\rho_{13} & \dots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \sigma_2\sigma_3\rho_{23} & \dots & \sigma_2\sigma_n\rho_{2n} \\ \sigma_1\sigma_3\rho_{13} & \sigma_2\sigma_3\rho_{23} & \sigma_3^2 & \dots & \sigma_3\sigma_n\rho_{3n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \sigma_1\sigma_n\rho_{1n} & \sigma_2\sigma_n\rho_{2n} & \sigma_3\sigma_n\rho_{3n} & \dots & \sigma_n^2 \end{vmatrix} \quad (3.1.20)$$

Многомерная нормальная плотность вероятности зависит от $n(n+3)/2$ параметров, из которых n являются средними значениями μ_i ($i=1, 2, \dots, n$); n — дисперсиями σ_i^2 ($i=1, 2, \dots, n$) и $n(n-1)/2$ — корреляциями ρ_{ij} ($i=1, 2, \dots, n, j=i+1, \dots, n$).

Если случайные величины независимы, то корреляции $\rho_{ij} = 0$, и матрица \mathbf{V} является диагональной, а совместная плотность ве-

роятности распадается согласно (3.1.18) на произведение n одномерных нормальных распределений.

Чтобы описать эмпирические данные с помощью многомерной нормальной плотности вероятности, необходимо оценить упомянутые выше $n(n+3)/2$ параметров. Этот вопрос обсуждается в гл. 4.

3.2. МОМЕНТЫ СЛУЧАЙНЫХ ВЕЛИЧИН

3.2.1. Моменты одномерных случайных величин

Если дано распределение вероятностей $p_X(x)$ дискретной случайной величины или плотность вероятности $f_X(x)$ непрерывной случайной величины, можно вычислить вероятность того, что случайная величина находится между двумя значениями x_1 и x_2 . Иногда невозможно найти распределение вероятностей или плотность вероятности точно, и в таких случаях возникает необходимость охарактеризовать распределение с помощью нескольких чисел. Самыми простыми из них являются *среднее значение* и *дисперсия*.

Среднее значение. Иногда полезно знать, какое значение случайная величина X принимает в среднем. В примере с контролем качества из разд. 3.1.1 это значение представляет собой среднее число дефектных изделий в выборке, которое можно было бы ожидать. Среднее число дефектных изделий, которое действительно наблюдалось в N выборках, равно

$$\bar{x} = \frac{1}{N} \sum_{x=0}^k x n_x = \sum_{x=0}^k x \left(\frac{n_x}{N} \right) \quad (3.2.1)$$

и называется *выборочным средним* частотного распределения. Для данных, приведенных на рис. 3.1, $\bar{x} = 7,1$, и эта величина показана в виде жирной горизонтальной линии, вокруг которой группируются значения x .

Так как отношения n_x/N являются оценками для вероятностей $p_X(x)$, *среднее значение* распределения вероятностей равно

$$\mu = \sum_{x=0}^k x p_X(x). \quad (3.2.2)$$

Величина μ обозначается обычно $E[X]$ и называется *математическим ожиданием* случайной величины X . Оно дает среднее, или ожидаемое, значение, которое будет принимать X в будущих экспериментах. Аналогично для непрерывной случайной величины

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (3.2.3)$$

Равенство (3.2.3) совпадает с выражением для центра тяжести неоднородного стержня с приходящейся на единицу длины удельной массой $f_X(x)$, расположенной на расстоянии x от его конца. Аналогичным образом $E[X]$ является центром тяжести плотности вероятности случайной величины X , и, следовательно, оно служит для характеристики *расположения* распределения.

Дисперсия. Найдя расположение распределения, естественно перейти к описанию следующего наглядного свойства — степени разброса распределения. Одной из мер этого разброса является *дисперсия*

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx = E[(X - \mu)^2], \quad (3.2.4)$$

которая характеризует рассеяние вокруг его среднего значения μ . Если $f_X(x)$ все более и более концентрируется около μ , то σ^2 будет уменьшаться. Обратно, если имеются значения x , удаленные от среднего, для которых $f_X(x)$ не слишком мало, то σ^2 будет большой. Возведение в квадрат и раскрытие скобок в (3.2.4) дает другую эквивалентную формулу для дисперсии

$$\sigma^2 = E[X^2] - \mu^2. \quad (3.2.5)$$

Выражение (3.2.4) аналогично формуле для момента инерции стержня с неравномерной плотностью относительно его центра тяжести. При этом формула (3.2.5) просто утверждает, что момент инерции относительно центра тяжести равен моменту инерции относительно начала координат минус момент полной массы стержня, сконцентрированной в центре тяжести, относительно начала координат. Табл. 3.3 дает среднее значение и дисперсию для некоторых важных дискретных и непрерывных распределений.

Дисперсию дискретного распределения вероятностей можно оценить с помощью выборочной дисперсии

$$s^2 = \frac{1}{N-1} \sum_{x=0}^k n_x (x - \bar{x})^2.$$

Аналогично среднее значение и дисперсию данных x_i ($i=1, 2, \dots, n$), соответствующих непрерывной случайной величине, можно оценить по формулам

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, \\ s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned} \quad (3.2.6)$$

Таблица 3.3

Некоторые важные функции распределения и их средние значения и дисперсии

Распределение	Распределение вероятностей $p_X(x)$	Среднее значение μ	Дисперсия σ^2
Биномиальное	$\binom{n}{x} p^x (1-p)^{n-x}$ $x=0, 1, 2, \dots, n$	np	$np(1-p)$
Пуассона	$e^{-\lambda} \frac{\lambda^x}{x!}, \lambda > 0$ $x=0, 1, 2, \dots, \infty$	λ	λ
	Плотность вероятности $f_X(x)$	Среднее значение μ	Дисперсия σ^2
Нормальное	$\frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right],$ $-\infty \leq x \leq \infty$	μ	σ^2
Прямоугольное (равномерное)	$\frac{1}{a+b}, -a \leq x \leq b$	$\frac{1}{2} (b-a)$	$\frac{(a+b)^2}{12}$
Отрицательное показательное (экспоненциаль- ное)	$\frac{1}{\mu} e^{-x/\mu}, 0 \leq x \leq \infty, \mu \geq 0$	μ	μ^2

Положительный квадратный корень σ из дисперсии σ^2 называется *стандартным отклонением*. Его можно использовать для нормировки распределения, как мы сейчас покажем.

Нормированное нормальное распределение. Нормальная плотность вероятности (3.1.9) обладает тем важным свойством, что она полностью задается параметрами μ и σ^2 , соответствующими среднему значению и дисперсии случайной величины. Следовательно, среднее значение μ и стандартное отклонение σ можно использовать для нормировки плотности вероятности. Так, если X распределена по закону $N(\mu, \sigma^2)$, то случайная величина

$$Y = \frac{X - \mu}{\sigma} \quad (3.2.7)$$

имеет плотность вероятности

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad -\infty \leq y \leq \infty. \quad (3.2.8)$$

Следовательно, Y распределена как $N(0, 1)$. Плотность вероятности (3.2.8) называется *нормированной нормальной плотностью вероятности*.

Из (3.2.7) получаем, что случайная величина X лежит внутри интервала $(\mu - \eta\sigma, \mu + \eta\sigma)$, когда случайная величина Y лежит внутри интервала $(-\eta, +\eta)$. Вероятность последнего события $\text{Pr}\{-\eta < Y \leq \eta\}$ можно найти в стандартных таблицах [1, 6*]. Некоторые полезные значения η приведены в табл. 3.4.

Старшие моменты. В общем случае одномерную плотность вероятности можно описать с помощью ее среднего μ и старших центральных моментов

$$\mu_k = E[(X - \mu)^k], \quad k = 2, 3, \dots, \quad (3.2.9)$$

так что дисперсия при этом соответствует $k=2$. Значения μ_k для $k > 2$ не имеют большой практической важности, поскольку, если некоторая плотность вероятности неадекватно описывается своим средним значением и дисперсией, то ее лучше представить с помощью соответствующей негауссовской плотности вероятности и затем оценить параметры этой плотности.

Моменты функций от случайных величин. Иногда нужно исследовать некоторую функцию $Y = g(X)$ от случайной величины X , например $Y = \ln X$. В этом случае моменты Y можно выразить через плотность вероятности X с помощью соотношений *)

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

$$\text{Var}[Y] = \int_{-\infty}^{\infty} \{g(x) - E[Y]\}^2 f_X(x) dx \quad (3.2.10)$$

и так далее.

3.2.2. Многомерные моменты

Результаты предыдущего раздела можно распространить на распределения более высокого порядка. Рассмотрим, например, функцию $g(X_1, X_2, \dots, X_n)$ от случайных величин X_1, X_2, \dots, X_n ,

*) Символом $\text{Var}[\dots]$ всюду в этой книге обозначается дисперсия случайной величины.— *Прим. перев.*

имеющих совместную плотность вероятности $f_{1,2,\dots,n}(x_1, x_2, \dots, x_n)$. Математическое ожидание $g(X_1, X_2, \dots, X_n)$ равно

$$E[g(X_1, X_2, \dots, X_n)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) \times \\ \times f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n, \quad (3.2.11)$$

что является многомерным аналогом равенства (3.2.10), упомянутого выше.

Если функция $g(X_1, X_2, \dots, X_n)$ распадается на множители $g(X_1, X_2, \dots, X_n) = g_1(X_1)g_2(X_2)\dots g_n(X_n)$ и в дополнение к этому случайные величины независимы, так что плотность вероятности также распадается на множители, то (3.2.11) переходит в

$$E[g_1(X_1)g_2(X_2)\dots g_n(X_n)] = \\ = E[g_1(X_1)]E[g_2(X_2)]\dots E[g_n(X_n)]. \quad (3.2.12)$$

Ковариация. Функциями $g(X_1, X_2, \dots, X_n)$, представляющими особую важность, являются произведения случайных величин, например

$$g(X_1, X_2) = (X_1 - \mu_1)(X_2 - \mu_2)$$

для двумерного случая. Математическое ожидание этого произведения называется *ковариацией* между X_1 и X_2 и записывается

$$\text{Cov}[X_1, X_2] = E[(X_1 - \mu_1)(X_2 - \mu_2)] = \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) f_{12}(x_1, x_2) dx_1 dx_2. \quad (3.2.13)$$

Заметим, что из определения (3.2.13) следует, что $\text{Cov}[X_1, X_2] = \text{Cov}[X_2, X_1]$ и что

$$\text{Var}[X_1] = \text{Cov}[X_1, X_1].$$

Если X_1 и X_2 независимы, то $f_{12}(x_1, x_2) = f_1(x_1)f_2(x_2)$ и, следовательно,

$$\text{Cov}[X_1, X_2] = E[X_1 - \mu_1]E[X_2 - \mu_2] = 0.$$

Таким образом, ковариация измеряет степень линейной зависимости двух случайных величин.

В спектральном анализе иногда приходится рассматривать ковариацию между функциями $g(X_1, \dots, X_n)$ и $h(X_1, \dots, X_n)$, а именно

$$\text{Cov}[g(X_1, \dots, X_n), h(X_1, \dots, X_n)] = \\ = E\{[g(X_1, \dots, X_n) - E[g(X_1, \dots, X_n)]] \times \\ \times [h(X_1, \dots, X_n) - E[h(X_1, \dots, X_n)]]\}.$$

Например, ковариация между $g(X_1, X_2) = X_1X_2$ и $h(X_3, X_4) = X_3X_4$ равна

$$\text{Cov} [X_1X_2, X_3X_4] = E [(X_1X_2 - E [X_1X_2])(X_3X_4 - E [X_3X_4])].$$

3.2.3. Моменты линейных функций от случайных величин

Рассмотрим произвольную линейную функцию $\lambda_1X_1 + \lambda_2X_2$ двух случайных величин X_1 и X_2 . Используя (3.2.11), получаем

$$\begin{aligned} E [\lambda_1X_1 + \lambda_2X_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\lambda_1x_1 + \lambda_2x_2) f_{12}(x_1, x_2) dx_1 dx_2 = \\ &= \lambda_1 E [X_1] + \lambda_2 E [X_2]. \end{aligned} \quad (3.2.14)$$

Следует отметить, что (3.2.14) справедливо, даже если X_1 и X_2 не являются независимыми. Вообще

$$E \left[\sum_{i=1}^n \lambda_i X_i \right] = \sum_{i=1}^n \lambda_i E [X_i]. \quad (3.2.15)$$

В качестве примера рассмотрим математическое ожидание среднего арифметического $\bar{X} = (1/n) \sum_{i=1}^n X_i$ набора случайных величин с одним и тем же средним значением μ . Равенство (3.2.15) показывает, что

$$E [\bar{X}] = E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \sum_{i=1}^n \frac{1}{n} \mu = \mu.$$

Следовательно, математическое ожидание среднего арифметического равно математическому ожиданию отдельной случайной величины.

Дисперсия линейных функций. Используя (3.2.13), получаем, что дисперсия линейной функции $\lambda_1X_1 + \lambda_2X_2$ равна

$$\begin{aligned} \text{Var} [\lambda_1X_1 + \lambda_2X_2] &= \lambda_1^2 \text{Var} [X_1] + \\ &+ \lambda_2^2 \text{Var} [X_2] + 2\lambda_1\lambda_2 \text{Cov} [X_1, X_2]. \end{aligned} \quad (3.2.16)$$

И вообще

$$\text{Var} \left[\sum_{i=1}^n \lambda_i X_i \right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov} [X_i, X_j], \quad (3.2.17)$$

где $\text{Cov} [X_i, X_i] = \text{Var} [X_i]$.

Если X_i независимы, то (3.2.17) сводится к

$$\text{Var} \left[\sum_{i=1}^n \lambda_i X_i \right] = \sum_{i=1}^n \lambda_i^2 \text{Var} [X_i]. \quad (3.2.18)$$

Рассмотрим, например, случайную величину $\bar{X} = (1/n) \sum_{i=1}^n X_i$, где X_i — независимые случайные величины с дисперсией σ^2 . Тогда

$$\text{Var} [\bar{X}] = \sum_{i=1}^n \left(\frac{1}{n}\right)^2 \text{Var} [X_i] = \frac{\sigma^2}{n}.$$

Используя (3.2.15) и (3.2.18) при $n=1$, получаем полезный результат: *нормированная* случайная величина

$$Y = \frac{X - \mu}{\sigma}$$

имеет нулевое среднее значение и единичную дисперсию. Дальнейший важный результат [2, 2*] состоит в том, что если случайные величины X_i являются нормальными, то плотность вероятности случайной величины

$$Y = \sum_{i=1}^n \lambda_i X_i$$

также является нормальной со средним значением (3.2.15) и дисперсией (3.2.17).

3.2.4. Коэффициент корреляции

Выражение (3.2.16) для дисперсии линейной функции двух случайных величин обязательно является положительным числом или нулем для любых действительных значений λ_1 и λ_2 .

Так как выражение в правой части является квадратным уравнением относительно λ_1 и λ_2 , то из положительности дисперсии следует, что его корни являются комплексными. Отсюда

$$\text{Var} [X_1] \text{Var} [X_2] \geq (\text{Cov} [X_1, X_2])^2,$$

что можно переписать в виде

$$\rho_{12}^2 = \frac{(\text{Cov} [X_1, X_2])^2}{\text{Var} [X_1] \text{Var} [X_2]} \leq 1. \quad (3.2.19)$$

Параметр ρ_{12} называется *коэффициентом корреляции* между X_1 и X_2 . Он заключен в интервале $-1 \leq \rho_{12} \leq +1$.

Мы уже отмечали, что если случайные величины независимы, то $\text{Cov} [X_1, X_2] = 0$ и, следовательно, $\rho_{12} = 0$. Для двумерной нормальной плотности вероятности было показано и обратное: если $\rho_{12} = 0$, то случайные величины независимы. Однако если $\rho_{12} = 0$ для распределения, отличного от нормального, то случайные величины не обязательно являются независимыми. В этом случае их называют *некоррелированными*.

Если $\rho_{12} = 0$, то диаграмма разброса для пар величин (x_1, x_2) , которые являются реализациями случайных величин (X_1, X_2) , была бы похожа на приведенную на рис. 3.8 а. Видно, что знание одного из членов пары никак не помогает в предсказании значения другого.

Для малых, но положительных значений ρ_{12} диаграмма разброса была бы похожа на показанную на рис. 3.8, б; этот рисунок

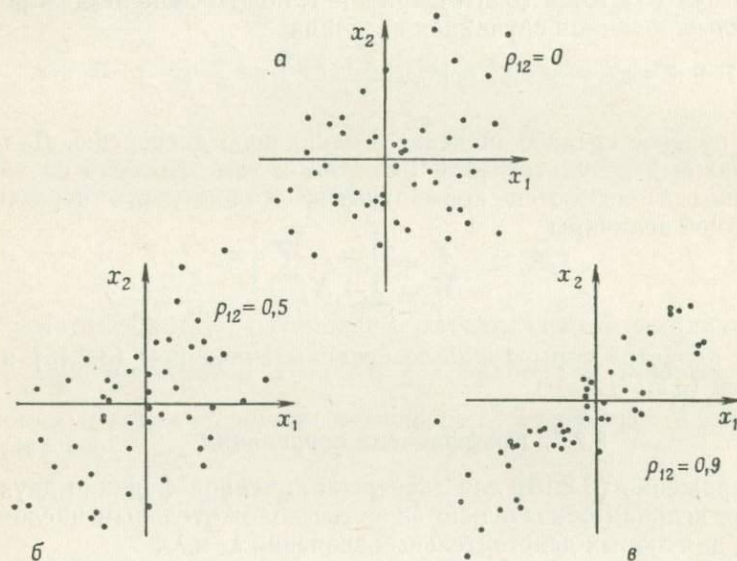


Рис. 3.8. Диаграммы разброса выборок двумерных нормальных случайных величин.

соответствует значению $\rho_{12} = +0,5$. Теперь уже заметна слабая тенденция к группированию значений вдоль прямой линии. Так, большие значения x_2 преимущественно соответствуют большим значениям x_1 , а малые значения x_2 — малым значениям x_1 . Если бы коэффициент корреляции был равен $-0,5$, то наклон прямой, вокруг которой группируются значения, был бы отрицательным. Следовательно, большим значениям x_2 преимущественно соответствовали бы малые значения x_1 , и наоборот. Для значений ρ_{12} , близких к единице, диаграмма разброса концентрируется около прямой линии, как показано на рис. 3,8, в; для этого рисунка $\rho_{12} = +0,9$. Следовательно, ρ_{12} является мерой *линейной зависимости* между случайными величинами X_1 и X_2 , и в предельном случае $\rho_{12} = 1$ имеется точное линейное соотношение вида $X_2 = a + bX_1$.

Сравнение диаграмм разброса рис. 3.8 с диаграммами разброса данных акселерометра, приведенных на рис. 3.7, показывает, что эти данные имеют коэффициент корреляции между 0,5 и 0,9.

Формула для выборочной оценки $\hat{\rho}_{12}$ коэффициента корреляции будет приведена в гл. 4; для акселерометрических данных она дает значение $\hat{\rho}_{12} = 0,78$. Эта величина достаточно мала и должна вызывать некоторое беспокойство относительно надежности визуального считывания показаний акселерометра пилотом!

Ковариация линейных функций. В качестве заключительного обобщения (3.2.17) рассмотрим линейные функции $\lambda_1 X_1 + \lambda_2 X_2$ и $\nu_1 X_1 + \nu_2 X_2$. Ковариация между ними равна

$$\text{Cov}[(\lambda_1 X_1 + \lambda_2 X_2), (\nu_1 X_1 + \nu_2 X_2)] = \lambda_1 \nu_1 \text{Cov}[X_1, X_1] + \lambda_1 \nu_2 \text{Cov}[X_1, X_2] + \lambda_2 \nu_1 \text{Cov}[X_2, X_1] + \lambda_2 \nu_2 \text{Cov}[X_2, X_2], \quad (3.2.20)$$

что сводится к (3.2.16) в случае $\lambda_1 = \nu_1$ и $\lambda_2 = \nu_2$.

Вообще имеем

$$\text{Cov} \left[\sum_{i=1}^n \lambda_i X_i, \sum_{j=1}^n \nu_j X_j \right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \nu_j \text{Cov}[X_i, X_j]. \quad (3.2.21)$$

Равенство (3.2.21) является важным результатом, который будет использован в гл. 6 для вычисления ковариации между сглаженными спектральными оценками.

Результаты этого раздела приведены в матричной форме в приложении ПЗ.1.

3.2.5. Моменты нелинейных функций случайных величин

Во многих практических статистических задачах необходимо рассматривать нелинейные функции от случайных величин. Например, большинство задач спектрального анализа являются нелинейными. За исключением некоторых специальных случаев, невозможно вывести точные плотности вероятности этих нелинейных функций, и, следовательно, нужно описывать эти плотности вероятности с помощью их моментов. В этом разделе показывается, как вывести приближенные выражения для среднего значения и дисперсии нелинейной функции от случайных величин.

Среднее значение нелинейной функции. Рассмотрим функцию $g(X_1, X_2, \dots, X_n)$ от случайных величин X_1, X_2, \dots, X_n , которые имеют средние μ_i и ковариации σ_{ij} ($i, j = 1, 2, \dots, n$). Рассмотрим далее разложение Тейлора функции $g(X_1, X_2, \dots, X_n)$ в точке $(a_1,$

a_2, \dots, a_n). Тогда с точностью до членов первого порядка

$$g(X_1, X_2, \dots, X_n) \approx g(a_1, a_2, \dots, a_n) + \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right)_a (X_i - a_i), \quad (3.2.22)$$

где $(\partial g / \partial X_i)_a$ обозначает частную производную по X_i в точке $(X_1, X_2, \dots, X_n) = (a_1, a_2, \dots, a_n)$. Взяв математическое ожидание от обеих частей (3.2.22), получаем

$$E[g(X_1, X_2, \dots, X_n)] \approx g(a_1, a_2, \dots, a_n) + \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right)_a (E[X_i] - a_i). \quad (3.2.23)$$

Если $a_i = \mu_i$, то второй член исчезает, так что

$$E[g(X_1, X_2, \dots, X_n)] \approx g(\mu_1, \mu_2, \dots, \mu_n). \quad (3.2.24)$$

Дисперсия нелинейной функции. Из (3.2.22) имеем

$$[g(X_1, X_2, \dots, X_n) - g(\mu_1, \mu_2, \dots, \mu_n)] \approx \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right)_\mu (X_i - \mu_i).$$

Последнее выражение линейно по X_i , поэтому, используя (3.2.17), получаем

$$\text{Var}[g(X_1, X_2, \dots, X_n)] \approx \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial g}{\partial X_i} \right)_\mu \left(\frac{\partial g}{\partial X_j} \right)_\mu \text{Cov}[X_i, X_j]. \quad (3.2.25)$$

Частные случаи формулы (3.2.25). Если $n=1$, то

$$\text{Var}[g(X)] \approx \left(\frac{dg}{dX} \right)_\mu^2 \text{Var}[X]. \quad (3.2.26)$$

Например, если $g(X) = X^2$, то

$$\text{Var}[X^2] \approx 4\mu^2 \text{Var}[X].$$

Если $n=2$, то

$$\begin{aligned} \text{Var}[g(X_1, X_2)] &\approx \left(\frac{\partial g}{\partial X_1} \right)_\mu^2 \text{Var}[X_1] + \left(\frac{\partial g}{\partial X_2} \right)_\mu^2 \text{Var}[X_2] + \\ &+ 2 \left(\frac{\partial g}{\partial X_1} \right)_\mu \left(\frac{\partial g}{\partial X_2} \right)_\mu \text{Cov}[X_1, X_2]. \end{aligned} \quad (3.2.27)$$

Например, если X_1 и X_2 — некоррелированные случайные величины и

$$g(X_1, X_2) = \sqrt{X_1^2 + X_2^2},$$

то (3.2.27) сводится к

$$\text{Var}[g(X_1, X_2)] = \frac{\mu_1^2 \text{Var}[X_1] + \mu_2^2 \text{Var}[X_2]}{\mu_1^2 + \mu_2^2}.$$

Преобразования, делающие дисперсию постоянной. В статистических задачах часто получается так, что дисперсия случайной величины является некоторой функцией от ее среднего значения μ , например $\text{Var}[X] = \mu^2$. В этом случае логичней рассматривать случайную величину $Y = X/\mu$, так как $\text{Var}[Y] = 1$ и, следовательно, масштаб измерения Y не зависит от ее среднего значения. Более общий подход состоит в том, что рассматривают такую функцию $g(X)$ от случайной величины, что $\text{Var}[g(X)]$ мало зависит от среднего значения X и, следовательно, от среднего значения $g(X)$.

Используя (3.2.26), получаем, что если потребовать, чтобы $\text{Var}[g(X)]$ была константой k_1 , то

$$k_1 \approx \left(\frac{dg}{dX} \right)_\mu^2 \varphi(\mu),$$

где $\text{Var}[X] = \varphi(\mu)$. Поэтому с точностью до аддитивной константы

$$g(\mu) \sim \int \frac{d\mu}{\sqrt{\varphi(\mu)}}. \quad (3.2.28)$$

а случайная величина $g(X)$ имеет дисперсию, которая мало зависит от среднего. В упоминавшемся выше примере $\text{Var}[X] = \varphi(\mu) = \mu^2$, так что $g(\mu) = \ln \mu$. Таким образом, случайная величина $\ln X$ имеет дисперсию, мало зависящую от среднего значения, и поэтому она дает более логичный масштаб измерения, чем сама X . Преобразования, делающие дисперсию постоянной, направлены на то, чтобы получить плотность вероятности для преобразованной случайной величины, более похожую на нормальную плотность, чем плотность величины X . Следовательно, плотность вероятности преобразованной случайной величины будет полнее охарактеризована с помощью ее среднего значения и дисперсии, чем плотность самой X .

3.3. ВЫБОРОЧНЫЕ РАСПРЕДЕЛЕНИЯ

Одна из важных задач теории вероятностей состоит в нахождении плотности вероятности $f_X(x)$ некоторой функции $X(X_1, X_2, \dots, X_n)$ от n случайных величин X_1, X_2, \dots, X_n , если дана их совместная плотность вероятности [2,2*]. Эти выводимые распределения

используются при статистическом анализе данных следующим образом.

Предположим, мы хотим собрать конечную выборку наблюдений x_1, x_2, \dots, x_n , по которым нам нужно сосчитать некоторую функцию $x(x_1, x_2, \dots, x_n)$, например среднее значение. Тогда, прежде чем данные собраны, можно описать все возможные наборы данных, которые можно было бы получить с помощью случайных величин X_1, X_2, \dots, X_n . Таким образом, полнота возможных экспериментов описывается n -мерным выборочным пространством, с которым можно связать совместную плотность вероятности $f_{12\dots n}(x_1, x_2, \dots, x_n)$. Используя методы, описанные, например, в [2], можно затем вычислить плотность вероятности $f_X(x)$ функции $X(X_1, X_2, \dots, X_n)$. Эту плотность вероятности можно использовать до сбора данных для предсказания частоты, с которой различные значения функции $x(x_1, x_2, \dots, x_n)$ будут попадать в интервал между двумя любыми пределами в повторяемых выборках объема n . Поэтому плотность вероятности $f_X(x)$ называется *выборочным распределением* случайной величины $X(X_1, X_2, \dots, X_n)$.

3.3.1. Выборочное распределение среднего значения в случае, когда дисперсия известна

Приведем простейший пример выборочного распределения. Пусть производится n независимых измерений некоторой переменной, например обратного коллекторного тока в транзисторе. В этом случае совместная плотность вероятности просто равна

$$f_{12\dots n}(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n). \quad (3.3.1)$$

Предположим, что нас интересует изменчивость выборочного среднего этих измерений. Тогда если предположить, что каждая X_i распределена как $N(\mu, \sigma^2)$, то можно показать [2], что плотность вероятности среднего арифметического значения случайных величин $\bar{X} = (1/n) \sum_{i=1}^n X_i$ будет распределена как $N(\mu, \sigma^2/n)$, т. е.

$$f_{\bar{X}}(\bar{x}) = \frac{1}{\sqrt{2\pi} (\sigma/\sqrt{n})} \exp \left[-\frac{n}{2} \left(\frac{\bar{x} - \mu}{\sigma} \right)^2 \right]; \quad (3.3.2)$$

(3.3.2) называется *выборочным распределением среднего* для нормальных случайных величин. Частотная интерпретация, которую можно применить к (3.3.2), заключается в следующем. Если представить себе очень большое число экспериментов, каждый из которых состоит из n независимых измерений, взятых из нормальной популяции с $N(\mu, \sigma^2)$, то гистограмма распределения \bar{x} стремилась бы к нормальному закону (3.3.2).

Выборочное распределение выборочного среднего обычно очень близко к нормальному, даже если отдельные распределения

$f_1(x_1), \dots, f_n(x_n)$ сами не являются нормальными. Этот важный результат следует из центральной предельной теоремы [2].

Выборочное распределение, подобно любому другому распределению, можно описать с помощью его моментов, обычно называемых *выборочными моментами*. Например, выборочное распределение среднего нормальных случайных величин (3.3.2) полностью описывается с помощью выборочных моментов

$$\begin{aligned} E[\bar{X}] &= \mu, \\ \text{Var}[\bar{X}] &= \frac{\sigma^2}{n}. \end{aligned} \quad (3.3.3)$$

Частотная интерпретация, которую можно дать этим моментам, состоит в том, что среднее из большого числа выборочных средних будет лежать очень близко к среднему значению популяции, или теоретическому значению μ , и что изменчивость выборочных средних от выборки к выборке характеризуется дисперсией σ^2/n .

Одно из основных применений выборочных распределений состоит в том, что они позволяют делать вероятностные утверждения относительно случайных величин, таких, как \bar{X} . Например, рассмотрим выборку из 9 значений x_i ($i = 1, 2, \dots, 9$) случайной величины X , про которую известно, что она распределена нормально с единичной дисперсией, но неизвестным средним значением μ . Из (3.3.2) и (3.3.3) получаем, что случайная величина \bar{X} распределена нормально с $E[\bar{X}] = \mu$ и $\text{Var}[\bar{X}] =$

$\frac{1}{9}$. Следовательно, воспользовавшись приводимыми в табл. 3.4 вероятностями, относящимися к нормальному закону, можно

Таблица 3.4

Вероятности, относящиеся к нормированной нормальной плотности

η	$\text{Pr}\{-\eta < Y \leq \eta\}$	$\alpha =$ площади вне интервала $(-\eta, +\eta)$
1,00	0,683	0,317
1,96	0,950	0,050
2,00	0,954	0,046
2,58	0,990	0,010
3,00	0,997	0,003

подсчитать вероятность того, что наблюдаемое значение \bar{x} случайной величины \bar{X} будет лежать в заданном интервале. Например,

$$\text{Pr}\left\{\mu - \frac{1,96}{\sqrt{9}} < \bar{X} \leq \mu + \frac{1,96}{\sqrt{9}}\right\} = 0,95$$

или

$$\text{Pr}\{\mu - 0,653 < \bar{X} \leq \mu + 0,653\} = 0,95.$$

Это означает, что если случайная величина X распределена нормально со средним значением μ и дисперсией 1, то с вероятностью 0,95 случайная величина \bar{X} будет лежать не дальше, чем на $\pm 0,653$ от μ . Частотная интерпретация этого факта состоит в том, что из большого числа выборок, каждая из которых состоит из 9 реализаций X , приблизительно одна из двадцати выборочных оценок \bar{x} будет отличаться от истинного значения μ больше, чем на 0,653. Обратная и более трудная задача получения выводов относительно μ по данному значению \bar{x} обсуждается в гл. 4.

3.3.2. Выборочное распределение дисперсии

Выборочное распределение среднего представляет собой распределение суммы случайных величин. Следующее простейшее выборочное распределение — распределение дисперсии нормальных случайных величин — представляет собой распределение суммы квадратов случайных величин $X_1^2 + X_2^2 + \dots + X_n^2$. Предположим, например, что имеется n независимых измерений из $N(0, 1)$ -популяции и требуется найти выборочное распределение случайной величины

$$\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2. \quad (3.3.4)$$

Распределение χ_n^2 называется χ^2 -распределением с n степенями свободы. Общий вид плотности вероятности χ^2 -распределения с ν степенями свободы следующий:

$$f_{\chi_\nu^2}(x) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{(\nu/2)-1} \exp\left(-\frac{x}{2}\right) \quad (0 \leq x < \infty), \quad (3.3.5)$$

где $\Gamma(\nu/2) = \int_0^\infty e^{-t} t^{(\nu/2)-1} dt$ — гамма-функция от аргумента $\nu/2$.

Графики зависимости $f_{\chi_\nu^2}(x)$ от x для $\nu = 1, 2, 3$ и 10 приведены на рис. 3.9. Для $\nu = 1$ плотность вероятности имеет бесконечную ординату при $x = 0$ и стремится к нулю, когда x стремится к бесконечности. Для $\nu = 2$ плотность вероятности является экспонентой, а для $\nu \geq 3$ плотность вероятности принимает унимодальную форму. Заметим, однако, что для малых ν распределение очень несимметрично. По мере того как ν возрастает, плотность вероятности начинает выглядеть все более и более похожей на нормальную, как это и предсказывается центральной предельной теоремой.

Первые два момента случайной величины χ_ν^2 , полученные из (3.3.5), равны

$$\begin{aligned} E[\chi_\nu^2] &= \nu, \\ \text{Var}[\chi_\nu^2] &= 2\nu. \end{aligned} \quad (3.3.6)$$

В гл. 4 будет показано, что выборочной оценкой дисперсии по выборке из n наблюдений является

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

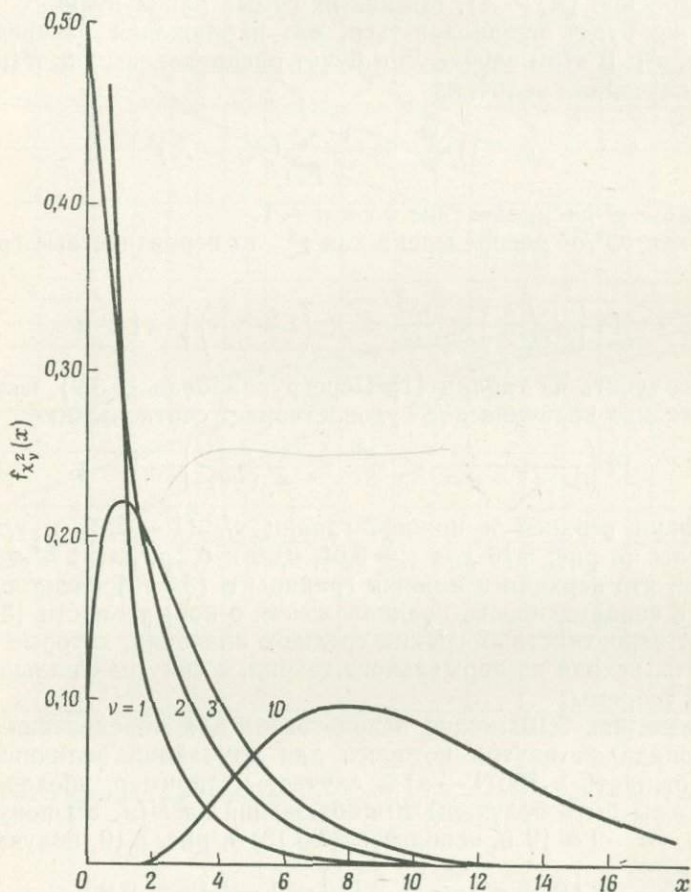


Рис. 3.9. Плотности вероятности χ^2 -распределения.

Чтобы описать изменчивость этой функции от одной выборки к другой, вводят соответствующую случайную величину S^2 , где

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (3.3.7)$$

Если X_i — независимые $N(\mu, 1)$ -случайные величины, то можно показать [2], что $(n-1)S^2$ распределена как χ^2 с $\nu = (n-1)$ степенями свободы. Термин «степени свободы» используется здесь в том же самом смысле, что и в статистической механике. Так, для любого множества из n наблюдений будет только $(n-1)$ независимое отклонение $(X_i - \bar{X})$, так как их сумма равна нулю.

Обычно будет предполагаться, что наблюдения распределены как $N(\mu, \sigma^2)$. В этом случае X_i/σ будут распределены как $N(\mu/\sigma, 1)$, так что случайная величина

$$(n-1) \frac{S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3.3.8)$$

будет иметь χ^2 -распределение с $\nu = n-1$.

Так как $\nu S^2/\sigma^2$ распределена как χ_ν^2 , то вероятностные границы вида

$$\Pr \left\{ x_\nu \left(\frac{\alpha}{2} \right) < \frac{\nu S^2}{\sigma^2} \leq x_\nu \left(1 - \frac{\alpha}{2} \right) \right\} = 1 - \alpha \quad (3.3.9)$$

можно получить из таблиц [1]. Перегруппировав (3.3.9), получаем, что случайная величина σ^2/S^2 удовлетворяет соотношению

$$\Pr \left\{ \frac{\nu}{x_\nu (1 - \alpha/2)} < \frac{\sigma^2}{S^2} \leq \frac{\nu}{x_\nu (\alpha/2)} \right\} = 1 - \alpha. \quad (3.3.10)$$

Графики верхней и нижней границ $\nu/x_\nu(1-\alpha/2)$ и $\nu/x_\nu(\alpha/2)$ приведены на рис. 3.10 для $\alpha = 0,01, 0,05$ и $0,2$ и для $3 \leq \nu \leq 100$. Отметим, что верхняя и нижняя границы в (3.3.10) очень чувствительны к справедливости предположения о нормальности [3], в отличие от вероятностных границ среднего значения, которые можно построить, исходя из нормального закона, в силу центральной предельной теоремы.

Кривые рис. 3.10 можно использовать для определения интервала, попадание внутрь которого для случайной величины S^2/σ^2 можно ожидать в $100(1-\alpha)\%$ случаев. Например, предположим, что должны быть получены 20 наблюдений из $N(\mu, \sigma^2)$ -популяции. Тогда $\nu = n-1 = 19$ и, используя (3.3.10) и рис. 3.10, получаем

$$\Pr \left\{ 0,58 < \frac{\sigma^2}{S^2} \leq 2,11 \right\} = 1 - 0,05 = 0,95.$$

Поэтому следовало бы ожидать, что в среднем в 19 случаях из 20 отношение σ^2/S^2 будет лежать в интервале от 0,58 до 2,11. Иначе говоря, значения S^2 будут лежать с вероятностью 0,95 в интервале $0,47\sigma^2 < S^2 \leq 1,72\sigma^2$, или же значение $\nu S^2/\sigma^2 = 19S^2/\sigma^2$ будет лежать в интервале $8,9 < 19S^2/\sigma^2 \leq 32,9$. Границы $8,9 = 19/2,11$ и $32,9 = 19/0,58$ для $\nu S^2/\sigma^2$ обычно приводятся в статистических таблицах.

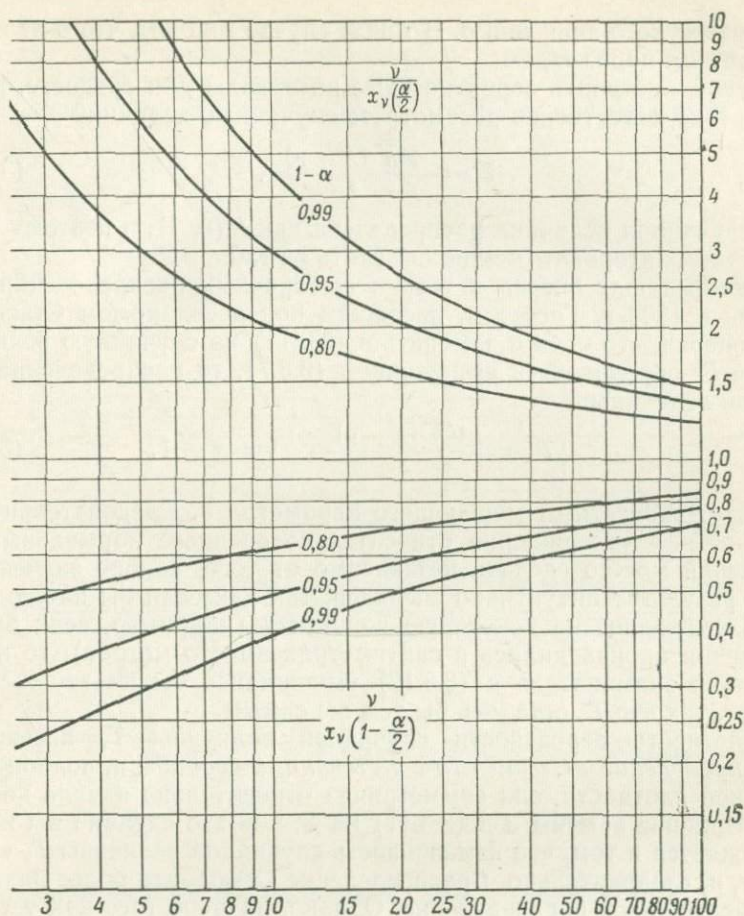


Рис. 3.10. Графики зависимости $\frac{v}{x_v(\frac{\alpha}{2})}$, $\frac{v}{x_v(1-\frac{\alpha}{2})}$ от v для $(1-\alpha) = 0,80; 0,95; 0,99$.

3.3.3. Выборочное распределение среднего в случае, когда дисперсия неизвестна

Для того чтобы определить вероятностные границы для среднего нормальных случайных величин, нужно знать σ — стандартное отклонение популяции. Если σ неизвестно, то невозможно сделать точные вероятностные утверждения, используя выборочное распределение \bar{X} , так как вероятностные границы будут зависеть

от неизвестного значения σ . В таком случае говорят, что σ является мешающим параметром.

Чтобы построить вероятностные интервалы для среднего, когда σ известно, естественно рассмотреть случайную величину

$$Y = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}. \quad (3.3.11)$$

Эта случайная величина распределена как $N(0, 1)$, и поэтому вероятностные интервалы можно получить из табл. 3.2.

Важный шаг вперед в теории выборочных распределений был сделан в 1908 г. Госсетом, писавшим под псевдонимом Стьюдент. Он показал, что если σ заменить в (3.3.11) на случайную величину S , где S^2 определяется выражением (3.3.7), то распределение случайной величины

$$T_v = \frac{\sqrt{n}(\bar{X} - \mu)}{S}, \quad v = n - 1, \quad (3.3.12)$$

не будет зависеть от мешающего параметра σ . Следовательно, вероятностные утверждения относительно среднего нормальных наблюдений можно сделать независимо от того, каково значение σ . Этот результат интуитивно очевиден, так как если бы наблюдения были умножены на некоторую константу (например, если бы наблюдения производились в сантиметрах вместо метров), то и числитель и знаменатель в (3.3.12) умножились бы на эту же константу, так что T_v осталось бы тем же самым.

Плотность вероятности случайной величины T_v называется *t-распределением Стьюдента с v степенями свободы* и, подобно нормальной плотности, она симметрична относительно начала координат. Влияние замены σ в (3.3.11) на S , как это сделано в (3.3.12), выражается в том, что изменчивость случайной величины T_v возрастает, и, следовательно, *t-распределение Стьюдента* более размыто, чем нормальное распределение. Однако, по мере того как v увеличивается, распределение S все более и более концентрируется около σ , и поэтому *t-распределение* стремится к стандартному нормальному распределению (3.2.8), как это вновь следует из центральной предельной теоремы.

t-распределение Стьюдента можно использовать для построения интервалов $t_v(\alpha/2)$, $t_v(1 - \alpha/2)$, в которые можно ожидать попадания случайной величины T_v в $(1 - \alpha)$ -й части всех случаев. Так как плотность вероятности симметрична, то $t_v(\alpha/2) = -t_v(1 - \alpha/2)$ и поэтому

$$\Pr \left\{ -t_v \left(1 - \frac{\alpha}{2} \right) < T_v \leq t_v \left(1 - \frac{\alpha}{2} \right) \right\} = 1 - \alpha.$$

Рис. 3.11 показывает кривые $t_v(1 - \alpha/2)$ в зависимости от v для $\alpha = 0,05$ и $\alpha = 0,01$. Заметим, что для больших v кривые стремятся

к значениям 1,96 и 2,58, являющимся 95%-ной и 99%-ной границами для нормированной нормальной плотности вероятности.

Чтобы проиллюстрировать использование кривых на рис. 3.11, предположим, что нужно произвести, как и в примере разд. 3.3.1,

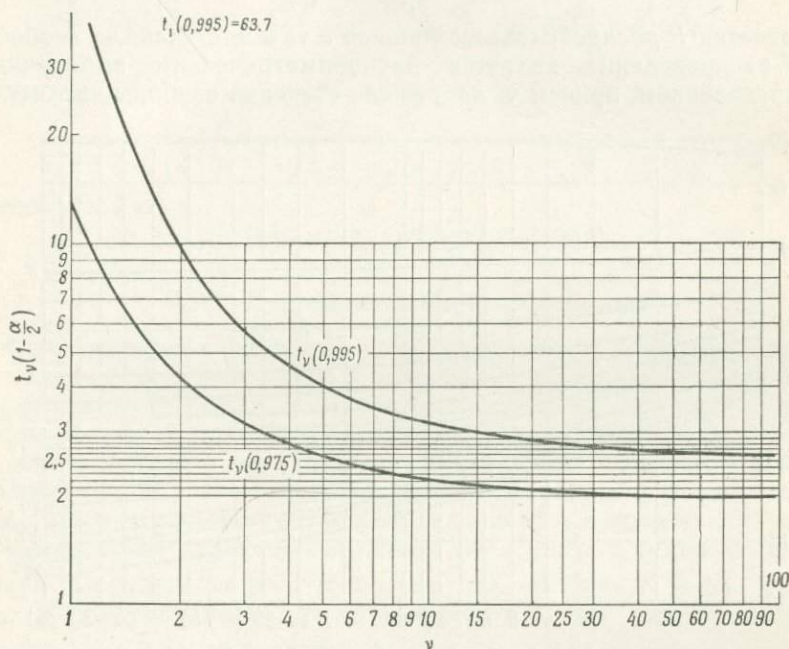


Рис. 3.11. Графики зависимости $t_{\nu}(1-\alpha/2)$ от ν для $(1-\alpha)=0,99; 0,95$.

9 измерений из $N(\mu, \sigma^2)$ -популяции. Тогда, согласно рис. 3.11, следует ожидать, что случайная величина $\sqrt{9}(X-\mu)/S$ будет лежать в интервале $(-2,3; +2,3)$ в 95% случаев. Заметим, что соответствующий интервал в случае известного σ , найденный из табл. 3.4, есть $(-1,96; +1,96)$. Этот интервал примерно на 15% уже.

3.3.4. Выборочное распределение отношения двух дисперсий

Другое важное выборочное распределение появляется, когда требуется сравнить выборочные оценки дисперсий s_1^2 и s_2^2 , полученные из двух независимых выборок объема n_1 и n_2 соответственно. Если выборки производятся из двух популяций, распределенных как $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$, то из разд. 3.3.2 следует, что $\nu_1 S_1^2 / \sigma_1^2$ есть случайная величина $\chi_{\nu_1}^2$ с $\nu_1 = n_1 - 1$ и аналогично $\nu_2 S_2^2 / \sigma_2^2$ есть

случайная величина $\chi_{v_2}^2$ с $v_2 = n_2 - 1$. Если $\chi_{v_1}^2$ и $\chi_{v_2}^2$ независимы, то плотность вероятности отношения

$$F_{v_1, v_2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} = \frac{v_2 \chi_{v_1}^2}{v_1 \chi_{v_2}^2} \quad (3.3.13)$$

называется *F-распределением Фишера с v_1 и v_2 степенями свободы*.

F-распределение является двухпараметрическим выборочным распределением, причем v_1 дает число степеней свободы числителя,

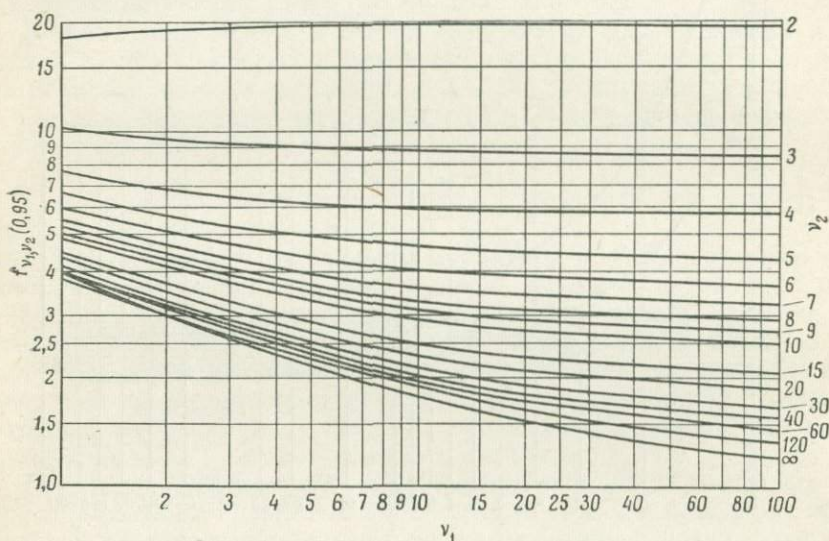


Рис. 3.12. Графики зависимости $f_{v_1, v_2}(0,95)$ от (v_1, v_2) .

а v_2 — знаменателя. Когда v_1 и v_2 оба велики, плотность вероятности случайной величины F_{v_1, v_2} концентрируется около единицы. Однако для малых значений v_1 или v_2 плотность распределяется до очень далеких от 1 значений.

На практике теоретические дисперсии σ_1^2 и σ_2^2 , которые появляются в (3.3.13), не будут известны. Однако если предположить, что $\sigma_1^2 = \sigma_2^2$, то из (3.3.13) следует, что S_1^2/S_2^2 распределено как F_{v_1, v_2} . Если же $\sigma_1^2 \neq \sigma_2^2$, то S_1^2/S_2^2 будет распределено как $(\sigma_1^2/\sigma_2^2)F_{v_1, v_2}$ и, следовательно, распределение будет концентрироваться около значения σ_1^2/σ_2^2 , а не 1.

Рис. 3.12 показывает 0,95-вероятностные точки для распределения F_{v_1, v_2} , т. е. значения $f_{v_1, v_2}(0,95)$ такие, что

$$\Pr \{ F_{v_1, v_2} \leq f_{v_1, v_2}(0,95) \} = 0,95.$$

Заметим, что так как $F_{v_1, v_2} = 1/F_{v_2, v_1}$, значения f_{v_1, v_2} и f_{v_2, v_1} можно использовать для построения вероятностных интервалов для случайной величины F_{v_1, v_2} . Таким образом,

$$\Pr \left\{ \frac{1}{f_{v_2, v_1}(0,95)} < F_{v_1, v_2} \leq f_{v_1, v_2}(0,95) \right\} = 0,90.$$

Например, если $v_1 = 4$, $v_2 = 20$, то из рис. 3.12 получаем

$$\Pr \{ F_{4, 20} \leq 2,9 \} = 0,95;$$

$$\Pr \{ F_{20, 4} \leq 5,9 \} = \Pr \{ F_{4, 20} > 0,17 \} = 0,95.$$

Следовательно,

$$\Pr \{ 0,17 < F_{4, 20} \leq 2,9 \} = 0,90.$$

3.3.5. Два свойства χ^2 -распределения

Приближение с помощью χ^2 -распределения. χ^2 -распределение занимает центральное место в вопросах приближения распределений сглаженных оценок спектральной плотности. Вообще, случайная величина χ^2 полезна для приближения случайной величины, скажем Y , принимающей только положительные значения. Предположим, например, что требуется аппроксимировать плотность вероятности положительной случайной величиной Y с помощью плотности вероятности случайной величины $a\chi_v^2$, где a и v пока не определены. Предполагается, что первые два момента Y даны. Тогда, если их приравнять первым двум моментам $a\chi_v^2$, которые можно вывести из (3.3.6), то получим

$$\begin{aligned} E[Y] &= av, \\ \text{Var}[Y] &= 2a^2v. \end{aligned}$$

Решая эти уравнения относительно a и v , получаем

$$v = \frac{2(E[Y])^2}{\text{Var}[Y]}, \quad (3.3.14)$$

$$a = \frac{E[Y]}{v}, \quad (3.3.15)$$

откуда получаем параметры аппроксимирующего χ^2 -распределения, выраженные через первые два момента Y .

Теорема разложения для случайных величин, подчиняющихся χ^2 -распределению. Предположим, что случайная величина χ_v^2 разлагается на k случайных величин $\chi_{v_i}^2$ в виде

$$\chi_v^2 = \chi_{v_1}^2 + \chi_{v_2}^2 + \dots + \chi_{v_k}^2.$$

Тогда можно показать [4, 7*], что если

$$v_1 + v_2 + \dots + v_k = v, \quad (3.3.16)$$

то $\chi_{v_i}^2$ взаимно независимы. Обратное, если $\chi_{v_i}^2$ независимы, то имеет место (3.3.16).

Простое приложение этого очень важного свойства состоит в следующем. Предположим, что X_1, X_2, \dots, X_n — n независимых случайных величин, распределенных как $N(0, 1)$. Тогда

$$\sum_{i=1}^n X_i^2 = (\sqrt{n} \bar{X})^2 + \sum_{i=1}^n (X_i - \bar{X})^2.$$

В левой части равенства стоит случайная величина χ_n^2 , а первый член в правой части, будучи квадратом величины, распределенной как $N(0, 1)$, является величиной χ_1^2 . Теперь можно проверить, что случайные величины \bar{X} и $X_i - \bar{X}$ независимы и, следовательно, две случайные величины в правой части независимы. Результат (3.3.16) в этом случае утверждает, что второй член распределен как χ_{n-1}^2 .

ЛИТЕРАТУРА

1. Fisher R. A., Yates F., Statistical Tables, Oliver and Boyd, London, 1938.
2. Parzen E., Modern Probability Theory and its Applications, John Wiley, New York, 1960.
3. Anderson S. L., Box G. E. P., J. Roy. Stat. Soc., **B17**, I (1955).
4. Hald A., Statistical Theory with Engineering Applications, John Wiley, New York, 1952. (Русский перевод: Хальд А., Математическая статистика с техническими приложениями, М., ИЛ, 1956).

МОМЕНТЫ ЛИНЕЙНЫХ ФУНКЦИЙ ОТ СЛУЧАЙНЫХ ВЕЛИЧИН

Результаты, выведенные в разд. 3.2.3 для линейных функций от случайных величин, можно записать изящнее в матричных обозначениях. Таким образом, если

$$\lambda' = (\lambda_1, \lambda_2, \dots, \lambda_n), \quad \mathbf{X}' = (X_1, X_2, \dots, X_n)$$

являются строками-векторами, а штрих обозначает транспонированную матрицу, то результат (3.2.15)

$$E \left[\sum_{i=1}^n \lambda_i X_i \right] = \sum_{i=1}^n \lambda_i E [X_i] \quad (\text{ПЗ.1.1})$$

можно записать в матричной форме в виде

$$E [\lambda' \mathbf{X}] = \lambda' E [\mathbf{X}] = \lambda' \boldsymbol{\mu}, \quad (\text{ПЗ.1.2})$$

где $\boldsymbol{\mu}' = E [\mathbf{X}'] = (E [X_1], E [X_2], \dots, E [X_n])$.

Аналогично результат (3.2.17)

$$\text{Var} \left[\sum_{i=1}^n \lambda_i X_i \right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov} [X_i, X_j] \quad (\text{ПЗ.1.3})$$

можно записать в матричной форме в виде

$$\text{Var} [\lambda' \mathbf{X}] = E [\lambda' (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \lambda] = \lambda' \mathbf{V} \lambda, \quad (\text{ПЗ.1.4})$$

где $\mathbf{V} = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']$ называется матрицей ковариаций случайных величин X_i . Эта матрица имеет вид

$$\mathbf{V} = \begin{pmatrix} \text{Var} [X_1] & \text{Cov} [X_1, X_2] & \dots & \text{Cov} [X_1, X_n] \\ \text{Cov} [X_2, X_1] & \text{Var} [X_2] & \dots & \text{Cov} [X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov} [X_n, X_1] & \text{Cov} [X_n, X_2] & \dots & \text{Var} [X_n] \end{pmatrix}. \quad (\text{ПЗ.1.5})$$

Матрица ковариаций обладает следующими свойствами: 1) так как $\text{Cov} [X_i, X_j] = \text{Cov} [X_j, X_i]$, то матрица \mathbf{V} симметрична, т. е. $\mathbf{V} = \mathbf{V}'$;

2) так как дисперсия случайной величины всегда неотрицательна, то выражение (ПЗ.1.4) будет всегда неотрицательным при любом выборе λ . Отсюда следует, что матрица \mathbf{V} является неотрицательно определенной, т. е. определители \mathbf{V} и всех ее главных миноров неотрицательны.

Наконец, результат (3.2.21) для ковариации между двумя различными линейными функциями от случайных величин X_i можно записать в виде

$$\text{Cov}[\lambda'X, \nu'X] = \lambda'V\nu. \quad (\text{ПЗ.1.6})$$

ВВЕДЕНИЕ В ТЕОРИЮ СТАТИСТИЧЕСКИХ ВЫВОДОВ

Статистика представляет собой науку обработки данных: как собрать нужный вид данных, как проанализировать их и как использовать результаты анализа для того, чтобы дать разумные практические рекомендации. Раздел статистики, имеющий дело с развитием общих методов анализа данных, называется *теорией статистических выводов*.

В свою очередь теория статистических выводов состоит из двух частей: критериев значимости и теории оценивания.

В критерии значимости имеющийся набор данных проверяется таким образом, чтобы можно было дать ответ, согласуется ли он с конкретной гипотезой относительно некоторой случайной величины, например является ли эта величина нормально распределенной с данным средним значением μ и данным стандартным отклонением σ . В теории оценивания данные используются для оценки значений параметров некоторой предполагаемой плотности вероятности этой случайной величины и для определения точности выборочных оценок. Последний подход обычно лучше соответствует практическим запросам, чем ограниченный ответ типа «да — нет», даваемый критерием значимости.

В этой главе мы будем различать два подхода к теории статистических выводов, а именно метод *выборочных распределений* (sampling distribution approach) и метод *правдоподобия*. Частным случаем метода правдоподобия, имеющим фундаментальную важность при оценивании спектров мощности, является *теория наименьших квадратов*, обсуждаемая в разд. 4.3. Метод правдоподобия идеально подходит для ситуаций, где по данным нужно оценить небольшой набор параметров. Обладая этим качеством, он не подходит непосредственно для оценивания спектров мощности, которые содержат по существу бесконечное число параметров. Единственный подход, который возможен в этом случае, заключается в использовании выборочного распределения. Однако мы включили метод правдоподобия в эту главу из-за его важности при оценивании параметров в параметрических моделях.

4.1. ИСТОРИЧЕСКОЕ РАЗВИТИЕ ТЕОРИИ СТАТИСТИЧЕСКИХ ВЫВОДОВ

Теория вероятностей развивалась, чтобы предсказывать *до проведения* эксперимента вероятность того, что случайная величина X лежит между двумя значениями x_1 и x_2 . По мере развития теории неизбежно стали появляться также и некоторые виды статистических выводов. Статистические выводы имеют дело с задачей, являющейся обратной по отношению к задаче теории вероятностей, а именно: как использовать данные x_1, x_2, \dots, x_n *после* эксперимента для того, чтобы сделать выводы о свойствах случайной величины X . Предположим, например, что в результате 15 бросаний монеты мы получили 12 гербов и требуется узнать, совместим ли этот результат с предположением о симметричности монеты. Классическое решение этой задачи представляет собой пример одного из ранних способов получения выводов, известного теперь под названием *критерия значимости*. Решение использует исключительно вероятностные понятия и состоит в вычислении вероятности получения 12 или более гербов при допущении гипотезы, что монета симметрична. Если эта вероятность мала, то она может рассматриваться как веский признак того, что предположение о симметричности монеты ложно; если вероятность велика, то этот результат не противоречит гипотезе о том, что монета симметрична. В упомянутом выше примере вероятность получить 12 или более гербов в 15 бросаниях в предположении, что монета симметрична, равна 0,018, из чего можно заключить, что монета несимметрична.

Другим давно известным способом получения выводов был метод наименьших квадратов, открытый Карлом Фридрихом Гауссом (1777—1855), когда он занимался определением орбит комет по данным наблюдений. В этой задаче положение орбиты дается принятой формой функциональной зависимости, включающей некоторые измеренные *величины* и некоторые фиксированные константы, или *параметры* орбиты. Задача оценивания, рассмотренная Гауссом, состояла в определении наилучших оценок этих параметров по данным наблюдений и в нахождении некоторой меры точности этих оценок.

За исключением работы Гаусса, положившей начало исследованиям в этом направлении, большая часть теории статистических выводов была развита в XX веке. В большинстве случаев она возникала в тех областях, которые обычно называют нефизическими науками, таких, например, как биология, генетика и сельское хозяйство. В этих областях экспериментальные единицы крайне изменчивы, например животные, на которых испытывают лекарства или корма, или земля, на которой сравниваются разные сорта пшеницы. Из-за этой большой изменчивости существенный прогресс в экспериментировании был невозможен без развития сложных ме-

тодов статистического анализа и планирования экспериментов, направленных на соби́рание информативных данных. С другой стороны, проникновение статистических методов в физические науки происходило медленно. Например, в экспериментальной физике можно с помощью значительных затрат и сложных методов снизить изменчивость от эксперимента к эксперименту настолько, что статистическими вопросами можно будет пренебречь.

В противоположность этому в инженерных исследованиях масштаб экспериментирования гораздо шире — от исследовательских работ в лаборатории через стадию опытных установок до промышленных экспериментов в большом масштабе. В такой ситуации адекватный контроль условий или невозможен, или же неэкономичен, вследствие чего применение статистических методов является жизненно необходимым.

Задачи, встречающиеся в этих экспериментальных областях, стимулировали развитие теории статистических выводов. В этой главе мы обсудим два важных подхода к этой теории. Первый из них имеет своим источником теорию вероятностей и называется *методом выборочных распределений*. Источником второго является теория наименьших квадратов, и он называется *методом правдоподобия*.

4.2. ПРИМЕНЕНИЕ МЕТОДА ВЫБОРОЧНЫХ РАСПРЕДЕЛЕНИЙ К СТАТИСТИЧЕСКИМ ВЫВОДАМ

В этом разделе будет показано, как метод выборочных распределений можно применить, во-первых, к задачам оценивания, а во-вторых, к критериям значимости.

4.2.1. Основной метод

В гл. 3 было показано, что *прежде* чем получить выборку наблюдений x_1, x_2, \dots, x_n , полезно посмотреть на них как на реализацию случайных величин X_1, X_2, \dots, X_n , определенных на n -мерном выборочном пространстве. С этим выборочным пространством связана плотность вероятности, называемая *выборочным распределением*, которая, вообще говоря, будет зависеть от набора неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$. Например, если случайные величины независимы и нормально распределены со средним значением θ_1 и дисперсией θ_2 , то выборочное распределение, связанное с данными, будет следующим:

$$f_{12\dots n}(x_1, x_2, \dots, x_n; \theta_1, \theta_2) = \\ = \frac{1}{(\sqrt{2\pi\theta_2})^n} \exp\left[-\frac{1}{2\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2\right], \quad (4.2.1)$$

где $\theta_1 = \mu$ и $\theta_2 = \sigma^2$ в использованных раньше обозначениях. Параметры включены в левую часть выражения (4.2.1) выборочной плотности вероятности для того, чтобы показать, что она является функцией не только от x , но также и от неизвестных параметров θ_1 и θ_2 .

Предположим, что даны наблюдения x_1, x_2, \dots, x_n и требуется оценить параметры θ_i совместной плотности вероятности случайных величин X_1, X_2, \dots, X_n . Применение метода выборочных рас-

пределений к задаче оценивания можно резюмировать в трех следующих разделах.

1. *Выбор формы выборочной плотности вероятности.* Сначала делается предположение о разумной форме совместной плотности вероятности наблюдений. Вид этой плотности будет зависеть от

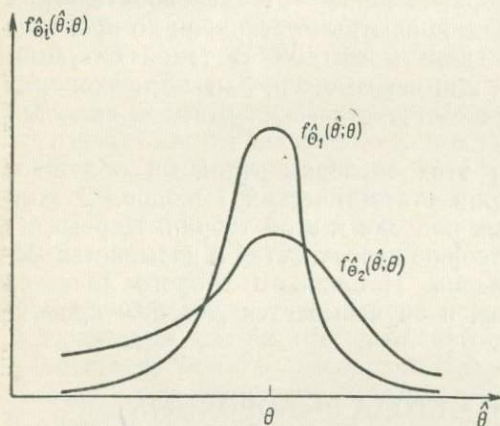


Рис. 4.1. Выборочные распределения для двух оценок.

различных предположений, таких, как независимость случайных величин X_i и вид функций $f_i(x_i)$. Ясно, что решения на этой стадии будут зависеть существенным образом от априорных сведений об изучаемой ситуации. Например, если предположение о независимости неоправдано, то некоторые из параметров совместной плотности вероятности могут описывать зависимость между случайными величинами X_i . В некоторых случаях выводы не очень существенно зависят от предположений, сделанных относительно математической формы совместной плотности вероятности. В других же случаях они могут сильно зависеть от этих предположений, следовательно, требуется некоторое статистическое умение и интуиция для установления точной формы модели.

2. *Выбор оценки.* Функции $\hat{\theta}(X_1, X_2, \dots, X_n)$ от случайных величин рассматриваются как возможные оценки параметра θ . Каждая такая функция, будучи сама случайной величиной, будет иметь выборочное распределение $f_{\hat{\theta}}(\hat{\theta}; \theta)$, зависящее от неизвестной величины θ , которое можно вывести из совместной плотности вероятности данных с помощью методов, описанных в [1]. *Выборочная оценка* $\hat{\theta}(x_1, x_2, \dots, x_n)$, которая получается в конкретном экс-

перименте, рассматривается в этом случае как реализация случайной величины $\hat{\Theta}(X_1, X_2, \dots, X_n)$.

Чтобы сделать выбор между различными оценками, нужно определить критерий оптимальности. Например, из двух оценок $\hat{\Theta}_1$ и $\hat{\Theta}_2$, имеющих выборочные распределения, изображенные на рис. 4.1, была бы выбрана, без сомнения, оценка $\hat{\Theta}_1$, так как $f_{\hat{\Theta}_1}(\hat{\theta}; \theta)$ теснее сосредоточена около истинного значения θ , чем $f_{\hat{\Theta}_2}(\hat{\theta}; \theta)$. Следовательно, для любой заданной выборки $\hat{\Theta}_1$ будет ближе к θ с большей вероятностью, чем $\hat{\Theta}_2$. Следовательно, если бы выборочные распределения двух оценок были известны точно, то выбор между ними можно было бы сделать, сравнивая вероятности того, что они находятся ближе к истинному значению θ . Однако в большинстве приложений невозможно вычислить точно выборочные распределения. В таких случаях нужно менее детально описывать оценку, например с помощью ее младших моментов.

Были предложены различные критерии, основанные на моментах. Они могут быть использованы для сравнения данных оценок. Важнейшим из этих критериев является критерий среднеквадратичной ошибки, обсуждаемый в разд. 4.2.3. *Оценки максимального правдоподобия*, обсуждаемые в разд. 4.2.4, образуют класс оценок, имеющих наименьшую среднеквадратичную ошибку для выборок большого объема.

3. *Доверительные интервалы*. Используя выборочное распределение отобранной оценки $\hat{\Theta}$ или приближение к ее выборочному распределению, основанное на младших моментах, можно делать вероятностные утверждения относительно $\hat{\Theta}$, такие, например, как

$$\Pr\{-l_1 < \hat{\Theta} - \theta \leq l_2\} = 1 - \alpha,$$

или, что то же самое,

$$\Pr\{(\hat{\Theta} - l_2 \leq \theta) \text{ и } (\hat{\Theta} + l_1 > \theta)\} = 1 - \alpha.$$

Следовательно, вероятность того, что случайный интервал между $\hat{\Theta} - l_2$ и $\hat{\Theta} + l_1$ накроет истинное значение θ , равна $1 - \alpha$. Соответствующий интервал, основанный на выборочной оценке, а именно $(\hat{\theta} - l_2, \hat{\theta} + l_1)$, называют в этом случае *доверительным интервалом* для θ с коэффициентом доверия $1 - \alpha$. Это означает, что такой интервал будет покрывать истинное значение в среднем в $100(1 - \alpha)\%$ всех случаев.

Построение доверительных интервалов является одной из важнейших задач процесса оценивания. Оно обсуждается в разд. 4.2.2. В тех случаях, когда невозможно построить точные доверительные интервалы, очень ценно получить хотя бы приближенные доверительные интервалы, определяющие грубо точность оценки. Метод получения приближенных доверительных интервалов приводится в разд. 4.2.4.

Обсуждение. Следует подчеркнуть логику метода выборочных распределений. Выборочное распределение $f_{\hat{\theta}}(\hat{\theta}; \theta)$ можно использовать для вычисления вероятности того, что значение $\hat{\theta}$ случайной величины $\hat{\theta}$ лежит между двумя пределами для всех возможных выборок объема n , предполагая, что параметр θ известен. Следовательно, как обсуждалось в гл. 3, распределение вероятностей дает нам возможность на основании общей модели судить о частной выборке. Однако цель теории оценивания состоит в том, чтобы использовать выборочную оценку $\hat{\theta}$ для получения утверждений относительно θ , т. е. судить на основании выборки о правильности модели. С этой точки зрения применение метода выборочных распределений в теории оценивания является искусственным в том смысле, что необходимо рассматривать не только конкретную доступную выборку, но и все другие выборки, которые могли бы быть получены. Тем не менее метод выборочных распределений важен и по своему историческому значению, и по следующим причинам.

1. Во многих случаях он приводит к заключениям, очень похожим на те, которые достигаются с помощью других способов получения выводов, таких, например, как метод правдоподобия, описываемый ниже.

2. В ситуациях, где имеет место повторная выборка, например при проверке промышленных деталей, метод, включающий рассмотрение всех возможных выборок, логичен. Однако это уже относится к области теории статистических решений, а не к теории статистических выводов.

3. В тех случаях, когда проблему нельзя свести к задаче оценивания небольшого набора параметров (как, например, в спектральном анализе, включающем оценивание большого числа параметров), метод выборочных распределений дает, по-видимому, единственно возможный подход к задаче.

4.2.2. Доверительные интервалы

Доверительный интервал для среднего значения. Чтобы проиллюстрировать метод выборочных распределений и продемонстри-

ровать, как строятся доверительные интервалы, предположим, что требуется оценить среднее значение μ для данных о токах транзисторов, приведенных на рис. 3.3, используя только девять наблюдений. Будем действовать в три этапа, как указывалось в разд. 4.2.1. На первом этапе нам нужно сделать предположение относительно формы плотности вероятности, которая должна быть связана с наблюдениями. Исходя из гистограммы рис. 3.6, разумно предположить, что наблюдения можно описать с помощью нормальной плотности вероятности. Кроме того, поскольку транзисторы выбирались в случайные моменты времени с поточной линии, разумно считать, что случайные величины независимы. Поэтому предполагалось, что выборочное распределение, связанное с наблюдениями, имеет вид (4.2.1), где $\mu = \theta_1$ и $\sigma^2 = \theta_2$ оба неизвестны.

Второй этап заключается в выборе оценки для среднего значения μ . В качестве оценок можно было бы выбрать среднее арифметическое

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n) \quad (4.2.2)$$

и медиану, которая является «средним» наблюдением в выборке. Например, медиана для приведенных ниже данных о транзисторах равна 3,12. Можно показать [5], что для выборочной плотности вероятности (4.2.1) «наилучшей» оценкой является среднее \bar{X} потому, что оно имеет наименьшую среднеквадратичную ошибку и вероятность его нахождения в заданной близости от μ является наибольшей.

Третий этап состоит в определении доверительного интервала для μ , основанного на выбранной оценке \bar{X} . Как показано в разд. 3.3.3, если σ неизвестно, следует использовать случайную величину

$$T_v = \frac{\sqrt{n}(\bar{X} - \mu)}{S},$$

выборочное распределение которой является t -распределением Стьюдента с $v = n - 1$ степенями свободы. Следовательно,

$$\Pr \left\{ -t_v \left(1 - \frac{\alpha}{2} \right) < \frac{\sqrt{n}(\bar{X} - \mu)}{S} \leq t_v \left(1 - \frac{\alpha}{2} \right) \right\} = 1 - \alpha$$

или

$$\Pr \left\{ \bar{X} - t_v \left(1 - \frac{\alpha}{2} \right) \frac{S}{\sqrt{n}} \leq \mu \text{ и } \bar{X} + t_v \left(1 - \frac{\alpha}{2} \right) \frac{S}{\sqrt{n}} > \mu \right\} = 1 - \alpha.$$

Таким образом, вероятность того, что интервал $\bar{X} \pm t_v [1 - (\alpha/2)] (S/\sqrt{n})$ накроет истинное значение μ , равна $1 - \alpha$. Следовательно, $100(1 - \alpha)\%$ -ный доверительный интервал для μ ,

основанный на выборочных оценках \bar{x} и s , полученных по данной выборке, имеет вид

$$\bar{x} \pm t_{\alpha/2} \left(1 - \frac{\alpha}{2}\right) \frac{s}{\sqrt{n}}. \quad (4.2.3)$$

Возвращаясь к нашему примеру, предположим, что значения тока в *мкА* для девяти транзисторов, выбранных случайно с поточной линии, равны

1,73	3,81	3,12
3,00	3,48	1,68
3,64	4,91	0,39

так что

$$\sum x = 1,73 + 3,00 + \dots + 0,39 = 25,76$$

и

$$\sum x^2 = (1,73)^2 + (3,00)^2 + \dots + (0,39)^2 = 88,6860.$$

Таким образом,

$$\bar{x} = \frac{25,76}{9} = 2,86$$

и

$$s^2 = \frac{88,686 - 9(2,86)^2}{8} = \frac{6,89}{8} = 0,86, \quad s = 0,93.$$

Отсюда, используя (4.2.3) при $\bar{x} = 2,86$; $s = 0,93$; $n = 9$ и значение $t_8(0,975) = 2,31$ с рис. 3.11, получаем, что 95%-ный доверительный интервал для μ имеет вид

$$(2,86 - 0,77 \cdot 0,93; \quad 2,86 + 0,77 \cdot 0,93),$$

т. е. (2,15; 3,57).

Интерпретация этого 95%-ного доверительного интервала заключается в том, что если для большого числа повторных выборок строить такие интервалы, то в 95% всех случаев они будут накрывать истинное значение μ .

Заметим, что для этого примера можно построить бесконечное число доверительных интервалов с коэффициентом доверия $1 - \alpha$. В этом случае, выбирая интервал, симметричный относительно выборочного среднего, мы получим самый короткий интервал.

Доверительные интервалы для дисперсии. Чтобы построить доверительный интервал для дисперсии σ^2 нормальной плотности вероятности, воспользуемся тем фактом, что выборочное распределение $(n - 1)S^2/\sigma^2$ совпадает с распределением случайной величины

χ_v^2 . Поэтому, воспользовавшись рис. 3.10, можно найти такие пределы l_1 и l_2 , что

$$\Pr \left\{ l_1 < \frac{\sigma^2}{S^2} \leq l_2 \right\} = 1 - \alpha$$

или

$$\Pr \{ (l_2 S^2 \geq \sigma^2) \text{ и } (l_1 S^2 < \sigma^2) \} = 1 - \alpha$$

Следовательно, $100(1 - \alpha)\%$ -ный доверительный интервал, основанный на выборочной оценке s^2 , имеет вид

$$(l_1 s^2, l_2 s^2). \quad (4.2.4)$$

Для данных о токе коллектора $s^2 = 0,86$; воспользовавшись рис. 3.10, получаем $l_1 = 0,36$; $l_2 = 5,95$, если $\alpha = 0,05$. Следовательно, 95% -ный доверительный интервал для σ^2 имеет вид $(0,36 \cdot 0,86; 5,95 \cdot 0,86)$, т. е. $(0,31; 5,11)$. Отсюда следует, что 95% -ный доверительный интервал для σ имеет вид $(0,56; 2,26)$.

Доверительные интервалы для отношения двух дисперсий. Если S_1^2 является оценкой для σ_1^2 с v_1 степенями свободы, а S_2^2 — независимая оценка σ_2^2 с v_2 степенями свободы, то, как показано в разд. 3.3.4, выборочное распределение случайной величины

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

является F_{v_1, v_2} -распределением Фишера. Следовательно, как показано в разд. 3.3.4,

$$\Pr \left\{ \frac{1}{f_{v_2, v_1} \left(1 - \frac{\alpha}{2}\right)} < F_{v_1, v_2} \leq f_{v_1, v_2} \left(1 - \frac{\alpha}{2}\right) \right\} = 1 - \alpha. \quad (4.2.5)$$

Подставив $F_{v_1, v_2} = \sigma_2^2 S_1^2 / \sigma_1^2 S_2^2$ в (4.2.5) и перегруппировав члены, получим

$$\Pr \left\{ \left(\frac{S_2^2}{S_1^2} \frac{1}{f_{v_2, v_1} \left(1 - \frac{\alpha}{2}\right)} < \frac{\sigma_2^2}{\sigma_1^2} \right) \text{ и } \left(\frac{S_2^2}{S_1^2} f_{v_1, v_2} \left(1 - \frac{\alpha}{2}\right) \geq \frac{\sigma_2^2}{\sigma_1^2} \right) \right\} = 1 - \alpha.$$

Следовательно, $100(1 - \alpha)\%$ -ный доверительный интервал, основанный на выборочных оценках s_1^2 и s_2^2 , полученных из двух независимых выборок, имеет вид

$$\left\{ \frac{s_2^2}{s_1^2} \frac{1}{f_{v_2, v_1} \left(1 - \frac{\alpha}{2}\right)}, \frac{s_2^2}{s_1^2} f_{v_1, v_2} \left(1 - \frac{\alpha}{2}\right) \right\}. \quad (4.2.6)$$

Например, для данных о токах транзисторов $s_1^2 = 0,86$, а число степеней свободы $\nu_1 = 8$. Для выборочной дисперсии другой выборки из 100 транзисторов была получена величина $s_2^2 = 1,025$, причем число степеней свободы $\nu_2 = 99$. Из рис. 3.12 находим, что $f_{8,99}(0,95) = 2,05$; $f_{99,8}(0,95) = 3,01$, и поэтому, подставив эти значения и отношение $s_2^2/s_1^2 = 1,025/0,86 = 1,16$ в (4.2.6), получаем 90%-ный доверительный интервал для σ_2^2/σ_1^2 :

$$\left(1,16 \frac{1}{3,01}; 1,16 \cdot 2,05\right),$$

или же (0,39; 2,38). Так как этот доверительный интервал включает отношение $\sigma_2^2/\sigma_1^2 = 1$, то возможность $\sigma_1^2 = \sigma_2^2$ не должна исключаться.

4.2.3. Свойства оценок

В разд. 4.2.1 было показано, что лучшую оценку параметра можно выявить, сравнивая выборочные распределения различных оценок. Иногда невозможно вывести точное выборочное распределение, и в таких случаях необходимо прибегнуть к помощи приближенных методов для отбора оценок. Эти методы используют свойства, определяемые младшими моментами оценок. Важнейшими из этих свойств являются смещение, дисперсия и среднеквадратичная ошибка.

Смещение. Смещение оценки $\hat{\theta}$ параметра θ определяется как

$$B = E[\hat{\theta}] - \theta. \quad (4.2.7)$$

Если $B = 0$, то плотность вероятности оценки имеет своим центром в точности истинное значение θ , и оценка называется *несмещенной*. Естественно выбирать оценку с малым или нулевым смещением, однако, как мы вскоре увидим, не всегда разумно настаивать на том, чтобы оценка была несмещенной.

Дисперсия. Дисперсия оценки

$$\text{Var}[\hat{\theta}] = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right] \quad (4.2.8)$$

измеряет рассеяние плотности вероятности случайной величины $\hat{\theta}$ относительно ее математического ожидания, и, следовательно, вообще говоря, дисперсия должна быть небольшой. Однако требования малого смещения и малой дисперсии не обязательно совместимы, и часто уменьшение одной из этих величин влечет за собой увеличение другой. Рассмотрим, например, оценки

$$S_k^2 = \frac{1}{k} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (4.2.9)$$

для дисперсии σ^2 нормальной плотности вероятности. Так как $(1/\sigma^2) \sum_{i=1}^n (X_i - \bar{X})^2$ является случайной величиной χ^2_ν с $\nu = n - 1$, то, воспользовавшись (4.2.7) и (3.3.6), получаем, что смещение этой оценки равно

$$B_k = \left(\frac{n-1-k}{k} \right) \sigma^2, \quad (4.2.10)$$

а из (3.3.6) получаем, что

$$\text{Var} [S_k^2] = \frac{2(n-1)}{k^2} \sigma^4. \quad (4.2.11)$$

Таким образом, несмещенная оценка для σ^2 получается при $k = n - 1$, и в этом случае

$$\text{Var} [S_{n-1}^2] = \frac{2\sigma^4}{n-1}.$$

С другой стороны, дисперсию оценки (4.2.9) можно уменьшить, сделав k большим. Однако увеличение k приводит к увеличению смещения, которое стремится к $-\sigma^2$, когда $k \rightarrow \infty$. Ясно, что необходимо найти компромисс между дисперсией и смещением.

Среднеквадратичная ошибка. Один из видов компромисса между дисперсией и смещением дает минимизация *среднеквадратичной ошибки* оценки, а именно

$$E [(\hat{\theta} - \theta)^2] = \text{Var} [\hat{\theta}] + B^2. \quad (4.2.12)$$

Для упомянутого выше примера среднеквадратичная ошибка равна

$$\frac{2(n-1)}{k^2} \sigma^4 + \frac{(n-1-k)^2}{k^2} \sigma^4.$$

Это выражение достигает минимального значения $2\sigma^4/(n+1)$ при $k = n+1$ по сравнению со среднеквадратичной ошибкой $2\sigma^4/(n-1)$ для несмещенной ($k = n-1$) оценки.

В некоторых случаях среднеквадратичная ошибка достигает минимума при нулевом смещении, т. е. одновременно с дисперсией. Такие оценки называются *несмещенными оценками с минимальной дисперсией*.

Одна из трудностей, связанных с использованием критерия среднеквадратичной ошибки, состоит в том, что он дает нам возможность лишь сравнить данные классы оценок, но он не говорит нам, как следует выбирать эти оценки. Впрочем, один класс оценок, удовлетворяющих свойству минимальности среднеквадратичной ошибки для больших выборок, можно найти из функции

правдоподобия, введенной Фишером. Эти оценки обсуждаются в разд. 4.2. Они сыграли важную роль в статистическом оценивании, так как становятся несмещенными для выборок большого объема и имеют также минимальную дисперсию среди всех возможных оценок. Следовательно, для выборок большого объема оценки максимального правдоподобия являются оценками с минимальной среднеквадратичной ошибкой.

Состоятельность. Другим свойством оценок, опирающихся на выборочное распределение, является *состоятельность*. Предположим, что смещение и дисперсия оценки стремятся к нулю, когда объем выборки n становится большим. Это означает, что выборочное распределение концентрируется вокруг θ и точность оценки безгранично возрастает. Оценка, обладающая этим свойством, называется *состоятельной оценкой*.

Например, если выборочное распределение стремится к нормальному, что обычно справедливо при довольно общих условиях, то оно будет для больших n близким к

$$f_{\hat{\theta}}(\hat{\theta}; \theta) = \frac{\sqrt{n}}{\sqrt{2\pi}c} \exp\left[-\frac{n}{2c^2} [(\hat{\theta} - \theta)^2]\right].$$

Когда n стремится к бесконечности, эта функция ведет себя подобно δ -функции, сосредоточенной в θ .

4.2.4. Оценки максимального правдоподобия

Функции правдоподобия, зависящие от одного переменного. Задача нахождения хорошей оценки для статистического параметра была решена для многих случаев Фишером [2, 3], который ввел класс *оценок максимального правдоподобия*. Чтобы проиллюстрировать этот метод, рассмотрим задачу оценки среднего срока службы партии осветительных ламп. Предполагается, что срок службы одной лампы хорошо описывается с помощью случайной величины X с плотностью вероятности

$$f_X(x; \lambda) = \lambda \exp(-\lambda x), \quad 0 \leq x < \infty.$$

Отсюда выборочная плотность вероятности для случайной выборки, состоящей из n ламп, будет иметь вид

$$f_{12 \dots n}(x_1, x_2, \dots, x_n) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right). \quad (4.2.13)$$

До того как произведен эксперимент, плотность вероятности (4.2.13) дает частоту получения различных выборок при условии, что λ задано. После того как эксперимент произведен, его можно интерпретировать по-разному. В нашем случае значения выборки

x_1, x_2, \dots, x_n известны, а параметр λ неизвестен. Зависящая от λ функция, которая получается при подстановке выборочных значений в плотность вероятности (4.2.13), называется *функцией правдоподобия* $L(\lambda)$ для параметра λ . Она выражает предпочтительность различных значений λ .

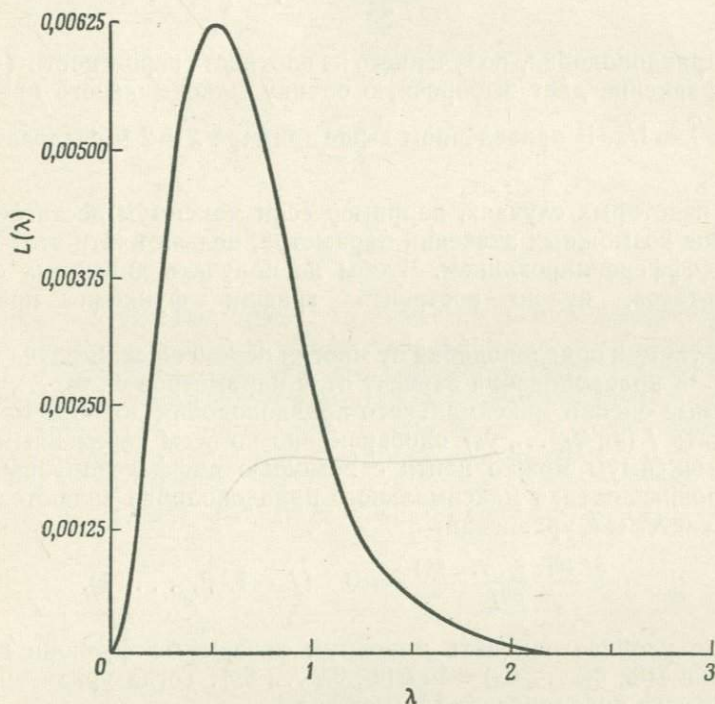


Рис. 4.2. Функция правдоподобия для выборки объема 3 из показательного распределения.

Например, предположим, что три лампы выбраны случайным образом из партии, проверены, и в результате проверки оказалось, что их сроки службы равны 2,6; 1,9 и 1,5 час соответственно. Так как $\sum x_i = 6$, то функция правдоподобия имеет вид

$$L(\lambda) = \lambda^3 \exp(-6\lambda). \quad (4.2.14)$$

График функции (4.2.14) приведен на рис. 4.2. Он представляет собой одновершинную кривую с максимумом при $\lambda = 0,5$. Значение $\hat{\lambda}$ параметра λ , которое максимизирует $L(\lambda)$, называется *выборочной оценкой максимального правдоподобия* параметра λ . Она дает

предпочтительное значение параметра λ , поскольку при этом значении вероятность получения данной выборки максимальна.

Как правило, для гладкой функции правдоподобия оценку максимального правдоподобия можно получить, решая уравнение

$$\frac{dL(\lambda)}{d\lambda} = 0.$$

Для правдоподобия, полученного из плотности вероятности (4.2.13), это уравнение дает выборочную оценку максимального правдоподобия $\hat{\lambda} = 1/\bar{x}$. В приведенном выше примере $\bar{x} = 2$ и, следовательно, $\hat{\lambda} = 0,5$.

В некоторых случаях, например если максимум достигается на границе возможных значений параметра, нельзя найти этот максимум дифференцированием. Чтобы не получать лишнего смысла результатов, нужно построить график функции правдоподобия.

Функции правдоподобия от многих переменных. В случае, когда функция правдоподобия зависит от k параметров $\theta_1, \theta_2, \dots, \theta_k$, выборочные оценки максимального правдоподобия должны максимизировать $L(\theta_1, \theta_2, \dots, \theta_k)$ одновременно по всем переменным. Если этот максимум можно найти с помощью дифференцирования, то выборочные оценки максимального правдоподобия являются решением системы k уравнений:

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_j} = 0 \quad (j=1, 2, \dots, k). \quad (4.2.15)$$

Иногда удобнее находить максимум логарифма функции правдоподобия $l(\theta_1, \theta_2, \dots, \theta_k) = \ln L(\theta_1, \theta_2, \dots, \theta_k)$. Тогда уравнения максимального правдоподобия имеют вид

$$\frac{\partial l}{\partial \theta_j} = \frac{\partial (\ln L)}{\partial \theta_j} = \frac{1}{L} \frac{\partial L}{\partial \theta_j} = 0 \quad (j=1, 2, \dots, k). \quad (4.2.16)$$

Пример 1. Рассмотрим функцию правдоподобия для среднего значения и дисперсии нормальной плотности вероятности, причем предполагается, что выборка состоит из n наблюдений:

$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi} \sigma)^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^n (x_t - \mu)^2 \right], \quad (4.2.17)$$

так что логарифмическая функция правдоподобия имеет вид

$$l(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{t=1}^n (x_t - \mu)^2. \quad (4.2.18)$$

Выборочные оценки максимального правдоподобия, получаемые из (4.2.18), являются решениями системы уравнений

$$\begin{aligned} \frac{1}{\hat{\sigma}^2} \sum_{t=1}^n (x_t - \hat{\mu}) &= 0, \\ -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{t=1}^n (x_t - \hat{\mu})^2 &= 0, \end{aligned} \quad (4.2.19)$$

т. е.

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{t=1}^n x_t = \bar{x}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2. \end{aligned} \quad (4.2.20)$$

Пример 2. Предположим, что имеется n пар измерений (x_{1i}, x_{2i}) , $i = 1, 2, \dots, n$, как это было для данных об акселерометре на рис. 3.7. Если предположить, что они могут быть описаны парой случайных величин, совместная плотность вероятности которых является двумерной нормальной плотностью, то логарифмическая функция правдоподобия для n пар наблюдений имеет вид

$$\begin{aligned} l(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_{12}) &= -n \ln 2\pi - n \ln \sigma_1 - n \ln \sigma_2 - \frac{n}{2} \ln(1 - \rho_{12}^2) - \\ &- \frac{1}{2(1 - \rho_{12}^2)} \sum_{i=1}^n \left\{ \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right)^2 - 2\rho_{12} \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right) \left(\frac{x_{2i} - \mu_2}{\sigma_2} \right) + \right. \\ &\quad \left. + \left(\frac{x_{2i} - \mu_2}{\sigma_2} \right)^2 \right\}. \end{aligned} \quad (4.2.21)$$

Функция правдоподобия (4.2.21) зависит от пяти параметров, и выборочные оценки максимального правдоподобия можно получить, дифференцируя эту функцию по очереди по всем пяти параметрам и решая полученные уравнения. Можно убедиться, что оценки среднего значения и дисперсии те же самые, что и полученные из правдоподобия (4.2.17), а выборочная оценка максимального правдоподобия для коэффициента корреляции ρ_{12} имеет вид

$$\hat{\rho}_{12} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\left\{ \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 \right\}^{1/2}} = r_{12}. \quad (4.2.22)$$

Заметим, что (4.2.22) можно переписать в виде

$$r_{12} = \frac{c_{12}}{\sqrt{c_{11}c_{22}}}, \quad (4.2.23)$$

где

$$c_{12} = \frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \quad (4.2.24)$$

является выборочной оценкой максимального правдоподобия ковариации γ_{12} между двумя рассматриваемыми случайными величинами; c_{11} и c_{22} — выборочные оценки максимального правдоподобия для дисперсий σ_1^2 и σ_2^2 соответственно.

Так как функция правдоподобия является функцией только θ , когда наблюдения известны, то выборочная оценка максимального правдоподобия $\hat{\theta}$ получается непосредственно как функция этих наблюдений. Обычно в этом месте в статистических работах оставляют функцию правдоподобия и возвращаются к методу выборочных распределений. При этом с выборочной оценкой $\hat{\theta}$ связывают оценку $\hat{\Theta}$ и находят ее выборочные свойства. Этот подход совместим с подходом выборочных распределений к оцениванию, но это не совпадает с использованием метода правдоподобия для выводов, о чем будет сказано в разд. 4.4.

Выборочные свойства оценок максимального правдоподобия приведены в работе [5]. Наиболее важное из них заключается в том, что для больших n оценки максимального правдоподобия приближенно несмещенные и распределены асимптотически нормально с дисперсией

$$\text{Var} [\hat{\theta}] \approx \frac{-1}{E [\partial^2 l(\theta) / \partial \theta^2]_{\hat{\theta}}}, \quad (4.2.25)$$

являющейся наименьшей дисперсией, которую может иметь любая несмещенная оценка. Поэтому можно построить приближенный доверительный интервал, используя выборочную оценку максимального правдоподобия, дисперсию (4.2.25) и табл. 3.4.

Результат (4.2.25) показывает, что дисперсия оценки максимального правдоподобия обратно пропорциональна второй производной (и, следовательно, кривизне) функции правдоподобия в точке ее максимума. Выражение

$$E \left[- \frac{\partial^2 l(\theta)}{\partial \theta^2} \right]_{\hat{\theta}} \quad (4.2.26)$$

называется *количеством информации Фишера* *). Его интерпретацию мы продолжим в разд. 4.4.

4.2.5. Критерии значимости

Другой вид выводов, включаемый в рамки метода выборочных распределений, представляет собой *критерий значимости*. Он дает возможность вынести решение о том, справедлива или нет некоторая гипотеза относительно статистических параметров. Например, иногда нужно проверить, совместима ли некоторая выборка наблюдений x_1, x_2, \dots, x_n с гипотезой о том, что они получены из нормальной плотности вероятности с некоторыми заданными значениями μ_0, σ_0^2 среднего и дисперсии.

Во многих случаях, когда применяют критерии значимости, лучший ответ на задачу можно было бы получить с помощью оценивания параметров и вычисления доверительных интервалов. В этом разделе мы приведем простой пример критерия значимости и затем покажем, как можно было бы получить несколько большую информацию, рассматривая нашу задачу как задачу оценивания.

Понятие критерия значимости восходит к первым работам по теории вероятностей. Систематическая теория критериев значимости была разработана до некоторой степени независимо, с одной стороны, Фишером, а с другой стороны, совместно Нейманом и Пирсоном. Двое последних включили идею критерия значимости в теорию, названную ими *теорией проверки гипотез*. Описание этой теории дается в [4].

Этапы построения критерия значимости. Проиллюстрируем этапы построения критерия значимости на примере с транзисторами из разд. 4.2.2.

1. Выдвигаем нулевую гипотезу H_0 , например, что ток коллектора для партии транзисторов распределен нормально со средним значением μ_0 , но с неизвестной дисперсией.

2. Определяем конкурирующие гипотезы. В нашем примере в качестве таких гипотез было бы естественно взять предположение $\mu > \mu_0$, поскольку желательно было бы забраковать партию, если средний ток коллектора был слишком высоким.

3. Решаем вопрос о наилучшей функции от наблюдаемых данных, или *статистике*, с помощью которой будем проверять гипотезу. Если дисперсия известна, то, как можно показать [4], наилучшей

* Точнее было бы в последней формуле и в (4.2.25) брать производную в точке θ — точке истинного значения параметра. — *Прим. перев.*

статистикой является среднее \bar{X} . Если дисперсия неизвестна, как в нашем примере, то наилучшей статистикой является

$$T_v = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}.$$

4. Выводим выборочное распределение этой статистики при условии, что нулевая гипотеза верна. В нашем примере это будет t -распределение Стьюдента с $v = n - 1$ степенями свободы.

5. Пользуясь (4) и (2), можно затем разделить выборочное пространство \mathcal{S} на две части: *критическую область* \mathcal{C} и *область принятия гипотезы* $\mathcal{S} - \mathcal{C}$, состоящую из всех точек выборочного пространства, не принадлежащих критической области \mathcal{C} . Критическая область выбирается так, что вероятность $\text{Pr}\{x_1, x_2, \dots, x_n \text{ лежит в } \mathcal{C} \mid H_0 \text{ верна}\} = \alpha$, где α мало, скажем 0,05 или 0,01. Вероятность α называется *уровнем значимости критерия*.

6. Наконец, критерий значимости заключается в том, что нулевая гипотеза отбрасывается, если наблюдаемая выборка $x_1, x_2, \dots, \dots, x_n$ попадает в \mathcal{C} , и не отбрасывается, если выборка попадает в $\mathcal{S} - \mathcal{C}$. Поскольку вероятность попадания выборочной точки в \mathcal{C} при условии, что H_0 верна, мала, то любой случай, когда она туда попадает, рассматривается как довод против нулевой гипотезы.

В нашем примере в силу того, что $\text{Pr}\{T_v > t_v(1 - \alpha)\} = \alpha$, критическая область определяется неравенством

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} > t_{n-1}(1 - \alpha)$$

или же

$$\bar{x} > \mu_0 + \frac{st_{n-1}(1 - \alpha)}{\sqrt{n}}.$$

Пример. Предположим, что $n = 4$, $\bar{x} = 10$ и $s = 2$ и нужно проверить гипотезу $\mu_0 = 8$ с уровнем значимости $\alpha = 0,025$. Из рис. 3.11 находим $t_3(0,975) = 3,18$, и, следовательно, критическая область имеет вид

$$\bar{x} > 8 + \frac{2(3,18)}{2} = 11,18.$$

Поскольку настоящее \bar{x} лежит вне критической области, нулевая гипотеза не отвергается с 2,5%-ным уровнем значимости.

Предположим, что конкурирующие гипотезы $\mu > \mu_0$ и $\mu < \mu_0$ одинаково важны. Например, если вес некоторого фасованного товара должен быть равен заданной величине μ_0 , то могли бы быть одинаково важными случаи недовеса и перевеса в конкретной вы-

борке. В таком случае разумно выбрать критическую область в виде

$$t > t_{n-1} \left(1 - \frac{\alpha}{2}\right), \quad t < -t_{n-1} \left(1 - \frac{\alpha}{2}\right),$$

т. е.

$$\bar{x} > \mu_0 + \frac{st_{n-1} \left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}, \quad \bar{x} < \mu_0 - \frac{st_{n-1} \left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}. \quad (4.2.27)$$

Для нашего примера при $\mu_0 = 8$ получаем критическую область

$$\bar{x} > 11,18, \quad \bar{x} < 4,72.$$

Так как наблюдаемая величина $\bar{x} = 10$ не лежит в критической области, то нулевая гипотеза не была бы отвергнута с 5%-ным уровнем значимости. Такой критерий называется двусторонним критерием значимости в противоположность упоминавшемуся выше одностороннему критерию.

Доверительные интервалы и критерии значимости. Чтобы продемонстрировать соотношение между критерием значимости и доверительным интервалом, заметим, что доверительный интервал (4.2.3) для μ имеет вид

$$\bar{x} \pm \frac{st_{n-1} \left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}.$$

Поэтому если μ_0 лежит внутри доверительного интервала, то, согласно (4.2.27), нулевая гипотеза не отвергается, а если μ_0 лежит вне доверительного интервала, то нулевая гипотеза отвергается.

В нашем примере 95%-ный доверительный интервал имеет вид

$$10 \pm \frac{2(3,18)}{\sqrt{4}} = (6,82; 13,18).$$

Так как $\mu_0 = 8$ попадает внутрь этого интервала, то нулевая гипотеза не отвергается с 5%-ным уровнем значимости. На самом деле, никакая нулевая гипотеза из интервала от 6,82 до 13,18 не была бы отвергнута с этим уровнем значимости. Теперь становится очевидной дополнительная информация, содержащаяся в доверительном интервале. Она показывает, что наш эксперимент был настолько неточным, что даже такие большие значения μ , как 13, правдоподобны. В этом случае единственное разумное заключение состоит в том, что требуется больше данных для того, чтобы оценить μ точнее.

4.3. ОЦЕНИВАНИЕ С ПОМОЩЬЮ НАИМЕНЬШИХ КВАДРАТОВ

4.3.1. Принцип наименьших квадратов

Принцип наименьших квадратов был открыт немецким математиком Карлом Фридрихом Гауссом, который опубликовал свою первую работу по этому вопросу в 1821 г. и затем возвращался к нему неоднократно в течение всей своей жизни. Его принцип наименьших квадратов представляет собой одно из первых крупных достижений в статистике, и даже на сегодняшний день он является одним из самых мощных методов, имеющихся в распоряжении статистиков.

Предположим, что выход η некоторой системы может быть предсказан по k входным переменным x_1, x_2, \dots, x_k с помощью некоторой предполагаемой линейной модели

$$\eta = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k. \quad (4.3.1)$$

Например, η могло бы быть выходом некоторого химического процесса, x — переменными процесса, такими, как температуры, давления и скорости потоков, а $\theta_1, \theta_2, \dots, \theta_k$ — неизвестными физическими параметрами, такими, как кинетические константы.

Линейная теория наименьших квадратов имеет дело с оцениванием параметров θ_r по данным, состоящим из одновременных измерений входных и выходных переменных. Значения, полученные в результате оценки параметров, можно подставить в (4.3.1) и полученное при этом выражение использовать для предсказания выхода при тех значениях входных переменных, которые появятся в будущем.

Заметим, что уравнение прогноза (4.3.1) не обязательно должно быть линейным по x_1, x_2, \dots, x_k , а лишь по параметрам θ . Например, если $x_1 = 1, x_2 = x, \dots, x_k = x^{k-1}$, то η является полиномом по x степени $k - 1$. Если же выход является нелинейной функцией параметров, то описываемые в этом разделе методы легко видоизменить [6] для оценивания параметров с помощью итераций линейного метода наименьших квадратов.

На практике можно наблюдать лишь отклик η , искаженный некоторой ошибкой z . Такое искажение неизбежно из-за ошибок измерения и из-за изменчивости, которую невозможно контролировать. Если модель не вполне соответствует действительности, то ошибка может иметь систематическую компоненту, обусловленную этим несовершенством модели. Поэтому окончательный вид модели следующий:

$$Y_i = \eta_i + Z_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_k x_{ik} + Z_i, \quad (4.3.2)$$

где

а) Y_i ($i = 1, 2, \dots, N$) — случайная величина, соответствующая измеренному отклику y_i в i -м эксперименте;

б) $x_{i1}, x_{i2}, \dots, x_{ik}$ — значения, принимаемые входными переменными x_1, x_2, \dots, x_k в i -м эксперименте;

в) Z_i — случайная величина, представляющая ошибку, причем $E[Z_i] = 0$.

Заметим, что если ошибки имеют отличное от нуля среднее значение θ_1 , то это можно учесть, считая $x_{i1} = 1$ в (4.3.2).

Теорема Гаусса. Подход с помощью метода наименьших квадратов к задаче оценивания содержится в фундаментальной теореме Гаусса. Она утверждает, что если ошибки Z_i некоррелированы, т. е. $\text{Cov}[Z_i, Z_j] = 0$ при $i \neq j$, и имеют нулевое среднее значение $E[Z_i] = 0$ и одинаковую дисперсию $E[Z_i^2] = \sigma^2$, то оптимальными выборочными оценками параметров θ_r являются значения $\hat{\theta}_r$, минимизирующие сумму квадратов расхождений между наблюдаемыми значениями и подбираемой моделью, т. е. сумму квадратов

$$S(\theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^N (y_i - \theta_1 x_{i1} - \theta_2 x_{i2} - \dots - \theta_k x_{ik})^2. \quad (4.3.3)$$

Как показано в приложении П4.1, выборочные оценки $\hat{\theta}_r$ оптимальны в том смысле, что для любой линейной функции

$$L = \lambda_1 \theta_1 + \lambda_2 \theta_2 + \dots + \lambda_k \theta_k$$

оценка

$$\hat{L} = \lambda_1 \hat{\theta}_1 + \lambda_2 \hat{\theta}_2 + \dots + \lambda_k \hat{\theta}_k$$

имеет наименьшую среднеквадратичную ошибку.

Выборочные оценки наименьших квадратов $\hat{\theta}_r$ ($r = 1, 2, \dots, k$) можно получить, дифференцируя $S(\theta_1, \theta_2, \dots, \theta_r, \dots, \theta_k)$ по θ_r и решая получившуюся систему из k уравнений:

$$\sum_{i=1}^N x_{ir} \{ y_i - \hat{\theta}_1 x_{i1} - \dots - \hat{\theta}_k x_{ik} \} = 0, \quad r = 1, 2, \dots, k, \quad (4.3.4)$$

которые обычно называются *нормальными уравнениями*.

Пример. Чтобы проиллюстрировать метод наименьших квадратов Гаусса, рассмотрим ускорение тела, начинающего движение из состояния покоя под действием постоянной силы. Модель в этом случае имеет вид

$$\eta = \theta x,$$

где η — скорость тела по истечении времени x . Был проведен эксперимент, в котором скорости y_i ($i = 1, 2, \dots, N$) тела замерялись в различные моменты времени x_i . Измерение моментов времени

x_i производилось очень точно, в то время как скорость измерялась с ошибкой. Поэтому в качестве вероятностной модели нашего эксперимента можно взять

$$Y_i = \theta x_i + Z_i. \quad (4.3.5)$$

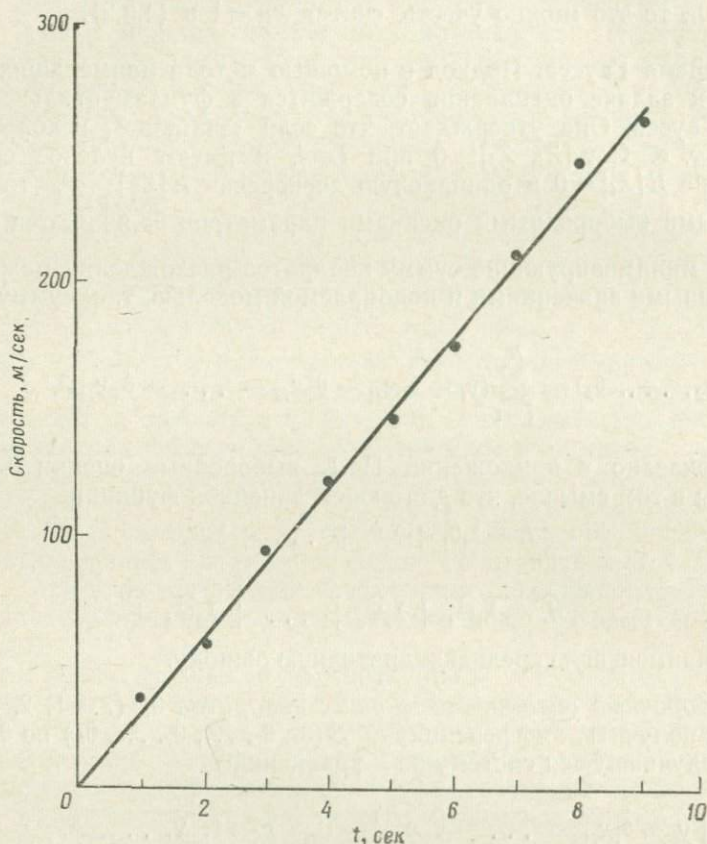


Рис. 4.3. Данные «скорость—время» и линия регрессии, полученная методом наименьших квадратов.

На рис. 4.3 и в табл. 4.1 приведены данные (x_i, y_i) , полученные в действительном эксперименте.

Для этого примера сумма квадратов (4.3.3) имеет вид

$$S(\theta) = \sum_{i=1}^N (y_i - \theta x_i)^2. \quad (4.3.6)$$

Таблица 4.1

Данные «скорость—время» для оценивания ускорения

x_i , сек	1	2	3	4	5	6	7	8	9
y_i , м/сек	35	58	94	121	147	175	212	247	264
Остатки $y_i - \hat{\theta}x_i$	5,0	-1,9	4,1	1,2	-2,8	-4,7	2,3	7,3	-5,6

Дифференцируя эту сумму по θ и приравнявая нулю производную, получаем единственное нормальное уравнение

$$\sum_{i=1}^N x_i (y_i - \hat{\theta}x_i) = 0.$$

Следовательно, выборочная оценка наименьших квадратов имеет вид

$$\hat{\theta} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}. \quad (4.3.7)$$

Для данных, помещенных в табл. 4.1, имеем $\sum x_i y_i = 8538$, $\sum x_i^2 = 285$ и, следовательно,

$$\hat{\theta} = \frac{8538}{285} = 29,96 \text{ м/сек}^2.$$

Подобранная линия $y = \hat{\theta}x$ показана на рис. 4.3. Она называется линией регрессии y на x . Теперь ее можно использовать для предсказания значения скорости y в заданный момент времени x в любых последующих экспериментах при тех же условиях.

4.3.2. Доверительные интервалы для одного параметра

Среднее значение и дисперсия оценки наименьших квадратов. Как отмечалось выше, существенно иметь меру точности оцениваемого параметра, например в виде доверительного интервала. Этот доверительный интервал можно использовать в свою очередь для построения доверительного интервала для прогноза, сделанного по подобранной модели.

В упоминавшемся выше примере доверительные интервалы для θ можно вывести, рассматривая выборочные свойства оценки $\hat{\theta}$,

соответствующей выборочной оценке наименьших квадратов (4.3.7).

Так как x_i являются фиксированными константами, то среднее значение оценки $\hat{\theta}$ равно

$$E[\hat{\theta}] = \frac{\sum_{i=1}^N x_i E[Y_i]}{\sum_{i=1}^N x_i^2} = \frac{\sum_{i=1}^N x_i^2 \theta}{\sum_{i=1}^N x_i^2} = \theta,$$

так что эта оценка несмещенная. Аналогично получаем из (3.2.18), что ее дисперсия равна

$$\text{Var}[\hat{\theta}] = \frac{\sigma^2 \sum_{i=1}^N x_i^2}{\left(\sum_{i=1}^N x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^N x_i^2}, \quad (4.3.8)$$

так как $\text{Var}[Y_i] = \text{Var}[Z_i] = \sigma^2$. Следовательно, если бы σ^2 было известно, то (4.3.8) можно было бы использовать для построения доверительных интервалов для θ , поскольку из того, что Y_i распре-

делены нормально, следует, что $\hat{\theta}$ также распределена нормально. Кроме того, если даже Z_i не являются нормально распределен-

ными, тем не менее $\hat{\theta}$ будет иметь распределение, близкое к нормальному в силу центральной предельной теоремы, и, таким образом, этот анализ будет устойчивым по отношению к предположениям, сделанным о распределении Z_i .

Оценивание остаточной дисперсии. В общем случае нам требуется оценивать σ^2 по данным. Чтобы увидеть, как это можно сделать, рассмотрим

$$\sum Z_i^2 = \sum (Y_i - \theta x_i)^2 = \sum \{Y_i - \hat{\theta} x_i + x_i (\hat{\theta} - \theta)\}^2, \quad (4.3.9)$$

где пределы суммирования временно опущены. Раскрытие скобок в (4.3.9) дает

$$\sum Z_i^2 = \sum (Y_i - \hat{\theta} x_i)^2 + 2(\hat{\theta} - \theta) \sum x_i (Y_i - \hat{\theta} x_i) + (\hat{\theta} - \theta)^2 \sum x_i^2,$$

и так как $\hat{\theta}$ является оценкой наименьших квадратов, средний член исчезает, что дает

$$\sum (Y_i - \theta x_i)^2 = \sum (Y_i - \hat{\theta} x_i)^2 + (\hat{\theta} - \theta)^2 \sum x_i^2. \quad (4.3.10)$$

Беря математическое ожидание от обеих частей (4.3.10), получаем

$$N\sigma^2 = E\left[\sum (Y_i - \hat{\theta} x_i)^2\right] + \text{Var}[\hat{\theta}] \sum x_i^2,$$

и отсюда, используя (4.3.8), получаем

$$E \left[\sum_{i=1}^N (Y_i - \hat{\theta} x_i)^2 \right] = (N-1) \sigma^2.$$

Таким образом, случайная величина

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \hat{\theta} x_i)^2$$

является несмещенной оценкой σ^2 . Поскольку $(N-1)S^2$ является квадратичной формой от нормальных случайных величин и $E[(N-1)S^2] = (N-1)\sigma^2$, отсюда следует, что эта величина распределена как $\sigma^2 \chi_{N-1}^2$.

Результат (4.3.10) является частным случаем теоремы (3.3.16) о разбиении χ^2 . Таким образом, из-за того, что случайные величины Y_i распределены как $N(\theta x_i, \sigma^2)$, левая часть (4.3.10) распределена как $\sigma^2 \chi_N^2$. Кроме того, случайная величина $(\hat{\theta} - \theta)$ распределена как $N(0, \sigma^2 / \sum x_i^2)$ и, следовательно, случайная величина $(\hat{\theta} - \theta)^2 \sum x_i^2$ распределена как $\sigma^2 \chi_1^2$. Можно показать также, что две случайные величины в правой части (4.3.10), имеющие χ^2 -распределение, независимы. Следовательно, случайную величину в левой части (4.3.10), имеющую χ^2 -распределение с N степенями свободы, можно разбить на две независимые случайные величины, имеющие χ^2 -распределение с $N-1$ и с одной степенью свободы соответственно.

Доверительные интервалы для θ . Так как $\hat{\theta} - \theta$ не зависит от S , то отсюда следует, что случайная величина

$$\frac{\sqrt{\sum x_i^2} (\hat{\theta} - \theta)}{S}$$

имеет t -распределение с $\nu = N-1$ степенями свободы. Отсюда $100(1-\alpha)\%$ -ный доверительный интервал для θ имеет вид

$$\hat{\theta} \pm t_{N-1} \left(1 - \frac{\alpha}{2}\right) \frac{s}{\sqrt{\sum x_i^2}}, \quad (4.3.11)$$

где $\hat{\theta}$ дается равенством (4.3.7), и

$$s^2 = \frac{1}{N-1} \sum (y_i - \hat{\theta} x_i)^2$$

является выборочной оценкой дисперсии. Заметим, что

$$\sum (y_i - \hat{\theta} x_i)^2 = \sum y_i^2 - \hat{\theta}^2 \sum x_i^2, \quad (4.3.12)$$

и, следовательно, поскольку $\hat{\theta}$ известно, для вычисления остаточной суммы квадратов и выборочной оценки дисперсии остается сосчитать лишь $\sum y_i^2$.

Для данных, приведенных в табл. 4.1, $\sum y^2 = 255\,949$. Следовательно,

$$s^2 = \frac{1}{8} \{ 255\,949 - (29,96)^2(258) \} = 21,06,$$

так что 95%-ный доверительный интервал для θ равен

$$29,96 \pm \frac{(2,306) \sqrt{21,06}}{\sqrt{285}} = (29,33; 30,58).$$

Полезно также проверить индивидуальные разности от подобранного уравнения регрессии, чтобы посмотреть, не является ли какое-нибудь наблюдение аномальным или же разности укладываются в рассматриваемую схему. Для нашего примера индивидуальные разности $y - \hat{\theta}x$ показаны в третьем ряду табл. 4.1. Мы видим, что они не содержат очевидных выбросов, которые могли бы вызвать сомнение в правильности модели.

Остаточную сумму квадратов (4.3.12) можно переписать также в виде

$$\sum (y_i - \hat{\theta}x_i)^2 = (1 - r^2) \sum y_i^2, \quad (4.3.13)$$

где

$$r = \frac{\sum x_i y_i}{(\sum x_i^2 \sum y_i^2)^{1/2}}$$

является выборочным коэффициентом корреляции между x_i и y_i (при условии, что линия регрессии проходит через начало координат^{*)}. Отсюда (4.3.12) можно записать в виде

$$\sum y_i^2 = (1 - r^2) \sum y_i^2 + r^2 \sum y_i^2. \quad (4.3.14)$$

Результат (4.3.14) показывает, что в этом примере сумма квадратов отклонений y от нуля может быть разбита на составляющую $r^2 \sum y_i^2$, равную сумме квадратов отклонений подобранной прямой от нуля, плюс сумму квадратов разностей между подобранными и наблюдаемыми величинами.

Результат (4.3.14) имеет много аналогов в спектральном анализе, как будет показано в последующих главах.

^{*)} Хотя величины x_i могут и не быть реализациями случайных величин, удобно тем не менее называть r выборочным коэффициентом корреляции между x_i и y_i . — Прим. перев.

Дисперсия прогноза. Если модель (4.3.5) используется для прогноза будущего значения скорости y , соответствующего данному моменту времени x , то наилучшей выборочной оценкой y будет

$$\hat{y} = \hat{\theta} x + \hat{z},$$

где $\hat{z} = 0$ является наилучшей выборочной оценкой ошибки. Соответствующая случайная величина имеет дисперсию

$$\text{Var} [\hat{Y}] = x^2 \text{Var} [\hat{\theta}] + \text{Var} [\hat{Z}] = \sigma^2 \left(\frac{x^2}{\sum x_i^2} + 1 \right).$$

Отсюда $100(1 - \alpha) \%$ -ный доверительный интервал для прогнозируемого значения имеет вид

$$\hat{y} \pm t_{n-1} \left(1 - \frac{\alpha}{2} \right) s \sqrt{1 + \frac{x^2}{\sum x_i^2}}. \quad (4.3.15)$$

Интервал (4.3.15) увеличивается с увеличением x , а также выявляет общее правило, заключающееся в том, что точность прогноза зависит от планирования эксперимента, т. е. от выбора x_i .

4.3.3. Доверительные области для нескольких параметров

Распространение результатов разд. 4.3.2 на случай оценки нескольких параметров наиболее быстро получается с помощью теории матриц. Эти результаты выведены в приложении П4.1, а в настоящем разделе лишь кратко резюмированы. В приложении П4.1 показано, что доверительный интервал заменяется в случае нескольких параметров доверительной областью в k -мерном пространстве параметров θ . Показано также, что еще одна интерпретация оптимальности оценок наименьших квадратов состоит в том, что они минимизируют объем доверительной области для параметров. Для любого отдельного параметра это означает, что оценка наименьших квадратов минимизирует длину доверительного интервала по координате, соответствующей этому параметру.

Для N измерений и k параметров результаты, выведенные в приложении П4.1, можно резюмировать следующим образом.

Нормальные уравнения:

$$(\mathbf{X}'\mathbf{X})\hat{\theta}' = \mathbf{X}'\mathbf{y}, \quad (\text{П4.1.7})$$

или в скалярной форме

$$p_{x_r y} = \hat{\theta}_1 p_{x_r x_1} + \hat{\theta}_2 p_{x_r x_2} + \dots + \hat{\theta}_k p_{x_r x_k} \quad (r = 1, 2, \dots, k),$$

где, например,

$$p_{x_r y} = \sum_{i=1}^N x_{ir} y_i.$$

Матрица ковариаций оценок:

$$C = (X'X)^{-1} \sigma^2, \quad (\text{П4.1.9})$$

или в скалярной форме

$$C = \begin{pmatrix} p_{x_1 x_1} & p_{x_1 x_2} & \cdots & p_{x_1 x_k} \\ p_{x_2 x_1} & p_{x_2 x_2} & \cdots & p_{x_2 x_k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{x_k x_1} & p_{x_k x_2} & \cdots & p_{x_k x_k} \end{pmatrix}^{-1} \sigma^2,$$

Выборочная оценка остаточной дисперсии:

$$s^2 = \frac{1}{N-k} (y'y - y'X\hat{\theta}) = \frac{1}{N-k} (y'y - \hat{\theta}'X'X\hat{\theta}), \quad (\text{П4.1.12})$$

или в скалярной форме

$$s^2 = \frac{1}{N-k} (p_{yy} - p_{x_1 y} \hat{\theta}_1 - \cdots - p_{x_k y} \hat{\theta}_k).$$

100(1 - α)⁰/₀-ная доверительная область:

$$(\theta - \hat{\theta})' X'X (\theta - \hat{\theta}) \leq k s^2 f_{k, N-k} (1 - \alpha), \quad (\text{П4.1.15})$$

или в скалярной форме

$$\sum_{r=1}^N \sum_{s=1}^N (\theta_r - \hat{\theta}_r) (\theta_s - \hat{\theta}_s) p_{rs} \leq k s^2 f_{k, N-k} (1 - \alpha),$$

где

$$\Pr \{ F_{k, N-k} \leq f_{k, N-k} (1 - \alpha) \} = 1 - \alpha.$$

Дисперсия прогноза:

$$\text{Var} [\hat{Y}] = \sigma^2 + x' C x, \quad (\text{П4.1.18})$$

или в скалярной форме

$$\text{Var} [\hat{Y}] = \sigma^2 + \sum_{j=1}^k \sum_{l=1}^k x_j x_l \text{Cov} (\hat{\theta}_j, \hat{\theta}_l).$$

Пример. Для иллюстрации приведенных выше результатов рассмотрим частный случай — двухпараметрическую модель:

$$Y_i = \theta_1 + \theta_2 x_i + Z_i. \quad (\text{4.3.16})$$

Выборочные оценки наименьших квадратов $\hat{\theta}_1$ и $\hat{\theta}_2$, получаемые из (П4.1.7), выведены в приложении П4.1.2. Они имеют вид

$$\begin{aligned}\hat{\theta}_1 &= \frac{\sum x^2 \sum y - \sum xy \sum x}{N \sum x^2 - (\sum x)^2}, \\ \hat{\theta}_2 &= \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2}.\end{aligned}\quad (4.3.17)$$

Из (П4.1.19) получаем выборочные оценки ковариаций оценок

$$\begin{aligned}\widehat{\text{Var}} [\hat{\theta}_1] &= \frac{s^2 \sum x^2}{N \sum (x - \bar{x})^2}, \\ \widehat{\text{Var}} [\hat{\theta}_2] &= \frac{s^2}{\sum (x - \bar{x})^2}, \\ \widehat{\text{Cov}} [\hat{\theta}_1, \hat{\theta}_2] &= -\frac{s^2 \sum x}{N \sum (x - \bar{x})^2},\end{aligned}\quad (4.3.18)$$

где выборочная дисперсия s^2 получается из (П4.1.12):

$$s^2 = \frac{1}{N-2} \left\{ \sum y_i^2 - \hat{\theta}_1 \sum y_i - \hat{\theta}_2 \sum y_i x_i \right\}. \quad (4.3.19)$$

Наконец, используя (П4.1.15), получаем $100(1 - \alpha)\%$ -ную доверительную область для θ_1, θ_2 :

$$\begin{aligned}N (\theta_1 - \hat{\theta}_1)^2 + (\theta_2 - \hat{\theta}_2)^2 \sum x_i^2 + 2 (\theta_1 - \hat{\theta}_1) (\theta_2 - \hat{\theta}_2) \times \\ \times \sum x_i \leq 2s^2 f_{2, N-2} (1 - \alpha).\end{aligned}\quad (4.3.20)$$

4.3.4. Ортогональность

В рассмотренном выше примере оценки наименьших квадратов для параметров модели (4.3.16) имели отличную от нуля ковариацию, а в уравнение доверительного эллипса для (θ_1, θ_2) входило произведение вида $(\theta_1 - \hat{\theta}_1)(\theta_2 - \hat{\theta}_2)$. Типичная доверительная область для модели (4.3.16) показана на рис. 4.4(а), где видно, что оси эллипса наклонены по отношению к осям (θ_1, θ_2) . Следовательно, нельзя определить доверительный интервал для θ_1 и θ_2 отдельно. В предельном случае очень высокой корреляции могло бы случиться, что очень широкий диапазон значений выборочных оценок был бы в согласии с данными.

Можно, однако, по-новому параметризовать эту задачу, так что полученные оценки будут некоррелированы, т. е. ортогональны.

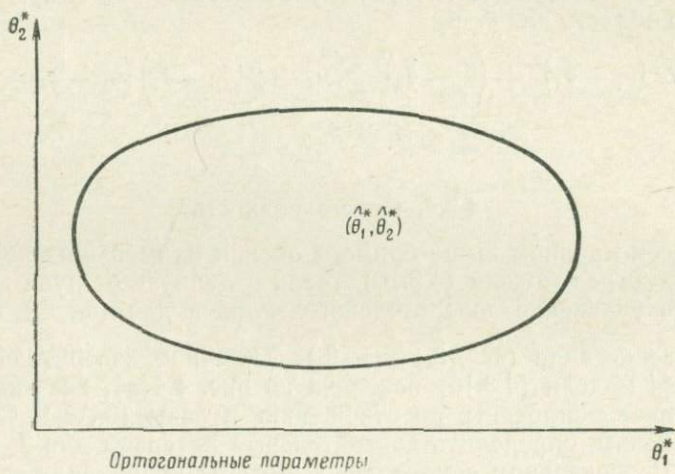
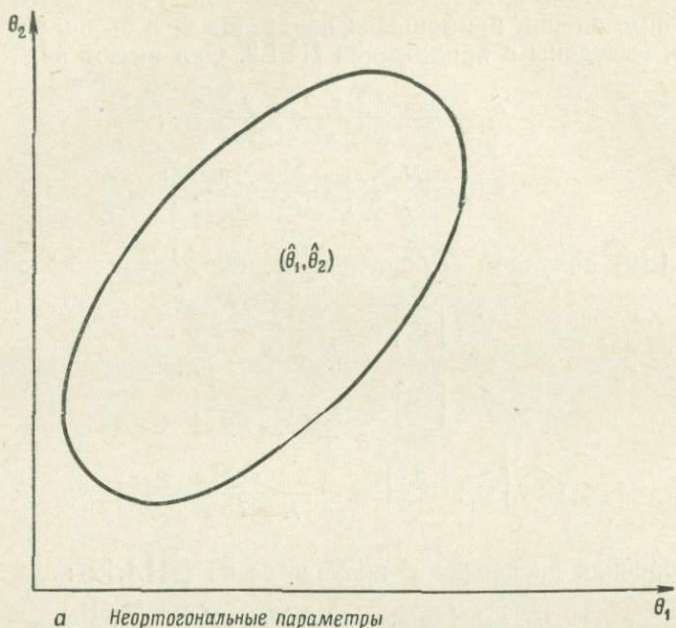


Рис. 4.4. Доверительные области для двух параметров.

Для двухпараметрической модели ортогональная параметризация имеет вид

$$Y_i = \theta_1^* + \theta_2^*(x_i - \bar{x}) + Z_i. \quad (4.3.21)$$

Из (П4.1.7) получаем выборочные оценки наименьших квадратов

$$\begin{aligned} \hat{\theta}_1^* &= \bar{y} = \frac{1}{N} \sum y_i, \\ \hat{\theta}_2^* &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \end{aligned} \quad (4.3.22)$$

и из (П4.1.9)

$$\begin{aligned} \widehat{\text{Var}} [\hat{\theta}_1^*] &= \frac{s^2}{N}, \\ \widehat{\text{Var}} [\hat{\theta}_2^*] &= \frac{s^2}{\sum (x_i - \bar{x})^2}, \\ \widehat{\text{Cov}} [\hat{\theta}_1^*, \hat{\theta}_2^*] &= 0, \end{aligned} \quad (4.3.23)$$

где

$$s^2 = \frac{1}{N-2} \left\{ \sum (y_i - \bar{y})^2 - (\hat{\theta}_2^*)^2 \sum (x_i - \bar{x})^2 \right\}. \quad (4.3.24)$$

100(1 - α) %-ная доверительная область для θ_1, θ_2 является эллипсом

$$N (\theta_1^* - \hat{\theta}_1^*)^2 + (\theta_2^* - \hat{\theta}_2^*)^2 \sum (x_i - \bar{x})^2 \leq 2s^2 f_{2, N-2} (1 - \alpha), \quad (4.3.25)$$

который не имеет члена с произведением переменных из-за отсутствия корреляции между оценками. Типичная доверительная область такого вида показана на рис. 4.4, б. Поскольку в этом случае оси эллипса параллельны осям параметров, можно определить отдельные доверительные интервалы для каждого из двух параметров.

Если $k > 2$, то способ вычитания среднего значения x_i , как это делалось в (4.3.21), не приводит к ортогональной параметризации. Однако при этом оценки становятся ближе к ортогональным, чем без вычитания средних, и, в частности, они будут ортогональны к постоянной составляющей модели. Поэтому лучше подбирать модель вида

$$Y_i = \theta_1 + \theta_2(x_{i2} - \bar{x}_2) + \theta_3(x_{i3} - \bar{x}_3) + \dots + \theta_k(x_{ik} - \bar{x}_k) + Z_i, \quad (4.3.26)$$

чем модель

$$Y_i = \theta_1 + \theta_2 x_{i2} + \theta_3 x_{i3} + \dots + \theta_k x_{ik} + Z_i.$$

В приложении П4.1 показано, что результаты этого раздела легко обобщаются на случай, когда Z_i имеют произвольную матрицу ковариаций.

4.4. ВЫВОДЫ, ОСНОВАННЫЕ НА ФУНКЦИИ ПРАВДОПОДОБИЯ

4.4.1. Основной метод

Функция правдоподобия была введена в статистику Фишером, но, как отмечалось в разд. 4.2, Фишер использовал ее главным образом для получения оценок максимального правдоподобия, которые можно было бы затем использовать для оценивания в методе выборочных распределений. Использование же метода правдоподобия для выводов ведет свое начало от работ Барнарда [7, 8] и представляет собой совершенно другой подход к статистическим выводам. Подход Барнарда можно коротко сформулировать в утверждении, что распределения вероятностей полезны при описании данных *до того*, как они собраны, в то время как функции правдоподобия полезны при описании данных *после того*, как они собраны.

Важнейшая отличительная черта выводов, основанных на правдоподобии, заключается в том, что они очень ясно показывают, что выборочное пространство не связано с оцениванием. Это логично, ибо свойства выборочной оценки должны, несомненно, зависеть от имеющихся данных, а не от данных, которые *могли бы быть* получены.

Способ получения выводов, основанных на правдоподобии, можно резюмировать в следующем виде:

1. Выборочная плотность вероятности наблюдений предполагается полностью известной, за исключением нескольких неизвестных значений параметров.

2. Функция правдоподобия получается подстановкой в плотность вероятности тех значений, которые получили наблюдения в данном эксперименте.

3. Функция правдоподобия строится как функция от неизвестных параметров.

4. Находятся подходящие способы извлечения и суммирования информации, содержащейся в функции правдоподобия.

В качестве простого примера применения метода правдоподобия рассмотрим несколько искусственную задачу оценки среднего значения μ нормальной плотности вероятности, дисперсия которой σ^2 известна. Выборочная плотность вероятности (4.2.1) выборки, до того как собраны данные, имеет вид

$$f_{12\dots n}(x_1, x_2, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi} \sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} =$$

$$= \left(\frac{1}{\sqrt{2\pi} \sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\}.$$

После того как данные получены, функция правдоподобия для μ оказывается пропорциональной экспоненте

$$L(\mu) = K \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right\}. \quad (4.4.1)$$

Отсюда функция правдоподобия, рассматриваемая как функция от μ , с точностью до множителя \bar{x} равна нормальной плотности вероятности со средним значением \bar{x} и дисперсией σ^2/n . В противоположность этому в методе выборочных распределений \bar{X} имеет нормальное распределение со средним значением μ и дисперсией σ^2/n .

Информация, даваемая функцией правдоподобия (4.4.1), по существу содержится в ее среднем значении \bar{x} (выборочная оценка максимального правдоподобия) и в ее дисперсии σ^2/n . Таким образом, точность, с которой оценивается параметр, сразу видна на графике функции правдоподобия. Если функция правдоподобия сплюснута, то параметр оценивается неточно, так как значения параметра, удаленные от выборочной оценки максимального правдоподобия, имеют правдоподобие ненамного меньше, чем правдоподобие самой оценки. Обратное, если функция правдоподобия сконцентрирована около выборочной оценки максимального правдоподобия, то параметр оценивается с большой точностью.

4.4.2. Свойства функций правдоподобия

В этом разделе мы рассмотрим интерпретацию функций правдоподобия и правила комбинирования этих функций.

Принцип правдоподобия. Принцип правдоподобия заключается в том, что если два эксперимента приводят к пропорциональным функциям правдоподобия, то выводы, получаемые из этих экспериментов, должны быть одинаковыми.

Предположим, например, что 8 транзисторов подвергаются проверке. До проведения эксперимента число дефектных транзисторов можно описать с помощью случайной величины R , выборочного пространства $r = 0, 1, 2, \dots, 8$ и биномиального распределения вероятностей

$$p_R(r) = \binom{8}{r} p^r (1-p)^{8-r}. \quad (4.4.2)$$

После проведения эксперимента, заключающегося в проверке 8 транзисторов, оказалось, что три транзистора дефектны. Функция

правдоподобия в этом случае имеет вид

$$L(p) = 56p^3(1-p)^5. \quad (4.4.3)$$

Теперь предположим, что был проведен другой эксперимент, в котором транзисторы проверялись до тех пор, пока не было обнаружено r дефектных. До проведения этого эксперимента число проверенных транзисторов можно описать с помощью случайной величины N , выборочного пространства $n = r, r+1, \dots, \infty$ и распределения вероятностей Паскаля

$$p_N(n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad (4.4.4)$$

которое дает вероятность того, что для получения r дефектных транзисторов нужно проверить всего n транзисторов.

Если оказалось, что для получения трех дефектных пришлось проверить $n=8$ транзисторов, то функция правдоподобия после проведения экспериментов окажется равной

$$L(p) = 21p^3(1-p)^5. \quad (4.4.5)$$

Равенство (4.4.5) пропорционально равенству (4.4.3), и, согласно принципу правдоподобия, информация относительно параметра p , содержащаяся в обоих экспериментах, одинакова. Если же принять метод выборочных распределений, то выводы, которые должны быть сделаны из этих двух экспериментов, будут разными, так как выборочные пространства и распределения вероятностей являются в них различными. Следовательно, доверительный интервал для p в первом эксперименте отличался бы от доверительного интервала во втором.

Отметим, что принцип правдоподобия является формальным выражением того факта, что выборочное пространство не связано с оценением p . Дальнейшее обсуждение принципа правдоподобия читатель может найти в [8, 9].

Перемножение правдоподобий. Если функцию правдоподобия для параметра θ , вычисленную по данным D_1 , обозначить $L(\theta | D_1)$ и если независимые данные D_2 дают функцию правдоподобия $L(\theta | D_2)$, то полная функция правдоподобия, основанная на данных D_1 и D_2 , имеет вид

$$L(\theta | D_1, D_2) = L(\theta | D_1) L(\theta | D_2).$$

Это сразу следует из того, что если два набора данных независимы, то полная выборочная плотность вероятности равна произведению отдельных выборочных плотностей вероятности.

В этом смысле функция правдоподобия подчиняется тому же мультипликативному закону, что и плотность независимых случайных величин. Правдоподобие, как и распределение вероятностей,

является неотрицательной величиной. Однако на этом их сходство кончается. Не существует закона для объединения правдоподобий, который был бы аналогичен правилу сложения для вероятности объединения двух взаимоисключающих событий.

Шансы, получаемые из отношения правдоподобия (likelihood odds). Рассмотрим функцию правдоподобия (4.4.3) для параметра p . Выборочной оценкой максимального правдоподобия является $\hat{p} = 3/8$, которая дает значение правдоподобия $L\left(\frac{3}{8}\right) = 0,282$. Для другой величины p , скажем $p = 1/8$, правдоподобие равно $L\left(\frac{1}{8}\right) = 0,0104$. Отсюда шансы, получаемые из отношения правдоподобия, за то, что $p = 3/8$ по сравнению с $p = 1/8$, равны 27 : 1.

Это утверждение можно интерпретировать следующим образом:

1) Шансы, получаемые из отношения правдоподобия, измеряют степень обоснованности, с которой по нашим данным параметру p можно приписать значение $3/8$, а не $1/8$. Для того чтобы эти два значения можно было считать одинаково правдоподобными, нужно, чтобы другой независимый эксперимент дал те же шансы в пользу значения $p = 1/8$:

$$\frac{L\left(\frac{1}{8}\right)}{L\left(\frac{3}{8}\right)} = 27.$$

2) Лицо, не имеющее никакой дальнейшей информации, было бы готово заключить пари в отношении 27 : 1 за то, что истинным значением окажется $p = 3/8$, а не $p = 1/8$.

4.4.3. Примеры функций правдоподобия

Оценивание параметра показательного распределения. Рассмотрим обсуждавшуюся в разд. 4.2.4 задачу оценивания среднего срока службы осветительных ламп. Первый шаг в методе правдоподобия заключается в том, что нужно выписать выборочную плотность вероятности для наблюдений. В нашей задаче соответствующей выборочной плотностью вероятности будет

$$f_{123}(x_1, x_2, x_3 | \lambda) = \lambda^3 \exp\left(-\lambda \sum_{i=1}^3 x_i\right). \quad (4.4.6)$$

Следующий шаг состоит в подстановке наблюдаемых значений $x_1 = 2,6$, $x_2 = 1,9$ и $x_3 = 1,5$ в (4.4.6), в результате чего получается функция правдоподобия.

$$L(\lambda) = \lambda^3 \exp(-6\lambda). \quad (4.4.7)$$

Третий шаг заключается в построении графика функции правдоподобия, который показан на рис. 4.2. На четвертом шаге из функции правдоподобия извлекается и кратко суммируется информация. В настоящем разделе мы опишем очень простой способ выполнения этого шага. Более детальные способы описания функций правдоподобия будут приведены в разд. 4.4.5.

В отличие от функции правдоподобия (4.4.1) функция правдоподобия (4.4.6) асимметрична. Рис. 4.2 показывает, что она круто возрастает от $\lambda = 0$ до своего максимума при $\lambda = 0,5$, а затем относительно медленно стремится к нулю при $\lambda \rightarrow \infty$. Значение $\lambda = 0,5$

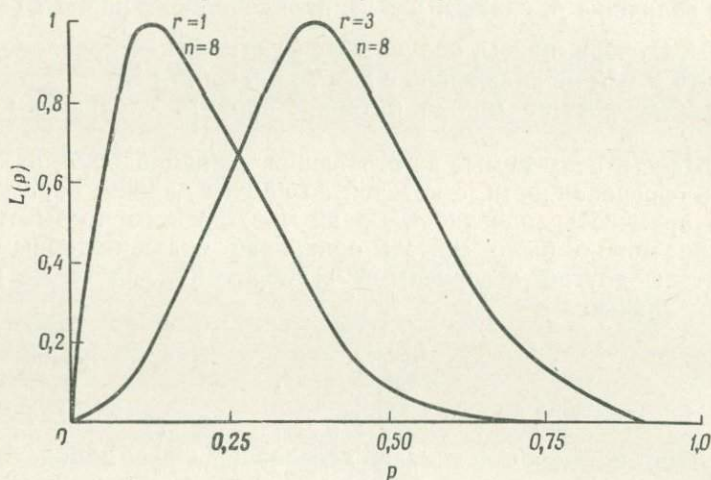


Рис. 4.5. Функции правдоподобия для биномиального распределения (нормированные).

является наиболее *правдоподобным*, или *вероятным*, значением для этих трех наблюдений. Оно называется *выборочной оценкой максимального правдоподобия* $\hat{\lambda}$ параметра λ . Значения $\lambda = 0,1$ и $\lambda = 1,4$

оба дают шансы правдоподобия 1:10 против $\hat{\lambda}$. Таким образом, шансы против того, что $\lambda \leq 0,1$ и $\lambda \geq 1,4$, не меньше чем 10:1. Поэтому область от $\lambda = 0,1$ до $\lambda = 1,4$ называют *вероятной областью* (credible region) с шансами, не меньшими чем 10:1, против любого значения не из этой области.

Оценивание биномиального параметра. Рассмотрим обсуждающуюся в разд. 4.4.2 задачу оценивания биномиального параметра. Используя (4.4.2) или (4.4.4), получаем, что при r успехах в n испытаниях функция правдоподобия имеет вид

$$L(p) = K p^r (1 - p)^{n-r}. \quad (4.4.8)$$

На рис. 4.5 показаны функции правдоподобия для двух случаев $r=1$, $n=8$ и $r=3$, $n=8$, причем обе кривые пронормированы так, что их максимум равен единице. Продифференцировав (4.4.8), находим, что выборочная оценка максимального правдоподобия имеет вид $\hat{p} = r/n$.

Для случая $r=1$ кривая правдоподобия похожа на кривую, изображенную на рис. 4.2, т. е. она резко возрастает до максимума при $p = \hat{p}$ и затем медленно убывает для p , больших чем \hat{p} . Вероятная область с шансами 10:1 простирается от 0,006 до 0,49, причем внутри нее в точке 0,125 находится выборочная оценка максимального правдоподобия.

Для случая $r=3$ кривая правдоподобия вполне симметрична относительно выборочной оценки максимального правдоподобия $\hat{p} = 0,375$. Вероятная область с шансами 10:1 простирается от 0,095 до 0,71.

4.4.4. Метод наименьших квадратов и оценивание с помощью правдоподобия

Оценивание с помощью наименьших квадратов эквивалентно оцениванию методом максимального правдоподобия при условии, что ошибки распределены по нормальному закону. Чтобы показать это, рассмотрим простую однопараметрическую модель

$$Y_i = \theta x_i + Z_i,$$

обсуждавшуюся в разд. 4.3. Если предположить, что ошибки Z_i независимы, имеют нулевое среднее значение и дисперсию σ^2 , то выборочная оценка наименьших квадратов получается при минимизации суммы квадратов

$$S(\theta) = \sum_{i=1}^N (y_i - \theta x_i)^2. \quad (4.4.9)$$

Если предположить, что ошибки независимы, имеют нулевое среднее значение и дисперсию σ^2 , а также распределены по нормальному закону, то плотность вероятности для данных до того, как проведен эксперимент, имеет вид

$$f_{12 \dots N}(y_1, y_2, \dots, y_N) = \frac{1}{(\sqrt{2\pi} \sigma)^N} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \theta x_i)^2 \right\}.$$

После того как данные собраны, логарифмическая функция правдоподобия равна

$$l(\theta) = -\frac{N}{2} \ln 2\pi - N \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \theta x_i)^2. \quad (4.4.10)$$

Таким образом, выборочная оценка θ , максимизирующая логарифмическую функцию правдоподобия (4.4.10), совпадает с выборочной оценкой, минимизирующей сумму квадратов (4.4.9). Следовательно, для нормально распределенных ошибок выборочные оценки наименьших квадратов и максимума правдоподобия совпадают.

Мы обосновали оценки наименьших квадратов в разд. 4.3.1, пользуясь критерием среднеквадратичной ошибки. Однако критерий среднеквадратичной ошибки нельзя использовать в теории правдоподобия, поскольку он включает усреднение по выборочному пространству. Следовательно, необходимо заново интерпретировать теорию наименьших квадратов с точки зрения метода правдоподобия.

Логарифмическую функцию правдоподобия (4.4.10) можно переписать в виде

$$l(\theta) = \ln K - \frac{1}{2\sigma^2} \sum x_i^2 (\theta - \hat{\theta})^2, \quad (4.4.11)$$

где $\hat{\theta} = \sum x_i y_i / \sum x_i^2$ является выборочной оценкой и наименьших квадратов, и максимума правдоподобия. Отсюда функция правдоподобия пропорциональна нормальной плотности вероятности со средним значением $\hat{\theta}$ и дисперсией

$$\frac{\sigma^2}{\sum x_i^2}. \quad (4.4.12)$$

Заметим, что выражение (4.4.12) в точности совпадает с выборочной дисперсией (4.3.8) оценки наименьших квадратов. Так как дисперсия (4.4.12) равна

$$\frac{-1}{[\partial^2 l(\theta) / \partial \theta^2]} = \frac{\sigma^2}{\sum x_i^2},$$

то отсюда следует, что количество информации Фишера $E[\partial^2 l / \partial \theta^2]_0$ заменяется в методе правдоподобия на фактически имеющееся значение второй производной логарифмической функции правдоподобия в точке ее максимума.

Вероятные области. В разд. 4.4.3 было показано, что понятие шансов, получаемых из отношения правдоподобия, можно использовать для определения вероятных областей для параметра. При этом, если сравнивать любое значение параметра внутри этой области с любым другим значением, то шансы правдоподобия не превосходят заданного отношения. Однако если функция является нормальной, то вероятная область, основанная на шансах правдоподобия, эквивалентна области, которую можно получить, набирая определенную долю площади под функцией правдоподобия. На-

пример, 7,5:1 — вероятная нормальная область эквивалентна охвату 95% площади под функцией правдоподобия. 95%-ная вероятная область для параметра θ в упомянутом выше примере имеет вид

$$\hat{\theta} \pm 1,96 \left(\frac{\sigma^2}{\sum x_i^2} \right)^{1/2},$$

и она является также вероятной областью с шансами 7,5:1. При построении интервала, исходя из площади, мы неявно считаем функцию правдоподобия распределением вероятностей. В байесовском подходе к выводам [10, 11, 2*] это делается явно.

Наименьшие квадраты в случае, когда независимые переменные содержат ошибки. При рассмотрении наименьших квадратов в разд. 4.3 предполагалось, что x_i не содержали ошибок. Однако во многих случаях невозможно осуществить какой-либо контроль над независимыми переменными (например, в рассматриваемых ниже задачах с временными рядами). В таких случаях x_i можно рассматривать как реализации случайных величин. Для однопараметрического случая совместную выборочную плотность вероятности наблюдений, до того как они произведены, можно записать в виде

$$f_{12\dots N}(y_1, y_2, \dots, y_N) = f_{12\dots N}(y_1, y_2, \dots, y_N | \theta, x_1, \dots, x_N) \times \tilde{f}_{12\dots N}(x_1, \dots, x_N).$$

Совместная плотность вероятности $f_{12\dots N}$ в правой части равенства представляет собой условное распределение y_i при условии, что x_i фиксированы, а плотность $\tilde{f}_{12\dots N}$ описывает плотность вероятности x_i .

После получения данных можно выписать функцию правдоподобия

$$L(\theta) = L(\theta | x_1, \dots, x_N) \tilde{L}(x_1, \dots, x_N).$$

Так как $\tilde{L}(x_1, \dots, x_N)$ не имеет среди своих аргументов θ , то функция правдоподобия для θ будет такой же (за исключением независимого от θ множителя), что и функция правдоподобия $L(\theta | x_1, \dots, x_N)$, полученная, когда x_i рассматриваются как фиксированные, или не содержащие ошибок. Таким образом, знание распределения x_i никак не помогает при оценивании θ . Отметим еще раз, что метод выборочных распределений дает другой ответ на эту задачу, так как дисперсия $\hat{\theta}$ равна

$$\text{Var} [\hat{\theta}] = E \left[\frac{\sigma^2}{\sum_{i=1}^N X_i^2} \right],$$

где математическое ожидание берется по выборочному пространству X_i^*). В методе правдоподобия выборочное пространство не имеет отношения к существу дела, а «дисперсия» функции правдоподобия дается все еще выражением (4.4.12) и, следовательно, зависит только от конкретных значений x_i , которые получились в данных измерениях.

4.4.5. Методы извлечения информации из функции правдоподобия

Квадратичные правдоподобия. Логарифмическая функция правдоподобия (4.4.11) квадратична по параметру θ . В более общем случае, если модель линейна по параметрам, а ошибки распределены по нормальному закону, логарифмическая функция правдоподобия является квадратичной формой от параметров θ_i . Следовательно, функция правдоподобия сама является многомерным распределением, и ее можно описать с помощью средних значений (выборочных оценок максимального правдоподобия) и матрицы ковариаций этого распределения. Из (3.1.19) мы видим, что матрица вторых производных

$$-\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}$$

является матрицей, обратной матрице ковариаций, соответствующей этому многомерному нормальному распределению.

Неквадратичные правдоподобия. Если модель нелинейна по параметрам или же выборочное распределение отличается от нормального, то функцию правдоподобия нельзя описать только с помощью ее первых двух производных. Как правило, для неквадратичной логарифмической функции правдоподобия лучше всего *построить график всей функции*. Задача получения выводов относительно θ сводится в этом случае к задаче описания, или *аппроксимации*, функции правдоподобия самым простым возможным способом. В некоторых случаях получаются функции правдоподобия с несколькими максимумами; извлечь информацию из такой функции и кратко описать ее трудно. Если же на графике функции пра-

* *) Рассуждения авторов в последнем подразделе не совсем точны. Если нам известны x_i , то вообще не имеет значения, как они получились и что они собой представляют, поскольку мы пользуемся условным распределением при фиксированных x_i . Если же в нашем распоряжении имеются лишь искаженные ошибками значения x_i , то мы не можем вычислить функцию $L(\theta | x_1 \dots x_N)$ и, следовательно, не можем получить из нее оценку для θ . Об оценках параметров функций, в случае когда независимые переменные содержат ошибки, см. подробнее в книге Клепикова Н. П. и Соколова С. Н. «Анализ и планирование экспериментов методом максимума правдоподобия», М., изд-во «Наука», 1964, гл. 3. — *Прим. перев.*

вдоподобия имеется один максимум, то можно использовать способы, приводимые ниже.

В первом из них функция правдоподобия приближается нормальной функцией правдоподобия, а во втором подбирается такое преобразование параметров, чтобы функция правдоподобия преобразованных переменных была ближе к нормальной, чем до применения преобразования.

Способ 1. Приближение с помощью нормального распределения. Предположим, что функция правдоподобия не является нормальной, все же разумно приблизить ее с помощью нормальной плотности вероятности по параметру θ . Поскольку функция правдоподобия определена с точностью до постоянного множителя, приближение будет иметь вид

$$L(\theta) \approx K \exp \left\{ -\frac{1}{2\sigma^2} (\theta - \bar{\theta})^2 \right\}, \quad (4.4.13)$$

где $\bar{\theta}$ — среднее значение аппроксимирующего распределения и σ^2 — его дисперсия. Если «моменты» функции правдоподобия определить с помощью соотношений

$$I_k = \int_{-\infty}^{\infty} \theta^k L(\theta) d\theta,$$

то, используя свойства нормальной плотности, можно найти константы K , $\bar{\theta}$ и σ^2 из (4.4.13):

$$\begin{aligned} K &= \frac{I_0}{\sqrt{2\pi} \sigma} = \frac{I_0^2}{\sqrt{2\pi} \sqrt{I_2 I_0 - I_1^2}}, \\ \bar{\theta} &= \frac{I_1}{I_0}, \\ \sigma^2 &= \frac{I_2}{I_0} - \bar{\theta}^2 = \frac{I_2 I_0 - I_1^2}{I_0^2}. \end{aligned} \quad (4.4.14)$$

Оценка среднего правдоподобия. Барнард [7] назвал

$$\bar{\theta} = \frac{\int_{-\infty}^{\infty} \theta L(\theta) d\theta}{\int_{-\infty}^{\infty} L(\theta) d\theta} \quad (4.4.15)$$

выборочной оценкой среднего правдоподобия. Если $L(\theta)$ — нормальная функция правдоподобия, то выборочная оценка среднего правдоподобия совпадает с выборочной оценкой максимального правдоподобия, но в общем случае они будут различны.

Преимущество выборочной оценки среднего правдоподобия над оценкой максимального правдоподобия состоит в том, что первая учитывает форму *всей* функции правдоподобия, в то время как

вторая характеризует только одну точку на кривой. Поэтому выборочная оценка максимального правдоподобия может вводить в заблуждение для малых выборок, если функция правдоподобия не является нормальной. Для больших выборок большинство функций правдоподобия стремится к нормальной плотности, так что выборочная оценка максимального правдоподобия вместе с ее дисперсией достаточны для описания всей функции правдоподобия.

Можно показать [4], что если нет никакой причины предполагать а priori, что какое-нибудь одно значение θ более вероятно, чем другие, то оценка \hat{p} , соответствующая выборочной оценке среднего правдоподобия (4.4.15), является оценкой с наименьшей среднеквадратичной ошибкой при любом объеме выборки. Это не означает, что для всех значений θ выборочная среднеквадратичная ошибка этой оценки равномерно меньше, чем для любой другой оценки. Это значит лишь, что после усреднения по всем значениям θ полученная полная среднеквадратичная ошибка будет наименьшей.

С точки зрения правдоподобия критерий среднеквадратичной ошибки не имеет отношения к делу, и, следовательно, выборочную оценку среднего правдоподобия лучше всего рассматривать как удобный способ описания центра расположения функции правдоподобия.

Пример. Чтобы проиллюстрировать этот способ извлечения информации из функции правдоподобия, рассмотрим пример с биномиальным параметром из разд. 4.2.2. Из функции правдоподобия (4.4.8) получается выборочная оценка максимального правдоподобия,

$$\hat{p} = \frac{r}{n},$$

в то время как выборочная оценка среднего правдоподобия равна

$$\bar{p} = \frac{r+1}{n+2},$$

а ее дисперсия имеет вид

$$\sigma^2 = \frac{r+1}{n+2} \frac{(n-r+1)}{(n+3)(n+2)}.$$

Отсюда для $r=3$, $n=8$ функцию правдоподобия (4.4.8) можно аппроксимировать нормальной плотностью вероятности со средним значением $\bar{p}=0,4$ и дисперсией $\sigma^2=0,022$. Следовательно, 95%-ная, или вероятная область с шансами 7,5:1 для p представляет собой интервал (0,11; 0,69). Для $r=1$, $n=8$ аппроксимирующая нормальная плотность будет иметь среднее значение $\bar{p}=0,2$ и дисперсию $\sigma^2=0,015$. Как можно увидеть из рис. 4.5, нормальное прибли-

*) Рассматриваемая как случайная величина. — Прим. перев.

жение для $r=3$ будет намного лучшим, чем для $r=1$, из-за асимметрии функции правдоподобия во втором случае. На самом деле 95%-ная вероятная область для $r=1$ имеет отрицательную левую границу, что говорит о том, что нормальное приближение не оправдано. В этом случае лучший способ состоял бы в следующем.

Способ 2. Преобразование параметров. Если логарифмическая функция правдоподобия не является квадратичной, то полезно найти такие преобразования $\varphi_i(\theta_1, \theta_2, \dots, \theta_k)$ параметров, что функция правдоподобия стала бы приближенно многомерной нормальной функцией от φ_i .

Как отмечалось выше, если функция правдоподобия является нормальной, то вторая производная ее логарифма постоянна, т. е. количество информации Фишера равно константе. Если функция правдоподобия не является нормальной, то $-d^2l/d\theta^2$ будет функцией от θ . Это нежелательно, так как в этом случае в различных точках шкалы θ получается различная информация относительно θ . Поэтому хотелось бы найти преобразование $\varphi = \varphi(\theta)$, такое, чтобы в масштабе φ производная $-d^2l/d\varphi^2$ была бы константой в окрестности выборочной оценки максимального правдоподобия $\hat{\varphi} = \varphi(\hat{\theta})$ параметра φ .

Сделав преобразование $\varphi = \varphi(\theta)$, имеем

$$\frac{dl}{d\varphi} = \frac{dl}{d\theta} \frac{d\theta}{d\varphi}$$

и

$$\frac{d^2l}{d\varphi^2} = \left(\frac{d^2l}{d\theta^2}\right) \left(\frac{d\theta}{d\varphi}\right)^2 + \left(\frac{dl}{d\theta}\right) \left(\frac{d^2\theta}{d\varphi^2}\right).$$

В точке выборочной оценки максимального правдоподобия $dl/d\varphi = 0$, так как $dl/d\theta = 0$, и, следовательно,

$$\frac{d^2l}{d\varphi^2} = \left(\frac{d^2l}{d\theta^2}\right)_0 \left(\frac{d\theta}{d\varphi}\right)^2.$$

Если потребовать, чтобы $-d^2l/d\varphi^2$ было положительной константой k , то отсюда получаем

$$\left(\frac{d\theta}{d\varphi}\right)^2 = \frac{k}{-\left(\frac{d^2l}{d\theta^2}\right)_0},$$

и, следовательно, с точностью до постоянного множителя желаемое преобразование $\varphi(\theta)$ имеет вид

$$\varphi(\theta) = \int \left(-\frac{d^2l}{d\theta^2}\right)^{1/2} d\theta. \quad (4.4.16)$$

Пример. Рассмотрим правдоподобие для биномиального распределения (4.4.8), обсуждавшееся выше. В этом случае получаем

$$\frac{d^2 l}{dp^2} = -\frac{r}{p^2} - \frac{n-r}{(1-p)^2} = -\frac{n}{\hat{p}(1-\hat{p})},$$

если производную брать в точке выборочной оценки максимального правдоподобия $\hat{p} = r/n$. Отсюда, используя (4.4.16), получаем

$$\varphi(p) = \int \frac{dp}{\sqrt{p(1-p)}} = \arcsin \sqrt{p}.$$

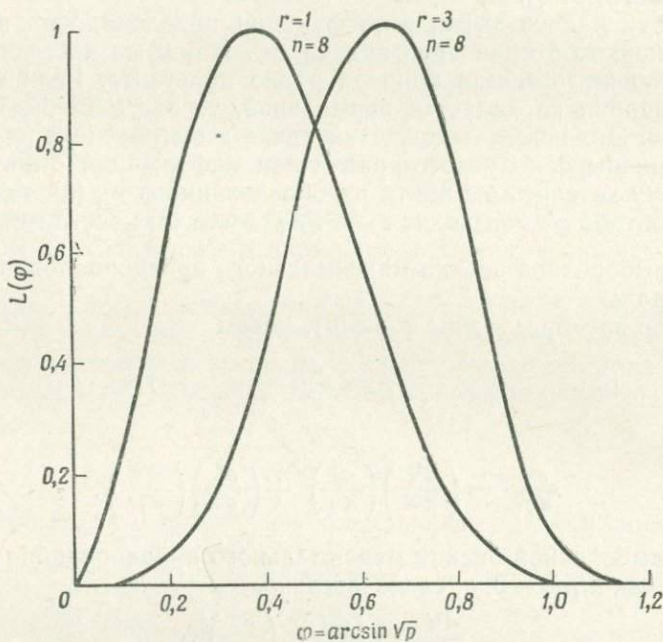


Рис. 4.6. Преобразованные функции правдоподобия для биномиального распределения (нормированные).

Таким образом, функция правдоподобия, у которой в качестве аргумента взят $\arcsin \sqrt{p}$, будет лучше аппроксимироваться нормальной плотностью вероятности со средним значением $\bar{\varphi}$ и дисперсией σ^2 , получаемой из (4.4.14).

На рис. 4.6 показаны правдоподобия после преобразования для случаев $r=1, n=8$ и $r=3, n=8$. В обоих случаях функции правдоподобия похожи на нормальную кривую, в то время как до преобразования кривая для $r=1$ очень сильно отличалась от нормальной, как видно из рис. 4.5.

В табл. 4.2 приведены среднее значение и дисперсия аппроксимирующего нормального распределения, а также 95%-ная, или вероятная область с шансами 7,5:1, для p до и после преобразования.

Таблица 4.2

Выборочные оценки среднего правдоподобия и вероятные области для биномиальных параметров, полученные из функций правдоподобия до и после преобразования

	Выборочная оценка максимального правдоподобия \hat{p}	До преобразования			После преобразования			
		\bar{p}	$\frac{\sigma^2}{p}$	95%-ная область	\bar{p}	$\bar{\varphi}$	$\frac{\sigma^2}{\varphi}$	95%-ная область
$n = 8, r = 3$	0,375	0,4	0,0217	(0,11; 0,69)	0,374	0,655	0,0256	(0,10; 0,68)
$n = 8, r = 1$	0,125	0,2	0,0145	(-0,04; 0,44)	0,148	0,394	0,0265	(0,05; 0,43)

Мы видим, что преобразование изменяет выборочную оценку среднего правдоподобия сильнее для асимметричного правдоподобия ($r = 1$), чем для более симметричного правдоподобия ($r = 3$).

4.4.6. Оценивание среднего значения и дисперсии нормального распределения

Чтобы проиллюстрировать описанные в предыдущих разделах способы получения выводов, основанных на правдоподобии, рассмотрим задачу оценивания среднего значения и дисперсии по выборке наблюдений, которые по предположению имеют нормальную плотность вероятности. Воспользовавшись (4.2.1), получаем функцию правдоподобия для μ и σ^2 в виде

$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi} \sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right\}. \quad (4.4.17)$$

Удобный способ описания двумерных правдоподобий состоит в построении на плоскости (μ, σ^2) контуров постоянного уровня функции правдоподобия. Если функция правдоподобия является двумерной нормальной функцией, то эти контуры будут эллипсами; в противном случае можно иногда так преобразовать параметры, что функция правдоподобия будет приблизительно двумерной нормальной функцией.

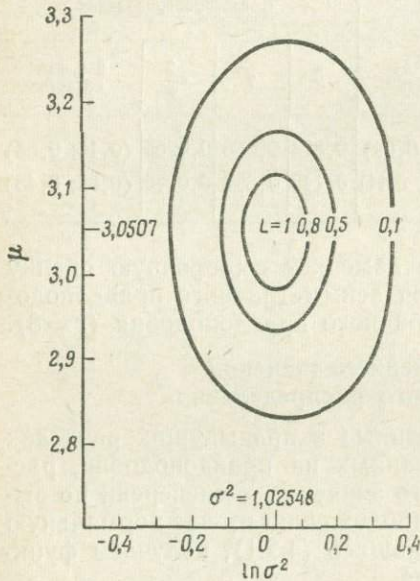
Так как функция правдоподобия (4.4.17) является нормальной по отношению к μ , то необязательно искать преобразование этого параметра. К тому же, так как оценки для μ и σ^2 независимы, то необходимо найти лишь преобразование для σ^2 .

Из (4.4.17) получаем

$$\frac{\partial l}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

и

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2. \quad (4.4.18)$$



Получаем выборочные оценки максимального правдоподобия

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

и в точке, координаты которой равны этим оценкам, выражение (4.4.18) становится равным

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = -\frac{n}{2\hat{\sigma}^4}.$$

Отсюда, пользуясь (4.4.16), получаем, что преобразование, приводящее к нормальному распределению, имеет вид $\varphi = \ln \sigma^2$. Функция

Рис. 4.7. Контуры линий уровня правдоподобия для среднего значения и дисперсии нормальных наблюдений при $n=100$.

правдоподобия для данных о транзисторах, изображенных на рис. 3.3, показана на рис. 4.7 как функция μ и $\ln \sigma^2$. Мы видим, что контуры функций правдоподобия очень близко аппроксимируются эллипсами в области, где функция существенна.

Маргинальные правдоподобия. Двумерная функция правдоподобия (4.4.17), если ее построить как функцию μ и $\ln \sigma^2$, ведет себя в сущности как произведение двух нормальных распределений. Проинтегрировав $L(\mu, \sigma^2)$ по μ , мы получим *маргинальное правдоподобие* для σ^2 , а именно

$$L_1(\sigma^2) = \int_{-\infty}^{\infty} L(\mu, \sigma^2) d\mu = \frac{1}{(\sqrt{2\pi} \sigma)^{n-1}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}. \quad (4.4.19)$$

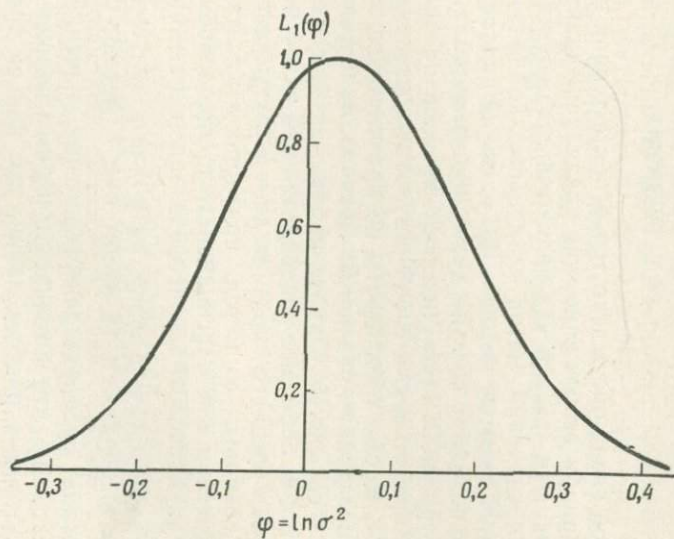


Рис. 4.8. Маргинальная функция правдоподобия для σ^2 при $n=100$.

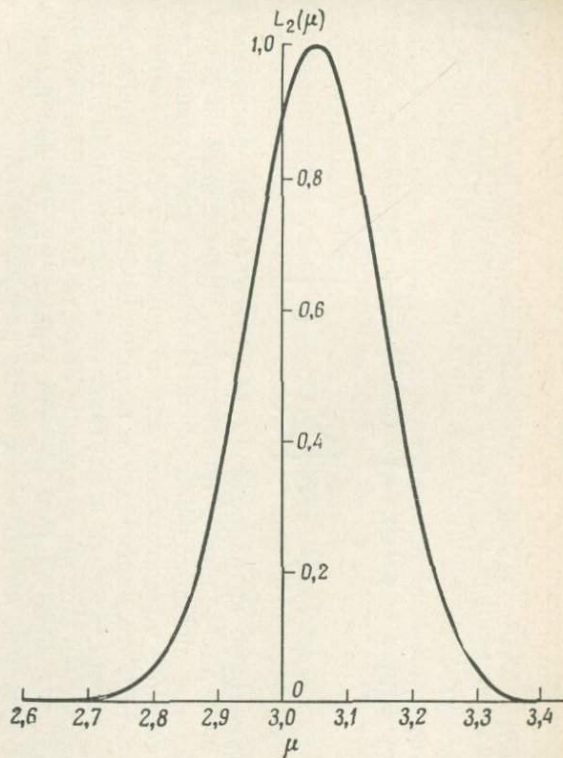


Рис. 4.9. Маргинальная функция правдоподобия для μ при $n=100$.

Далее, двумерная функция правдоподобия приблизительно нормальна по переменным μ и $\varphi = \ln \sigma^2$, так что маргинальное правдоподобие для μ можно получить, проинтегрировав $L(\mu, \varphi)$ по φ , т. е.

$$L_2(\mu) = \int_{-\infty}^{\infty} L(\mu, \sigma^2 = e^\varphi) d\varphi = \int_0^{\infty} L(\mu, \sigma^2) \frac{d(\sigma^2)}{\sigma^2} = \\ = K \left[\sum_{i=1}^n (x_i - \mu)^2 \right]^{-n/2}. \quad (4.4.20)$$

Маргинальные правдоподобия (4.4.19) и (4.4.20) показаны на рис. 4.8 и 4.9 для данных о транзисторах, приведенных на рис. 3.3. Маргинальное правдоподобие для μ построено как функция от μ , а маргинальное правдоподобие для σ^2 построено как функция от $\ln \sigma^2$. Отметим, что маргинальное правдоподобие для σ^2 пропорционально χ^2 -распределению, а маргинальное правдоподобие для μ — t -распределению. Отсюда вероятные области для μ и σ^2 в этом примере были бы точно такими же, как в разд. 4.2, где они были получены с помощью метода выборочных распределений.

4.5. РЕЗЮМЕ

В этой главе обсуждено три аспекта теории статистических выводов, причем особое внимание уделялось задачам оценивания параметров. Эти три аспекта являются следующими: метод выборочных распределений, метод наименьших квадратов и метод правдоподобия. Четвертый метод — Байесовский подход — был опущен, но он очень похож по виду на метод правдоподобия.

Эти три вида статистических выводов не являются разрозненными, а представляют собой результат постепенного исторического развития. Кроме того, ответы на практические задачи, полученные при использовании различных методов, не будут существенно отличаться, а во многих случаях вообще не будут отличаться. Например, метод выборочных распределений в качестве выборочного распределения среднего значения дает t -распределение с $(n-1)$ степенью свободы, а метод правдоподобия дает то же самое распределение для маргинального правдоподобия. В методе выборочных распределений t -распределение с $(n-1)$ степенью свободы представляет распределение возможных значений \bar{x} около μ в повторных выборках, в то время как в методе правдоподобия оно представляет распределение вероятных значений μ около \bar{x} .

Естественно, что исторически первым должен был появиться метод выборочных распределений, так как он требовал лишь непосредственного применения существовавшей теории вероятностей к задачам статистических выводов. Например, выборочное распре-

деление некоторой оценки является распределением вероятностей, дающим относительную частоту появления значений оценки в повторных выборках объема n . По плотности вероятности этой оценки можно сосчитать область, покрывающую истинные значения параметров с вероятностью $1 - \alpha$. Заменяя оценки на выборочные оценки, полученные по данной выборке, мы получим $100(1 - \alpha)\%$ -ную доверительную область для параметров.

Теория наименьших квадратов также развивалась в рамках метода выборочных распределений. Так, оценки наименьших квадратов обладают тем свойством, что они минимизируют среднеквадратичную ошибку, или, что эквивалентно, минимизируют ожидаемый объем доверительной области для параметров.

Метод правдоподобия, хотя он часто и дает ответы, аналогичные тем, которые получаются из метода выборочных распределений, имеет совершенно иную отправную точку. В то время как выборочное распределение описывает все возможные значения наблюдений при данных значениях параметров, функция правдоподобия описывает все возможные значения параметров при данных значениях наблюдений.

Метод правдоподобия дает возможность по-новому интерпретировать теорию наименьших квадратов. Например, функция правдоподобия является по существу поверхностью суммы квадратов $S(\theta_1, \theta_2, \dots, \theta_h)$, если ошибки Z нормальны и независимы. Так как эта сумма является квадратичной формой от θ_i , то функцию правдоподобия можно просто описать с помощью выборочных оценок

наименьших квадратов ($\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_h$) и вторых производных S . Эти производные можно интерпретировать как ковариации оценок в методе выборочных распределений или как меры рассеяния функции правдоподобия в методе правдоподобия. Наиболее важной стороной метода правдоподобия является построение функции правдоподобия в таких переменных, для которых имеется примерно одинаковая информация относительно всех параметров. Тогда информация, заключенная в функции правдоподобия, по существу содержится в ее выборочной оценке среднего правдоподобия и в вероятной области.

Существуют как различия, так и общие стороны у этих методов. Метод правдоподобия совершенно справедливо фокусирует внимание на множестве доступных наблюдений, а не на других множествах наблюдений, которые могли бы получиться. В некоторых случаях метод правдоподобия приводит к более разумным ответам, чем метод выборочных распределений. В разд. 4.4.5 приводился пример, где было показано, что сведения о распределении ошибок в независимых переменных не дают никакой информации для оценивания параметров в моделях наименьших квадратов. Другие примеры, когда метод выборочных распределений является

неудачным, получаются, если оценки, выбранные из-за того, что они хороши в среднем, явно абсурдны в применении к данной выборке. В таких случаях построение функции правдоподобия покажет, что данная конкретная выборка содержит мало информации. Как правило, функция правдоподобия никогда не обманывает.

ЛИТЕРАТУРА

1. Parzen E., *Modern Probability Theory and its Applications*, John Wiley, New York, 1960.
2. Fisher R. A., *Phil. Trans.*, **A222**, 309 (1922).
3. Fisher R. A., *Proc. Cambridge Phil. Soc.*, **22**, 700 (1925).
4. Lehman E. L., *Testing Statistical Hypotheses*, John Wiley, New York, 1959. (Русский перевод: Леман Э., Проверка статистических гипотез, М., изд-во «Наука», 1964.)
5. Hald A., *Statistical Theory with Engineering Applications*, John Wiley, New York, 1952. (Русский перевод: Хальд А., Математическая статистика с техническими приложениями, М., ИЛ, 1956.)
6. Vox G. E. P., *Ann N. Y. Acad. Sci.*, **86**, 3 (1960).
7. Barnard G. A., *Jour. Roy. Stat. Soc.*, **B11**, 116 (1949).
8. Barnard G. A. и др., *J. Roy. Stat. Soc.*, **A125**, 321 (1962).
9. Birnbaum A., *J. Amer. Stat. Assoc.*, **57**, 269 (1962).
10. Jeffreys H., *Theory of Probability*, 3rd ed., Clarendon Press, Oxford, 1961.
11. Savage L. J. и др., *The Foundations of Statistical Inference*, Methuen, London, 1962.

ЛИНЕЙНАЯ ТЕОРИЯ НАИМЕНЬШИХ КВАДРАТОВ

Выборочные ошибки с минимальной среднеквадратичной ошибкой. В этом разделе содержатся доказательства некоторых общих результатов линейной теории наименьших квадратов. Частными случаями этих результатов являются результаты, упоминавшиеся в разд. 4.3.

Предполагается, что модель эксперимента имеет вид

$$Y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_k x_{ik} + Z_i \quad (i=1, 2, \dots, N), \quad (\text{П4.1.1})$$

или в матричной форме

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{Z}, \quad (\text{П4.1.2})$$

где векторы-столбцы \mathbf{Y} , $\boldsymbol{\theta}$ и \mathbf{Z} получаются транспонированием из векторов-строк

$$\mathbf{Y}' = (Y_1, Y_2, \dots, Y_N),$$

$$\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_k),$$

$$\mathbf{Z}' = (Z_1, Z_2, \dots, Z_N)$$

соответственно, а

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{pmatrix}$$

есть матрица наборов значений, принимаемых k выходными переменными x_1, x_2, \dots, x_k в N экспериментах. Предполагается, что ошибки \mathbf{Z} имеют нулевое среднее значение и матрицу ковариаций \mathbf{V} , элементы которой равны $V_{ij} = \text{Cov}[Z_i, Z_j]$. Кроме этого, о совместной плотности вероятности ошибок ничего не известно.

Если нам даны наблюдаемые в N экспериментах отклики y , то *обобщенные выборочные оценки наименьших квадратов* $\boldsymbol{\theta}$ определяются как те значения $\boldsymbol{\theta}$, которые минимизируют квадратичную форму

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}). \quad (\text{П4.1.3})$$

Дифференцирование (П4.1.3) по θ и приравнивание производных нулю дает следующие линейные уравнения для этих оценок:

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\hat{\theta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \quad (\text{П4.1.4})$$

Критерий (П4.1.3) может быть обоснован с двух точек зрения.

а) Используя (3.1.19) и (П4.1.2) и предполагая, что ошибки имеют многомерное нормальное распределение с матрицей ковариаций \mathbf{V} , получаем, что логарифмическая функция правдоподобия для параметров θ равна с точностью до аддитивной константы выражению (П4.1.3). Следовательно, при дополнительном предположении о том, что ошибки имеют многомерное нормальное распределение, обобщенные выборочные оценки наименьших квадратов совпадают с выборочными оценками максимального правдоподобия.

б) Предположим, что в (П4.1.4) выборочные оценки $\hat{\theta}$ заменены на соответствующие оценки $\hat{\Theta}$. Тогда обобщенный принцип наименьших квадратов утверждает, что оценки $\hat{\Theta}$ таковы, что средний квадрат разности двух линейных комбинаций

$$L = \lambda_1\theta_1 + \lambda_2\theta_2 + \dots + \lambda_k\theta_k = \lambda'\theta$$

и

$$\hat{L} = \lambda_1\hat{\theta}_1 + \lambda_2\hat{\theta}_2 + \dots + \lambda_k\hat{\theta}_k = \lambda'\hat{\Theta}$$

принимает свое минимальное значение. Следовательно, произвольная линейная функция от параметров оценивается с минимальной среднеквадратичной ошибкой.

Доказательство обобщенного принципа наименьших квадратов.

Чтобы доказать, что оценки наименьших квадратов, получаемые из (П4.1.4), минимизируют среднеквадратичную ошибку между L и \hat{L} , рассмотрим оценку \hat{L} линейной комбинации L , являющуюся линейной функцией общего вида от случайных величин Y_i , т. е.

$$\hat{L} = l_0 + l_1Y_1 + l_2Y_2 + \dots + l_NY_N = l_0 + \mathbf{l}'\mathbf{Y}.$$

Из (П4.1.2) получаем $E[\mathbf{Y}] = \mathbf{X}\theta$, так как $E[\mathbf{Z}] = 0$ и, следовательно,

$$E[\hat{L}] = l_0 + \mathbf{l}'\mathbf{X}\theta.$$

Далее, так как Y_i имеют ту же самую матрицу ковариаций, что и Z_i , то

$$\text{Var}[\hat{L}] = \mathbf{l}'\mathbf{V}\mathbf{l}.$$

Отсюда среднеквадратичная ошибка оценки \hat{L} равна

$$\text{Var}[\hat{L}] + (E[\hat{L}] - L)^2 = \mathbf{l}'\mathbf{V}\mathbf{l} + (l_0 + \mathbf{l}'\mathbf{X}\theta - \lambda'\theta)^2.$$

Если теперь линейная комбинация $L = \lambda' \theta$ может принимать неограниченные значения, то и среднеквадратичная ошибка будет неограниченно возрастать всегда, за исключением только случая, когда $\mathbf{1}'\mathbf{X} = \lambda'$. Отсюда для достижения минимума среднеквадратичной ошибки надо положить $l_0 = 0$ и минимизировать квадратичную форму $\mathbf{1}'\mathbf{V}\mathbf{1}$ при следующем ограничении на $\mathbf{1}$:

$$\mathbf{1}'\mathbf{X} = \lambda'. \quad (\text{П4.1.5})$$

Это эквивалентно нахождению безусловного минимума квадратичной формы

$$\mathbf{1}'\mathbf{V}\mathbf{1} - (\mathbf{1}'\mathbf{X} - \lambda')\mu,$$

где $\mu' = (\mu_1, \mu_2, \dots, \mu_k)$ — вектор множителей Лагранжа. Приравнивание нулю производных по $\mathbf{1}'$ дает

$$\mathbf{V}\mathbf{1} = \mathbf{X}\mu. \quad (\text{П4.1.6})$$

Решая (П4.1.5) и (П4.1.6) относительно μ' и $\mathbf{1}'$, получаем

$$\mu' = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

и

$$\mathbf{1}' = \lambda' (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}.$$

Отсюда оценка линейной комбинации L с минимальной среднеквадратичной ошибкой имеет вид

$$\widehat{L} = \mathbf{1}'\mathbf{y} = \lambda' (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}).$$

Но из (П4.1.4) получаем, что это выражение в точности совпадает с

$$\widehat{L} = \lambda' \widehat{\theta},$$

где $\widehat{\theta}$ — оценка, соответствующая выборочной оценке (П4.1.4). Приведенное выше доказательство является обобщением доказательства, приведенного Барнардом [1] для случая некоррелированных ошибок Z_i . Если ошибки некоррелированы и имеют одинаковую дисперсию σ^2 , то $\mathbf{V} = \sigma^2 \mathbf{I}$, где \mathbf{I} — единичная матрица. Равенство (П4.1.4) переходит при этом в

$$(\mathbf{X}'\mathbf{X})\widehat{\theta} = \mathbf{X}'\mathbf{y}. \quad (\text{П4.1.7})$$

Пример. Чтобы проиллюстрировать применение формулы (П4.1.7), рассмотрим простой двухпараметрический вариант модели (П4.1.1):

$$Y_i = \theta_1 + \theta_2 x_i + Z_i, \quad i = 1, 2, \dots, N,$$

и предположим, что ошибки Z_i некоррелированы и имеют нулевое среднее значение и дисперсию σ^2 . Тогда (П4.1.7) сводится к

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix},$$

т. е.

$$\begin{pmatrix} N & \sum x \\ \sum x & \sum x^2 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix},$$

где суммирование всюду производится от $i=1$ до $i=N$.

Отсюда выборочные оценки наименьших квадратов имеют вид

$$\hat{\theta}_1 = \frac{\sum x^2 \sum y - \sum xy \sum x}{N \sum x^2 - (\sum x)^2}, \quad \hat{\theta}_2 = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2}.$$

Для ортогональной параметризации разд. 4.3.4, а именно для

$$Y_i = \theta_1^* + \theta_2^*(x_i - \bar{x}) + Z_i,$$

матричное уравнение (П4.1.7) сводится к

$$\begin{pmatrix} N & 0 \\ 0 & \sum (x - \bar{x})^2 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1^* \\ \hat{\theta}_2^* \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum y(x - \bar{x}) \end{pmatrix}.$$

Отсюда выборочные оценки наименьших квадратов равны

$$\hat{\theta}_1^* = \bar{y} = \frac{1}{N} \sum y, \quad \hat{\theta}_2^* = \frac{\sum y(x - \bar{x})}{\sum (x - \bar{x})^2} = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sum (x - \bar{x})^2}.$$

Матрица ковариаций оценок. Чтобы оценить точность выборочных оценок параметров, нужно вычислить матрицу ковариаций соответствующих оценок. Диагональные элементы этой матрицы дают дисперсии каждой из оценок, а недиагональные элементы дают ковариации каждой пары оценок.

Мы имеем

$$\hat{\theta} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}$$

и, воспользовавшись (П3.1.2), получаем

$$E[\hat{\theta}] = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} E[\mathbf{Y}].$$

Отсюда матрица ковариаций оценок равна

$$\begin{aligned} \mathbf{C} &= E \left[\left(\hat{\boldsymbol{\theta}} - E \left[\hat{\boldsymbol{\theta}} \right] \right) \left(\hat{\boldsymbol{\theta}} - E \left[\hat{\boldsymbol{\theta}} \right] \right)' \right] = \\ &= E \left[\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \left(\mathbf{Y} - E \left[\mathbf{Y} \right] \right) \left(\mathbf{Y} - E \left[\mathbf{Y} \right] \right)' \mathbf{V}^{-1} \mathbf{X} \left(\mathbf{X}\mathbf{V}^{-1}\mathbf{X}' \right)^{-1} \right] = \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1}. \end{aligned} \quad (\text{П4.1.8})$$

Если $\mathbf{V} = \sigma^2 \mathbf{I}$, то (П4.1.8) сводится к

$$\mathbf{C} = \left(\mathbf{X}'\mathbf{X} \right)^{-1} \sigma^2. \quad (\text{П4.1.9})$$

Для приводившегося выше примера с двухпараметрической моделью имеем

$$\left(\mathbf{X}'\mathbf{X} \right) = \begin{pmatrix} N & \sum x \\ \sum x & \sum x^2 \end{pmatrix},$$

так что

$$\mathbf{C} = \left(\mathbf{X}'\mathbf{X} \right)^{-1} \sigma^2 = \frac{\sigma^2}{N \sum (x - \bar{x})^2} \begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & N \end{pmatrix}.$$

Следовательно,

$$\begin{aligned} \text{Var} \left[\hat{\theta}_1 \right] &= \frac{\sigma^2 \sum x^2}{N \sum (x - \bar{x})^2}, \\ \text{Var} \left[\hat{\theta}_2 \right] &= \frac{\sigma^2}{\sum (x - \bar{x})^2}, \\ \text{Cov} \left[\hat{\theta}_1, \hat{\theta}_2 \right] &= -\frac{\sigma^2 \sum x}{N \sum (x - \bar{x})^2}. \end{aligned}$$

Для ортогональной параметризации

$$\left(\mathbf{X}'\mathbf{X} \right) = \begin{pmatrix} N & 0 \\ 0 & \sum (x - \bar{x})^2 \end{pmatrix},$$

так что

$$\mathbf{C} = \frac{\sigma^2}{N \sum (x - \bar{x})^2} \begin{pmatrix} \sum (x - \bar{x})^2 & 0 \\ 0 & N \end{pmatrix}.$$

Следовательно,

$$\begin{aligned} \text{Var} \left[\hat{\theta}_1^* \right] &= \frac{\sigma^2}{N}, \\ \text{Var} \left[\hat{\theta}_2^* \right] &= \frac{\sigma^2}{\sum (x - \bar{x})^2}, \\ \text{Cov} \left[\hat{\theta}_1^*, \hat{\theta}_2^* \right] &= 0. \end{aligned}$$

Ранее было показано, что оценки наименьших квадратов минимизируют среднеквадратичную ошибку (т. е. дисперсию, так как оценки несмещенные) линейной функции $\lambda' \boldsymbol{\theta}$ параметров $\boldsymbol{\theta}$.

Так как

$$\text{Var} [\lambda' \hat{\theta}] = \lambda' C \lambda, \quad (\text{П4.1.10})$$

то отсюда следует, что оценки наименьших квадратов минимизируют определитель $|C|$ матрицы ковариаций оценок $\hat{\theta}$.

Оценивание остаточной дисперсии. Эта задача в ее наиболее общей постановке включала бы оценивание всех элементов матрицы ковариаций ошибок V . В этом разделе мы рассмотрим лишь частный случай $V = \sigma^2 I$, так что оценивание V сводится к оцениванию σ^2 , являющейся дисперсией каждой из Z_i .

Пусть $S(\theta)$ обозначает квадратичную форму

$$S(\theta) = \sigma^2 z' V^{-1} z = \sigma^2 (y - X\theta)' V^{-1} (y - X\theta),$$

которая сводится для случая $V = \sigma^2 I$ к сумме квадратов

$$S(\theta) = (y - X\theta)' (y - X\theta).$$

Подставив выборочные оценки $\theta = \hat{\theta}$, получим

$$S(\hat{\theta}) = y'y - \hat{\theta}' X'y - y'X\hat{\theta} + \hat{\theta}' X'X\hat{\theta},$$

но так как, согласно (П4.1.7), $\hat{\theta}' X'y = \hat{\theta}' X'X\hat{\theta}$, то

$$S(\hat{\theta}) = y'y - y'X\hat{\theta}$$

и, заменяя $y'X$ на $\hat{\theta}' X'X$, получаем

$$S(\hat{\theta}) = y'y - \hat{\theta}' (X'X) \hat{\theta}. \quad (\text{П4.1.11})$$

Заменяя выборочные оценки в (П4.1.11) оценками и беря математическое ожидание, получаем [2, 3*]

$$E [S(\hat{\theta})] = N\sigma^2 - k\sigma^2 = (N - k)\sigma^2,$$

так что

$$s^2 = \frac{y'y - \hat{\theta}' (X'X) \hat{\theta}}{N - k} \quad (\text{П4.1.12})$$

является несмещенной оценкой σ^2 . Для однопараметрического случая равенство (П4.1.12) сводится к

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \hat{\theta}_1 x_i)^2,$$

как и получалось в разд. 4.3.2. Для двухпараметрического случая равенство (П4.1.12) имеет вид

$$\begin{aligned}
 s^2 &= \frac{1}{N-2} \sum_{i=1}^N (y_i - \hat{\theta}_1 - \hat{\theta}_2 x_i)^2 = \\
 &= \frac{1}{N-2} \left(\sum_{i=1}^N y_i^2 - N\hat{\theta}_1^2 - 2\hat{\theta}_1 \hat{\theta}_2 \sum_{i=1}^N x_i - \hat{\theta}_2^2 \sum_{i=1}^N x_i^2 \right),
 \end{aligned}$$

и для ортогонального двухпараметрического случая

$$\begin{aligned}
 s^2 &= \frac{1}{N-2} \sum_{i=1}^N [y_i - \hat{\theta}_1^* - \hat{\theta}_2^* (x_i - \bar{x})]^2 = \\
 &= \frac{1}{N-2} \left[\sum_{i=1}^N (y_i - \bar{y})^2 - (\hat{\theta}_2^*)^2 \sum_{i=1}^N (x_i - \bar{x})^2 \right].
 \end{aligned}$$

С помощью рассуждений, аналогичных рассуждениям, использовавшимся в разд. 4.2.3, можно показать, что оценка, имеющая $N+2-k$ в знаменателе, дает наименьшую среднеквадратичную ошибку, и, следовательно, она предпочтительней, чем (П4.1.12). Однако на практике чаще всего используют выборочную оценку (П4.1.12).

Доверительные области. Чтобы вывести доверительные области для θ , рассмотрим тождество

$$y - X\theta = (y - X\hat{\theta}) - X(\theta - \hat{\theta}).$$

Отсюда

$$\begin{aligned}
 \frac{S(\theta)}{\sigma^2} &= (y - X\theta)' V^{-1} (y - X\theta) = \\
 &= (y - X\hat{\theta})' V^{-1} (y - X\hat{\theta}) + (\theta - \hat{\theta})' X' V^{-1} X (\theta - \hat{\theta}) - \\
 &- (y - X\hat{\theta})' V^{-1} X (\theta - \hat{\theta}) - (\theta - \hat{\theta})' X' V^{-1} (y - X\hat{\theta}).
 \end{aligned}$$

Из нормальных уравнений (П4.1.4) следует, что два последних члена тождественно равны нулю. Исчезновение этих членов со смешанными произведениями обусловлено тем, что векторы $y - X\hat{\theta}$ и $X(\hat{\theta} - \theta)$ ортогональны в N -мерном выборочном пространстве. Отбрасывая эти исчезающие члены и заменяя y на Y и $\hat{\theta}$ на $\hat{\Theta}$, получим

$$S(\theta) = S(\hat{\Theta}) + (\theta - \hat{\Theta})' X' V^{-1} X (\theta - \hat{\Theta}) \sigma^2. \quad (\text{П4.1.13})$$

Предполагая, что ошибки Z_i распределены нормально, получаем отсюда, что $S(\theta)$ является квадратичной формой от N нормаль-

ных случайных величин и, следовательно, является случайной величиной χ^2_N . Эта случайная величина, согласно (П4.1.13), разлагается на χ^2_{N-k} и χ^2_k . Отсюда случайная величина

$$\sigma^2 \frac{(\theta - \hat{\theta})' \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} (\theta - \hat{\theta})}{S(\hat{\theta})} \frac{N-k}{k}$$

распределена, как $F_{k, N-k}$. Следовательно, область, вероятность попадания в которую есть $(1 - \alpha)$, имеет вид

$$\sigma^2 (\theta - \hat{\theta})' \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} (\theta - \hat{\theta}) \leq \frac{k}{N-k} f_{k, N-k} (1 - \alpha) S(\hat{\theta}). \quad (\text{П4.1.14})$$

Заменяя $\hat{\theta}$ на $\hat{\theta}$ в (П4.1.14), получаем $100(1 - \alpha)\%$ -ную доверительную область для параметров θ . Область (П4.1.14) является эллипсоидом в k -мерном пространстве параметров θ , и ее объем, как нетрудно проверить, обратно пропорционален определителю $|\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}|$. Но $\mathbf{C} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1}$, и так как выборочные оценки наименьших квадратов минимизируют определитель $|\mathbf{C}|$, то они, следовательно, минимизируют также и объем доверительного эллипсоида для параметров.

Подставляя $\mathbf{V} = \mathbf{I}\sigma^2$ в (П4.1.14) и замечая, что из (П4.1.12) следует, что $s^2 = S(\hat{\theta}) / (N - k)$ является выборочной оценкой σ^2 , получаем $100(1 - \alpha)\%$ -ную доверительную область для θ

$$(\theta - \hat{\theta})' \mathbf{X}' \mathbf{X} (\theta - \hat{\theta}) \leq k s^2 f_{k, N-k} (1 - \alpha), \quad (\text{П4.1.15})$$

в случае когда $\mathbf{V} = \mathbf{I}\sigma^2$.

Для одного параметра (П4.1.15) имеет вид

$$(\theta - \hat{\theta})^2 \leq \frac{s^2}{\sum x^2} f_{1, N-1} (1 - \alpha), \quad (\text{П4.1.16})$$

что является другой записью доверительного интервала (4.3.11), так как $t^2_{N-1} (1 - \alpha/2) = f_{1, N-1} (1 - \alpha)$. Для двухпараметрического примера неравенство (П4.1.15) принимает вид

$$\begin{aligned} ((\theta_1 - \hat{\theta}_1), (\theta_2 - \hat{\theta}_2)) \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \begin{pmatrix} \theta_1 - \hat{\theta}_1 \\ \theta_2 - \hat{\theta}_2 \end{pmatrix} \leq \\ \leq 2s^2 f_{2, N-2} (1 - \alpha). \end{aligned}$$

Это сводится к неравенству

$$\begin{aligned} N (\theta_1 - \hat{\theta}_1)^2 + 2 (\theta_1 - \hat{\theta}_1) (\theta_2 - \hat{\theta}_2) \sum_{i=1}^N x_i + (\theta_2 - \hat{\theta}_2)^2 \sum_{i=1}^N x_i^2 \leq \\ \leq 2s^2 f_{2, N-2} (1 - \alpha), \end{aligned}$$

которое представляет собой уравнение эллипса на плоскости (θ_1, θ_2) .

Для ортогональной двухпараметрической модели (П4.1.15) сводится к

$$N(\theta_1^* - \hat{\theta}_1^*)^2 + (\theta_2^* - \hat{\theta}_2^*)^2 \sum_{i=1}^N (x_i - \bar{x})^2 \leq 2s^2 f_{2, N-2}(1-\alpha),$$

что также является уравнением эллипса на плоскости (θ_1^*, θ_2^*) , но в этом случае оси эллипса параллельны осям координат.

Вывод доверительных областей непосредственно по контурам, образуемым линиями уровня суммы квадратов. В нелинейных задачах невозможно вывести явные выражения для выборочных оценок наименьших квадратов и матрицы $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. Примеры таких задач приводятся в разд. 5.4.4. В этом случае разложение (П4.1.13) можно записать в виде

$$S(\theta) = S(\hat{\theta}) + S(\theta - \hat{\theta}).$$

Используя те же рассуждения, что и при выводе (П4.1.14), получаем, что случайная величина

$$\frac{S(\theta) - S(\hat{\theta})}{S(\hat{\theta})} \frac{N-k}{k}$$

распределена, как $F_{k, N-k}$. Следовательно, область

$$S(\theta) \leq S(\hat{\theta}) \left[1 + \frac{k}{N-k} f_{k, N-k}(1-\alpha) \right] \quad (\text{П4.1.17})$$

является $100(1-\alpha)\%$ -ной доверительной областью для параметров. Если имеются контуры функции $S(\theta)$, то $100(1-\alpha)\%$ -ный контур соответствует константе (уровню), полученной умножением остаточной суммы квадратов $S(\hat{\theta})$ на константу в квадратных скобках в (П4.1.17).

Дисперсия прогноза. Если для предсказания отклика в будущем эксперименте используется модель (П4.1.1), то значение прогнозируемой величины будет иметь вид

$$\begin{aligned} \hat{y} &= \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 + \dots + \hat{\theta}_k x_k + \hat{z} = \\ &= \hat{\theta}' \mathbf{x} + \hat{z} = \hat{\theta}' \mathbf{x}, \end{aligned}$$

так как $\hat{z} = 0$. Дисперсия соответствующей оценки равна

$$\text{Var}[\hat{Y}] = \text{Var}[\hat{\theta}' \mathbf{x}] + \text{Var}[\hat{Z}] = \mathbf{x}' \mathbf{C} \mathbf{x} + \sigma^2, \quad (\text{П4.1.18})$$

где мы использовали (П3.1.4). Если $\mathbf{V} = \sigma^2 \mathbf{I}$, то (П4.1.18) сводится к

$$\text{Var} [\hat{Y}] = [\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x} + 1] \sigma^2. \quad (\text{П4.1.19})$$

Отсюда $100(1 - \alpha)\%$ -ный доверительный интервал, основанный на предсказываемом значении \hat{y} и выборочной оценке s^2 дисперсии σ^2 , имеет вид

$$\hat{y} \pm t_{N-k} \left(1 - \frac{\alpha}{2}\right) s \sqrt{1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}}. \quad (\text{П4.1.20})$$

ЛИТЕРАТУРА

1. Barnard G. A., J. Roy. Stat. Soc., **B25**, 124 (1963)
 2. Blackett R. L., Principles of Regression Analysis. Clarendon Press, Oxford, 1960.

ВВЕДЕНИЕ В АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

В этой главе мы рассмотрим основные понятия теории временных рядов. Наиболее важными среди них являются понятия случайного процесса, стационарного процесса, линейного стационарного процесса и ковариационной функции стационарного процесса. В разд. 5.1 показано, что для описания статистической природы наблюдаемого временного ряда нужно рассматривать его как элемент абстрактного множества функций, называемого случайным процессом. Простейшим типом случайного процесса является линейный процесс, который можно получить в результате линейной операции над чисто случайным процессом. Большое практическое значение имеют два частных случая линейного процесса: процесс авторегрессии и процесс скользящего среднего. В разд. 5.2 показано, что стационарный случайный процесс общего типа удобно описывать с помощью его ковариационной функции, в то время как линейный стационарный процесс лучше всего описывается его параметрами. В разд. 5.3 рассматривается оценивание ковариационной функции по наблюдаемому временному ряду, а в разд. 5.4 — оценивание параметров процессов авторегрессии и скользящего среднего.

5.1. СТАЦИОНАРНЫЕ И НЕСТАЦИОНАРНЫЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ

5.1.1. Определение и классификация временных рядов

Под статистическим временным рядом понимают сигнал, или функцию времени $x(t)$, проявляющую свойства *случайности*, или *нерегулярного изменения*. Имея запись такого ряда, невозможно точно предсказать его будущие значения в отличие от детерминированного сигнала, как указывалось в гл. 2. Следовательно, такой ряд может быть описан только с помощью статистических законов, или моделей, которые можно было бы использовать, например, для прогноза будущих значений ряда. Примерами таких статистических рядов являются:

а) напряжение в сети, флуктуирующее из-за случайного движения электронов, которое называют обычно тепловым шумом;

- б) флуктуирующий выход продукции химического реактора, измеряемый непрерывно с помощью инфракрасного спектрометра;
 в) выходное напряжение приемного устройства радиолокатора. Типичный отраженный радиолокационный сигнал показан на рис. 5.1.

Дискретные и непрерывные ряды. Временные ряды в примерах (а), (б) и (в) являются непрерывными измерениями и называются *непрерывными временными рядами*. Другой тип рядов представляют собой *дискретные временные ряды*, значения которых заданы только в определенные моменты времени. Один из способов, с помощью которых может быть получен дискретный временной



Рис. 5.1. Отраженный радиолокационный сигнал.

ряд, состоит в отсчете значений непрерывного временного ряда через равные промежутки времени, скажем Δ . Записывая $x(t=r\Delta) = x_r$, мы получим последовательность $\{x_r\}$, $r = \dots, -2, -1, 0, 1, 2$, образующую дискретный временной ряд.

Непрерывные временные ряды в примерах (а), (б) и (в) должны быть записаны с помощью физического инструмента, обладающего инерцией. Поэтому такие ряды имеют *ограниченную полосу частот*, т. е. они не содержат частот выше некоторой максимальной частоты, определяемой частотной характеристикой инструмента. Таким образом, используя теорию гл. 2, можно определить интервал отсчета Δ так, чтобы дискретный временной ряд $\{x_t\}$, полученный из значений непрерывного временного ряда $x(t)$, содержал бы всю информацию, имевшуюся в исходном ряде $x(t)$. Следовательно, непрерывный временной ряд можно анализировать либо в аналоговой (непрерывной), либо в цифровой (дискретной) форме.

Дискретный временной ряд может также получаться, когда физическая величина не имеет мгновенных значений, а приобретает смысл лишь в *накопленном*, или проинтегрированном по соответствующему временному интервалу, виде. Примерами таких накопленных рядов являются цифры суточных осадков, даваемые метеостанцией, или же выход продукции в последовательных партиях некоторого промышленного процесса. Пример дискретного временного ряда приведен на рис. 5.2, где показаны значения накопленной выходной продукции в 70 последовательных партиях, получен-

ных на дистилляционной колонке, в зависимости от номера партии. Данные, по которым построен этот рисунок, даны в табл. 5.1. Заметим, что, хотя данные и подчиняются определенной высокочастотной структуре, предсказать точное значение следующей партии невозможно.

Многомерные временные ряды. Во многих случаях представляет интерес вектор

$$\mathbf{x}(t) = \{x_1(t), x_2(t), \dots, x_q(t)\},$$

состоящий из временных рядов. В этом случае $\mathbf{x}(t)$ называют *многомерным временным рядом*. Так, например, мы имеем двумерный

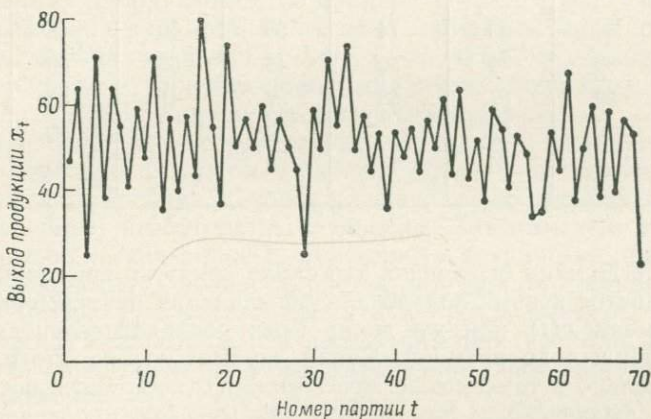


Рис. 5.2. Выход продукции в 70 последовательных партиях промышленного процесса.

временной ряд в радиолокации, когда $x_1(t)$ является «курсовым углом цели» радиолокатора по азимуту, а $x_2(t)$ — «курсовым углом цели» по возвышению. Заметим, что эти два временных ряда равноправны в том смысле, что ни один из них не влияет на другой, ибо они характеризуют различные виды движения радиолокационной антенны.

Иногда $x_1(t), \dots, x_l(t)$ являются входными сигналами некоторой физической системы, а $x_{l+1}(t), \dots, x_q(t)$ — соответствующими выходными сигналами той же системы. Например, $x_1(t)$ могло бы характеризовать скорость потока холодной воды у впускного отверстия в водонагреватель, а $x_2(t)$ — температуру потока у выпускного отверстия. Зная устройство водонагревателя, можно сделать разумный прогноз ряда $x_2(t)$ по ряду $x_1(t)$. В этом случае временные ряды неравноправны, поскольку изменения $x_1(t)$ могут вызвать изменения в $x_2(t)$, но не наоборот.

Таблица 5.1

Выход продукции в 70 последовательных партиях промышленного процесса

1—10	11—18	19—28	29—36	37—46	47—54	55—64	65—70
47	71	37	25	45	44	53	39
64	35	74	59	54	64	49	59
23	56	51	50	36	43	34	40
71	40	58	71	54	52	35	57
38	58	50	56	48	38	54	54
65	44	60	74	55	60	45	23
55	80	44	50	45	55	68	
41	55	57	58	57	41	38	
59		50		50		50	
48		45		62		60	

Теперь должно быть ясно, что слово «ряд» во временных рядах употребляется весьма вольно для обозначения непрерывных функций времени $x(t)$ или же дискретных последовательностей $\{x_t\}$, упорядоченных во времени. Слово «время» также употребляется весьма вольно в том смысле, что t может относиться к некоторому другому физическому параметру, такому, как *пространственная координата*. Например, при изучении вибрации самолета иногда производят эксперименты, в которых датчики деформации прикрепляются к крылу или к какой-нибудь другой части самолета, и флуктуирующие напряжения в этой структуре измеряются на различных высотах полета. Хотя самолет летит в течение некоторого промежутка времени, полученная запись является скорее функцией области в пространстве, пересеканной самолетом, чем функцией времени.

Временные ряды, зависящие от нескольких переменных. В некоторых случаях ряд является функцией $x(t_1, t_2, \dots, t_n)$ нескольких физических параметров t_1, t_2, \dots, t_n . В этом случае он называется *случайным полем*. Например, $x(t_1, t_2)$ может представлять измерения локальных флуктуаций магнитного поля Земли в точке с координатами t_1, t_2 . В других случаях изучаемый процесс может быть многомерным, а также зависеть от нескольких переменных. Например, геофизики заинтересованы в исследовании соотношения между магнитным полем Земли $x_1(t_1, t_2)$ и глубиной океана $x_2(t_1, t_2)$.

5.1.2. Описание случайного процесса

Определение случайного процесса. При анализе данных в виде временных рядов возникает необходимость выполнять различные операции над фактическими числами, полученными из некоторого эксперимента. До того как данные собраны, удобно рассматривать их, как это делается во всех статистических работах, как один из многих наборов данных, которые могли бы быть получены из этого эксперимента. Это достигается тем, что с каждым моментом времени t в интервале $(-\infty \leq t \leq \infty)$ связывается некоторая случайная величина $X(t)$, имеющая выборочное пространство $\{-\infty \leq X(t) \leq \infty\}$ и плотность вероятности $f_{X(t)}(x)$. Кроме того, нужно задать совместные плотности вероятности, относящиеся к любому произвольному набору моментов времени (t_1, t_2, \dots, t_n) . Таким образом, временной ряд можно описать с помощью упорядоченного множества случайных величин $X(t)$ $(-\infty \leq t \leq \infty)$ в случае непрерывного ряда и упорядоченного множества случайных величин $\{X_t\}$, $t = 0, \pm 1, \pm 2, \dots$, в случае дискретного ряда.

Упорядоченное множество случайных величин называется *случайным процессом*. Случайный процесс дает вероятностное описание физического явления, которое развивается во времени по вполне определенным вероятностным законам. Заметим, что выборочное пространство, связанное с одномерным случайным процессом, дважды бесконечно. Оно простирается для каждого момента времени от $-\infty$ до $+\infty$, и само время изменяется также от $-\infty$ до $+\infty$. Дважды бесконечное множество функций времени, которые могут быть определены на этом выборочном пространстве, называется *ансамблем*.

Ансамбль. Наблюденный временной ряд $x(t)$ рассматривается как одна реализация бесконечного ансамбля функций, которые могли бы быть наблюдаемы. Этот один или несколько временных рядов, предназначенных для анализа, рассматриваются как выбранные случайным образом из такого ансамбля функций, наподобие того, как производится выборка лиц среди населения для проведения обследования. Такая выборка должна быть представительной для всего описываемого населения, что следует учитывать, начиная обследование. Точно так же при работе с временными рядами для ясного понимания вопроса полезно в начале исследования описать точный характер ансамбля, или популяции, типичным представителем которого считается наблюдаемый временной ряд.

Предположим, например, что нужно измерить высоту волн с помощью передатчика, прикрепленного к бую. Если буй бросают в море случайным образом, то наблюдаемый временной ряд можно было бы рассматривать как один из многих рядов, которые,

возможно, наблюдались бы при слегка отличающихся местоположениях буя. Более тщательная проверка, возможно, обнаружила бы, что этот временной ряд типичен, если принять во внимание конкретное время дня или время года или же конкретный район в океане. Чем больше факторов влияют на эксперимент, тем шире становится ансамбль описываемых временных рядов и, следовательно, тем большая осторожность требуется при интерпретации результатов.

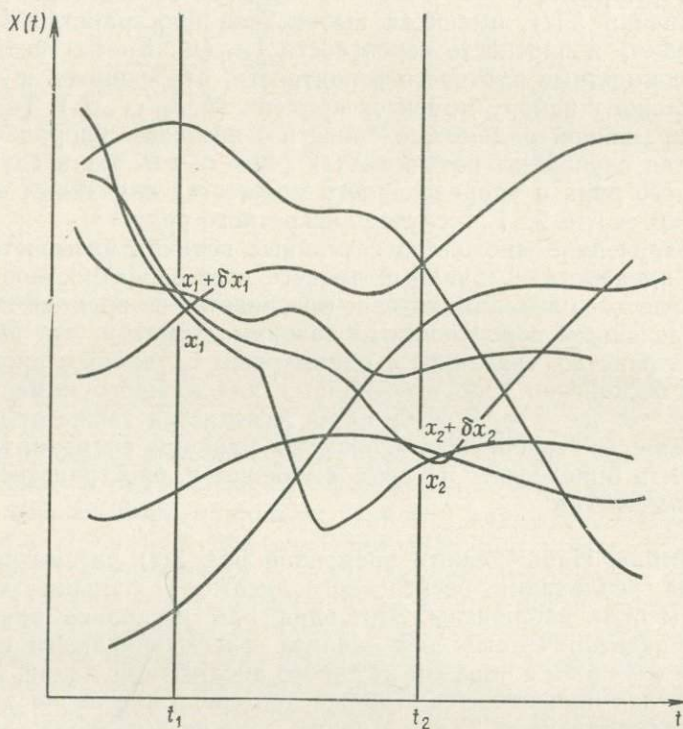


Рис. 5.3. Представители ансамбля, образованного случайным процессом.

Во многих практических задачах интересно знать, как изменяются свойства временного ряда, когда некоторые внешние условия намеренно изменяются по плану эксперимента. В других случаях невозможно осуществлять контроль над внешними факторами. Например, нельзя управлять солнечной радиацией при изучении ее влияния на статистические свойства атмосферной турбулентности. Тем не менее корреляция статистических свойств временных рядов с этими неконтролируемыми факторами может ока-

заться важнейшим выводом, полученным из анализа рядов. Основная цель этого обсуждения заключается в том, чтобы показать следующее: вопрос с том, из чего должен состоять ансамбль возможных временных рядов в любой конкретной задаче, должен решаться на основе разумных научных заключений, а не на основе чисто статистических аспектов.

Распределения вероятностей, связанные со случайным процессом. Если ансамбль ясно определен, то поведение временного ряда в данный момент времени можно описать до сбора данных с помощью случайной величины $X(t)$ и ее плотности вероятности $f_{X(t)}(x)$ *). Как подчеркивалось в гл. 3, выбор функции $f_{X(t)}(x)$ является делом здравого суждения или опыта.

Аналогично случайные величины $X(t_1)$ и $X(t_2)$, соответствующие двум моментам времени t_1 и t_2 , можно описать с помощью их совместной плотности вероятности $f_{X(t_1)X(t_2)}(x_1, x_2)$, которую сокращенно обозначим $f_{12}(x_1, x_2)$. Один из способов интерпретации этих плотностей вероятности состоит в следующем: через отверстие шириной δx_1 , расположенное около точки x_1 , в момент времени t_1 проходят члены ансамбля функций, составляющие от него долю $f_1(x_1) \delta x_1$, как показано на рис. 5.3; аналогично через отверстие от x_1 до $x_1 + \delta x_1$ в момент времени t_1 и отверстие от x_2 до $x_2 + \delta x_2$ в момент времени t_2 проходят члены ансамбля, составляющие долю $f_{12}(x_1, x_2) \delta x_1 \delta x_2$.

Другой полезной функцией является *условная плотность вероятности*

$$f_{2|1}(x_1, x_2) = \frac{f_{12}(x_1, x_2)}{f_1(x_1)}, \quad (5.1.1)$$

которая читается как «плотность вероятности X_2 при условии $X_1 = x_1$ ». Таким образом, из проходящих через отверстие от x_1 до $x_1 + \delta x_1$ в момент t_1 членов ансамбля, составляющих от него долю $f_1(x_1) \delta x_1$, только часть функций, составляющая от прошедших долю $f_{2|1}(x_1, x_2) \delta x_2$, пройдет через отверстие от x_2 до $x_2 + \delta x_2$ в момент t_2 .

В общем случае одномерный случайный процесс может быть описан с помощью совместных плотностей вероятности

$$f_{X(t_1) X(t_2) \dots X(t_n)}(x_1, x_2, \dots, x_n)$$

для произвольно выбранного набора моментов времени t_1, t_2, \dots, t_n , но такое описание могло бы быть довольно сложным. На практике необходим более простой подход, основанный на младших моментах.

*) Строго говоря, случайные величины $X(t)$ следует описывать их функциями распределения, так как плотности вероятности могут не существовать, но мы не будем обращать на это внимания, поскольку это несущественно с практической точки зрения.

Простые моменты случайного процесса. Для любого t можно определить одномерные моменты вида

$$E [(X(t))^k] = \int_{-\infty}^{\infty} x^k f_{X(t)}(x) dx. \quad (5.1.2)$$

Отсюда очень просто описать случайный процесс, построив функцию среднего значения $\mu(t)$ и функцию дисперсии $\sigma^2(t)$ в зависимости от времени. Аналогично двумерные моменты

$$E [(X(t_1))^k (X(t_2))^l] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^k x_2^l f_{X(t_1) X(t_2)}(x_1, x_2) dx_1 dx_2 \quad (5.1.3)$$

можно было бы использовать для описания зависимости между значениями временного ряда в двух соседних точках t_1 и t_2 . Простейшим из моментов (5.1.3) и наиболее важным на практике является автоковариационная функция *)

$$\gamma_{XX}(t_1, t_2) = E [(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))]. \quad (5.1.4)$$

Отметим, что $\gamma_{XX}(t_1, t_1) = \sigma^2(t_1)$. Заметим также, что ковариационная функция временного ряда имеет те же свойства, что и ковариация между двумя случайными величинами X_1 и X_2 (эти свойства мы перечислили в разд. 3.2.2).

Поскольку $\gamma_{XX}(t_1, t_2)$ зависит от масштаба измерения X , удобно при сравнении двух рядов, которые могут иметь различные масштабы измерения, определить нормированную величину, называемую автокорреляционной функцией *)

$$\rho_{XX}(t_1, t_2) = \frac{\gamma_{XX}(t_1, t_2)}{\sigma(t_1) \sigma(t_2)}.$$

Подобно обычному коэффициенту корреляции (3.2.19), $\rho_{XX}(t_1, t_2)$ лежит между крайними значениями -1 и $+1$, соответствующими полной отрицательной и положительной линейным зависимостям.

Вообще, случайный процесс можно было бы описывать с помощью его старших моментов

$$E [(X(t_1))^{k_1} (X(t_2))^{k_2} \dots (X(t_n))^{k_n}], \quad (5.1.5)$$

но они не очень полезны на практике.

5.1.3. Стационарность и ковариационная функция

Стационарность. В общем случае свойства случайного процесса будут зависеть от времени. Часто ради упрощения предполагают, что ряд достиг некоторой формы устойчивого состояния, или равновесия, в том смысле, что статистические свойства ряда не зависят

*) Там, где это не может привести к недоразумениям, мы будем использовать более простые термины: ковариационная функция и корреляционная функция. — Прим. перев.

от абсолютного времени. Например, было бы разумно предположить для данных о партиях продукта, приведенных на рис. 5.2, что если бы контроль за процессом осуществлялся достаточно хорошо, то статистические свойства ряда оставались бы довольно стабильными во времени. Наименьшее требование для того, чтобы это было верно, состоит в том, что плотность вероятности $f_{X(t)}(x)$ не должна зависеть от времени и, следовательно, стационарный временной ряд имеет постоянное среднее значение μ и постоянную дисперсию σ^2 . Поэтому одинаковую для всех моментов времени плотность вероятности $f_X(x)$ можно оценить, построив гистограмму данных так, как это описывалось в гл. 3. Например, гистограмма данных табл. 5.1 показана на рис. 5.4.

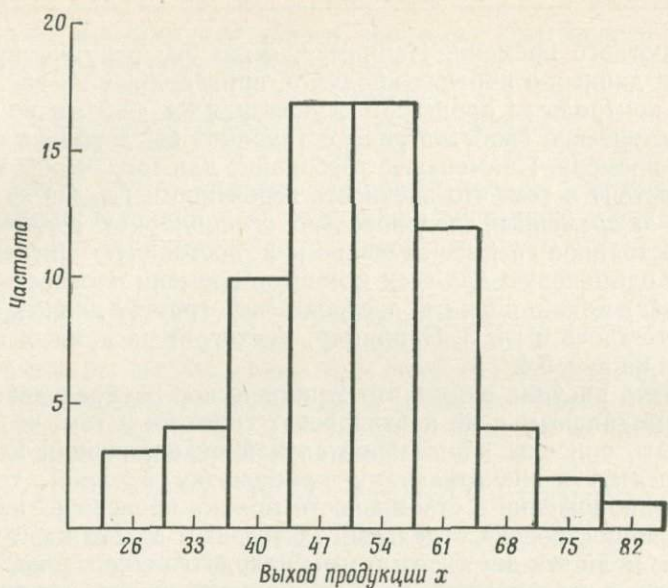
Из этого рисунка видно, что эмпирическое распределение является унимодальным и не противоречит гипотезе о том, что данные могут быть описаны нормальной случайной величиной. Если данные относятся к значительному промежутку времени, то разумность предположения о стабильности можно проверить, например, с помощью построения отдельных гистограмм для каждой из половин ряда. Если эти две гистограммы находятся в согласии, то предположение о независимости от времени вероятности, по-видимому, оправдано.

Из предположения о том, что процесс находится в состоянии равновесия, вытекает и другое следствие: совместная плотность вероятности $f_{12}(x_1, x_2)$ зависит только от разности моментов времени $t_2 - t_1$, а не от абсолютных значений t_1 и t_2 . Предположим, что временной ряд — дискретный и что наблюдаемыми значениями являются x_1, x_2, \dots, x_n . Тогда пары точек $(x_1, x_{k+1}), (x_2, x_{k+2}), \dots, (x_{n-k}, x_n)$ можно рассматривать как $(n-k)$ наблюдений, имеющих совместную плотность вероятности $f_{12}(x_1, x_2)$, которая в этом случае одинакова для всех моментов времени, отстоящих друг от друга на $k\Delta$.

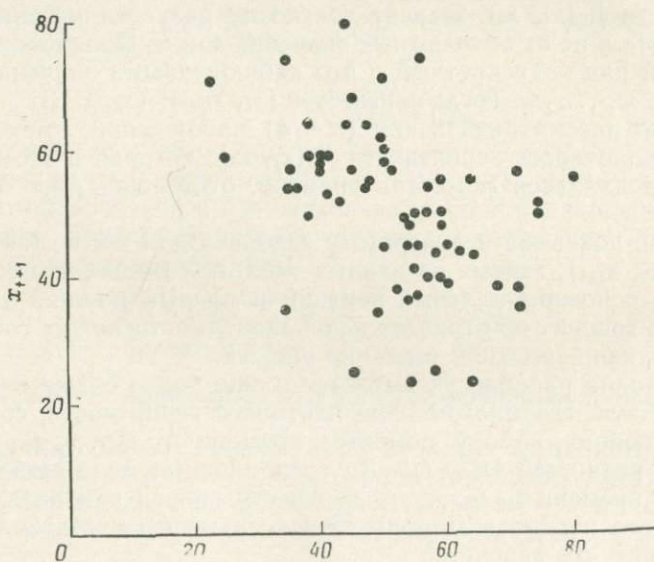
Рис. 5.5 показывает диаграмму разброса для последовательных партий (x_t, x_{t+1}) , взятых из данных табл. 5.1. Видно, что точки падают в основном в левый верхний и правый нижний углы рисунка, что говорит об отрицательной зависимости между соседними партиями, явно заметной также на рис. 5.2.

Из условия равновесия вытекает и еще более общее следствие, а именно свойства многомерной плотности вероятности, соответствующей любому набору моментов времени t_1, t_2, \dots, t_n , зависят только от разностей $|t_i - t_j|$. Другими словами, если любой набор моментов времени t_1, t_2, \dots, t_n перенести вперед или назад на величину k , то плотность вероятности не изменится. Математически это означает, что равенство

$$\begin{aligned} f_{X(t_1) X(t_2) \dots X(t_n)}(x_1, x_2, \dots, x_n) &= \\ = f_{X(t_1+k) \dots X(t_n+k)}(x_1, x_2, \dots, x_n), & \quad (5.1.6) \end{aligned}$$



Р и с. 5.4. Гистограмма данных о последовательных партиях продукции, приведенных на рис. 5.2.



Р и с. 5.5. Диаграмма разброса для пар последовательных партий, приведенных на рис. 5.2.

справедливо для любых наборов моментов времени и для всех смещений k . Случайный процесс, удовлетворяющий условиям (5.1.6), называется *строго стационарным* *).

Ковариационная функция. Из предположения стационарности сразу следует, что ковариационная функция $\gamma_{XX}(t_1, t_2)$ зависит только от $u = t_2 - t_1$ и, следовательно, ее можно записать в виде

$$\gamma_{XX}(u) = E [(X(t) - \mu)(X(t+u) - \mu)] = \text{Cov} [X(t), X(t+u)]. \quad (5.1.7)$$

Смещение u называется *запаздыванием*. Ковариационная функция показывает, как изменяется зависимость между соседними значениями случайного процесса в зависимости от запаздывания u . Если $X(t)$ имеют многомерную нормальную плотность, то ковариационная функция и среднее значение полностью характеризуют процесс, как отмечалось в разд. 3.1.5.

Корреляционная функция. Для стационарного процесса корреляционная функция

$$\rho_{XX}(u) = \frac{\gamma_{XX}(u)}{\gamma_{XX}(0)} \quad (5.1.8)$$

зависит только от запаздывания u . Хотя методы оценивания ковариационных и корреляционных функций будут рассмотрены лишь в разд. 5.3, мы проиллюстрируем сейчас на рис. 5.6 выборочную корреляционную функцию для данных о партиях продукта, изображенных на рис. 5.2. Видно, что корреляции затухают очень быстро и что практически нет корреляции для запаздываний больше 10. Кроме того, корреляции меняют знак. Это говорит о том, что за высоким выходом продукта в одной партии в среднем получается низкий выход в следующей партии, и наоборот.

Слабая стационарность. Более слабое, чем (5.1.6), предположение, которое иногда принимают, состоит в том, что многомерные моменты вида (5.1.5) вплоть до порядка

$$k_1 + k_2 + \dots + k_n = K$$

зависят только от разностей моментов времени $|t_i - t_j|$. Случайный процесс с таким свойством называется стационарным процессом K -го порядка. Например, если $K = 2$, то только среднее значение, дисперсия и ковариационная функция (5.1.7) зависят от разностей моментов времени, и процесс является стационарным

*) Такие процессы называют также стационарными в узком смысле, или вполне стационарными. — *Прим. перев.*

второго порядка^{*)}. Впрочем, если многомерная плотность вероятности в (5.1.6) является нормальной (так что она полностью задается ее средними значениями и ковариациями), то из стационарности второго порядка следует строгая стационарность.

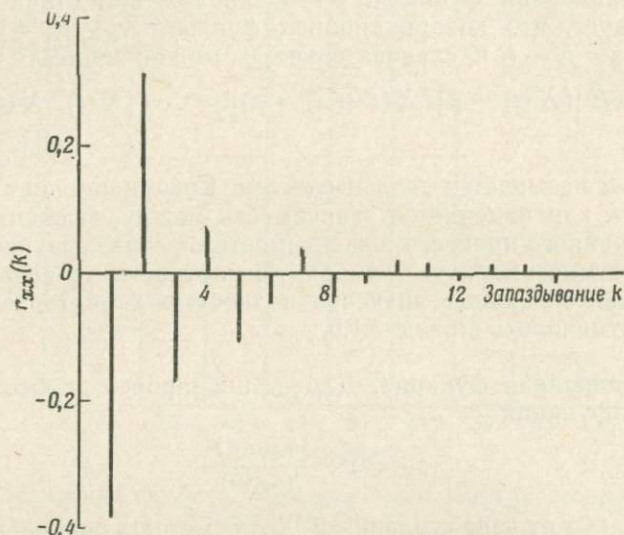


Рис. 5.6. Выборочная корреляционная функция для данных, приведенных на рис. 5.2, $r_{xx}(0)=1$.

Чисто случайный процесс. Простейшим примером стационарного процесса является дискретный процесс Z_t , такой, что случайные величины Z_t взаимно независимы и одинаково распределены. В этом случае из (5.1.7) следует, что $\gamma_{zz}(k)=0$ для всех $k \neq 0$. Такой процесс статистики называют *чисто случайным процессом*, а инженеры — *белым шумом с ограниченной полосой частот*.

5.1.4. Классификация временных рядов, встречающихся на практике

В большинстве работ, посвященных анализу временных рядов, исследователи рассматривают свойства стационарных процессов. Если эмпирический ряд не является стационарным, то можно применить различные приемы для устранения очевидных трендов, после чего оставшийся ряд будет разумно считать стационарным. Например, данные можно приблизить некоторыми математическими

^{*)} Такие процессы называются также стационарными в широком смысле, или слабо стационарными. — Прим. перев.

функциями, такими, как экспоненты, ряды Фурье или многочлены. Другим приемом, который будет широко использоваться в последующих главах, является фильтрация или устранение низкочастотных трендов с помощью соответствующим образом рассчитанного фильтра.

Временные ряды, встречающиеся обычно на практике, можно разделить на три большие категории.

а) Ряды, являющиеся стационарными в течение относительно больших промежутков времени, благодаря некоторой форме контроля над внешними условиями. Примерами могут служить ряды, получаемые из генераторов случайного шума, таких, как электронные лампы, температура которых поддерживается постоянной, и вольферовский ряд чисел солнечных пятен, зарегистрированных в течение нескольких столетий. С практической точки зрения невероятно долго, так что важным является вопрос, насколько длинную запись можно взять для анализа, чтобы при этом не нарушилось предположение о стационарности. Для геофизических рядов, таких, как числа солнечных пятен, условия могут оставаться стабильными в течение столетий. Однако в других случаях условия могут оставаться стабильными лишь в течение часов или минут или, возможно, вообще быть совершенно нестабильными.

б) Ряды, с которыми можно обращаться как со стационарными лишь при условии, что рассматриваются достаточно короткие реализации. Ошибки, допускаемые оператором при слежении, рассматриваются как стационарные, если только характеристики прослеживаемого сигнала поддерживаются неизменными, а измерения производятся в течение достаточно коротких промежутков времени так, чтобы оператор не успел устать. Если измеряют напряжение в некоторой точке самолета, летящего через турбулентную среду, то хорошо известно, что в течение коротких промежутков времени, скажем до получаса, ряд можно рассматривать как стационарный. Для больших периодов времени дисперсия ряда может измениться заметным образом из-за изменения уровня, или интенсивности, турбулентности.

в) Ряды, которые, совершенно очевидно, являются нестационарными как по своему виду, так и из-за априорных сведений об изучаемом явлении.

Формы проявления нестационарности рядов могут быть различными. Приведем несколько примеров простейших форм нестационарности.

Нестационарность среднего значения. Многие ряды проявляют нестационарность лишь в виде тренда среднего значения, не обнаруживая видимым образом каких-либо более сложных форм отклонения от стационарности. Например, большинство

увеличивающейся амплитудой. Таким образом, процесс случайного блуждания имеет как нестационарное среднее, так и нестационарную дисперсию.

Используя (5.1.4), получаем, что ковариационная функция процесса X_t равна

$$\gamma_{XX}(t_1, t_2) = \min(t_1, t_2) \sigma_Z^2.$$

Процесс, определяемый равенствами (5.1.9), называют также процессом с *некоррелированными*, или *ортогональными*, приращениями, так как приращения

$$Z_t = X_t - X_{t-1}$$

образуют чисто случайный процесс и, следовательно, Z_t некоррелировано (имеет нулевую ковариацию) с другими приращениями, такими, как $Z_{t-1} = X_{t-1} - X_{t-2}$, и т. д.

5.1.5. Анализ систем на основе критерия минимума среднеквадратичной ошибки

В предыдущих разделах обсуждены простые способы описания временных рядов с помощью их младших моментов. Важнейшим из этих моментов является корреляционная функция. Одно из многих применений корреляционной функции состоит в том, что она служит источником исходных идей при построении вероятностной модели механизма, породившего временной ряд. В следующей главе будет показано, что временной ряд можно описать совершенно эквивалентным образом с помощью его спектральной плотности, являющейся преобразованием Фурье от ковариационной функции.

Широкое применение ковариационной функции или спектральной плотности в технических задачах основано на том, что знание любой из этих функций достаточно для синтеза линейных фильтров или линейных систем регулирования с минимальной среднеквадратичной ошибкой для случаев, когда рассматриваемые сигналы искажаются шумом. Теория синтеза систем с минимальной среднеквадратичной ошибкой была впервые разработана Винером [4]. Она сыграла важную роль в развитии современной теории управления и теории связи.

Синтез следящих систем. Одно из первых инженерных применений анализа на основе критерия минимума среднеквадратичной ошибки было сделано при синтезировании следящих систем для зенитных орудий и в радиолокационных следящих системах [5]. Например, от радиолокационной следящей системы требуется, чтобы она следила за самолетом несмотря на возмущения отра-

женного радиосигнала, обусловленные вариациями полного коэффициента отражения из-за вращения пропеллера, вибрации моторов и изменений относительного положения самолета, вызванных рысканием и покачиванием во всех направлениях. Понятно, нельзя ожидать от следящей системы, чтобы она сопровождала самолет абсолютно точно при таких неблагоприятных условиях. Следовательно, нужно было бы исследовать характеристику работы системы в среднем и ее вероятный разброс, а не точную характеристику. Один из способов описания этих свойств использует среднеквадратичную ошибку между желаемым и действительным выходными сигналами системы. В свою очередь, среднеквадратичную ошибку можно выразить через ковариационную функцию входного и желаемого выходного сигналов. Поэтому знание ковариационных функций достаточно для синтеза систем с минимальной среднеквадратичной ошибкой.

Оценивание отклика линейной системы на единичный импульс. Другое применение критерий минимума среднеквадратичной ошибки находит в задаче об «идентификации системы». В этом случае в распоряжении имеются входной сигнал и соответствующий ему выходной сигнал от некоторой системы; требуется вывести линейное приближение к этой системе для дальнейшего его использования при управлении или моделировании. Предположим, например, что система представляет собой «черный ящик» (рис. 5.7). Если вход является реализацией случайного процесса $X(t)$, то выход можно рассматривать как реализацию случайного процесса $Y(t)$, где

$$Y(t) - \mu_Y = \int_0^{\infty} h(u) [X(t-u) - \mu_X] du + Z(t). \quad (5.1.10)$$

Равенство (5.1.10) утверждает, что выход можно рассчитать, беря взвешенное среднее от входного сигнала, причем весовая функция должна равняться $h(u)$. В (5.1.10) $Z(t)$ является шумом, или членом ошибки, содержащим систематическую компоненту (обусловленную несовершенством аппроксимации линейной системы) и случайную компоненту, обусловленную ошибками измерения и недостаточным контролем над переменными, управляющими выходом.

Если ковариационные функции процессов $X(t)$ и $Y(t)$ известны точно, то можно воспользоваться винеровским критерием минимума среднеквадратичной ошибки. Этот критерий утверждает, что функция $h(u)$ должна быть выбрана так, чтобы минимизировать среднеквадратичную ошибку шумовой компоненты, т. е.

$$E [Z^2(t)] = E \left[\left\{ (Y(t) - \mu_Y) - \int_0^{\infty} h(u) (X(t-u) - \mu_X) du \right\}^2 \right]. \quad (5.1.11)$$

Целесообразность использования критерия (5.1.11) мы обсудим полнее в разд. 5.3.1, где будет рассмотрена задача идентификации системы по записям конечной длины.

Если предположить, что процессы $X(t)$ и $Y(t)$ стационарны, то (5.1.11) можно записать в виде

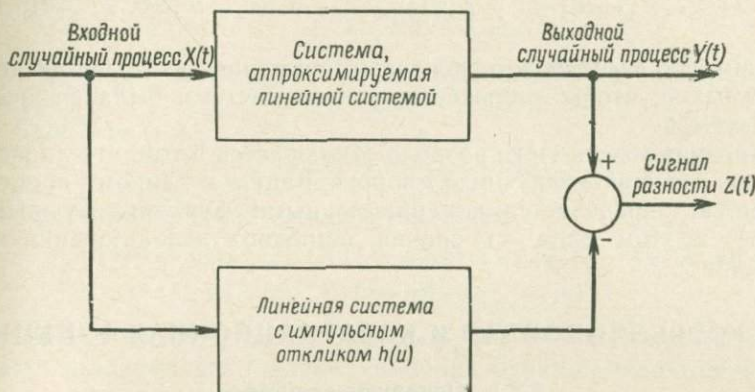


Рис. 5.7. Определение отклика на единичный импульс на основе минимума среднеквадратичной ошибки.

$$\begin{aligned}
 E[Z^2(t)] &= E[(Y(t) - \mu_Y)^2] - 2E\left[(Y(t) - \mu_Y) \int_0^{\infty} h(u) \times \right. \\
 &\quad \left. \times (X(t-u) - \mu_X) du\right] + E\left[\int_0^{\infty} \int_0^{\infty} h(u) h(v) \times \right. \\
 &\quad \left. \times (X(t-u) - \mu_X)(X(t-v) - \mu_X) du dv\right] = \\
 &= \gamma_{YY}(0) - 2 \int_0^{\infty} h(u) \gamma_{XY}(u) du + \int_0^{\infty} \int_0^{\infty} h(u) \times \\
 &\quad \times h(v) \gamma_{XX}(u-v) du dv, \tag{5.1.12}
 \end{aligned}$$

где $\gamma_{YY}(0) = \sigma_Y^2$ — дисперсия $Y(t)$;

$$\gamma_{XY}(u) = E[(X(t) - \mu_X)(Y(t+u) - \mu_Y)]$$

есть взаимная ковариационная функция между $X(t)$ и $Y(t+u)$ и

$$\gamma_{XX}(u) = E[(X(t) - \mu_X)(X(t+u) - \mu_X)]$$

есть автоковариационная функция $X(t)$.

Отсюда среднеквадратичная ошибка полностью определяется ковариационными функциями $\gamma_{YY}(0)$, $\gamma_{XY}(u)$, $\gamma_{XX}(u)$ и откликом на единичный импульс $h(u)$.

Функцию $h(u)$, дающую минимальную среднеквадратичную ошибку, можно получить с помощью вариационного исчисления, как показано в приложении П5.1, откуда следует, что $h(u)$ должна удовлетворять интегральному уравнению Винера—Хопфа

$$\gamma_{XY}(u) = \int_{-\infty}^{\infty} h(v) \gamma_{XX}(u-v) dv, \quad u \geq 0. \quad (5.1.13)$$

Заметим, что $h(v)$ должна тождественно равняться нулю при отрицательных v , чтобы аппроксимирующая система была физически реализуемой.

Основная мысль этого раздела заключается в том, что линейная система, дающая наилучшую аппроксимацию к данному процессу, полностью определяется ковариационными функциями $\gamma_{XX}(u)$ и $\gamma_{XY}(u)$. В этом одна из причин широкого использования этих функций.

5.2. КОРРЕЛЯЦИОННАЯ И КОВАРИАЦИОННАЯ ФУНКЦИИ

5.2.1. Основные свойства

В этом разделе выводятся свойства корреляционной и ковариационной функций. Взаимную ковариационную функцию $\gamma_{XY}(u)$, введенную в разд. 5.1.5, мы будем подробно обсуждать в гл. 8.

В общем случае случайный процесс $X(t)$ имеет ковариационную функцию

$$\text{Cov}[X(t), X(t+u)] = \gamma_{XX}(t, t+u) \quad (5.2.1)$$

и корреляционную функцию

$$\begin{aligned} \rho_{XX}(t, t+u) &= \frac{\text{Cov}[X(t), X(t+u)]}{\{\text{Var}[X(t)] \text{Var}[X(t+u)]\}^{1/2}} = \\ &= \frac{\gamma_{XX}(t, t+u)}{\{\gamma_{XX}(t, t) \gamma_{XX}(t+u, t+u)\}^{1/2}} = \frac{\gamma_{XX}(t, t+u)}{\sigma_X(t) \sigma_X(t+u)}. \end{aligned} \quad (5.2.2)$$

Если $X(t)$ — стационарный, то (5.2.1) и (5.2.2) сводятся к

$$\text{Cov}[X(t), X(t+u)] = \gamma_{XX}(u) \quad (5.2.3)$$

и

$$\rho_{XX}(u) = \frac{\gamma_{XX}(u)}{\gamma_{XX}(0)} = \frac{\gamma_{XX}(u)}{\sigma_X^2} \quad (5.2.4)$$

соответственно. Отсюда

$$\gamma_{XX}(u) = \sigma_X^2 \rho_{XX}(u).$$

Функция $\rho_{XX}(u)$, зависящая от запаздывания u , называется *корреляционной функцией* стационарного процесса $X(t)$. Если процесс

непрерывный, u может принимать любое значение от $-\infty$ до $+\infty$; для дискретного же процесса $\rho_{XX}(u)$ будет определена лишь для целых значений u .

Ниже перечислены и коротко объяснены свойства корреляционной функции (5.2.4).

Свойство 1.

$$\rho_{XX}(0) = 1.$$

Это немедленно следует из определения (5.2.4), если положить $u = 0$.

Свойство 2.

$$\rho_{XX}(u) = \rho_{XX}(-u).$$

Из-за стационарности процесса мы имеем

$$\begin{aligned} \gamma_{XX}(u) &= \text{Cov}[X(t), X(t+u)] = \text{Cov}[X(t-u), X(t)] = \\ &= \text{Cov}[X(t), X(t-u)] = \gamma_{XX}(-u). \end{aligned}$$

Из (5.2.4) следует, что $\rho_{XX}(u) = \rho_{XX}(-u)$. Следовательно, как ковариационная, так и корреляционная функции являются четными функциями от запаздывания u . Поэтому их нужно вычислять лишь для неотрицательных u .

Свойство 3.

$$|\rho_{XX}(u)| \leq 1 \text{ для всех } u.$$

Это можно получить из того факта, что дисперсия случайной величины

$$Y = \lambda_1 X(t) + \lambda_2 X(t+u)$$

неотрицательна, с помощью рассуждений, аналогичных приведенным в разд. 3.2.4.

Свойство 4. Корреляционная матрица является положительно полуопределенной, т. е. определитель

$$\begin{vmatrix} 1 & \rho_{XX}(t_2 - t_1) & \rho_{XX}(t_3 - t_1) & \dots & \rho_{XX}(t_n - t_1) \\ \rho_{XX}(t_2 - t_1) & 1 & \rho_{XX}(t_3 - t_2) & \dots & \rho_{XX}(t_n - t_2) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{XX}(t_n - t_1) & \rho_{XX}(t_n - t_2) & \rho_{XX}(t_n - t_3) & \dots & 1 \end{vmatrix}$$

и все его главные миноры неотрицательны. Этот результат является более общим, чем свойство 3. Он следует из того, что дисперсия случайной величины

$$Y = \lambda_1 X(t_1) + \dots + \lambda_n X(t_n)$$

неотрицательна. Из свойства 4 вытекает, что корреляции стационарного процесса не могут быть произвольными, но должны удовлетворять некоторым соотношениям. Заметим, что при $n=2$ свойство 4 сводится к свойству 3. Свойство 4 положительной полуопределенности приводит к понятию спектра мощности процесса, которое будет обсуждаться подробнее в гл. 6 и 11.

Свойство 5. Если случайный процесс является непрерывным, то $\rho_{xx}(u)$ должна быть непрерывной функцией от запаздывания u . Это условие непрерывности требуется для того, чтобы можно было построить разумную математическую теорию для непрерывного времени. На самом деле достаточно потребовать лишь, чтобы $\rho_{xx}(u)$ была непрерывной при $u=0$, так как из этого вытекает непрерывность во всех других точках.

Белый шум. Одно из следствий свойства 5 состоит в том, что невозможно определить непрерывный по времени случайный процесс, являющийся аналогом чисто случайного процесса с дискретным временем, введенного в разд. 5.1.3. Для такого непрерывного случайного процесса потребовалось бы, чтобы $\rho_{zz}(0)=1$ и $\rho_{zz}(u)=0$ при $u \neq 0$, но такая корреляционная функция была бы разрывной при $u=0$.

Один выход из этой трудности заключается в том, чтобы определить чисто случайный процесс для непрерывного времени, или *белый шум*, как процесс, который состоит целиком из некоррелированных смежных импульсов. Таким образом, его ковариационная функция будет равна

$$\gamma_{zz}(u) = \sigma_z^2 \delta(u), \quad (5.2.5)$$

где $\delta(u)$ — дельта-функция Дирака. Поскольку $\delta(u)$ можно рассматривать как функцию, равную нулю при $u \neq 0$ и бесконечную при $u=0$, то мы добились того, что ковариация между соседними точками равна нулю, хотя для этого пришлось сделать бесконечной дисперсию процесса $\gamma_{zz}(0)$. Ниже мы покажем, что бесконечная величина дисперсии получается неизбежно.

5.2.2. Линейный процесс и его ковариационная функция

Важный класс процессов можно получить с помощью пропуска чистого случайного процесса через линейную систему, или фильтр. Для непрерывного времени соотношение между выходным процессом $X(t)$ и входным процессом $Z(t)$ можно записать в виде

$$X(t) - \mu = \int_0^{\infty} h(v) Z(t-v) dv, \quad (5.2.6)$$

а для дискретного времени — в виде

$$X_t - \mu = \sum_{k=0}^{\infty} h_k Z_{t-k}, \quad (5.2.7)$$

где $E[Z(t)] = 0$ и $E[Z_t] = 0$. Случайный процесс, который получается из белого шума с помощью выражений (5.2.6) или (5.2.7), называется *линейным процессом*.

Беря математическое ожидание от обеих частей в (5.2.6), получим

$$E[X(t) - \mu] = \int_0^{\infty} h(v) E[Z(t-v)] dv = 0,$$

т. е.

$$E[X(t)] = \mu.$$

Отсюда ковариационная функция выхода равна

$$\begin{aligned} \gamma_{XX}(u) &= E[(X(t) - \mu)(X(t+u) - \mu)] = \\ &= \int_0^{\infty} \int_0^{\infty} h(v) h(v') E[Z(t-v)Z(t+u-v')] dv dv'. \end{aligned} \quad (5.2.8)$$

Если процесс $Z(t)$ стационарен и имеет ковариационную функцию $\gamma_{ZZ}(u)$, то (5.2.8) сводится к

$$\gamma_{XX}(u) = \int_0^{\infty} \int_0^{\infty} h(v) h(v') \gamma_{ZZ}(u+v-v') dv dv'. \quad (5.2.9)$$

Подставляя в (5.2.9) вместо $\gamma_{ZZ}(u)$ ковариационную функцию белого шума (5.2.5), получим

$$\gamma_{XX}(u) = \sigma_Z^2 \int_0^{\infty} h(v) h(v+u) dv. \quad (5.2.10)$$

Отсюда корреляционная функция линейного процесса $X(t)$ равна

$$\rho_{XX}(u) = \frac{\int_0^{\infty} h(v) h(v+u) dv}{\int_0^{\infty} h^2(v) dv}. \quad (5.2.11)$$

В гл. 6 будет показано, что процесс $X(t)$ является стационарным, если

$$\int_0^{\infty} |h(v)| dv < M, \quad (5.2.12)$$

где M — конечное число. Заметим, что условие (5.2.12) совпадает с условием (2.3.11) устойчивости линейной системы.

В приложении П5.2 как обобщение результата (5.2.10) получены следующие выражения для третьего и четвертого моментов и для четвертого кумулянта:

$$E [(X(t) - \mu)(X(t+u_1) - \mu)(X(t+u_2) - \mu)] = \\ = E [Z^3] \int_0^{\infty} h(v) h(v+u_1) h(v+u_2) dv, \quad (5.2.13)$$

$$E [(X(t) - \mu)(X(t+u_1) - \mu)(X(t+u_2) - \mu)(X(t+u_3) - \mu)] = \\ = E [Z^4] \int_0^{\infty} h(v) h(v+u_1) h(v+u_2) h(v+u_3) dv \quad (5.2.14)$$

и

$$K_X(u_1, u_2, u_3) = K_4(Z) \int_0^{\infty} h(v) h(v+u_1) h(v+u_2) h(v+u_3) dv, \quad (5.2.15)$$

где

$$K_X(u_1, u_2, u_3) = E [(X(t) - \mu)(X(t+u_1) - \mu)(X(t+u_2) - \mu) \times \\ \times (X(t+u_3) - \mu)] - \gamma_{XX}(u_1)\gamma_{XX}(u_3 - u_2) - \\ - \gamma_{XX}(u_2)\gamma_{XX}(u_3 - u_1) - \gamma_{XX}(u_3)\gamma_{XX}(u_2 - u_1) \quad (5.2.16)$$

и

$$K_4(Z) = E [Z^4(t)] - 3\sigma_Z^4.$$

Соответствующие формулы для дискретного линейного процесса (5.2.7) можно получить, заменяя интегралы на суммы. Например, (5.2.10) переходит в

$$\gamma_{XX}(k) = \sigma_Z^2 \sum_{j=0}^{\infty} h_j h_{j+k}, \quad (5.2.17)$$

а условие стационарности, или стабильности, соответствующее условию (5.2.12), имеет вид

$$\sum_{j=0}^{\infty} |h_j| < M, \quad (5.2.18)$$

где M — конечное число. Неравенство (5.2.18) совпадает с условием устойчивости (2.3.39) для дискретных систем.

Пример линейного процесса. Как частный случай линейного процесса рассмотрим среднее значение процесса $Z(t)$, сосчитанное по конечному интервалу T , т. е.

$$X(t) = \frac{1}{T} \int_{t-T}^t Z(v) dv.$$

Отсюда

$$h(v) = \begin{cases} 0, & v < 0, \\ \frac{1}{T}, & 0 \leq v \leq T, \\ 0, & v > T. \end{cases}$$

Математическое ожидание $X(t)$ равно

$$E[X(t)] = \frac{1}{T} \int_{t-T}^t E[Z(v)] dv = 0$$

при условии, что $E[Z(t)] = 0$. Если $Z(t)$ — произвольный стационарный процесс, то, используя (5.2.9), получаем ковариационную функцию процесса $X(t)$:

$$\gamma_{XX}(u) = \frac{1}{T^2} \int_0^T \int_0^T \gamma_{ZZ}(u + v - v') dv dv'.$$

Сделав замену $y = v - v'$ и положив $u = 0$, получим

$$\text{Var}[X(t)] = \frac{1}{T} \int_{-T}^{+T} \left(1 - \frac{|y|}{T}\right) \gamma_{ZZ}(y) dy. \quad (5.2.19)$$

Если $Z(t)$ — белый шум, то (5.2.19) сводится к

$$\text{Var}[X(t)] = \sigma_Z^2 = \frac{\sigma_Z^2}{T}. \quad (5.2.20)$$

Заметим, что (5.2.20) совпадает в дискретном случае с выражением для дисперсии среднего арифметического, состоящего из T независимых случайных величин, а именно

$$\text{Var}[\bar{Z}] = \text{Var}\left[\frac{1}{T} \sum_{k=1}^T Z_k\right] = \frac{\sigma_Z^2}{T}.$$

Таким образом, для дискретного белого шума дисперсия среднего арифметического равна дисперсии сигнала Z , деленной на число наблюдений, но для белого шума с непрерывным временем конечная величина σ_Z^2/T получается при делении *бесконечной* дисперсии на *бесконечное* число независимых наблюдений. Этот пример достаточно хорошо показывает, что интерпретацию и получение результатов с помощью белого шума нужно проводить очень осторожно.

Следует отметить, что дельта-функция в выражении (5.2.5) для ковариационной функции белого шума является существенной частью, а не просто служит параметром «расположения». Это означает, что дисперсия действительно бесконечна и ковариация между сколь угодно близкими значениями действительно равна нулю. Для того чтобы физически сколь угодно близкие значения

процесса были некоррелированными, т. е. чтобы процесс мог без ограничений флуктуировать от момента к моменту, он должен иметь бесконечную дисперсию.

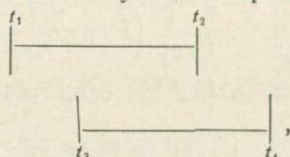
Процесс Башелье—Винера. По аналогии с дискретным процессом (5.1.9) для непрерывного времени также можно построить процесс, имеющий некоррелированные приращения. Формально непрерывный аналог случайного блуждания можно записать в виде

$$X(t) = \int_0^t Z(v) dv \quad (5.2.21)$$

или же

$$X(t_2) - X(t_1) = \int_{t_1}^{t_2} Z(v) dv.$$

Если $X(t)$ — непрерывный процесс с некоррелированными приращениями, то $E[(X(t_2) - X(t_1))(X(t_4) - X(t_3))]$ будет равно нулю для неперекрывающихся интервалов (t_1, t_2) и (t_3, t_4) . Если же интервалы перекрываются следующим образом:



то, записав

$$X(t_2) - X(t_1) = X(t_2) - X(t_3) + X(t_3) - X(t_1)$$

и

$$X(t_4) - X(t_3) = X(t_4) - X(t_2) + X(t_2) - X(t_3),$$

получим

$$E[(X(t_2) - X(t_1))(X(t_4) - X(t_3))] = E[(X(t_2) - X(t_3))^2].$$

Можно показать далее [6], что это математическое ожидание должно иметь вид $\sigma_Z^2 |t_2 - t_3|$ для любого разумного процесса с некоррелированными приращениями *).

Рассмотрим теперь производный процесс

$$Y(t) = \frac{X(t + \tau) - X(t)}{\tau}.$$

Из (5.2.11) при фиксированном τ получаем, что ковариационная функция процесса $Y(t)$ равна

$$\gamma_{YY}(u) = \begin{cases} 0, & |u| > \tau, \\ \frac{\sigma_Z^2}{\tau^2} (\tau - |u|), & |u| \leq \tau. \end{cases} \quad (5.2.22)$$

* Точнее, с некоррелированными стационарными приращениями. — Прим. перев.

Если $\tau \rightarrow 0$, то $\gamma_{XX}(u)$ стремится к $\sigma_Z^2 \delta(u)$ — ковариационной функции белого шума. Следовательно, белый шум можно представлять себе как несобственный случайный процесс, являющийся производной $\dot{X}(t)$ процесса с некоррелированными приращениями.

Если в дополнение к этому разность $X(t+\tau) - X(t)$ распределена по нормальному закону со средним значением $\tau\mu$ и дисперсией $\tau\sigma_Z^2$, то $Y(t)$ будет нормальным, или гауссовским, белым шумом, состоящим из некоррелированных импульсов, площадь которых имеет среднее значение μ и дисперсию σ_Z^2 . Этот процесс был использован Винером и другими для описания броуновского движения частицы, которая находится во взвешенном состоянии в некоторой среде и испытывает случайные соударения с частицами среды.

5.2.3. Процесс скользящего среднего конечного порядка

Предположим, что веса h_k линейного процесса (5.2.7) равны нулю при $k > l$, т. е.

$$X_t - \mu = h_0 Z_t + h_1 Z_{t-1} + \dots + h_l Z_{t-l}. \quad (5.2.23)$$

Тогда если Z_t — чисто случайный процесс, то X_t называется процессом *скользящего среднего* конечного порядка l .

Процессы скользящего среднего конечного порядка полезны во многих областях, например при прогнозе поведения эконометрических систем и систем управления. Однако наиболее полезны они в сочетании с процессами авторегрессии, которые будут введены в следующем разделе. Из (5.2.17) получаем, что ковариационная функция процесса скользящего среднего конечного порядка (5.2.23) равна нулю при $k > l$. Рассмотрим, например, процесс скользящего среднего второго порядка:

$$X_t - \mu = Z_t + 0,5Z_{t-1} + 0,5Z_{t-2}.$$

Из (5.2.17) получаем, что ковариационная функция процесса X_t равна

$$\begin{aligned} \gamma_{XX}(0) &= \sigma_Z^2 [1 + (0,5)^2 + (0,5)^2] = 1,50\sigma_Z^2, \\ \gamma_{XX}(1) &= \sigma_Z^2 [1(0,5) + (0,5)^2] = 0,75\sigma_Z^2, \\ \gamma_{XX}(2) &= \sigma_Z^2 [1(0,5)] = 0,5\sigma_Z^2, \\ \gamma_{XX}(k) &= 0, \quad k \geq 3. \end{aligned}$$

Отсюда корреляционная функция равна

$$\rho_{XX}(k) = \begin{cases} 1, & k=0, \\ 0,5, & k=1, \\ 0,333, & k=2, \\ 0, & k \geq 3. \end{cases}$$

Примером непрерывного процесса скользящего среднего конечного порядка является процесс $Y(t)$, использованный при выводе процесса Башелье—Винера, для которого ковариационная функция (5.2.22) равна нулю при $|u| > \tau$.

5.2.4. Процессы авторегрессии

Непрерывный процесс первого порядка. Рассмотрим линейную систему первого порядка, описываемую дифференциальным уравнением вида (2.3.2), а именно

$$T \frac{dx}{dt} + x(t) = z(t),$$

где $z(t)$ — вход системы, а $x(t)$ — ее выход. Если в эту систему вводится белый шум $Z(t)$, то выход $X(t)$ будет линейным процессом (5.2.6) с $h(v) = (1/T) \exp\{-v/T\}$. Процесс $X(t)$, определяемый уравнением

$$T \frac{dX(t)}{dt} + [X(t) - \mu] = Z(t), \quad (5.2.24)$$

называется *процессом авторегрессии* первого порядка. Из (5.2.11) следует, что корреляционная функция выхода равна

$$\rho_{XX}(u) = e^{-|u|/T}. \quad (5.2.25)$$

Условие устойчивости (5.2.12) требует, чтобы временная константа T была положительной. Это является также условием того, что процесс $X(t)$ — стационарный, а его дисперсия конечна.

Дискретный процесс первого порядка. Дискретный процесс авторегрессии первого порядка получается из чисто случайного процесса Z_t с помощью уравнения

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + Z_t. \quad (5.2.26)$$

Используя z -преобразование, (5.2.26) можно записать в виде

$$(1 - \alpha_1 z^{-1})(X_t - \mu) = Z_t.$$

Следовательно,

$$\begin{aligned} (X_t - \mu) &= \left(\frac{1}{1 - \alpha_1 z^{-1}} \right) Z_t = Z_t + \alpha_1 z^{-1} Z_t + \alpha_1^2 z^{-2} Z_t + \dots = \\ &= Z_t + \alpha_1 Z_{t-1} + \alpha_1^2 Z_{t-2} + \dots \end{aligned}$$

Используя (5.2.17), получаем отсюда, что корреляционная функция процесса авторегрессии X_t равна

$$\rho_{XX}(k) = \alpha_1^{|k|}, \quad k = 0, \pm 1, \pm 2, \dots \quad (5.2.27)$$

Условие устойчивости, или стационарности, (5.2.18) сводится теперь к условию $|\alpha_1| < 1$, так как $\sum_{k=0}^{\infty} |h_k| = 1/(1 - |\alpha_1|)$.

Пример. На рис. 5.8, а показан ряд из 40 членов, полученный согласно уравнению (5.2.26) при $\alpha_1 = 0,9$. Значения чисто случайного процесса Z_t брались из таблиц независимых нормальных чисел [7]. При $\alpha_1 = 0,9$ корреляционная функция равна $\rho_{XX}(k) = (0,9)^{|k|}$. Эта функция становится близкой к нулю лишь при больших значениях k . Таким образом, соседние точки процесса имеют

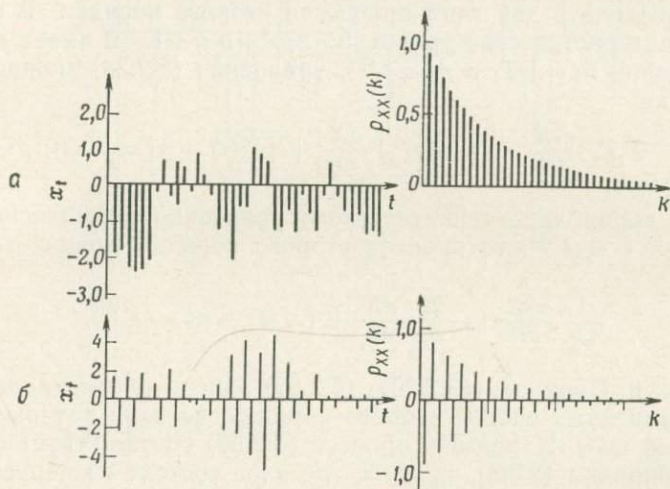


Рис. 5.8. Выборки процессов авторегрессии первого порядка и их теоретические корреляционные функции; а) $\alpha_1 = +0,9$; б) $\alpha_1 = -0,9$.

большую положительную корреляцию, например $\rho_{XX}(1) = 0,9$, и плавный характер ряда отражается в плавности корреляционной функции. Ряд, показанный на рис. 5.8, б, соответствует случаю $\alpha_1 = -0,9$. Соседние точки теперь имеют высокую отрицательную корреляцию, так как $\rho_{XX}(k) = (-0,9)^{|k|}$, и корреляционная функция осциллирует от положительных до отрицательных значений, отражая осциллирующий характер ряда. Заметим, что непрерывный процесс авторегрессии первого порядка (5.2.24) может приводить лишь к положительным корреляциям, и поэтому он соответствует дискретному случаю $\alpha_1 \geq 0$.

Процесс авторегрессии первого порядка иногда называют марковским процессом первого порядка. Это обусловлено тем, что случайная величина X_t при фиксированной X_{t-1} не зависит от предшествующих величин X_{t-2} , X_{t-3} и т. д. Из (5.2.26) видно, что если Z_t — нормальный процесс со средним значением 0 и

дисперсией σ_z^2 , то условная плотность вероятности $f_{X(t)|X(t-1)}(x_t, x_{t-1})$ является нормальной со средним значением $\alpha_1 x_{t-1}$ и дисперсией σ_z^2 .

Непрерывные процессы второго порядка. Непрерывный процесс авторегрессии второго порядка можно записать в виде

$$a_2 \frac{d^2 X}{dt^2} + a_1 \frac{dX}{dt} + [X(t) - \mu] = Z(t). \quad (5.2.28)$$

Можно различать два типа процессов второго порядка. В случае, если характеристическое уравнение $a_2 p^2 + a_1 p + 1 = 0$ имеет вещественные корни $\pi_1 = 1/T_1$ и $\pi_2 = 1/T_2$, уравнение (5.2.28) можно записать в виде

$$T_1 T_2 \frac{d^2 X}{dt^2} - (T_1 + T_2) \frac{dX}{dt} + [X(t) - \mu] = Z(t). \quad (5.2.29)$$

Если же корни характеристического уравнения комплексны $\pi_1 = \omega_n e^{j\theta}$, $\pi_2 = \omega_n e^{-j\theta}$, то процесс второго порядка можно записать в виде

$$\frac{1}{\omega_n^2} \frac{d^2 X}{dt^2} - \frac{2\zeta}{\omega_n} \frac{dX}{dt} + [X(t) - \mu] = Z(t), \quad (5.2.30)$$

где $\zeta = \cos \theta$. Процессы (5.2.29), (5.2.30) можно рассматривать как выходы линейных систем второго порядка, на вход которых подается белый шум. Например, процесс (5.2.30) соответствует системе второго порядка (2.3.8) из гл. 2, где вход состоит из непрерывной последовательности случайных импульсов. Отсюда выход $X(t)$ является непрерывной искаженной периодической функцией. Для того чтобы (5.2.30) имело смысл, необходимо принять, что изменения $Z(t)$ создают разрывные изменения ускорения выхода $\ddot{X}(t)$.

Дискретные процессы второго порядка. Для дискретного времени процесс авторегрессии второго порядка имеет вид

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \alpha_2 (X_{t-2} - \mu) + Z_t. \quad (5.2.31)$$

Моделью (5.2.31) пользовался в 1921 г. математик Юл. Он утверждал, что при $Z_t = 0$ в (5.2.31) эта модель описывает поведение простого маятника, демпфированного сопротивлением воздуха, пропорциональным его скорости. Если Z_t является чисто случайным процессом, то маятник подвергается случайным толчкам через равные промежутки времени. Вместо затухающих колебаний маятник теперь совершает возмущенное периодическое движение.

На рис. 5.9 показан ряд из 40 членов, полученный по схеме дискретного процесса авторегрессии второго порядка (5.2.31) при $\alpha_1 = 1,0$ и $\alpha_2 = -0,5$. Видно, что ряд имеет определенную периоди-

ческую структуру. Однако период и фазы постоянно изменяются благодаря воздействию случайной компоненты Z_t .

Процесс (5.2.31) можно рассматривать как выход дискретной линейной системы, на вход которой подается чисто случайный про-

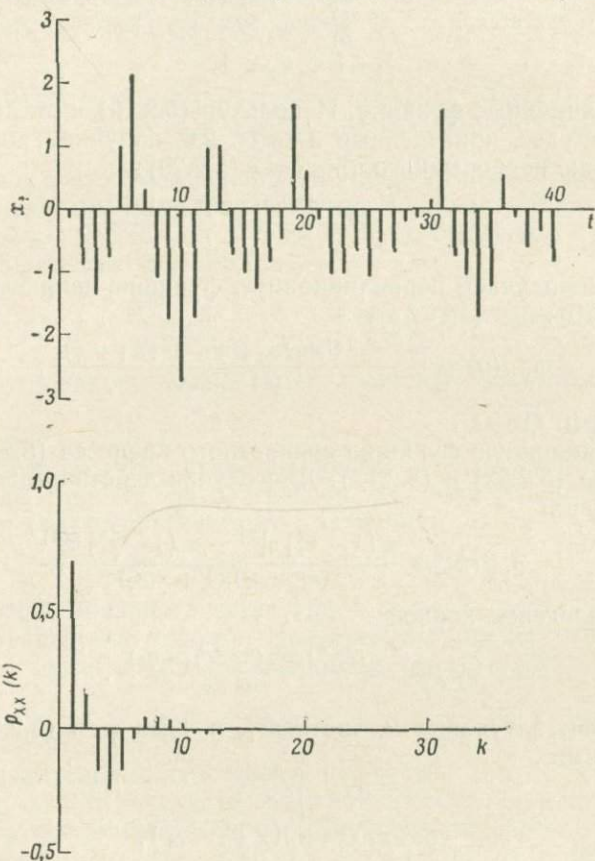


Рис. 5.9. Выборка процесса авторегрессии второго порядка и теоретическая корреляционная функция.

цесс Z_t . Функция отклика этой системы на единичный импульс была введена в разд. 2.3.5. Она равна

$$h_k = \frac{\pi_1^{k+1} - \pi_2^{k+2}}{\pi_1 - \pi_2} \quad (5.2.32)$$

для случая $\alpha_1^2 \geq -4\alpha_2$. Если $\alpha_1^2 < -4\alpha_2$, то

$$h_k = \frac{R^k [\sin 2\pi f_0 (k+1)]}{\sin 2\pi f_0}, \quad (5.2.33)$$

где $\pi_1 = \operatorname{Re} j^{2\pi f_0}$, $\pi_2 = \operatorname{Re} -j^{2\pi f_0}$. В гл. 2 было показано также, что для стационарности X_t нужно, чтобы параметры α_1 и α_2 из (5.2.31) лежали в треугольной области

$$\begin{aligned} \alpha_1 + \alpha_2 &< 1, \\ \alpha_1 - \alpha_2 &> -1, \\ -1 &< \alpha_2 < 1. \end{aligned} \quad (5.2.34)$$

Корреляционные функции. Используя (5.2.10) и отклики на единичный импульс, приведенные в табл. 2.6, получаем корреляционную функцию непрерывного процесса (5.2.29):

$$\rho_{XX}(u) = \frac{T_1 e^{-|u|/T_1} - T_2 e^{-|u|/T_2}}{T_1 - T_2}. \quad (5.2.35)$$

Аналогично получаем корреляционную функцию непрерывного процесса (5.2.30):

$$\rho_{XX}(u) = \frac{e^{-\zeta \omega_n |u|} \sin(\omega_n \sqrt{1 - \zeta^2} |u| + \varphi)}{\sqrt{1 - \zeta^2}}, \quad (5.2.36)$$

где $\varphi = \arcsin \sqrt{1 - \zeta^2}$.

Корреляционную функцию дискретного процесса (5.2.31) можно получить из (5.2.32) и (5.2.17). Для случая действительных корней она имеет вид

$$\rho_{XX}(k) = \frac{\pi_1 (1 - \pi_2^2) \pi_1^{|k|} - \pi_2 (1 - \pi_1^2) \pi_2^{|k|}}{(\pi_1 - \pi_2) (1 + \pi_1 \pi_2)} \quad (5.2.37)$$

и для комплексных корней

$$\rho_{XX}(k) = \frac{R^{|k|} \cos(2\pi f_0 k - \varphi_0)}{\cos \varphi_0}. \quad (5.2.38)$$

Коэффициент затухания R , частота f_0 и фаза φ_0 в (5.2.38) даются выражениями

$$\begin{aligned} R &= \sqrt{-\alpha_2}, \\ \cos 2\pi f_0 &= \alpha_1 / (2 \sqrt{-\alpha_2}), \\ \operatorname{tg} \varphi_0 &= \frac{1 - R^2}{1 + R^2} \operatorname{tg} 2\pi f_0. \end{aligned}$$

Для ряда, изображенного на рис. 5.9, где $\alpha_1 = 1,0$ и $\alpha_2 = -0,5$, коэффициент затухания $R = 0,71$, частота $f_0 = 0,125$ и фаза $\varphi_0 = 18^\circ 30'$. Корреляционная функция этого ряда построена под самим рядом на рис. 5.9. Видно, что она затухает очень быстро.

Из-за большого разнообразия корреляционных функций, порождаемых процессами авторегрессии, они находят широкое применение в качестве моделей для анализа стационарных временных рядов. Задача оценивания параметров процессов авторегрессии будет обсуждена в разд. 5.4.

5.2.5. Общие процессы скользящего среднего — авторегрессии

Этот раздел содержит краткую сводку наиболее важных свойств процессов авторегрессии и скользящего среднего. Общий процесс авторегрессии порядка m для дискретного времени порождается чисто случайным процессом Z_t с помощью разностного уравнения

$$(X_t - \mu) = \alpha_1 (X_{t-1} - \mu) + \alpha_2 (X_{t-2} - \mu) + \dots \\ \dots + \alpha_m (X_{t-m} - \mu) + Z_t. \quad (5.2.39)$$

Для непрерывного времени общий процесс авторегрессии определяется как выход линейного фильтра, на вход которого подается белый шум, а соотношение между входом и выходом определяется дифференциальным уравнением

$$a_m \frac{d^m X}{dt^m} + a_{m-1} \frac{d^{m-1} X}{dt^{m-1}} + \dots + a_0 [X(t) - \mu] = Z(t), \quad (5.2.40)$$

где, как отмечалось выше, $Z(t)$ изменяет разрывным образом $d^m X/dt^m$.

Устойчивость, или стационарность. 1) *Дискретный процесс.* Дискретный процесс авторегрессии X_t является стационарным, если корни характеристического уравнения

$$p^m - \alpha_1 p^{m-1} - \dots - \alpha_m = 0 \quad (5.2.41)$$

лежат внутри единичного круга $|p| = 1$.

2) *Непрерывный процесс.* Непрерывный процесс авторегрессии $X(t)$ будет стационарным, если корни характеристического уравнения

$$a_m p^m + a_{m-1} p^{m-1} + \dots + a_0 = 0 \quad (5.2.42)$$

имеют отрицательные действительные части.

В разд. 5.2.2 отмечалось, что условие стационарности совпадает с условием устойчивости соответствующей линейной системы. Поэтому условия (5.2.41) и (5.2.42) получаются из условий (2.3.38) и (2.3.20).

Корреляционная функция. 1) *Дискретный процесс.* Корреляционная функция дискретного процесса X_t удовлетворяет разностному уравнению

$$\rho_{XX}(k) = \alpha_1 \rho_{XX}(k-1) + \alpha_2 \rho_{XX}(k-2) + \dots \\ \dots + \alpha_m \rho_{XX}(k-m), \quad k \geq 1, \\ 1 = \alpha_1 \rho_{XX}(1) + \dots + \alpha_m \rho_{XX}(m) + \frac{\sigma_Z^2}{\gamma_{XX}(0)}, \quad k=0. \quad (5.2.43)$$

Общее решение этого разностного уравнения имеет вид

$$\rho_{XX}(k) = A_1 \pi_1^{|k|} + A_2 \pi_2^{|k|} + \dots + A_m \pi_m^{|k|}, \quad (5.2.44)$$

где π_i — корни (возможно, комплексные) уравнения (5.2.41). Если имеются комплексные корни, то они скомбинированы так, что в (5.2.44) получаются члены вида $R_i^{|k|} \cos(f_i k + \varphi_i)$. Поэтому, вообще говоря, корреляционная функция $\rho_{XX}(k)$ будет содержать показательные члены и затухающие синусоидальные волны. Константы A_i в (5.2.44) можно найти, решая первые m уравнений (5.2.43) относительно α_i , как показано ниже.

2) *Непрерывный процесс.* Корреляционная функция непрерывного процесса $X(t)$ удовлетворяет дифференциальному уравнению

$$a_m \frac{d^m \rho_{XX}(u)}{du^m} + a_{m-1} \frac{d^{m-1} \rho_{XX}(u)}{du^{m-1}} + \dots + \alpha_0 \rho_{XX}(u) = 0 \quad (u > 0). \quad (5.2.45)$$

Это уравнение имеет общее решение

$$\rho_{XX}(u) = A_1 e^{-\pi_1 |u|} + A_2 e^{-\pi_2 |u|} + \dots + A_m e^{-\pi_m |u|}, \quad (5.2.46)$$

где π_i — корни уравнения (5.2.42). Если имеются комплексные корни, то они скомбинированы так, что получаются члены вида $e^{-k|u|} \cos(k_2 u + \varphi)$.

Доказательство. Мы докажем упомянутые выше результаты только для дискретного случая. Если обе части равенства

$$(X_t - \mu) = \alpha_1 (X_{t-1} - \mu) + \dots + \alpha_m (X_{t-m} - \mu) + Z_t$$

умножить на $(X_{t-k} - \mu)$, то получим

$$(X_t - \mu)(X_{t-k} - \mu) = \alpha_1 (X_{t-1} - \mu)(X_{t-k} - \mu) + \dots \\ \dots + \alpha_m (X_{t-m} - \mu)(X_{t-k} - \mu) + Z_t (X_{t-k} - \mu).$$

Беря математическое ожидание от обеих частей, получаем

$$\sigma_X^2 \rho_{XX}(k) = \alpha_1 \sigma_X^2 \rho_{XX}(k-1) + \dots + \alpha_m \sigma_X^2 \rho_{XX}(k-m) + \\ + E[Z_t (X_{t-k} - \mu)].$$

Поскольку случайную величину X_{t-k} можно выразить в виде

$$X_{t-k} - \mu = \sum_{i=0}^{\infty} h_i Z_{t-k-i}$$

и так как это выражение не содержит Z_t , то $E[Z_t (X_{t-k} - \mu)] = 0$. Отсюда получается результат (5.2.43).

Пример. Корреляционная функция дискретного процесса авторегрессии второго порядка удовлетворяет рекуррентному уравнению

$$\rho_{XX}(k) = \alpha_1 \rho_{XX}(k-1) + \alpha_2 \rho_{XX}(k-2), \quad k > 0. \quad (5.2.47)$$

Это уравнение имеет решение

$$\rho_{XX}(k) = A_1 \pi_1^{k-1} + A_2 \pi_2^{k-1}, \quad (5.2.48)$$

где π_1, π_2 — корни характеристического уравнения $\rho^2 - \alpha_1 \rho - \alpha_2 = 0$. Отсюда

$$\pi_1 + \pi_2 = \alpha_1, \quad \pi_1 \pi_2 = -\alpha_2.$$

Далее, уравнение (5.2.47) при $k=1$ имеет вид

$$\rho_{XX}(1) = \alpha_1 \rho_{XX}(0) + \alpha_2 \rho_{XX}(-1).$$

Отсюда

$$\rho_{XX}(1) = \frac{\alpha_1}{1 - \alpha_2} = \frac{\pi_1 + \pi_2}{1 + \pi_1 \pi_2},$$

так как $\rho_{XX}(0) = 1, \rho_{XX}(-1) = \rho_{XX}(1)$.

Из (5.2.48) получаем

$$\begin{aligned} \rho_{XX}(0) &= 1 = A_1 + A_2, \\ \rho_{XX}(1) &= \frac{\pi_1 + \pi_2}{1 + \pi_1 \pi_2} = A_1 \pi_1 + A_2 \pi_2. \end{aligned}$$

Отсюда

$$A_1 = \frac{\pi_1(1 - \pi_2^2)}{(1 + \pi_1 \pi_2)(\pi_1 - \pi_2)}, \quad A_2 = \frac{-\pi_2(1 - \pi_1^2)}{(1 + \pi_1 \pi_2)(\pi_1 - \pi_2)}$$

и, таким образом,

$$\rho_{XX}(k) = \frac{(1 - \pi_2^2) \pi_1^{k+1} - (1 - \pi_1^2) \pi_2^{k+1}}{(1 + \pi_1 \pi_2)(\pi_1 - \pi_2)},$$

что согласуется с (5.2.37) для $k > 0$.

Свойство дискретизации по времени. Если значения непрерывного процесса авторегрессии (5.2.40) измерять через равные промежутки времени Δ , то получится дискретный процесс

$$\begin{aligned} (X_t - \mu) &= \alpha_1 (X_{t-1} - \mu) + \dots + \alpha_m (X_{t-m} - \mu) + \\ &+ Z_t + \beta_1 Z_{t-1} + \dots + \beta_{m-1} Z_{t-m+1}, \end{aligned} \quad (5.2.49)$$

где Z_t — чисто случайный процесс. Уравнение (5.2.49) представляет собой смесь дискретного процесса скользящего среднего (5.2.23) и дискретного процесса авторегрессии (5.2.39). Отметим интересную особенность (5.2.49): в то время как исходный непрерывный процесс имел в качестве входа белый шум, дискретный процесс

авторегрессии имеет в качестве входа процесс скользящего среднего, порядок которого на единицу меньше порядка дифференциального уравнения, описывающего систему. Следовательно, этот вход будет иметь ненулевые корреляции лишь для первых $(m - 1)$ запаздываний. Результат (5.2.49) получен в [8.]

Общие смешанные процессы авторегрессии — скользящего среднего. Более общим образом можно определить смешанный дискретный процесс авторегрессии — скользящего среднего в виде

$$(X_t - \mu) = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_m(X_{t-m} - \mu) + Z_t + \beta_1 Z_{t-1} + \dots + \beta_l Z_{t-l}, \quad (5.2.50)$$

где l не связано с m . Для стационарности требуется, чтобы корни характеристического уравнения авторегрессионной компоненты лежали внутри единичного круга.

Для непрерывного времени смешанный процесс принимает вид

$$\begin{aligned} \alpha_m \frac{d^m X}{dt^m} + \alpha_{m-1} \frac{d^{m-1} X}{dt^{m-1}} + \dots + \alpha_0 [X(t) - \mu] = \\ = b_l \frac{d^l Z}{dt^l} + b_{l-1} \frac{d^{l-1} Z}{dt^{l-1}} + \dots + b_0 Z(t). \end{aligned} \quad (5.2.51)$$

Из (2.3.19) следует, что условия стационарности, или устойчивости, непрерывного процесса (5.2.51) заключаются в том, что $l \leq m - 1$ и корни характеристического уравнения авторегрессионной компоненты имеют отрицательные действительные части.

Важность модели (5.2.50) состоит в следующем: в то время как модель, основанная на чисто авторегрессионном процессе или на чистом процессе скользящего среднего, может потребовать большого числа параметров, для смешанной модели (5.2.50) их может потребоваться относительно немного.

5.2.6. Интерпретация корреляционной функции

Случайный процесс называется гауссовским, или нормальным, если многомерное распределение, связанное с произвольным набором значений времени, является многомерным нормальным распределением. В этом случае процесс полностью определяется своим средним значением, дисперсией и корреляционной функцией. Однако существует обширный класс негауссовских процессов, имеющих ту же самую корреляционную функцию, что и заданный гауссовский процесс, но заметно отличающихся от него в других отношениях. Например, в разд. 5.2.4 было показано, что модель (5.2.24) приводит к показательной корреляционной функции $\rho_{XX}(u) = e^{-|u|/T}$. Если входной процесс системы первого порядка

(5.2.24) является нормальным, то можно показать, что выход также будет нормальным и, таким образом, полностью задается своей корреляционной функцией.

Сейчас мы построим другой процесс, имеющий показательную корреляционную функцию, но в других отношениях сильно отличающийся от нормального процесса. Этот процесс называется *случайным телеграфным сигналом* и описан в [9]. Альфа-частицы радиоактивного источника служат для запуска триггерного устройства, принимающего попеременно значения $+1$ и -1 . Моменты t_i ,

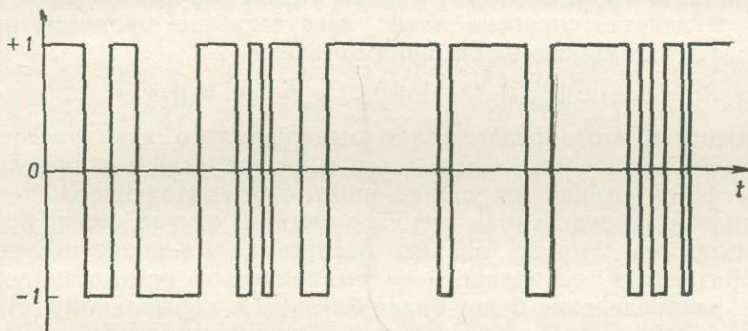


Рис. 5.10. Реализация случайного телеграфного сигнала.

в которые происходят изменения процесса, образуют пуассоновский процесс с параметром λ , так что типичная реализация процесса могла бы быть такой, как показано на рис. 5.10. Предполагая, что процесс начался при $t = -\infty$, мы получим

$$\Pr \{X(t) = 1\} = \Pr \{X(t) = -1\} = \frac{1}{2},$$

откуда $E[X(t)] = 0$. Следовательно, ковариационная функция равна

$$\gamma_{XX}(u) = E[X(t)X(t+u)] = \begin{cases} 1 \times \Pr \{ \text{четное число изменений} \\ \text{в } (t, t+u) \}, \\ -1 \times \Pr \{ \text{нечетное число изме-} \\ \text{нений в } (t, t+u) \}, \end{cases}$$

$$\gamma_{XX}(u) = \sum_{k=0}^{\infty} (p_{2k} - p_{2k+1}),$$

где

$$p_k = \frac{e^{-\lambda|u|} (\lambda|u|)^k}{k!}.$$

Просуммировав эти ряды, получим

$$\begin{aligned} \gamma_{XX}(u) &= e^{-\lambda|u|} \left[\frac{e^{\lambda|u|} + e^{-\lambda|u|}}{2} - \frac{e^{\lambda|u|} - e^{-\lambda|u|}}{2} \right] = \\ &= e^{-2\lambda|u|} = \rho_{XX}(u), \end{aligned}$$

так как $\gamma_{XX}(0) = 1$. Если $\lambda = T/2$, то эта функция совпадает с корреляционной функцией (5.2.25). Так как распределение $X(t)$ сосредоточено в двух точках ± 1 , поведение этого процесса заметно отличается от нормального с той же самой корреляционной функцией. В действительности такие негауссовские процессы нужно описывать с помощью их старших моментов

$$E [X(t)X(t+u_1) \dots X(t+u_k)].$$

Важность этого раздела для эмпирического анализа временных рядов заключается в том, что при интерпретации корреляционной функции (и, как мы увидим ниже, соответствующего спектра) необходима определенная осторожность в случае, если процесс негауссовский. Может, однако, оказаться, что после некоторого преобразования, основанного на эмпирической плотности вероятности, распределение будет более близким к нормальному. Например, неотрицательная величина, такая, как температура или давление, возможно, стала бы более близкой к нормальной, если бы был использован логарифмический масштаб. Заметим, однако, что если даже такое преобразование и приближает одномерную плотность к нормальной, оно не обязательно оказывает такое же действие и на многомерные распределения.

5.3. ОЦЕНИВАНИЕ КОВАРИАЦИОННЫХ ФУНКЦИЙ

В разд. 5.1.5 было показано, что обладающую наименьшей среднеквадратичной ошибкой оценку функции отклика некоторой системы на единичный импульс можно было бы выразить через ковариационные функции входа и выхода. На практике невозможно знать эти ковариационные функции точно, и, следовательно, необходимо оценивать их по записям конечной длины.

В разд. 5.3.1 будут выведены выборочные оценки наименьших квадратов для функции отклика на единичный импульс в случае, когда в распоряжении имеются конечные записи входа и выхода. Будет показано, что результаты получаются аналогичные тем, которые были выведены в разд. 5.1.5, с той разницей, что теоретические ковариационные функции заменяются их выборочными оценками. Кроме того, будет показано, что этот подход приводит к вычислению по данным таких функций, которые являются естественными выборочными оценками авто- и взаимных ковариационных функций. В разд. 5.3.2 определяются другие выборочные оценки

автоковариационной функции, а в разд. 5.3.3 выводятся их выборочные свойства. Разд. 5.3.4 и 5.3.5 состоят из обсуждения некоторых практических вопросов, возникающих при оценивании автоковариационной функции.

5.3.1. Анализ систем методом наименьших квадратов

Предположим, что вместо случайных процессов $X(t)$ и $Y(t)$, являющихся входом и выходом системы на рис. 5.7, имеются лишь реализации конечной длины T . Тогда модель (5.1.10) можно переписать в виде

$$y(t) - \bar{y} = \int_0^{\infty} h(u) [x(t-u) - \bar{x}] du + z(t), \quad 0 \leq t \leq T, \quad (5.3.1)$$

где \bar{x} , \bar{y} — выборочные средние, например

$$\bar{x} = \frac{1}{T} \int_0^T x(t) dt.$$

Если предположить, что $Z(t)$ — белый шум, то для непрерывного времени выборочная оценка наименьших квадратов для функции $h(u)$ получается с помощью минимизации интеграла от квадрата ошибки:

$$S = \int_0^T z^2(t) dt = \int_0^T \left\{ y(t) - \bar{y} - \int_0^{\infty} h(u) [x(t-u) - \bar{x}] du \right\}^2 dt. \quad (5.3.2)$$

Если быть более точным, следовало бы оценить и параметр μ_Y , входящий в (5.1.10). Однако с высокой степенью точности он будет равен величине \bar{y} , и поэтому для облегчения изложения мы заменим $\hat{\mu}_Y$ на \bar{y} до минимизации по функции $h(u)$.

Ясно, что $h(u)$ можно оценить лишь для $0 \leq u \leq T$, но на практике $h(u)$ затухает на довольно коротком участке записи по сравнению со всей длиной. Таким образом, обычно интересуются оценкой $h(u)$ в интервале $0 \leq u \leq T_0$, где T_0 значительно меньше, чем T . Заметим, что, хотя $x(t)$ является реализацией случайного процесса $X(t)$, принцип наименьших квадратов все же применим, если рассматривать $x(t)$ как фиксированную функцию. Как отмечалось в разд. 4.4.4, знание совместного распределения случайных величин $X(t)$ не дает ничего для оценивания $h(u)$ *).

*). См. примечание переводчика на стр. 154. — Прим. ред.

Поступая, как и в разд. 5.1.5, величину S можно разложить следующим образом:

$$S = \int_0^T [y(t) - \bar{y}]^2 dt - 2 \int_0^T \left\{ [y(t) - \bar{y}] \int_0^\infty [x(t-u) - \bar{x}] h(u) du \right\} dt + \\ + \int_0^T \left\{ \int_0^\infty \int_0^\infty [x(t-u) - \bar{x}] [x(t-v) - \bar{x}] h(u) h(v) du dv \right\} dt.$$

Меняя порядок интегрирования, получаем

$$S = \int_0^T [y(t) - \bar{y}]^2 dt - 2T \int_0^\infty h(u) \left\{ \frac{1}{T} \int_0^T [x(t-u) - \bar{x}] \times \right. \\ \times [y(t) - \bar{y}] dt \left. \right\} du + T \int_0^\infty \int_0^\infty h(u) h(v) \left\{ \frac{1}{T} \int_0^T [x(t-u) - \bar{x}] \times \right. \\ \times [x(t-v) - \bar{x}] dt \left. \right\} du dv. \quad (5.3.3)$$

Сравнивая (5.3.3) и (5.1.12), мы видим, что в (5.3.3) член

$$\frac{1}{T} \int_0^T [x(t-u) - \bar{x}] [y(t) - \bar{y}] dt$$

аналогичен в (5.1.12) члену

$$E [(X(t-u) - \mu_X)(Y(t) - \mu_Y)] = \gamma_{XY}(u).$$

Это наводит на мысль определить выборочную оценку взаимной ковариационной функции следующим образом:

$$c_{xy}(u) = \frac{1}{T} \int_0^T [x(t-u) - \bar{x}] [y(t) - \bar{y}] dt, \quad -T \leq u \leq T. \quad (5.3.4)$$

Аналогично выборочная оценка автоковариационной функции определяется равенством

$$c_{xx}(u) = \frac{1}{T} \int_0^T [x(t) - \bar{x}] [x(t+u) - \bar{x}] dt = \\ = \frac{1}{T} \int_0^{T-|u|} [x(t) - \bar{x}] [x(t+u) - \bar{x}] dt, \quad -T \leq u \leq T, \quad (5.3.5)$$

так как $x(t) = 0$ для $t < 0$, $t > T$.

Равенство (5.3.3) можно переписать в следующей форме:

$$S = T c_{yy}(0) - 2T \int_0^{\infty} h(u) c_{xy}(u) du + T \int_0^{\infty} \int_0^{\infty} c_{xx}(u-v) h(u) h(v) du dv, \quad (5.3.6)$$

которая соответствует (5.1.12). Таким образом, интеграл от квадрата ошибки полностью определяется, если даны выборочные оценки ковариационных функций и отклик на единичный импульс $h(u)$, точно так же как среднеквадратичная ошибка определялась полностью теоретическими ковариационными функциями и откликом на единичный импульс. Заметим, однако, что, в то время как в подходе со среднеквадратичной ошибкой требовалась стационарность процессов $X(t)$ и $Y(t)$, метод наименьших квадратов не зависит от этого предположения. Функции $x(t)$ и $y(t)$ могут быть реализациями нестационарных случайных процессов.

После того как S выражена через выборочные оценки ковариационных функций или выборочные ковариационные функции, выборочная оценка наименьших квадратов $\hat{h}(u)$ получается с помощью вариационного исчисления, как описано в приложении П.5.1. Там показано, что $\hat{h}(u)$ должна удовлетворять интегральному уравнению

$$c_{xy}(u) = \int_{-\infty}^{\infty} c_{xx}(u-v) \hat{h}(v) dv, \quad u \geq 0, \quad (5.3.7)$$

которое в точности совпадает с интегральным уравнением Винера—Хопфа (5.1.13), с тем лишь отличием, что функции γ_{xx} , γ_{xy} заменены на c_{xx} , c_{xy} . Как и прежде, для физической реализуемости $h(v)$ нужно, чтобы $h(v) = 0$ при $v < 0$.

5.3.2. Выборочные ковариационные функции

В предыдущем разделе мы видели, что выборочная ковариационная функция $c_{xx}(u)$ появилась совершенно естественно в качестве выборочной оценки теоретической ковариационной функции $\gamma_{xx}(u)$. Оценку, соответствующую (5.3.5), можно записать в виде

$$c_{xx}(u) = \begin{cases} \frac{1}{T} \int_0^{T-|u|} [X(t) - \bar{X}] [X(t+|u|) - \bar{X}] dt, & 0 \leq |u| \leq T, \\ 0, & |u| > T, \end{cases} \quad (5.3.8)$$

где явно подчеркнут тот факт, что $x(t) = 0$ вне $(0, T)$. Другой оценкой, которая также широко используется, является

$$c'_{XX}(u) = \begin{cases} \frac{1}{T-|u|} \int_0^{T-|u|} [X(t) - \bar{X}] [X(t+|u|) - \bar{X}] dt, & 0 \leq |u| \leq T, \\ 0, & |u| > T. \end{cases} \quad (5.3.9)$$

Оценки $c_{XX}(u)$ и $c'_{XX}(u)$, широко используемые главным образом в статистических работах, выбраны по интуиции, а не из-за того, что они являются наилучшими в каком-нибудь известном смысле. Конечно, в идеальном случае при выборе оценки ковариационной функции нужно было бы выписать функцию правдоподобия наблюдаемого временного ряда. Дифференцируя эту функцию правдоподобия, мы получили бы систему уравнений для выборочных оценок максимального правдоподобия этих ковариаций. Предполагая, что плотность вероятности нормальная, нетрудно выписать функцию правдоподобия, но, к сожалению, полученные в результате дифференцирования уравнения поддаются решению лишь с большим трудом. Таким образом, приходится пользоваться такими оценками, как $c_{XX}(u)$ и $c'_{XX}(u)$, которые, как допускается многими, основаны на интуиции. Однако эти оценки можно сравнить, пользуясь некоторым критерием, таким, как минимальная среднеквадратичная ошибка, и затем выбрать наилучшую из доступных оценок. Такой подход принят нами в следующем разделе.

5.3.3. Свойства оценок ковариационных функций

Сейчас мы выведем свойства оценок ковариационных функций $c_{XX}(u)$ и $c'_{XX}(u)$, связанные с первым и вторым моментами, предполагая, что сигнал $x(t)$ ($0 \leq t \leq T$) является реализацией стационарного случайного процесса $X(t)$, обладающего следующими свойствами:

$$E[X(t)] = 0, \quad (5.3.10)$$

$$\text{Cov}[X(t), X(t+u)] = \gamma_{XX}(u), \quad (5.3.11)$$

$$\begin{aligned} \text{Cov}[X(t)X(t+u_1), X(v)X(v+u_2)] = \\ = \gamma_{XX}(v-t)\gamma_{XX}(v-t+u_2-u_1) + \\ + \gamma_{XX}(v-t+u_2)\gamma_{XX}(v-t-u_1) + K_4(v-t, u_1, u_2). \end{aligned} \quad (5.3.12)$$

Функция $K_4(v-t, u_1, u_2)$ в (5.3.12) является четвертым совместным кумулянтom случайного процесса $X(t)$, так что для нормального процесса $K_4 = 0$. Для других процессов можно показать [8], что при выводе свойств оценок ковариаций вкладом этого члена

можно пренебречь. Поэтому далее мы отбросим этот член. Заметим также, что сейчас предполагается $E[X(t)] = 0$. Эффекты, возникающие, когда допускается ненулевое среднее значение, обсуждаются лишь вкратце.

Среднее значение оценок ковариаций. Используя (5.3.11), получаем среднее значение оценки ковариации (5.3.8)

$$E[c_{XX}(u)] = E\left[\frac{1}{T} \int_0^{T-|u|} X(t)X(t+u) dt\right] = \frac{1}{T} \int_0^{T-|u|} \gamma_{XX}(u) dt = \\ = \begin{cases} \gamma_{XX}(u) \left(1 - \frac{|u|}{T}\right), & 0 \leq |u| \leq T, \\ 0, & |u| > T. \end{cases}$$

Отсюда

$$E[c_{XX}(u)] = \begin{cases} \gamma_{XX}(u) \left(1 - \frac{|u|}{T}\right), & |u| \leq T, \\ 0, & |u| > T. \end{cases} \quad (5.3.13)$$

Аналогично

$$E[c'_{XX}(u)] = \begin{cases} \gamma_{XX}(u), & |u| \leq T, \\ 0, & |u| > T. \end{cases} \quad (5.3.14)$$

Таким образом, $c'_{XX}(u)$ является несмещенной оценкой $\gamma_{XX}(u)$, в то время как $c_{XX}(u)$ только асимптотически несмещенная, когда длина записи T стремится к бесконечности. Однако ниже будет показано, что смещенная оценка имеет меньшую среднеквадратичную ошибку.

Ковариация оценок ковариаций. Свойства оценок $c_{XX}(u)$ и $c'_{XX}(u)$, связанные со вторыми моментами, можно вывести, используя (5.3.12), где мы отбросим член K_4 . Подробный вывод этого результата с объяснением всех приближений дан в приложении П9.1*). Здесь дается лишь краткий набросок вывода и результаты иллюстрируются примерами.

Ковариация двух оценок**) $c_{XX}(u_1)$ и $c_{XX}(u_2)$, где аргументами взяты запаздывания u_1 и u_2 (причем предполагается $u_2 \geq u_1 \geq 0$), равна

$$\text{Cov}[c_{XX}(u_1), c_{XX}(u_2)] = \text{Cov}\left[\frac{1}{T} \int_0^{T-u_1} X(t)X(t+u_1) dt, \right.$$

*) Во второй части книги. — Прим. перев.

**) Результаты для оценок $c'_{XX}(u_1)$ и $c'_{XX}(u_2)$ получаются из результатов для $c_{XX}(u_1)$ и $c_{XX}(u_2)$ с помощью замены T на $T - |u_1|$ и $T - |u_2|$ в знаменателе за знаком интеграла.

$$\left. \frac{1}{T} \int_0^{T-u_1} X(v) X(v+u_2) dv \right] = \frac{1}{T^2} \int_0^{T-u_1} \int_0^{T-u_2} \text{Cov} [X(t) X(t+u_1), X(v) X(v+u_2)] dt dv. \quad (5.3.15)$$

(Условие $u_2 \geq u_1 \geq 0$ не является никоим образом ограничительным, как показано в приложении П9.1.) Подставляя (5.3.12) в интеграл (5.3.15), получим

$$\text{Cov} [c_{XX}(u_1), c_{XX}(u_2)] = \frac{1}{T^2} \int_0^{T-u_1} \int_0^{T-u_2} [\gamma_{XX}(v-t) \gamma_{XX}(v-t+u_2-u_1) + \gamma_{XX}(v-t+u_2) \gamma_{XX}(v-t-u_1)] dt dv. \quad (5.3.16)$$

Замена переменных $v-t=r$, $t=s$ преобразует область интегрирования из квадрата на плоскости (t, v) в параллелограмм на плоскости (r, s) , как показано на рис. 5.11. После этого интегрирование в (5.3.16) сводится к

$$\text{Cov} [c_{XX}(u_1), c_{XX}(u_2)] = \frac{1}{T^2} \int_{-(T-u_1)}^{T-u_2} [\gamma_{XX}(r) \gamma_{XX}(r+u_2-u_1) + \gamma_{XX}(r+u_2) \gamma_{XX}(r-u_1)] dr \int ds, \quad (5.3.17)$$

где пределы интегрирования определяются из параллелограмма на рис. 5.11. Так как подынтегральное выражение не зависит от s , интегрирование по s дает длину $\varphi(r)$ отрезка на высоте r , а именно

$$\varphi(r) = \begin{cases} T-u_2-r, & r \geq 0, \\ T-u_2, & -(u_2-u_1) \leq r \leq 0, \\ T-u_1+r, & -(T-u_1) \leq r \leq -(u_2-u_1). \end{cases} \quad (5.3.18)$$

Поэтому из (5.3.17) и (5.3.18) получаем

$$\begin{aligned} & \text{Cov} [c_{XX}(u_1), c_{XX}(u_2)] = \\ & = \frac{1}{T^2} \int_{-(T-u_1)}^{T-u_2} \varphi(r) [\gamma_{XX}(r) \gamma_{XX}(r+u_2-u_1) + \\ & + \gamma_{XX}(r+u_2) \gamma_{XX}(r-u_1)] dr. \end{aligned} \quad (5.3.19)$$

Результат (5.3.19) является точным. Первоначально он был получен в [8]. При $u_1 = u_2$ (5.3.19) приводится к симметричной форме

$$\begin{aligned} \text{Var} [c_{XX}(u)] & = \frac{1}{T^2} \int_{-(T-u)}^{T-u} (T-u-|r|) \times \\ & \times [\gamma_{XX}^2(r) + \gamma_{XX}(r+u) \gamma_{XX}(r-u)] dr. \end{aligned} \quad (5.3.20)$$

Для несмещенной оценки $c'_{XX}(u)$ результат, соответствующий (5.3.20), выглядит следующим образом:

$$\text{Var} [c'_{XX}(u)] = \frac{1}{(T-|u|)^2} \int_{-(T-u)}^{T-u} (T-u-|r|) \times \\ \times [\gamma_{XX}^2(r) + \gamma_{XX}(r+u)\gamma_{XX}(r-u)] dr. \quad (5.3.21)$$

Равенство (5.3.19) показывает, что в общем случае соседние значения оценок ковариационных функций будут сильно коррелиро-

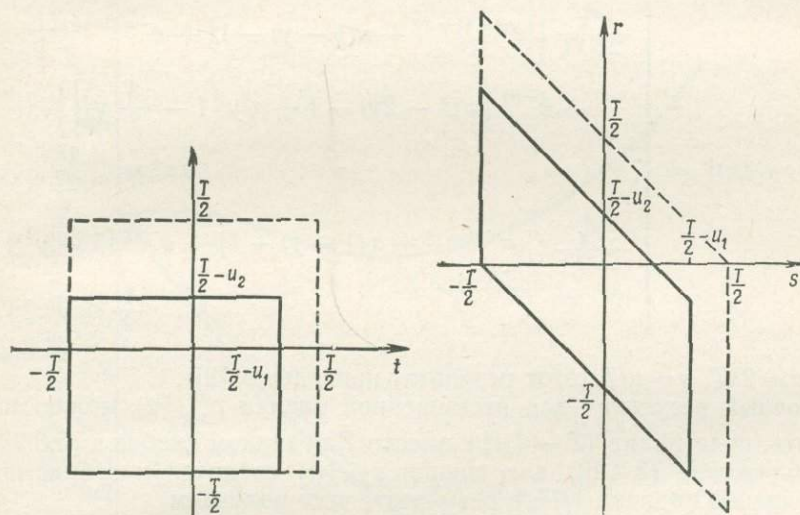


Рис. 5.11. Области интегрирования для вычисления ковариационной функции.

ваны, и, следовательно, выборочные ковариационные функции не всегда затухают с такой же быстротой, как их математические ожидания. Этот эффект проиллюстрирован в разд. 5.3.5.

Одно полезное приближение. Вычисление ковариации по формуле (5.3.19) обычно очень трудно проводить, если только не сделать простых предположений о форме ковариационных функций. Одно полезное приближение для больших T предложено в [8]. Оно связано с тем, что

$$\lim_{T \rightarrow \infty} \{T \text{Cov} [c_{XX}(u_1), c_{XX}(u_2)]\} = \\ = \int_{-\infty}^{\infty} [\gamma_{XX}(r)\gamma_{XX}(r+u_2-u_1) + \gamma_{XX}(r+u_2)\gamma_{XX}(r-u_1)] dr$$

и, следовательно, для больших T

$$\begin{aligned} & \text{Cov} [c_{XX}(u_1), c_{XX}(u_2)] \approx \\ & \approx \frac{1}{T} \int_{-\infty}^{\infty} [\gamma_{XX}(r) \gamma_{XX}(r+u_2-u_1) + \gamma_{XX}(r+u_2) \gamma_{XX}(r-u_1)] dr. \end{aligned} \quad (5.3.22)$$

Пример. Рассмотрим непрерывный процесс авторегрессии первого порядка, у которого $\gamma_{XX}(u) = \sigma_X^2 e^{-\lambda|u|}$ (эта функция обсуждалась в разд. 5.2.4). Подставляя эту $\gamma_{XX}(u)$ в (5.3.20), получаем

$$\text{Var} [c_{XX}(u)] = \begin{cases} \frac{\sigma_X^4}{2\lambda^2 T^2} \left\{ [e^{-\alpha(1-y)} + \alpha(1-y) - 1] + e^{-\alpha(1-y)} + \right. \\ \left. + e^{-\alpha y} \left[\alpha(1-2y) - 1 + \alpha^2 y \left(1 - \frac{3}{2} y \right) \right] \right\}, & 0 \leq y \leq \frac{1}{2}, \\ \frac{\sigma_X^4}{2\lambda^2 T^2} \left\{ [e^{-\alpha(1-y)} + \alpha(1-y) - 1] + e^{-\alpha y} (1-y^2) \right\}, & \frac{1}{2} \leq y \leq 1, \end{cases} \quad (5.3.23)$$

где $\alpha = 2\lambda T$, $y = u/T$ (этот результат приведен в [9]).

Точный результат для несмещенной оценки $c'_{XX}(u)$ можно получить, подставляя $(T - |u|)$ вместо T за знаком скобок в (5.3.23). Приближение (5.3.22) для оценки $c_{XX}(u)$ сводится к отбрасыванию членов порядка $1/T^2$, в результате чего получаем

$$\text{Var} [c_{XX}(u)] \approx \frac{\sigma_X^4}{\lambda T} (1 + e^{-\alpha y} + \alpha y e^{-\alpha y}). \quad (5.3.24)$$

Дисперсии двух оценок $c_{XX}(u)$ и $c'_{XX}(u)$ в зависимости от запаздывания u изображены на рис. 5.12 для случая $\lambda T = 2,5$. Видно, что эти дисперсии совпадают при $u=0$, но при $u \rightarrow T$ дисперсия смещенной оценки стремится к нулю, в то время как дисперсия несмещенной оценки стремится к бесконечности. Именно это свойство несмещенной оценки $c'_{XX}(u)$ и делает ее такой неудобной.

Среднеквадратичная ошибка оценок ковариаций. Для того, чтобы сравнение двух оценок было справедливым, нужно сравнивать их среднеквадратичные ошибки. Используя выражение (4.2.12) для среднеквадратичной ошибки, а именно

$$E [(c_{XX}(u) - \gamma_{XX}(u))^2] = \text{Var} [c_{XX}(u)] + B^2 [c_{XX}(u)],$$

и выражение (5.3.13), из которого можно получить смещение $B[c_{XX}(u)]$, находим среднеквадратичные ошибки смещенной и несмещенной оценок:

$$E[(c_{XX}(u) - \gamma_{XX}(u))^2] = \text{Var}[c_{XX}(u)] + \sigma_X^4 \left(\frac{u}{T}\right)^2 e^{-2\lambda u}$$

и

$$E[(c'_{XX}(u) - \gamma'_{XX}(u))^2] = \text{Var}[c'_{XX}(u)].$$

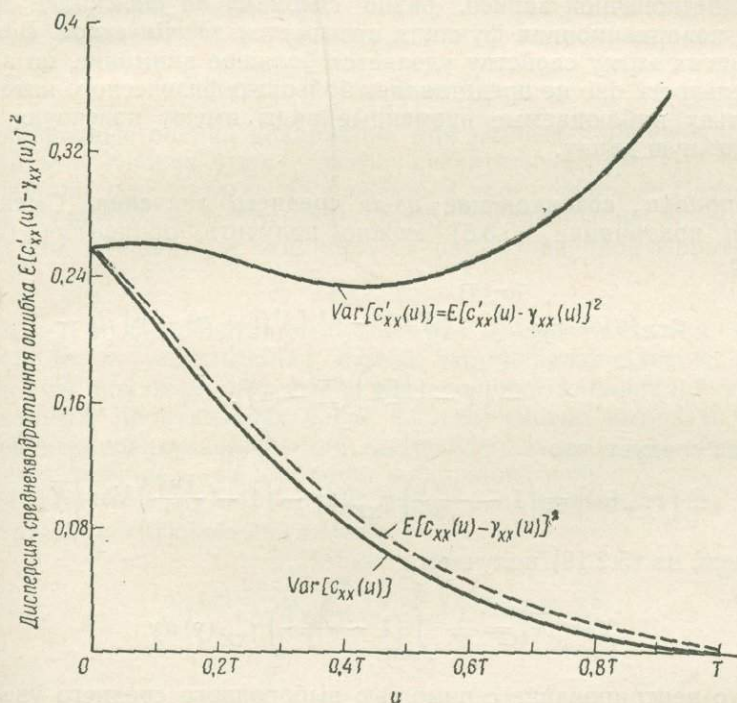


Рис. 5.12. Дисперсии и среднеквадратичные ошибки оценок ковариационной функции для непрерывного процесса первого порядка.

Эти среднеквадратичные ошибки показаны на рис. 5.12 вместе с дисперсиями для непрерывного процесса авторегрессии первого порядка с $\lambda T = 2,5$. Мы видим, что среднеквадратичная ошибка для $c'_{XX}(u)$ устойчиво держится выше, чем для $c_{XX}(u)$ (этот результат отмечен в [10]). Мы доказали здесь это утверждение для упомянутой выше ковариационной функции, однако есть предположение, что оно справедливо и для большинства других ковариационных функций [11].

Эргодичность. Из (5.3.13), (5.3.14) и (5.3.22) следует, что для больших T математические ожидания $c_{XX}(u)$ и $c'_{XX}(u)$ равны $\gamma_{XX}(u)$, а дисперсии пропорциональны $1/T$. Следовательно, эти две оценки ковариационных функций являются асимптотически состоятельными. Таким образом, ковариационную функцию $E[X(t)X(t+u)]$ процесса $X(t)$ можно оценить с произвольно малой ошибкой с помощью единственной достаточно длинной записи. В таком случае для ковариационной функции *среднее по времени*, взятое по одной бесконечной записи, равно *среднему по ансамблю*, и поэтому ковариационная функция называется *эргодической*. Во многих книгах этому свойству уделяется большое внимание, но в действительности оно не представляет большого физического интереса, поскольку наблюдаемые временные ряды имеют конечную, а не бесконечную длину.

Поправки, возникающие из-за среднего значения. Смещение оценки ковариации (5.3.8) можно получить, записывая (5.3.8) в виде

$$c_{XX}(u) = \frac{1}{T} \int_0^{T-|u|} [X(t) - \mu][X(t+u) - \mu] dt - \left(1 - \frac{|u|}{T}\right)(\bar{X} - \mu)^2.$$

Отсюда следует, что

$$E[c_{XX}(u)] = \left(1 - \frac{|u|}{T}\right)\gamma_{XX}(u) - \left(1 - \frac{|u|}{T}\right)\text{Var}[\bar{X}].$$

Наконец, из (5.2.19) получаем

$$\text{Var}[\bar{X}] = \frac{1}{T} \int_{-T}^T \left(1 - \frac{|y|}{T}\right)\gamma_{XX}(y) dy,$$

так что центрирование с помощью выборочного среднего увеличивает смещение еще больше на члены порядка $1/T$ и более высокого.

5.3.4. Выборочные оценки ковариаций для случая дискретного времени

Если наблюдения x_1, x_2, \dots, x_N получены из дискретного временного ряда, то дискретная выборочная оценка, соответствующая непрерывной оценке (5.3.5), равна

$$c_{xx}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}), \quad k=0, 1, \dots, N-1, \quad (5.3.25)$$

где

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$$

является выборочным средним значением всего ряда.

Приближенное значение ковариации оценок, соответствующих выборочным оценкам (5.3.25), можно получить, заменяя интегралы в (5.3.22) суммами, а именно

$$\begin{aligned} \text{Cov} [c_{xx}(k), c_{xx}(l)] \approx \frac{1}{N} \sum_{r=-\infty}^{\infty} [\gamma_{xx}(r) \gamma_{xx}(r+l-k) + \\ + \gamma_{xx}(r+l) \gamma_{xx}(r-k)]. \end{aligned} \quad (5.3.26)$$

Выборочные оценки ковариаций для данных, пропущенных через фильтр. Иногда бывает нужно сосчитать выборочные оценки ковариаций для данных, *пропущенных через фильтр*. Например, может оказаться желательным устранить тренды из x_t , образуя новый ряд данных y_t с помощью операции линейной фильтрации

$$y_t = x_t - ax_{t-1}. \quad (5.3.27)$$

Если $a = 1$, то y_t представляет собой ряд первых разностей, а если $a = -1$, то y_t — скользящая сумма пар первоначального ряда. Сейчас мы покажем, что выборочные оценки ковариаций $c_{yy}(k)$ для данных, пропущенных через фильтр, можно выразить через выборочные оценки ковариаций $c_{xx}(k)$ исходных данных. Этим достигается значительная экономия машинного времени.

Согласно (5.3.25), имеем для выборочной оценки $c_{yy}(k)$ после фильтрации следующее выражение:

$$c_{yy}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y}), \quad (5.3.28)$$

где

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N (x_t - ax_{t-1}) \approx (1-a)\bar{x}.$$

Подставляя (5.3.27) в (5.3.28), получаем

$$\begin{aligned} c_{yy}(k) &\approx \frac{1}{N} \sum_{t=1}^{N-k} \{ [x_t - ax_{t-1} - (1-a)\bar{x}] \times \\ &\times [x_{t+k} - ax_{t+k-1} - (1-a)\bar{x}] \} = \\ &= \frac{1}{N} \sum_{t=1}^{N-k} [(x_t - \bar{x}) - a(x_{t-1} - \bar{x})] \times \\ &\times [(x_{t+k} - \bar{x}) - a(x_{t+k-1} - \bar{x})]. \end{aligned}$$

Раскрывая это выражение, находим

$$\begin{aligned} c_{yy}(k) &\approx c_{xx}(k) - ac_{xx}(k-1) - ac_{xx}(k+1) + a^2c_{xx}(k) = \\ &= -ac_{xx}(k-1) + (1+a^2)c_{xx}(k) - ac_{xx}(k+1). \end{aligned} \quad (5.3.29)$$

Таким образом, $c_{yy}(k)$ можно получить с помощью простой операции линейной фильтрации, примененной к $c_{xx}(k)$. В качестве примера для фильтра первых разностей $y_t = x_t - x_{t-1}$ имеем

$$c_{yy}(k) \approx -c_{xx}(k-1) + 2c_{xx}(k) - c_{xx}(k+1). \quad (5.3.30)$$

Так что выборочные ковариации $c_{yy}(k)$ первых разностей, взятых от исходных данных, равны центральным вторым разностям от выборочных ковариаций $c_{xx}(k)$.

Для операции фильтрации

$$y_t = x_t - ax_{t-1} - bx_{t-2}, \quad (5.3.31)$$

которая может быть применена для устранения осциллирующих компонент временного ряда, выборочные ковариации $c_{yy}(k)$ равны

$$\begin{aligned} c_{yy}(k) &\approx (1+a^2+b^2)c_{xx}(k) - a(1-b)[c_{xx}(k-1) + c_{xx}(k+1)] - \\ &\quad - b[c_{xx}(k-2) + c_{xx}(k+2)]. \end{aligned} \quad (5.3.32)$$

Оценивание корреляционной функции. Иногда требуется сравнить два временных ряда, масштабы измерения которых могут быть различными, так что больше подходят выборочные оценки корреляционных, а не ковариационных функций. Выборочные оценки корреляционных функций можно получить, разделив рассмотренные выше выборочные оценки ковариаций на выборочную оценку дисперсии. Таким образом, получаем

$$r_{xx}(k) = \frac{c_{xx}(k)}{c_{xx}(0)}, \quad (5.3.33)$$

где $c_{xx}(k)$ определяются равенствами (5.3.25).

Еще одна выборочная оценка корреляций. Другой выборочной оценкой корреляционной функции, часто используемой статистиками, является

$$r_{xx}(k) = \frac{\sum (x_t - \bar{x}_1)(x_{t+k} - \bar{x}_2)}{[\sum (x_t - \bar{x}_1)^2 \sum (x_{t+k} - \bar{x}_2)^2]^{1/2}}, \quad (5.3.34)$$

где \bar{x}_1 и \bar{x}_2 — средние значения первых $N-k$ и последних $N-k$ наблюдений соответственно, а суммирование происходит от $t=1$ до $t=N-k$. Равенство (5.3.34) основано на диаграмме разброса пар (x_t, x_{t+k}) для $t=1, 2, \dots, N-k$. Рис. 5.5, например, соответствует случаю $k=1$. Если предположить, что совместная плотность вероятности случайных величин X_t и X_{t+k} есть двумерная нормальная плотность с коэффициентом корреляции $\rho_{xx}(k)$, то (5.3.34)

является выборочной оценкой максимального правдоподобия для $\rho_{xx}(k)$ *).

Пользоваться оценкой (5.3.4) не рекомендуется на том основании, что хотя для отдельного значения корреляционной функции $\rho_{xx}(k)$, рассматриваемого изолированно от других значений, она и является разумной выборочной оценкой, но ее нежелательно применять в случае, когда нужна *совокупность выборочных оценок* $r_{xx}(1)$, $r_{xx}(2)$, ..., $r_{xx}(m)$ для первых m корреляций $\rho_{xx}(1)$, $\rho_{xx}(2)$, ..., $\rho_{xx}(m)$.

Основной недостаток (5.3.34) состоит в том, что для коррекции среднего значения она использует две величины, которые зависят от запаздывания: в результате с изменением запаздывания k меняется нормирующий множитель. Эти изменения приводят к тому, что выборочные оценки не являются положительно определенными в смысле свойства 4 из разд. 5.2.1. Это может привести к нежелательному поведению выборочных оценок спектра, которые будут рассмотрены ниже.

5.3.5. Практические аспекты оценивания ковариационных функций

В разд. 5.1.5 было указано, почему нужно изучать ковариационные функции: во-первых, они входят в уравнения для синтеза линейных систем и, во-вторых, их можно использовать при оценивании функций отклика на единичный импульс. С более общей статистической точки зрения одна из важных причин изучения временных рядов заключается в том, чтобы дать возможность построить модель для лежащего в основе явления случайного процесса. Эту модель можно затем использовать для прогноза, синтеза систем или для других целей, таких, как имитация систем. В таких случаях эмпирический анализ ковариационной функции или спектра может дать полезные наводящие идеи относительно того, какие модели должны были бы соответствовать временному ряду.

Пример. Чтобы проиллюстрировать, как можно использовать корреляционную функцию для того, чтобы в сжатом виде выразить информацию, содержащуюся в исходном ряде, рассмотрим выборочную оценку корреляционной функции для данных о партиях продукта, приведенных на рис. 5.2. Первые пятнадцать значений этой оценки, полученные по формулам (5.3.33) и (5.3.25), даны в табл. 5.2; их график построен на рис. 5.6. Из табл. 5.2 видно, что корреляции меняют знак. Это является следствием того, что за высоким выходом продукта в одной партии следует, как правило,

*) Строго говоря, для этого нужно было бы еще потребовать, чтобы пары (X_t, X_{t+k}) были независимы при разных t . — *Прим. перев.*

пониженный выход продукта в следующей партии, и наоборот. Видно также, что корреляции почти полностью затухают, начиная с запаздывания, равного 6, и показывают последовательное ослабление зависимости между наблюдениями при увеличении запаздывания между ними.

Таблица 5.2

Первые 15 значений выборочной корреляционной функции для данных из табл. 5.1

k	$r_{xx}(k)$	k	$r_{xx}(k)$	k	$r_{xx}(k)$
1	-0,39	6	-0,05	11	0,11
2	0,30	7	0,04	12	-0,07
3	-0,17	8	-0,04	13	0,15
4	0,07	9	-0,01	14	0,04
5	-0,10	10	0,01	15	-0,01

Основной механизм, порождающий этот вид корреляционной функции, хорошо известен для приведенных данных. Промышленная установка, на которой были получены наблюдения, представляла собой дистилляционную колонку, где содержимое перегонного куба подогревалось в течение некоторого времени, и продукт перегонки накапливался и сцеживался. Во время перегонки дегтеобразные остатки скапливаются в перегонном кубе и сцеживаются в конце каждой партии. Однако это сцеживание не является полным, так что некоторое количество дегтеобразного вещества остается в перегонном кубе. Это оказывает неблагоприятное действие на выход продукта в следующей партии, так что производится меньше продукта перегонки и, следовательно, меньше остается дегтеобразных веществ. Этим объясняется отрицательная корреляция между партиями.

Вычисление выборочной оценки корреляционной функции. Для вычисления выборочной оценки корреляционной функции необходима вычислительная машина. Программа вычислений для этой цели описана в приложении П5.3. Однако для того, чтобы лучше понять последовательные стадии вычислений, желательно, чтобы читатель просчитал один пример на настольной счетной машине. Для иллюстрации рассмотрим вычисление $r_{xx}(2)$ для данных о партиях продукта из табл. 5.1. Выборочную оценку (5.3.25) можно записать в виде

$$c_{xx}(k) = \frac{1}{N} \left\{ \sum_{t=1}^{N-k} x_t x_{t+k} - \bar{x} \left[\sum_{t=1}^{N-k} (x_t + x_{t+k}) \right] + (N-k) \bar{x}^2 \right\}, \quad (5.3.35)$$

и, следовательно, большая часть времени счета идет на вычисление сумм сдвинутых произведений $\sum_{i=1}^{N-k} x_i x_{i+h}$. При работе с настольной вычислительной машиной следует иметь в виду, что если из каждого наблюдения вычесть произвольную константу, то отклонения $x_i - \bar{x}$ не изменятся, а следовательно, и $c_{xx}(k)$ не изменится. Поэтому для снижения порядка чисел, которые требуется перемножить, удобно вычесть из каждого наблюдения константу, по возможности близкую к среднему значению. Данные в табл. 5.1 изменяются от 20 до 80. Поэтому подходящей константой для вычитания является 50. В таком случае получим

$$\sum_{i=1}^{68} (x_i - 50)(x_{i+2} - 50) = (-3)(-27) + (14)(21) + \dots \\ \dots + (7)(-27) = 3084.$$

Аналогично имеем

$$\sum_{i=1}^{70} (x_i - 50) = 79, \quad \sum_{i=1}^{68} (x_i - 50) = 102, \quad \sum_{i=1}^{68} (x_{i+2} - 50) = 68.$$

Следовательно,

$$c_{xx}(2) = \frac{1}{70} \left[3084 - \frac{79}{70} (102 + 68) + 68 \left(\frac{79}{70} \right)^2 \right] = 42,55.$$

Дисперсия ряда $c_{xx}(0)$ равна 139,8. Следовательно,

$$r_{xx}(2) = \frac{c_{xx}(2)}{c_{xx}(0)} = \frac{42,55}{139,8} = 0,30,$$

что совпадает со вторым значением в табл. 5.2.

Резюме. Ниже мы резюмируем важные моменты, на которые следует обратить внимание при оценивании корреляционной функции.

а) Среднеквадратичная ошибка оценок, имеющих нормирующий множитель $1/T$, обычно меньше, чем у оценок с множителем $1/(T-u)$. Кроме того, первые являются положительно определенными, а вторые нет.

б) Необходимо провести в той или иной форме коррекцию данных для исключения влияния низкочастотных трендов. В простых случаях, как, например, в (5.3.25), этого можно добиться с помощью устранения постоянной составляющей. Эта коррекция среднего значения сохраняет положительную определенность выборочной оценки. В других случаях, таких, как (5.3.27), тренды должны устраняться с помощью операции фильтрации, а автоковариации надо считать по формуле (5.3.29).

в) Равенство (5.3.19) показывает, что если корреляции в исходном ряде достаточно сильны, то будут и сильные корреляции оценок автоковариаций. Выборочную ковариационную функцию, аргументом которой является запаздывание, можно рассматривать как новый временной ряд, полученный из первоначального временного ряда $x(t)$; в таком случае (5.3.19) показывает, что, вообще говоря, этот новый временной ряд будет сильнее коррелирован, чем исходный.

г) Одно из следствий корреляции соседних ординат оценки ковариационной функции заключается в том, что ее выборочная оценка не всегда затухает так же быстро, как математическое ожидание оценки. Чтобы проиллюстрировать этот эффект, на рис. 5.13 приведена теоретическая корреляционная функция дискретного процесса авторегрессии второго порядка:

$$X_t = X_{t-1} - 0,5X_{t-2} + Z_t. \quad (5.3.36)$$

Значения этой корреляционной функции можно получать из рекуррентного соотношения (5.2.43) при $m=2$, а именно

$$\rho_{XX}(k) = \rho_{XX}(k-1) - 0,5\rho_{XX}(k-2).$$

Корреляционная функция представляет собой затухающую периодическую функцию вида (5.2.38) и имеет период, равный 8. На рис. 5.13 приведены две выборочные корреляционные функции искусственного ряда, полученного по формуле (5.3.36), причем в качестве Z_t брались случайные нормальные числа из таблицы [7]. Верхняя функция сосчитана по 100 наблюдениям, а нижняя по 400. Характерной особенностью выборочной корреляционной функции, сосчитанной по 100 наблюдениям, являются большие осцилляции, которые сохраняются даже там, где теоретическая функция уже близка к нулю. Дело в том, что из-за большой положительной корреляции соседних значений выборочных ковариаций за большим положительным значением корреляции следует, как правило, другое большое положительное значение. В результате этого искажается вид корреляционной функции. Выборочная корреляционная функция, сосчитанная по 400 наблюдениям, затухает быстрее, но все еще значительно отличается от теоретической корреляционной функции.

Главный вывод, который следует из проведенного обсуждения, состоит в том, что иногда опасно придавать слишком большое значение видимым особенностям выборочной корреляционной функции, особенно сосчитанной по коротким рядам. В настоящей книге мы будем использовать корреляционную функцию главным образом как промежуточную ступень при оценивании спектральной плотности, а также для получения рекомендаций при спектральном анализе.

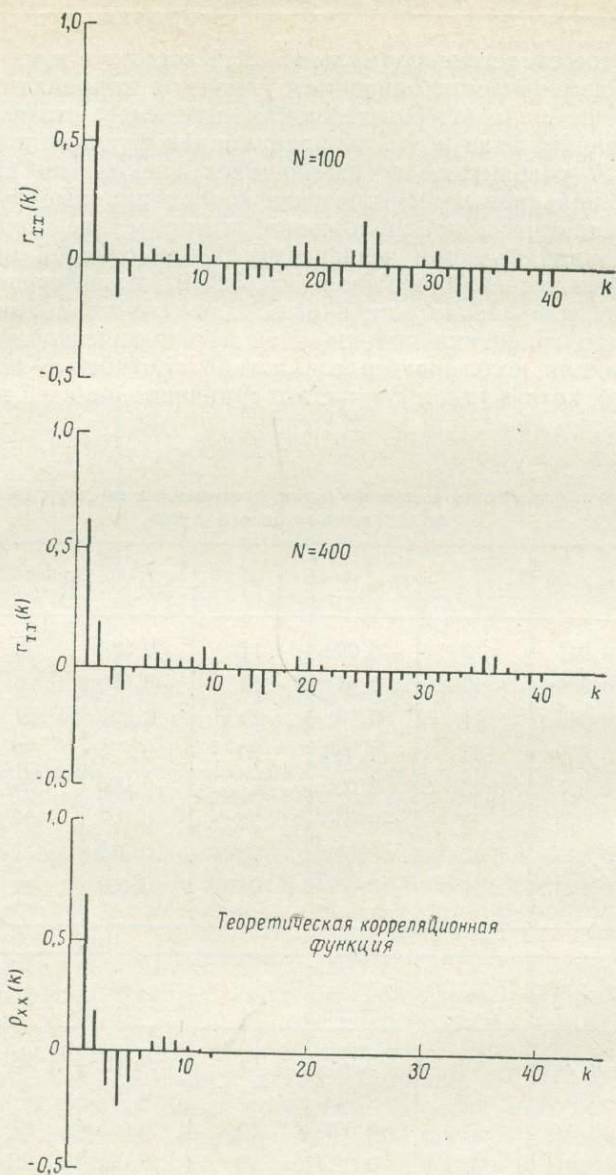


Рис. 5.13. Теоретическая и выборочные корреляционные функции для процесса авторегрессии второго порядка.

д) Другое следствие формулы (5.3.19) состоит в том, что нельзя судить об изменчивости одиночного значения корреляции, не учитывая других значений. Предположим, например, что имеется модель временного ряда и что корреляционная функция этой модели известна. В учебниках, не являющихся специально статистическими, наблюдаемая и теоретическая корреляционные функции часто сравниваются в предположении, что соседние точки оценки корреляционной функции независимы. Из-за сильной корреляции этих соседних значений, что видно из (5.3.19), такое предположение может быть совершенно ошибочным. Для точного анализа нужно было бы при сравнении наблюдаемой и теоретической корреляционных функций пользоваться совместной плотностью вероятности корреляций, хотя в таком случае это сравнение, по всей видимости,

Таблица 5.3

Выборочные корреляционные функции, построенные по двум выборкам искусственного белого шума

Ряд 1 k	$r_{xx}(k)$	k	$r_{xx}(k)$	k	$r_{xx}(k)$	k	$r_{xx}(k)$
1	0,041	9	-0,009	17	0,025	25	0,010
2	0,024	10	0,047	18	-0,020	26	0,029
3	0,045	11	0,061	19	0,032	27	0,011
4	0,330	12	0,083	20	0,075	28	0,068
5	0,007	13	0,026	21	-0,000	29	-0,004
6	0,012	14	-0,030	22	0,027	30	0,016
7	0,025	15	0,019	23	0,012	31	0,025
8	0,102	16	0,099	24	0,033	32	0,035

Ряд 2 k	$r_{xx}(k)$	k	$r_{xx}(k)$	k	$r_{xx}(k)$	k	$r_{xx}(k)$
1	-0,014	9	0,020	17	-0,047	25	0,039
2	-0,008	10	0,013	18	-0,012	26	-0,016
3	-0,038	11	0,007	19	0,025	27	0,025
4	0,011	12	-0,022	20	0,001	28	0,031
5	-0,047	13	0,017	21	0,009	29	-0,071
6	-0,051	14	-0,020	22	0,059	30	0,040
7	0,000	15	0,017	23	0,018	31	0,012
8	-0,041	16	-0,047	24	0,031	32	-0,025

было бы очень сложным. Когда задана параметрическая модель, гораздо лучше использовать методы правдоподобия или наименьших квадратов, которые описаны в гл. 4.

Критерий для проверки гипотезы о том, что шум белый. Есть один случай, когда соседние точки выборочной корреляционной функции действительно являются некоррелированными. Это имеет место для чисто случайного временного ряда, или белого шума. В этом случае из (5.3.19) следует, что при отсутствии коррекции среднего значения ковариация корреляционных оценок равна нулю.



Рис. 5.14. Выборочная корреляционная функция для выборки, образованной случайными нормальными числами, $r_{xx}(0) = 1$.

Коррекция среднего значения вносит в ковариацию члены порядка $1/T^2$, поэтому этими членами можно пренебречь. Можно показать, [12], что, когда число членов ряда достаточно велико, допустимо считать, что $r_{xx}(k)$ распределено по нормальному закону с нулевым средним значением и дисперсией $1/N$.

В качестве примера в табл. 5.3 приведены выборочные корреляционные функции, сосчитанные по случайным нормальным числам, выданным вычислительной машиной. Результаты некоторого эксперимента по имитации заставили предположить, что эти числа на самом деле были очень непохожи на случайные. Поэтому были взяты массивы чисел, примерно по 1000 штук в массиве, и по ним сосчитаны выборочные корреляционные функции. Типичная такая функция, сосчитанная по 900 числам, частично приведена в табл. 5.3 под заголовком «Ряд 1». Поскольку стандартное отклонение выборочной оценки одиночного значения корреляционной функции равно $1/\sqrt{900} = 0,033$, то 95%-ные доверительные границы для одиночной корреляции $r_{xx}(k)$ приблизительно равны $r_{xx}(k) \pm 0,033 \cdot 1,96 =$

$= r_{xx}(k) \pm 0,065$. 95%-ный доверительный интервал изображен рядом с выборочной корреляционной функцией на рис. 5.14. Видно, что 7 из 32 доверительных интервалов не покрывают нуль. Исходя из доверительного уровня, следовало бы ожидать, что примерно 5% от общего числа доверительных интервалов, т. е. 1 или 2, не накроют нуль. На самом деле, функция на рис. 5.14 обнаруживает систематическую компоненту с периодом, равным 4, из-за несовершенства метода получения случайных нормальных чисел.

Под заголовком «Ряд 2» в табл. 5.3 приведена типичная выборочная корреляционная функция, сосчитанная после того, как метод получения случайных чисел был улучшен. Заметим, что лишь для $\rho_{xx}(29)$ доверительный интервал не покрывает нуль. Это находится в согласии с гипотезой о том, что временной ряд является чисто случайным.

5.4. ОЦЕНИВАНИЕ ПАРАМЕТРОВ ЛИНЕЙНОГО ПРОЦЕССА

В этом разделе мы применим методы гл. 4 к оцениванию параметров процессов авторегрессии и скользящего среднего, введенных в разд. 5.2. Предположим, например, что требуется подобрать авторегрессионную модель

$$(X_t - \mu) = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_m(X_{t-m} - \mu) + Z_t \quad (5.4.1)$$

к наблюдаемому временному ряду x_1, x_2, \dots, x_N . Процедура подгонки состоит в следующем:

- 1) вынесение решения о порядке m процесса;
- 2) для заданного m оценивание параметров $\mu, \alpha_1, \dots, \alpha_m$.

Поскольку решение о том, каков порядок m процесса, можно вынести, лишь подгоняя процессы различных порядков, сначала необходимо рассмотреть оценивание параметров.

5.4.1. Оценивание параметров авторегрессии методом максимального правдоподобия

Приближенная функция правдоподобия. Предполагая, что процесс Z_t является нормальным, можно получить логарифмическую функцию правдоподобия для фиксированного m следующим образом. Во-первых, заметим, что совместную плотность вероятности случайных величин $Z_{m+1}, Z_{m+2}, \dots, Z_N$ можно записать в виде

$$f_{m+1, \dots, N}(z_{m+1}, z_{m+2}, \dots, z_N) = \frac{1}{(\sqrt{2\pi}\sigma_Z)^{N-m}} \times \\ \times \exp\left(-\frac{1}{2\sigma_Z^2} \sum_{t=m+1}^N z_t^2\right),$$

где $E[Z_t] = 0$, $E[Z_t^2] = \sigma_z^2$. Если перейти от переменных z к x_i согласно формуле (5.4.1), то, учитывая, что якобиан преобразования равен единице, получим

$$f_{m+1, \dots, N}(x_{m+1}, x_{m+2}, \dots, x_N | x_1, \dots, x_m) = \\ = \frac{1}{(\sqrt{2\pi} \sigma_z)^{N-m}} \exp \left\{ -\frac{1}{2\sigma_z^2} \sum_{t=m+1}^N [(x_t - \mu) - \alpha_1(x_{t-1} - \mu) - \dots \right. \\ \left. \dots - \alpha_m(x_{t-m} - \mu)]^2 \right\}. \quad (5.4.2)$$

Обозначения в левой части равенства (5.4.2) подчеркивают, что оно изображает условную совместную плотность случайных величин X_{m+1}, \dots, X_N при условии, что величины X_1, \dots, X_m фиксированы и равны своим выборочным значениям. Чтобы получить полную плотность вероятности, нужно было бы умножить (5.4.2) на совместную плотность величин X_1, \dots, X_m . Так как обычно m мало, результат такой «концевой поправки» будет несущественным, и, поскольку она значительно усложняет функцию правдоподобия, мы ее опустим. Если x_i известны, то (5.4.2) рассматривается как функция $\mu, \alpha_1, \dots, \alpha_m$ и дает условную функцию правдоподобия этих параметров при фиксированных x_1, \dots, x_m . Логарифмическая функция правдоподобия, таким образом, равна

$$l(\mu, \alpha_1, \dots, \alpha_m | x_1, \dots, x_m) = -(N-m) \ln \sqrt{2\pi} - (N-m) \ln \sigma_z - \\ - \frac{1}{2\sigma_z^2} \sum_{t=m+1}^N [(x_t - \mu) - \alpha_1(x_{t-1} - \mu) - \dots - \alpha_m(x_{t-m} - \mu)]^2. \quad (5.4.3)$$

При оценивании параметров $\mu, \alpha_1, \dots, \alpha_m$ важной величиной является сумма квадратов

$$S(\mu, \alpha_1, \dots, \alpha_m | x_1, \dots, x_m) = \\ = \sum_{t=m+1}^N [(x_t - \mu) - \alpha_1(x_{t-1} - \mu) - \dots - \alpha_m(x_{t-m} - \mu)]^2. \quad (5.4.4)$$

Теперь выборочные оценки максимального правдоподобия, или наименьших квадратов, можно получить, дифференцируя (5.4.4). Рассмотрим некоторые частные случаи.

Процесс авторегрессии первого порядка. Дифференцирование суммы квадратов

$$S(\mu, \alpha_1) = \sum_{t=2}^N [(x_t - \mu) - \alpha_1(x_{t-1} - \mu)]^2$$

приводит к нормальным уравнениям, аналогичным тем, которые получались в разд. 4.3.3. Таким образом, имеем

$$\begin{aligned} \bar{x}_2 - \hat{\mu} - \hat{\alpha}_1 (\bar{x}_1 - \hat{\mu}) &= 0, \\ \sum_{t=2}^N \{ (x_{t-1} - \hat{\mu}) [(x_t - \hat{\mu}) - \hat{\alpha}_1 (x_{t-1} - \hat{\mu})] \} &= 0, \end{aligned}$$

где \bar{x}_1 , \bar{x}_2 — средние арифметические первых и последних $(N-1)$ наблюдений соответственно. Отсюда

$$\begin{aligned} \hat{\mu} &= \frac{\bar{x}_2 - \hat{\alpha}_1 \bar{x}_1}{1 - \hat{\alpha}_1}, \\ \alpha_1 &= \frac{\sum_{t=2}^N (x_{t-1} - \hat{\mu})(x_t - \hat{\mu})}{\sum_{t=2}^N (x_{t-1} - \hat{\mu})^2}. \end{aligned} \quad (5.4.5)$$

Поскольку \bar{x}_1 и \bar{x}_2 очень близки к полному среднему \bar{x} , выборочную оценку $\hat{\mu}$ можно считать приближенно равной \bar{x} и, следовательно, выборочную оценку $\hat{\alpha}_1$ — равной $r_{xx}(1)$. Остаточную сумму квадра-

$$S(\hat{\mu}, \hat{\alpha}_1) = \sum_{t=2}^N [(x_t - \hat{\mu}) - \hat{\alpha}_1 (x_{t-1} - \hat{\mu})]^2$$

тов можно упростить, используя (П4.1.11):

$$S(\hat{\mu}, \hat{\alpha}_1) = \sum_{t=2}^N (x_t - \hat{\mu})^2 - \hat{\alpha}_1 \sum_{t=2}^N (x_t - \hat{\mu})(x_{t-1} - \hat{\mu}). \quad (5.4.6)$$

Аппроксимируя (5.4.6), точно так же как это делалось выше для $\hat{\alpha}_1$, получаем простое выражение

$$S(\hat{\mu}, \hat{\alpha}_1) \approx (N-1) [c_{xx}(0) - \hat{\alpha}_1 c_{xx}(1)] = (N-1) c_{xx}(0) [1 - r_{xx}^2(1)]. \quad (5.4.7)$$

Поскольку в $S(\mu, \alpha_1)$ фактически входят $(N-1)$ наблюдений и две степени свободы потеряны при подгонке констант μ , α_1 , дисперсию процесса Z_t можно оценить с помощью

$$s_z^2 = \frac{1}{N-3} S(\hat{\mu}, \hat{\alpha}_1).$$

Используя (П4.1.15) и те же самые приближения, что и выше, получаем $100(1 - \alpha)\%$ -ный доверительный интервал для α_1 :

$$(\alpha_1 - \hat{\alpha}_1)^2 \leq \frac{s_z^2 f_{1, N-3} (1 - \alpha)}{N \sum_{t=1}^N (x_t - \bar{x})^2}. \quad (5.4.8)$$

Процесс авторегрессии второго порядка. Выборочные оценки максимального правдоподобия можно получить, дифференцируя (5.4.3) по μ , α_1 и α_2 и приравнявая эти производные нулю. Это приводит к уравнениям

$$\begin{aligned} (\bar{x}_3 - \hat{\mu}) &= \hat{\alpha}_1 (\hat{x}_2 - \hat{\mu}) + \hat{\alpha}_2 (\bar{x}_1 - \hat{\mu}), \\ \sum (x_t - \hat{\mu})(x_{t-1} - \hat{\mu}) &= \hat{\alpha}_1 \sum (x_{t-1} - \hat{\mu})^2 + \\ &\quad + \hat{\alpha}_2 \sum (x_{t-1} - \hat{\mu})(x_{t-2} - \hat{\mu}), \\ \sum (x_t - \hat{\mu})(x_{t-2} - \hat{\mu}) &= \hat{\alpha}_1 \sum (x_{t-1} - \hat{\mu})(x_{t-2} - \hat{\mu}) + \\ &\quad + \hat{\alpha}_2 \sum (x_{t-2} - \hat{\mu})^2, \end{aligned} \quad (5.4.9)$$

где

$$\bar{x}_j = \frac{1}{N-3} \sum x_{t-3+j}$$

и суммирование распространяется от $t=3$ до $t=N$. Если заменить \bar{x}_j полным средним значением \bar{x} , то шесть функций от наблюдений, входящих в эти уравнения, можно объединить в пары и положить $\hat{\mu} = \bar{x}$. Например, две функции

$$\sum (x_t - \hat{\mu})(x_{t-1} - \hat{\mu})$$

и

$$\sum (x_{t-1} - \hat{\mu})(x_{t-2} - \hat{\mu})$$

имеют $N-3$ общих члена и отличаются только на один член в начале и в конце. С достаточной степенью точности их можно заменить на $Nc_{xx}(1)$, где $c_{xx}(1)$ является выборочной оценкой ковариации (5.3.25). Тогда уравнения (5.4.9) можно приближенно переписать в виде

$$c_{xx}(1) = \hat{\alpha}_1 c_{xx}(0) + \hat{\alpha}_2 c_{xx}(1),$$

$$c_{xx}(2) = \hat{\alpha}_1 c_{xx}(1) + \hat{\alpha}_2 c_{xx}(0).$$

Отсюда, вводя выборочные оценки корреляции $r_{xx}(k) = c_{xx}(k)/c_{xx}(0)$, получаем

$$\begin{aligned}\hat{a}_1 &\approx \frac{r_{xx}(1)[1 - r_{xx}(2)]}{1 - r_{xx}^2(1)}, \\ \hat{a}_2 &\approx \frac{r_{xx}(2) - r_{xx}^2(1)}{1 - r_{xx}^2(1)}.\end{aligned}\quad (5.4.10)$$

Используя те же самые приближения, что и выше, остаточную сумму квадратов $S(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2)$ можно записать в виде

$$S(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2) = (N - 2) [c_{xx}(0) - \hat{\alpha}_1 c_{xx}(1) - \hat{\alpha}_2 c_{xx}(2)]. \quad (5.4.11)$$

Остаточная дисперсия равна

$$s_z^2 = \frac{1}{N-5} S(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2)$$

и имеет $N - 5$ степеней свободы, так как исходное правдоподобие (5.4.13) содержит $N - 2$ наблюдения и еще 3 степени свободы потеряны при подгонке трех параметров μ , α_1 и α_2 .

Снова используя то же приближение, что и в (П4.1.15), получаем совместную $100(1 - \alpha)\%$ -ную доверительную область для параметров (α_1, α_2) :

$$\begin{aligned}(\alpha_1 - \hat{\alpha}_1)^2 + 2r_{xx}(1)(\alpha_1 - \hat{\alpha}_1)(\alpha_2 - \hat{\alpha}_2) + (\alpha_2 - \hat{\alpha}_2)^2 &\leq \\ &\leq 2 \frac{s_z^2 f_{2, N-5}(1 - \alpha)}{\sum_{t=1}^N (x_t - \bar{x})^2}.\end{aligned}$$

В качестве примера рассмотрим данные о партиях продукта, приведенные на рис. 5.2. В разд. 5.4.3 будет показано, что к этим данным вполне подходит процесс авторегрессии второго порядка. Используя значения $r_{xx}(1)$ и $r_{xx}(2)$ из табл. 5.2 и формулу (5.4.10),

получаем выборочные оценки параметров $\hat{\alpha}_1 = -0,32$ и $\hat{\alpha}_2 = +0,18$.

Остаточная сумма квадратов $S(\hat{\alpha}_1, \hat{\alpha}_2)$ равна 7768,5, так что $s_z^2 = 7768,5/65 = 119,6$. Следовательно, приближенная 95%-ная доверительная область имеет вид

$$(\alpha_1 + 0,32)^2 - 0,78(\alpha_1 + 0,32)(\alpha_2 - 0,18) + (\alpha_2 - 0,18)^2 \leq 0,077.$$

На рис. 5.15 показаны линии уровня точной суммы квадратов, изображенные на плоскости (α_1, α_2) в области, где процесс стационарен. Заштрихованная область является 95%-ной доверительной областью. Видно, что она лежит целиком внутри области стационарности.

Общий процесс авторегрессии. Действуя так же, как и выше, уравнения правдоподобия можно приближенно записать в виде

$$c_{xx}(j) = \hat{\alpha}_1 c_{xx}(j-1) + \hat{\alpha}_2 c_{xx}(j-2) + \dots + \hat{\alpha}_m c_{xx}(j-m), \quad (5.4.12)$$

где $j=1, 2, \dots, m$. Аналогично для остаточной суммы квадратов получаем приближенное выражение

$$S(\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_m) = (N-m) [c_{xx}(0) - \hat{\alpha}_1 c_{xx}(1) - \dots - \hat{\alpha}_m c_{xx}(m)]. \quad (5.4.13)$$

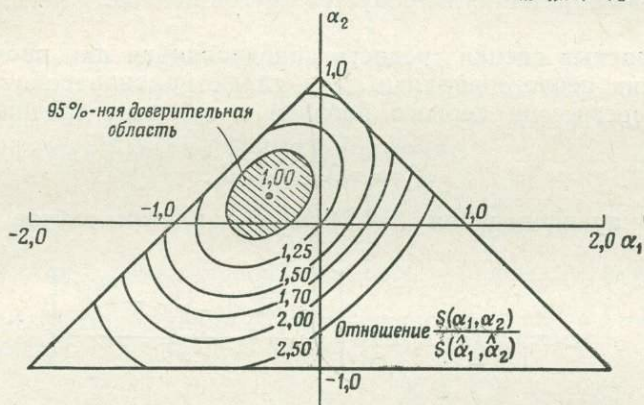


Рис. 5.15. Линии уровня суммы квадратов для процесса авторегрессии второго порядка, подобранного к данным о партиях продукта, изображенным на рис. 5.2.

Из (5.4.13) получаем приближенную выборочную оценку остаточной дисперсии

$$\hat{s}_z^2 = \frac{1}{N-2m-1} S(\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_m).$$

Наконец, используя (П4.1.15), можно написать приближенную доверительную область в матричных обозначениях

$$(\alpha - \hat{\alpha})' C (\alpha - \hat{\alpha}) \leq \frac{m}{N-2m-1} f_{m, N-2m-1}(1-\alpha) \hat{s}_z^2,$$

где $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_m)$ и

$$C = \begin{pmatrix} c_{xx}(0) & c_{xx}(1) & \dots & c_{xx}(m-1) \\ c_{xx}(1) & c_{xx}(0) & \dots & c_{xx}(m-2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{xx}(m-1) & c_{xx}(m-2) & \dots & c_{xx}(0) \end{pmatrix}.$$

5.4.2. Выборочные оценки среднего правдоподобия для параметров авторегрессии

Так как функция правдоподобия (5.4.2) является с точностью до множителя многомерной нормальной плотностью, то с первого взгляда могло бы показаться, что ее можно адекватно описать с помощью средних значений и ковариаций, как указано в разд. 4.4.1. Однако если выборочные оценки максимального правдоподобия лежат близко к границам области стационарности, то функция правдоподобия обрезается и требуется другой подход.

Выборочные оценки среднего правдоподобия для процесса авторегрессии первого порядка. Для иллюстрации рассмотрим процесс авторегрессии первого порядка с нулевым средним значением

$$X_t = \alpha_1 X_{t-1} + Z_t.$$

Функцию правдоподобия (5.4.2) можно в этом случае записать в виде

$$L(\alpha_1 | x_1) = K_1 \exp \left[- \frac{\sum_{t=2}^N x_{t-1}^2 (\alpha_1 - \hat{\alpha}_1)^2}{2\sigma_Z^2} \right], \quad (5.4.14)$$

где

$$\hat{\alpha}_1 = \frac{\sum_{t=2}^N x_{t-1} x_t}{\sum_{t=2}^N x_{t-1}^2} \quad (5.4.15)$$

является выборочной оценкой максимального правдоподобия для α_1 . Отсюда видно, что при условии, если σ_Z^2 известна, функция правдоподобия является с точностью до постоянного множителя нормальной со средним значением α_1 и дисперсией

$$\frac{\sigma_Z^2}{\sum_{t=2}^N x_{t-1}^2}.$$

Описание функции правдоподобия (5.4.14) с помощью ее среднего значения и дисперсии было бы адекватным при условии, что область изменения α_1 была бы от $-\infty$ до $+\infty$. Однако в силу того, что модель является стационарной лишь для $|\alpha_1| < 1$, при описании с помощью нормальной плотности возникают трудности, когда функция правдоподобия имеет максимум вблизи $|\alpha_1| = 1$. В таком случае функция правдоподобия резко отсекалась бы в одной из

точек $\alpha_1 = 1$, $\alpha_1 = -1$, так что аппроксимация с помощью нормального распределения была бы несправедливой.

Методы преобразования правдоподобий, обсуждавшиеся в разд. 4.4.5, также неприменимы, так как не существует преобразования, дающего нормальное распределение, если функция правдоподобия заканчивается ненулевым значением. В этом случае наилучший способ состоит в вычислении выборочной оценки среднего правдоподобия в интервале $(-1, 1)$, определяемой равенством

$$\bar{\alpha}_1 = \frac{\int_{-1}^1 \alpha_1 L(\alpha_1) d\alpha_1}{\int_{-1}^1 L(\alpha_1) d\alpha_1}.$$

Подставляя сюда $L(\alpha_1)$ из (5.4.14), получаем

$$\bar{\alpha}_1 = \hat{\alpha}_1 - \frac{\sigma_z}{(\sum x_{t-1}^2)^{1/2}} \frac{f(L_2) - f(L_1)}{F(L_2) - F(L_1)}, \quad (5.4.16)$$

где $f(x)$ и $F(x)$ — нормальная плотность вероятности и нормальная функция распределения соответственно, $\hat{\alpha}_1$ — выборочная оценка максимального правдоподобия и

$$L_2 = \frac{\sum x_{t-1}^2}{\sigma_z^2} (1 - \hat{\alpha}_1), \quad L_1 = \frac{\sum x_{t-1}^2}{\sigma_z^2} (-1 - \hat{\alpha}_1).$$

Если σ_z^2 неизвестна, то маргинальное правдоподобие для α_1 можно получить, интегрируя (5.4.4) с дифференциалом $d(\sigma_z^2)/\sigma_z^2$, как указывалось в разд. 4.4.6. Можно проверить, что после такого интегрирования получается следующее маргинальное правдоподобие для α_1 :

$$L_1(\alpha_1) = K \left[(N-2)s_z^2 + \sum_{t=2}^N x_{t-1}^2 (\alpha_1 - \hat{\alpha}_1)^2 \right]^{-(N-1)/2}. \quad (5.4.17)$$

Случайная величина, соответствующая (5.4.17), имеет с точностью до постоянного множителя t -распределение с $(N-2)$ степенями свободы. Величина s_z^2 в (5.4.17) является обычной выборочной оценкой σ_z^2 , полученной по остаточной сумме квадратов, а именно

$$s_z^2 = \frac{1}{N-2} \left[\sum_{t=2}^N x_t^2 - \hat{\alpha}_1^2 \sum_{t=2}^N x_{t-1}^2 \right]. \quad (5.4.18)$$

Можно проверить, что и теперь выборочная оценка среднего правдоподобия задается формулой (5.4.16), но при этом σ_z^2

заменяется на s_z^2 , а $f(x)$ и $F(x)$ относятся к t -распределению с $(N - 2)$ степенями свободы.

Предположим, например, что из временного ряда длины $N = 20$ мы получили значения $s_z^2 = 1,44$, $\sum x_{t-1}^2 = 64$, $\hat{\alpha}_1 = 0,9$. Тогда маргинальная функция правдоподобия для α_1 представляет собой усеченное t -распределение, как показано на рис. 5.16. Используя (5.4.16),

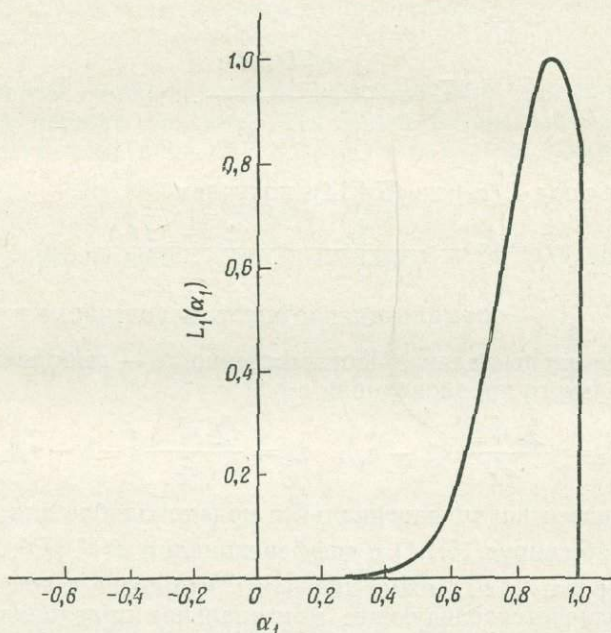


Рис. 5.16. Маргинальная функция правдоподобия для процесса авторегрессии первого порядка.

где $f(x)$ и $F(x)$ относятся к t -распределению с 18 степенями свободы, получаем, что выборочная оценка среднего правдоподобия $\bar{\alpha}_1 = 0,86$. Заметим, что операция усреднения правдоподобия отодвинула выборочную оценку максимального правдоподобия $\hat{\alpha}_1 = 0,9$ от границы стационарности.

Для простоты среднее значение μ временного ряда было положено равным нулю в предыдущих рассуждениях. Можно допустить и ненулевое среднее значение, подставляя $x_t - \mu$ вместо x_t . Интегрирование по μ , а также по σ_z^2 дает маргинальное правдоподобие для α_1 , которое совпадает с (5.4.17), за исключением того, что x_t

в (5.4.17) заменяется на отклонение от среднего $(x_t - \bar{x})$, а N — на $N - 1$, т. е.

$$L_1(\alpha_1) = K \left[(N-3)s_z^2 + \sum_{t=2}^N (x_{t-1} - \bar{x})^2 (\alpha_1 - \hat{\alpha}_1)^2 \right]^{-(N-2)/2}.$$

Выборочные оценки среднего правдоподобия для процесса авторегрессии второго порядка. Рассмотрим процесс второго порядка:

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + Z_t.$$

Функция правдоподобия равна

$$\begin{aligned} L(\alpha_1, \alpha_2, \mu, \sigma_z^2 | x_1, x_2) = \\ = \frac{1}{(\sqrt{2\pi}\sigma_z)^{N-2}} \exp \left\{ -\frac{1}{2\sigma_z^2} \sum_{t=3}^N [(x_t - \mu) - \alpha_1(x_{t-1} - \mu) - \right. \\ \left. - \alpha_2(x_{t-2} - \mu)]^2 \right\}. \end{aligned}$$

Интегрируя по μ , получаем с точностью до малых «концевых» по правок

$$\begin{aligned} L(\alpha_1, \alpha_2, \sigma_z^2 | x_1, x_2) = \\ = \frac{1}{(\sqrt{2\pi}\sigma_z)^{N-3}} \exp \left\{ -\frac{1}{2\sigma_z^2} \sum_{t=3}^N [(x_t - \bar{x}) - \alpha_1(x_{t-1} - \bar{x}) - \right. \\ \left. - \alpha_2(x_{t-2} - \bar{x})]^2 \right\}. \quad (5.4.19) \end{aligned}$$

Следующее интегрирование с дифференциалом $d(\sigma_z^2)/\sigma_z^2$ дает совместное правдоподобие для α_1, α_2 :

$$\begin{aligned} L(\alpha_1, \alpha_2 | x_1, x_2) = K \left\{ \sum_{t=3}^N [(x_t - \bar{x}) - \alpha_1(x_{t-1} - \bar{x}) - \right. \\ \left. - \alpha_2(x_{t-2} - \bar{x})]^2 \right\}^{-(N-3)/2}. \quad (5.4.20) \end{aligned}$$

В случае когда 95%-ная доверительная область лежит полностью в области стационарности, как, например, на рис. 5.15, функция правдоподобия адекватно описывается своими средними значениями и ковариациями. Если же выборочные оценки максимального правдоподобия лежат близко к границе стационарности, то единственный надежный метод заключается в нанесении линий уровня

правдоподобия. Вместо оценок максимального правдоподобия в этом случае лучше вычислить выборочные оценки среднего правдоподобия. Впрочем, сначала удобно сделать преобразование

$$\varphi_1 = \frac{\alpha_1}{1 - \alpha_2}, \quad \varphi_2 = \alpha_2. \quad (5.4.21)$$

С помощью (5.4.20) можно проверить, что в окрестности максимума правдоподобия

$$\frac{d^2 l}{d\varphi_1 d\varphi_2} \approx 0$$

и, следовательно, параметры φ_1 и φ_2 ортогональны.

Преобразование (5.4.21) переводит треугольную область стационарности в квадратную область $|\varphi_1| < 1$, $|\varphi_2| < 1$. Теперь можно получить численным методом выборочные оценки среднего правдоподобия, например

$$\bar{\varphi}_1 = \frac{\int_{-1}^1 \int_{-1}^1 \varphi_1 L(\varphi_1, \varphi_2) d\varphi_1 d\varphi_2}{\int_{-1}^1 \int_{-1}^1 L(\varphi_1, \varphi_2) d\varphi_1 d\varphi_2},$$

где

$$L(\varphi_1, \varphi_2) = K \left\{ \sum_{t=3}^N [(x_t - \bar{x}) - \varphi_1 (1 - \varphi_2) (x_{t-1} - \bar{x}) - \varphi_2 (x_{t-2} - \bar{x})]^2 \right\}^{-(N-3)/2}.$$

Выборочные оценки среднего правдоподобия для α_1 и α_2 можно затем получить из $\bar{\varphi}_1$ и $\bar{\varphi}_2$ с помощью обратного преобразования

$$\bar{\alpha}_1 = \bar{\varphi}_1 (1 - \bar{\varphi}_2), \quad \bar{\alpha}_2 = \bar{\varphi}_2.$$

5.4.3. Определение порядка процесса авторегрессии

В этом разделе рассматривается задача определения порядка m процесса авторегрессии. Простой метод основан на том, что если в подбираемой модели (5.4.1) взято недостаточное число членов, то выборочная оценка дисперсии σ_Z^2 будет завышена за счет тех членов, которые еще не включены в модель. Лишь когда правильное число членов включено в модель, получается правильная оценка σ_Z^2 .

Это наводит на мысль о том, что если выборочную оценку

$$s_z^2(m) = \frac{1}{N - 2m - 1} S(\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_m)$$

остаточной дисперсии построить в зависимости от m , то кривая будет иметь минимум или станет пологой в точке, соответствующей правильной степени процесса. На рис. 5.17 показан график $s_z^2(m)$, в зависимости от m для данных о партиях продукта, изображенных

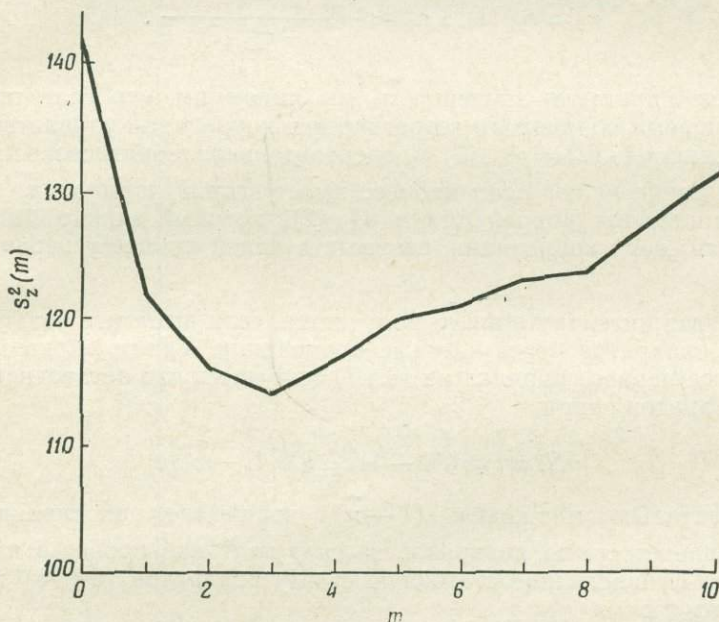


Рис. 5.17. Остаточные дисперсии для моделей авторегрессии, подобранных к данным о партиях продукта, изображенным на рис. 5.2.

на рис. 5.2. Видно, что кривая становится пологой около $m = 2$ и 3. Значит, для этих данных подошел бы процесс авторегрессии второго или третьего порядка.

Частная корреляционная функция. Один из недостатков метода, основанного на $s_z^2(m)$, состоит в том, что он не всегда может уверенно указывать, какое требуется значение m . Например, достаточно ли уменьшение $s_z^2(m)$ на рис. 5.17 при переходе от $m = 2$ к $m = 3$, чтобы гарантировать справедливость модели третьего порядка? Более чувствительный критерий получается с помощью определения для каждого значения m выборочной оценки $\hat{\pi}_m = \hat{\alpha}_m$

последнего коэффициента α_m в подбираемой модели, а также доверительной области для него, как это делалось в разд. 5.4.1. По причинам, которые будут объяснены в гл. 11, график $\hat{\pi}_m$ в зависимости от m называется *частной корреляционной функцией*. Используя результаты разд. 5.4.1, можно выписать следующие приближенные выражения для первых двух значений $\hat{\pi}_m$:

$$\hat{\pi}_1 \approx r_{xx}(1), \quad \hat{\pi}_2 \approx \frac{r_{xx}(2) - r_{xx}^2(1)}{1 - r_{xx}^2(1)}.$$

Для интерпретации значения $\hat{\pi}_2$ мы напомним, что если процесс имеет первый порядок, то теоретические корреляции удовлетворяют уравнению $\rho_{xx}(2) = \rho_{xx}^2(1)$ и, следовательно, теоретическое значение π_2 равно нулю. Если процесс имеет второй порядок, то π_2 измеряет *избыток* корреляции в $\rho_{xx}(2)$, который можно было бы ожидать сверх корреляции, соответствующей процессу первого порядка.

Другая интерпретация $\hat{\pi}_2$ получается, если выразить остаточную сумму квадратов через частные корреляции. Таким образом, для процесса первого порядка из (5.4.7) получаем, что остаточная сумма квадратов равна

$$S(\hat{\alpha}_1) \approx (N-1)c_{xx}(0)(1 - \hat{\pi}_1^2).$$

Следовательно, множитель $(1 - \hat{\pi}_1^2)$ показывает, во сколько раз уменьшается сумма квадратов за счет подгонки процесса первого порядка. Аналогично остаточную сумму квадратов (5.4.11) можно записать в виде

$$S(\hat{\alpha}_1, \hat{\alpha}_2) \approx (N-2)c_{xx}(0)(1 - \hat{\pi}_1^2)(1 - \hat{\pi}_2^2).$$

Следовательно, $(1 - \hat{\pi}_2^2)$ дает *дополнительный* уменьшающий множитель для суммы квадратов, получающийся за счет увеличения порядка модели до второго.

На рис. 5.18 показана частная корреляционная функция для данных о партиях продукта, изображенных на рис. 5.2. Доверительные интервалы с уровнем доверия 95% на рис. 5.18 центрированы около нуля, чтобы выделить те коэффициенты, которые можно считать отличными от нуля. Видно, что $\hat{\pi}_1$ лежит вне этой полосы, $\hat{\pi}_2$ лежит вблизи границы, а значения $\hat{\pi}_m$ при $m > 2$ лежат глубоко внутри полосы. Это указывает на то, что для описания этих данных подходящим является процесс первого порядка, а не третьего, как

это следовало из рис. 5.17. Однако, учитывая, что $\hat{\pi}_2$ лежит вблизи границы доверительного интервала и что $s_z^2(2)$ значительно меньше, чем $s_z^2(1)$, как видно из рис. 5.17, можно заключить, что для правильного соответствия этим данным требуется модель второго порядка.

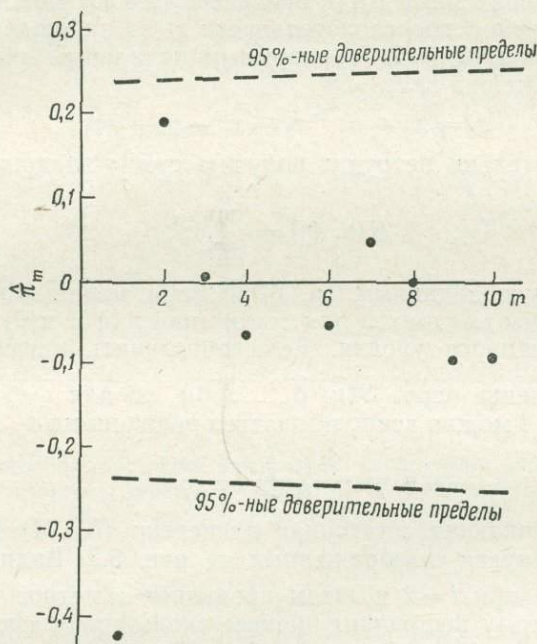


Рис. 5.18. Частные корреляции для данных о партиях продукта, изображенных на рис. 5.2.

5.4.4. Оценивание параметров процесса скользящего среднего

Первый вопрос, который надо решить при подборе процесса скользящего среднего

$$X_t = \mu + Z_t + \beta_1 Z_{t-1} + \dots + \beta_l Z_{t-l}, \quad (5.4.22)$$

заключается в выборе подходящего порядка модели l . Метод анализа здесь более сложный, чем для процесса авторегрессии, и ради простоты приходится определять наилучшее значение l по выборочной оценке остаточной дисперсии $s_z^2(l)$. Это происходит из-за того, что трудно в явном виде выписать функцию правдоподобия процесса (5.4.22), хотя для частного случая это было сделано в [13]. Впрочем, можно использовать простые численные способы для

рекуррентного вычисления логарифмической функции правдоподобия [1].

Для иллюстрации этого подхода рассмотрим процесс скользящего среднего первого порядка

$$X_t = \mu + Z_t + \beta_1 Z_{t-1}. \quad (5.4.23)$$

При заданных значениях μ и β_1 равенство (5.4.23) можно использовать для получения последовательности z_t из наблюдаемых значений x_t . Так как $E[Z_t] = 0$, то разумным начальным значением является $z_0 = 0$. Отсюда получаем

$$z_1 = x_1 - \mu, \quad z_2 = x_2 - \mu - \beta_1 z_1$$

и т. д. Следовательно, нетрудно получить сумму квадратов

$$S(\mu, \beta_1) = \sum_{i=1}^N z_i^2,$$

соответствующую заданным (μ, β_1) . Затем можно построить поверхность суммы квадратов для сетки значений μ и β_1 и наметить контуры постоянного уровня. Если обозначить наименьшее для данного l значение через $S(\hat{\mu}, \hat{\beta}_1, \dots, \hat{\beta}_l)$, то для выбора наилучшего значения l можно воспользоваться величинами

$$s_z^2(l) = \frac{S(\hat{\mu}, \hat{\beta}_1, \dots, \hat{\beta}_l)}{N - (l + 1)}. \quad (5.4.24)$$

На рис. 5.19 показана остаточная дисперсия (5.4.24) для данных о партиях продукта, изображенных на рис. 5.2. Видно, что $s_z^2(l)$ выравнивается при $l = 2$ и затем проявляет заметное уменьшение при $l = 8$. Поэтому необходим процесс скользящего среднего восьмого порядка, чтобы получить приблизительно такое же согласие с данными, что и у процесса авторегрессии второго порядка. Ясно, что более простой процесс авторегрессии является и более реалистичной моделью.

Поскольку трудно выписать в явном виде сумму квадратов, приходится рассмотреть и другой способ получения доверительных областей. Если контуры линий уровня суммы квадратов построены, то доверительную область можно получить, выбирая согласно (П4.1.17) тот контур, для которого

$$S(\mu, \beta_1, \dots, \beta_l) = S(\hat{\mu}, \hat{\beta}_1, \dots, \hat{\beta}_l) \times \left[1 + \frac{l+1}{N - (l+1)} f_{l+1, N-l-1} (1 - \alpha) \right]. \quad (5.4.25)$$

Для иллюстрации равенства (5.4.25) с помощью случайных гауссовских чисел было получено 50 членов процесса

$$X_t = 5 + Z_t + 0,5Z_{t-1}.$$

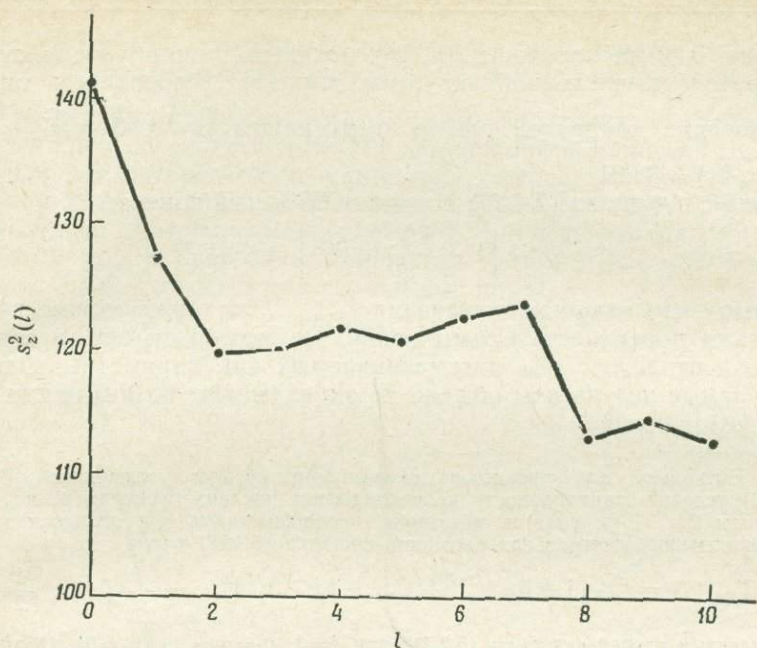


Рис. 5.19. Остаточные дисперсии для моделей скользящего среднего, подобранных к данным о партиях продукта, изображенным на рис. 5.2.

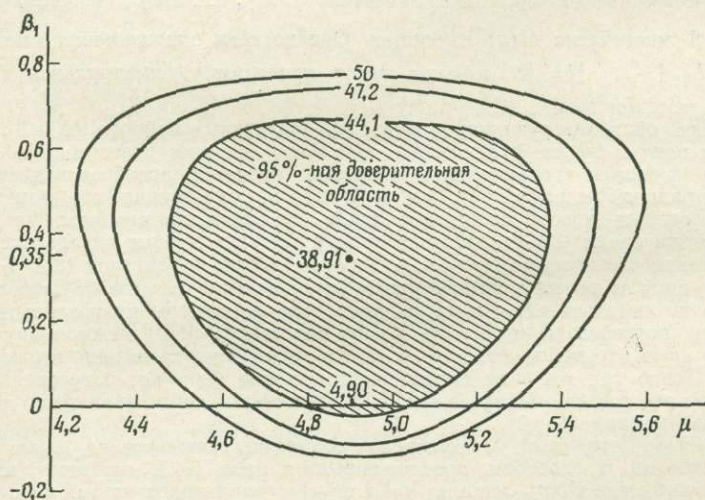


Рис. 5.20. Линии уровня суммы квадратов для процесса скользящего среднего первого порядка.

На рис. 5.20 на плоскости (μ, β_1) показаны линии уровня суммы квадратов, вычисленной по этим данным. Выборочные оценки наименьших квадратов для μ и β_1 равны $\hat{\mu} = 4,90$ и $\hat{\beta}_1 = 0,35$, а $S(\hat{\mu}, \hat{\beta}_1) = 38,91$.

Отсюда с помощью (5.4.25) получаем 95%-ный контур;

$$38,91 \left[1 + \frac{2}{48} (3,20) \right] = 44,1.$$

Отметим, что начальные значения z_{1-l}, \dots, z_0 также можно варьировать, а поверхность суммы квадратов можно строить в зависимости и от z_{1-l}, \dots, z_0 , рассматриваемых как параметры. Однако получаемые результаты обычно не оправдывают возникающих при этом осложнений*).

*) Поскольку для процессов скользящего среднего конечного порядка (5.4.22) условие стационарности не накладывает никаких ограничений на коэффициенты β_1, \dots, β_l , имеется некоторая неоднозначность при оценивании этих коэффициентов. Спектральная плотность процесса (5.4.22) равна

$$\Gamma_{XX}(f) = \sigma_z^2 \cdot \left| 1 + \beta_1 e^{-j2\pi f} + \dots + \beta_l e^{-j2\pi fl} \right|^2, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}$$

(это частный случай формулы (6.2.24) при $\Delta=1, \alpha_1=\alpha_2=\dots=\alpha_m=0$). Обозначим корни многочлена

$$M(p) = 1 + \beta_1 p + \dots + \beta_l p^l$$

через b_1, \dots, b_l . Если заменить любой набор из этих корней b_{k_1}, \dots, b_{k_r} на комплексно-сопряженные обратные значения $1/b_{k_1}^*, \dots, 1/b_{k_r}^*$, то коэффициенты β_1, \dots, β_l многочлена $M(p)$ изменятся. Однако если одновременно заменить σ_z^2 на $\sigma_z^2 |b_{k_1}|^2 \cdot \dots \cdot |b_{k_r}|^2$, то, как легко проверить, спектральная плотность $\Gamma_{XX}(f)$ не изменится.

Таким образом, существует несколько различных наборов $(\sigma_z, \beta_1, \dots, \beta_l)$, дающих одну и ту же функцию $\Gamma_{XX}(f)$. В гл. 6 станет известно, что ковариационная функция стационарного процесса является преобразованием Фурье от спектральной плотности и, таким образом, однозначно ею определяется. В свою очередь ковариационная функция гауссовского процесса (с нулевым средним значением) однозначно определяет все многомерные распределения процесса. Таким образом, существуют различные наборы параметров $(\sigma_z, \beta_1, \dots, \beta_l)$, дающие одни и те же конечные распределения процесса. Следовательно, безуспешно пытаться однозначно оценить эти параметры по реализации. Если, например, потребовать, чтобы все корни многочлена $M(p)$ лежали внутри единичного круга, то набор $(\sigma_z, \beta_1, \dots, \beta_l)$ и спектр будут связаны взаимно однозначно. Точно так же ради однозначности можно было бы потребовать, чтобы все корни многочлена $M(p)$ лежали вне единичного круга (при этом дисперсия σ_z^2 была бы наименьшей).

Отметим, что для устойчивости процедуры определения реализации z_t по реализации x_t способом, предложенным в разд. 5.4.4, требуется, чтобы все корни многочлена $M(p)$ лежали либо внутри единичного круга, либо вне его. Во втором случае нужно задать значения z_N, \dots, z_{N-l+1} и из рекуррентного соотношения (5.4.22) при $t=N, N-1, \dots, 1$ последовательно определять значения $z_{N-l}, z_{N-l-1}, \dots, z_{1-l}$. — Прим. перев.

5.4.5. Оценивание параметров смешанного процесса авторегрессии — скользящего среднего

Так как при дискретизации непрерывного процесса авторегрессии порядка m , согласно (5.2.49), получается смешанный дискретный процесс авторегрессии — скользящего среднего, было бы естественно ожидать, что смешанные модели окажутся полезными при подгонке ко многим временным рядам. Для иллюстрации того, как при этом можно построить поверхность логарифмической функции правдоподобия, рассмотрим смешанный процесс

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + Z_t + \beta_1 Z_{t-1}.$$

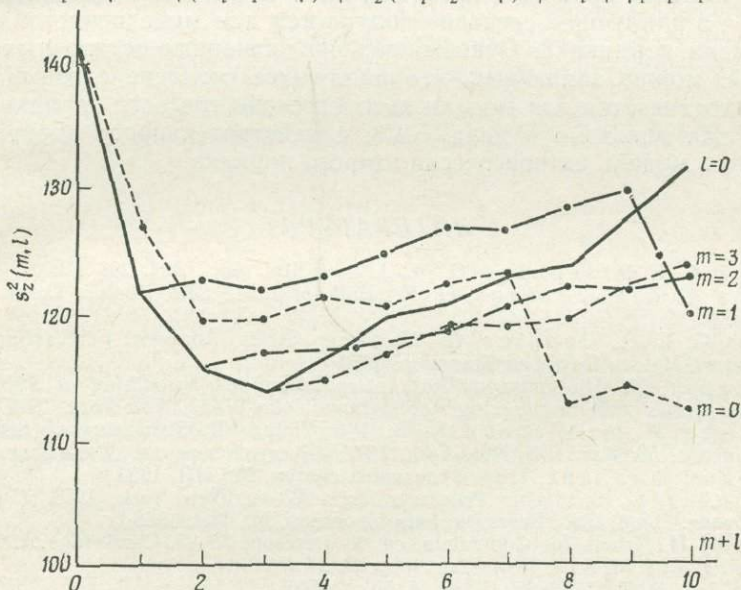


Рис. 5.21. Остаточные дисперсии для смешанных моделей, подобранных к данным о партиях продукта, изображенным на рис. 5.2.

При фиксированных значениях параметров μ , α_1 , α_2 и β_1 последовательность z можно получить из равенств

$$z_3 = (x_3 - \mu) - \alpha_1(x_2 - \mu) - \alpha_2(x_1 - \mu),$$

$$z_4 = (x_4 - \mu) - \alpha_1(x_3 - \mu) - \alpha_2(x_2 - \mu) - \beta_1 z_3$$

и т. д. Поверхность суммы квадратов в таком случае можно представить, строя

$$S(\mu, \alpha_1, \alpha_2, \beta_1 | x_1, x_2, z_1=0, z_2=0) = \sum_{i=3}^N z_i^2$$

как функцию от μ , α_1 , α_2 и β_1 . С помощью вычислительной машины нетрудно произвести перебор этих моделей, сначала фиксируя m — порядок процесса авторегрессии, а затем меняя l — порядок процесса скользящего среднего. Затем можно построить остаточную дисперсию

$$s_z^2(m, l) = \frac{1}{N-l-2m-1} S(\hat{\mu}, \dots, \hat{a}_m, \hat{\beta}_1, \dots, \hat{\beta}_l)$$

как функцию l и m .

На рис. 5.21 показаны $s_z^2(m, l)$ в зависимости от $m+l$ для данных о партиях продукта, изображенных на рис. 5.2. Видно, что при $l+m < 8$ наилучшее согласие получается для модели чистой авторегрессии порядка 3. Основываясь на значениях остаточных дисперсий, можно заключить, что наилучшее согласие с этими данными достигается для модели авторегрессии третьего порядка. Однако, как показано в разд. 5.4.3, в действительности адекватной является модель авторегрессии второго порядка.

ЛИТЕРАТУРА

1. Box G. E. P., Jenkins G. M., J. Roy. Stat. Soc., **B24**, 297 (1962).
2. Box G. E. P., Jenkins G. M., Bull. of I.S.I., 24th session, Ottawa, 943 (1963).
3. Box G. E. P., Jenkins G. M., Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, 1970.
4. Wiener N., The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, John Wiley, New York, 1949.
5. James H. M., Nichols N. B., Phillips R. S., Theory of Servomechanisms, McGraw-Hill, New York, 1947. (Русский перевод: Джеймс, Никольс, Филлипс, Теория следящих систем, М., ИЛ, 1953.)
6. Doob J. L., Stochastic Processes, John Wiley, New York, 1953. (Русский перевод: Дуб Дж., Вероятностные процессы, М., ИЛ, 1956.)
7. Wold H., Tracts for Computers, ed. K. Pearson, № 25, Cambridge, 1948.
8. Bartlett M. S., J. Roy. Stat. Soc., **B8**, 27 (1946).
9. Fuller A. T., J. Electr. Contr., **4**, 551 (1958).
10. Parzen E., Technometrics, **3**, 167 (1961).
11. Schaerf M. G., Stanford Univ. Tech. Rep., 12 (1964).
12. Anderson R. L., Ann. Math. Stat., **13**, 1 (1942).
13. Barnard G. A., Jenkins G. M., Winsten C. B., J. Roy. Stat. Soc., **A125**, 321 (1962).

ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ

Линейную систему $\hat{h}(u)$, дающую минимальную среднеквадратичную ошибку (5.1.12), можно найти очень просто с помощью вариационного исчисления. Мы будем предполагать, что ковариационные функции $\gamma_{XX}(u)$ и $\gamma_{XY}(u)$ известны для всех значений запаздывания u .

Прежде всего отметим, что (5.1.12) можно переписать в виде

$$\begin{aligned} \varepsilon[h(u)] = & \gamma_{YY}(0) - 2 \int_0^{\infty} h(u) \gamma_{XY}(u) du + \\ & + \int_0^{\infty} \int_0^{\infty} h(u) h(v) \gamma_{XX}(u-v) du dv, \quad (\text{П5.1.1}) \end{aligned}$$

так как единственной неизвестной функцией является $h(u)$. Общий метод, используемый в вариационном исчислении, состоит в следующем. Предполагают, что ответ известен, и затем находят условия, вытекающие из того, что этот ответ правильный. Таким образом, мы предположим, что конкретная функция $\hat{h}(u)$ является именно той функцией, которая минимизирует $\varepsilon[h(u)]$, т. е.

$$\varepsilon[\hat{h}(u)] < \varepsilon[h(u)] \text{ для всех } h(u) \neq \hat{h}(u). \quad (\text{П5.1.2})$$

Далее, для любой функции $\tilde{h}(u) = \hat{h}(u) + g(u)$, где $g(u)$ — произвольная функция от u , удовлетворяющая граничным условиям на $h(u)$, мы будем иметь $\varepsilon[\tilde{h}(u)] > \varepsilon[\hat{h}(u)]$, если $g(u)$ не равна тождественно нулю.

Вообще если

$$h(u) = \hat{h}(u) + bg(u), \quad (\text{П5.1.3})$$

то $\varepsilon[h(u)]$ достигает минимума, когда $h(u) = \hat{h}(u)$, т. е. при $b = 0$. Выражая это условие минимума через производные, получаем

$$\frac{\partial \varepsilon[h(u)]}{\partial b} = 0 \quad (b=0), \quad \frac{\partial^2 \varepsilon[h(u)]}{\partial b^2} > 0 \quad (b=0). \quad (\text{П5.1.4})$$

Подставляя в нашем частном случае в (П5.1.1) вместо $h(u)$ выражение (П5.1.3), получаем

$$\begin{aligned} \varepsilon [h(u)] = & \gamma_{YY}(0) - 2 \int_0^{\infty} [\hat{h}(u) + bg(u)] \gamma_{XY}(u) du + \\ & + \int_0^{\infty} \int_0^{\infty} [\hat{h}(u) + bg(u)] [\hat{h}(v) + bg(v)] \gamma_{XX}(u-v) du dv \end{aligned} \quad (\text{П5.1.5})$$

и

$$\begin{aligned} \frac{\partial \varepsilon}{\partial b} [h(u)] = & -2 \int_0^{\infty} g(u) \gamma_{XY}(u) du + \\ & + \int_0^{\infty} \int_0^{\infty} [\hat{h}(u) g(v) + g(u) \hat{h}(v) + 2bg(u)g(v)] \times \\ & \times \gamma_{XX}(u-v) du dv. \end{aligned} \quad (\text{П5.1.6})$$

Приравнивая $b=0$ в (П5.1.6) и используя первое из условий (П5.1.4), получаем

$$0 = -2 \int_0^{\infty} g(u) \left[\gamma_{XY}(u) - \int_0^{\infty} \hat{h}(v) \gamma_{XX}(u-v) dv \right] du, \quad (\text{П5.1.7})$$

поскольку $\gamma_{XX}(u)$ — четная функция, как показано в разд. 5.2.1.

Так как равенство (П5.1.7) должно выполняться для любой функции $g(u)$, то $\hat{h}(u)$ должна удовлетворять условию

$$\gamma_{XY}(u) = \int_0^{\infty} \hat{h}(v) \gamma_{XX}(u-v) dv, \quad u \geq 0. \quad (\text{П5.1.8})$$

Можно проверить, что вторая производная по b в точке $b=0$ положительна, так что это решение действительно соответствует минимуму. Таким образом, $\hat{h}(u)$ должна удовлетворять интегральному уравнению (П5.1.8), которое называется интегральным уравнением Винера—Хопфа.

Ограничение, требующее, чтобы соотношение (П5.1.8) было верным лишь при $u \geq 0$, появляется из-за условия физической реализуемости фильтра, а именно $h(u) = 0$ при $u < 0$.

ПРИЛОЖЕНИЕ П5.2

МОМЕНТЫ ЛИНЕЙНОГО ПРОЦЕССА

Рассмотрим общий линейный процесс (5.2.6), а именно

$$X(t) - \mu = \int_0^{\infty} h(v) Z(t-v) dv, \quad (\text{П5.2.1})$$

где $Z(t)$ — белый шум, обладающий следующими свойствами:

$$E [Z(t)] = 0, \quad (\text{П5.2.2})$$

$$\text{Cov} [Z(t), Z(t+u)] = \sigma_Z^2 \delta(u), \quad (\text{П5.2.3})$$

$$E [Z(t) Z(t+u_1) Z(t+u_2)] = \mu_3 \delta(u_1) \delta(u_2), \quad (\text{П5.2.4})$$

$$\begin{aligned} E [Z(t) Z(t+u_1) Z(v) Z(v+u_2)] &= \sigma_Z^4 [\delta(u_1) \delta(u_2) + \\ &+ \delta(v-t) \delta(v-t+u_2-u_1) + \\ &+ \delta(v-t+u_2) \delta(v-t-u_1)] + \\ &+ K_4(Z) \delta(u_1) \delta(v-t) \delta(v-t+u_2). \end{aligned} \quad (\text{П5.2.5})$$

Как и прежде, $\delta(u)$ обозначает дельта-функцию Дирака. Сейчас мы выведем младшие моменты случайного процесса $X(t)$, считая, что процесс $Z(t)$ обладает указанными свойствами.

Из (П5.2.1) и (П5.2.2) получаем

$$\begin{aligned} E [X(t) - \mu] &= E \left[\int_0^{\infty} h(v) Z(t-v) dv \right] = \\ &= \int_0^{\infty} h(v) E [Z(t-v)] dv = 0. \end{aligned} \quad (\text{П5.2.6})$$

Аналогично из (П5.2.1) и (П5.2.3) имеем

$$\begin{aligned} \text{Cov} [X(t), X(t+u)] &= \text{Cov} \left[\int_0^{\infty} h(v) Z(t-v) dv, \right. \\ &\quad \left. \int_0^{\infty} h(v') Z(t+u-v') dv' \right] = \\ &= \int_0^{\infty} \int_0^{\infty} h(v) h(v') \text{Cov} [Z(t-v), \\ &\quad Z(t+u-v')] dv dv' = \\ &= \int_0^{\infty} \int_0^{\infty} h(v) h(v') \sigma_Z^2 \delta(u-v'+v) dv dv' = \\ &= \sigma_Z^2 \int_0^{\infty} h(v) h(v+u) dv = \gamma_{XX}(u). \end{aligned} \quad (\text{П5.2.7})$$

При $u = 0$ это сводится к

$$\text{Var} [X(t)] = \sigma_Z^2 \int_0^{\infty} h^2(v) dv.$$

Поэтому если интеграл $\int_0^{\infty} h^2(v) dv$ конечен, то $X(t)$ является стационарным процессом второго порядка, так как ковариационная функция $\gamma_{XX}(t, t+u)$ зависит только от запаздывания u .

Аналогично получаем

$$\begin{aligned} E[(X(t) - \mu)(X(t+u_1) - \mu)(X(t+u_2) - \mu)] &= \\ &= \mu_3 \int_0^{\infty} h(v) h(v+u_1) h(v+u_2) dv \quad (\text{П5.2.8}) \end{aligned}$$

и

$$\begin{aligned} \text{Cov}[(X(t) - \mu)(X(t+u_1) - \mu), (X(v) - \mu)(X(v+u_2) - \mu)] &= \\ &= \gamma_{XX}(v-t) \gamma_{XX}(v-t+u_2-u_1) + \\ &+ \gamma_{XX}(v-t+u_2) \gamma_{XX}(v-t-u_1) + \\ &+ K_4(Z) \int_0^{\infty} h(v') h(v'+u_1) h(v'+v-t) \times \\ &\times h(v'+v-t+u_2) dv', \quad (\text{П5.2.9}) \end{aligned}$$

где $\gamma_{XX}(u)$ дается формулой (П5.2.7). Формула (П5.2.9) использована в разд. 5.3.3 при выводе выражения для ковариации оценок ковариационной функции.

Отметим, что из (П5.2.9) следует, что четвертый кумулянт процесса $X(t)$ равен четвертому кумулянту процесса $Z(t)$, умноженному на интеграл от произведения четырех весовых функций, т. е.

$$K_4(X) = K_4(Z) \int_0^{\infty} h(v) h(v+u_1) h(v+u_2) h(v+u_3) dv.$$

Для нормального белого шума $K_4(Z)$ тождественно равен нулю, и, следовательно, $K_4(X)$ также является тождественным нулем. Для негауссовского белого шума интеграл

$$\int_0^{\infty} h(v) h(v+u_1) h(v+u_2) h(v+u_3) dv,$$

вообще говоря, мал по сравнению с интегралами вида $\int_0^{\infty} h(v) h(v+u_1) dv$, и поэтому кумулянтным членом в (П5.2.9) можно пренебречь по сравнению с членами, содержащими γ_{XX} . Это приближение используется при выводе моментов оценок выборочной ковариационной функции в разд. 5.3.3.

ПРИЛОЖЕНИЕ П5.3

ЛОГИЧЕСКАЯ СХЕМА ПРОГРАММЫ ВЫЧИСЛЕНИЯ
КОВАРИАЦИЙ

Ниже приводится логическая схема вычислительной программы, предназначенной для обработки NS рядов, каждый из которых состоит из N точек *). Программа вычисляет NS автоковариаций и NS(NS — 1) взаимных ковариаций, причем максимальное запаздывание, до которого производятся вычисления, равно MAXM. Программа также считает приближенные авто- и взаимные ковариации для первых разностей от входных данных. Выход состоит из печати всех ковариаций и разностных ковариаций, графиков всех корреляций, над которыми построены графики разностных корреляций, и записанных на магнитную ленту или на перфокарты значений всех ковариаций и разностных ковариаций для последующего использования в спектральных программах. Печатный выход используется главным образом в качестве повторного контроля, когда ковариации являются входом для следующей программы.

Программа MULTICOR

1. Считывать NS, N, MAXM.
2. Считывать IDENT(J), X(I, J), I = 1, M, J = 1, NS.
3. Вычислить средние значения

$$XM(J) = \frac{1}{N} \sum_{I=1}^N X(I, J), \quad J = 1, NS.$$

4. Запомнить отклонения

$$X(I, J) - XM(J) \rightarrow X(I, J).$$

5. Вычислить ковариации

$$COV(K, J, L) = \frac{1}{N} \sum_{I=1}^{N-k} X(I, J) * X(I+K, L),$$

$$K=0, MAXM+1; \quad J=1, NS; \quad L=1, NS.$$

6. Вычислить разностные ковариации

$$DCOV(K, J, L) = -COV(K-1, J, L) + 2 * COV(K, J, L) -$$

$$- COV(K+1, J, L).$$

$$K=0, MAXM; \quad J=1, NS; \quad L=1, NS.$$

) Авторы используют в приложении П5.3 некоторые стандартные обозначения и символы, употребляемые в международном алгоритмическом языке ФОРТРАН. Подробнее о ФОРТРАНе см. [11]. — Прим. перев.

7. Вывести ковариации и разностные ковариации с помощью печатающего устройства и перфоратора или магнитофона.

8. Вычислить корреляции

$$\text{COR}(K, J, L) = \text{COV}(K, J, L) / \sqrt{\text{COV}(0, J, J) * \text{COV}(0, L, L)}$$

$$\text{DCOR}(K, J, L) = \text{DCOV}(K, J, L) / \sqrt{\text{DCOV}(0, J, J) * \text{DCOV}(0, L, L)}$$

$$K = 0, \text{ MAXM}; \quad J = 1, \text{ NS}; \quad L = 1, \text{ NS}.$$

9. Построить графики корреляций, а над ними графики разностных корреляций.

В гл. 5 было показано, что стационарный случайный процесс просто описывается с помощью ковариационной функции. Точно такое же описание дается его спектром мощности, который является преобразованием Фурье ковариационной функции. Спектр мощности показывает, как дисперсия случайного процесса распределена по частоте.

В разд. 6.1 говорится о том, что классический анализ Фурье не применим к временным рядам. Так, оценка спектра, полученная по формулам анализа Фурье, а именно выборочный спектр, обладает тем нежелательным свойством, что ее дисперсия не уменьшается при увеличении длины временного ряда. Поэтому для временных рядов методы гл. 2 нужно видоизменить. В результате мы приходим в разд. 6.2 к такому определению спектра, которое подходит для случайных процессов. В этом разделе рассматриваются также спектры процессов авторегрессии и скользящего среднего.

В разд. 6.3 показано, что с помощью сглаживания выборочного спектра можно получить улучшенную оценку спектра. Чем сильнее сглаживание, тем меньше дисперсия этой оценки, однако при этом возрастает смещение, или систематическое искажение. Поэтому нужно выбирать некоторый компромисс между смещением и дисперсией.

В разд. 6.4 выводятся дальнейшие свойства сглаженных оценок, в том числе свойства, связанные с понятием ширины полосы частот. Показано также, что доверительные интервалы для каждой частоты легко получить, используя логарифм выборочной оценки спектра.

6.1. ВЫБОРОЧНЫЙ СПЕКТР

6.1.1. Применение методов Фурье к временным рядам

Анализ Фурье. В гл. 2 было показано, что дисперсию, или среднюю мощность, сигнала $x(t)$ на отрезке $-T/2 \leq t \leq T/2$ можно разложить на вклады от гармоник $f_m = m/T$ основной частоты $f_1 = 1/T$ согласно формуле

$$s_T^2 = \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt = \sum_{m=-\infty}^{\infty} |X_m|^2. \quad (6.1.1)$$

X_m называется комплексной амплитудой гармоники $f_m = m/T$. Она дает амплитуды синусоидального и косинусоидального членов сигнала $x(t)$ на частоте f_m . Комплексную амплитуду можно вычислить по формуле

$$X_m = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi mt/T} dt, \quad (6.1.2)$$

подставляя в нее

$$e^{-j2\pi mt/T} = \cos \frac{2\pi mt}{T} - j \sin \frac{2\pi mt}{T}.$$

Напомним, что разложение $x(t)$ в ряд Фурье имеет вид

$$x(t) = \sum_{m=-\infty}^{\infty} X_m e^{+j2\pi mt/T}.$$

Аналогично для дискретного сигнала, наблюдаемого в моменты времени $t = -n\Delta, -(n-1)\Delta, \dots, (n-1)\Delta$, среднюю мощность можно разложить на вклады *конечного* числа гармоник основной частоты $f_1 = 1/N\Delta$ ($N = 2n$), а равенства, соответствующие (6.1.1) и (6.1.2), имеют вид

$$s_T^2 = \frac{1}{N} \sum_{t=-n}^{n-1} x_t^2 = \sum_{m=-n}^{n-1} |X_m|^2, \quad (6.1.3)$$

$$X_m = \frac{1}{N} \sum_{t=-n}^{n-1} x_t e^{-j2\pi mt \Delta/N\Delta} = \frac{1}{N} \sum_{t=-n}^{n-1} x_t e^{-j2\pi mt/N}. \quad (6.1.4)$$

Вклад $|X_m|^2$ в среднюю мощность на частоте f_m называется *интенсивностью* сигнала на этой частоте, а график величин $|X_m|^2$ в зависимости от m называется *линейчатым спектром Фурье*. Пример такого спектра приведен на рис. 2.2.

Спектр мощности детерминированных сигналов. Главное различие в анализе детерминированных и случайных сигналов выявляется как раз тогда, когда длина записи неограниченно возрастает. Во многих технических учебниках это различие не объясняется, а используются рассуждения следующего характера. Из (6.1.1) дисперсия бесконечной записи равна

$$\sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt = \lim_{T \rightarrow \infty} \sum_{m=-\infty}^{\infty} (T |X_m|^2) \frac{1}{T} = \int_{-\infty}^{\infty} \Gamma(f) df,$$

где функция

$$\Gamma(f) = \lim_{T \rightarrow \infty} T |X_m|^2 \quad (6.1.5)$$

называется «спектром мощности» Фурье. Воспользовавшись формулой (6.1.2), функцию $T|X_m|^2$ можно записать в виде

$$T|X_m|^2 = C_{xx}(f) = \frac{1}{T} \left| \int_{-T/2}^{T/2} x(t) e^{-j2\pi ft} dt \right|^2. \quad (6.1.6)$$

Отметим, что функция $C_{xx}(f)$ определена на непрерывном интервале частот $-\infty \leq f \leq \infty$. Она называется *выборочным спектром*, или *выборочной спектральной плотностью* *). Для дискретного случая выборочный спектр равен

$$\begin{aligned} C_{xx}(f) &= \frac{\Delta}{N} \left| \sum_{t=-n}^{n-1} x_t e^{-j2\pi ft\Delta} \right|^2 = \\ &= \frac{\Delta}{N} \left[\left(\sum_{t=-n}^{n-1} x_t \cos 2\pi ft\Delta \right)^2 + \left(\sum_{t=-n}^{n-1} x_t \sin 2\pi ft\Delta \right)^2 \right], \\ &\qquad\qquad\qquad \frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.1.7) \end{aligned}$$

Частота $1/2\Delta$ в (6.1.7) называется *наиквистовой*. Мы обсуждали ее в гл. 2; это — наивысшая из частот, которую можно обнаружить по данным, отсчитываемым через Δ секунд.

Заметим, что если преобразование Фурье сигнала $x(t)$ является регулярной функцией, то предел (6.1.5) для $\Gamma(f)$ равен нулю. Это происходит потому, что если преобразование Фурье функции $x(t)$ существует, то сама она должна стремиться к нулю при $t \rightarrow \pm\infty$. Однако если $x(t)$ не затухает на бесконечности, то функция $C_{xx}(f)$ будет обычно стремиться к вполне определенному пределу $\Gamma(f)$. Для детерминированных сигналов $C_{xx}(f)$ сходится к $\Gamma(f)$ плавно в том смысле, что функция $C'_{xx}(f)$, полученная при увеличении длины записи от T до T' , является сглаженным вариантом функции $C_{xx}(f)$, вычисленной по записи длины T .

В следующем разделе будет показано, что определение (6.1.5) не подходит для случая, когда $x(t)$ является реализацией случайного процесса. Основное различие в анализе Фурье детерминированных и случайных сигналов состоит в том, что во втором случае при увеличении длины записи от T до $T' > T$ функция $C'_{xx}(f)$ не становится более устойчивой, т. е. $C_{xx}(f)$ не сходится в каком-либо статистическом смысле к предельному значению при $T \rightarrow \infty$.

*) В оригинале *sample spectrum*. Более точно было бы называть функцию $C_{xx}(f)$ *выборочной спектральной плотностью*, однако ради краткости мы будем использовать и термин «выборочный спектр» там, где это не приводит к неясности. Для дискретного времени выборочный спектр (6.1.7) часто называют *периодограммой*. — *Прим. перев.*

6.1.2. Выборочный спектр белого шума

Чтобы проиллюстрировать результат применения анализа Фурье к случайному процессу, был взят ряд из 400 случайных нормальных чисел (гауссовский белый шум). Выборочный спектр $C_{zz}(t)$ вычислялся для четырех рядов, состоящих из первых 50, 100, 200 и 400 членов соответственно. На рис. 6.1 приведены значения выборочных спектров $C_{zz}(f)$, сосчитанные по формуле (6.1.7), на частотах $f = 0,02; 0,04, \dots, 0,50$ гц для случаев $N = 50$ и $N = 100$ при $\Delta = 1$ сек. На рисунке изображен также теоретический спектр, который, как показано в разд. 6.2.3, равен константе в интервале $-1/2 \leq f < 1/2$.

Как видно из рис. 6.1, функции $C_{zz}(f)$ сильно колеблются, и на основании этих графиков трудно предположить, что истинный спектр равен константе, т. е. что временной ряд является белым шумом. Отметим также, что отклонения $C_{zz}(f)$ от истинного спектра для $N = 100$ такие же, как и для $N = 50$, что указывает на отсутствие статистической сходимости какого-либо типа.

В табл. 6.1 представлены характеристики, полученные из выборочных спектров, сосчитанных по 50, 100, 200 и 400 членам. Поскольку теоретический спектр равен константе, флуктуации $C_{zz}(f)$ можно охарактеризовать, сосчитав среднее значение, дисперсию и среднеквадратичную ошибку величин $C_{zz}(f)$ при изменении частоты. Видно, что для каждого из рядов среднее значение близко к единице — теоретическому спектру. Следовательно, значения $C_{zz}(f)$ группируются около некоторой центральной величины. Однако, как видно из табл. 6.1, дисперсии не уменьшаются с ростом N , что говорит о том, что выборочные оценки спектра, сосчитанные по 100, 200 или 400 членам, не лучше оценки, сосчитанной по 50 членам.

Таблица 6.1

Поведение выборочных спектров белого шума по мере возрастания длины записи

N	50	100	200	400
Среднее	0,85	1,07	1,00	0,95
Дисперсия	0,630	0,777	0,886	0,826
Среднеквадратичная ошибка	0,652	0,782	0,886	0,828

В гл. 4 мы видели, что хорошие оценки обладают тем свойством, что их дисперсия убывает с ростом N . Отсюда можно заклю-

чить, что $C_{zz}(f)$ не является хорошей выборочной оценкой спектра, по крайней мере в том виде, в каком она здесь приведена.

Чтобы показать, что выборочный спектр не сходится в каком-либо статистическом смысле и для процессов, отличных от белого

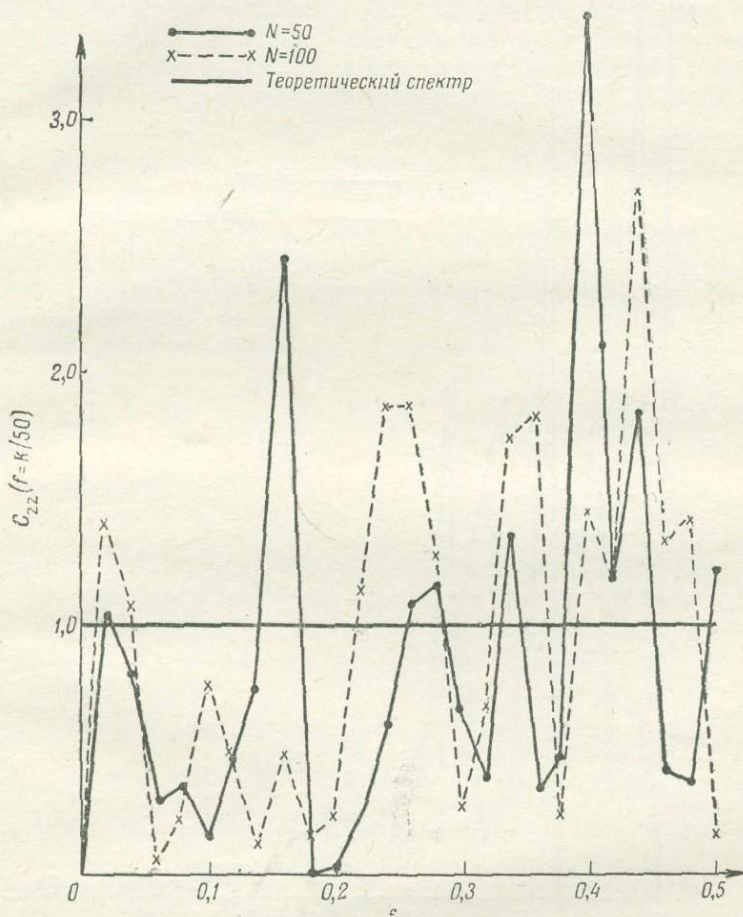


Рис. 6.1. Выборочные спектры для первой половины ($N=50$) и для всей реализации ($N=100$) дискретного нормального белого шума.

шума, рассмотрим процесс авторегрессии, построенный по формуле (5.3.36). Теоретическая корреляционная функция и соответствующая выборочная функция, сосчитанные по реализации из 400 членов, показаны на рис. 5.13. Теоретический спектр и выборочный спектр, сосчитанные по той же самой реализации, приведены на

рис. 6.2. Как и в примере с белым шумом, выборочный спектр очень сильно колеблется и мало похож на теоретический.

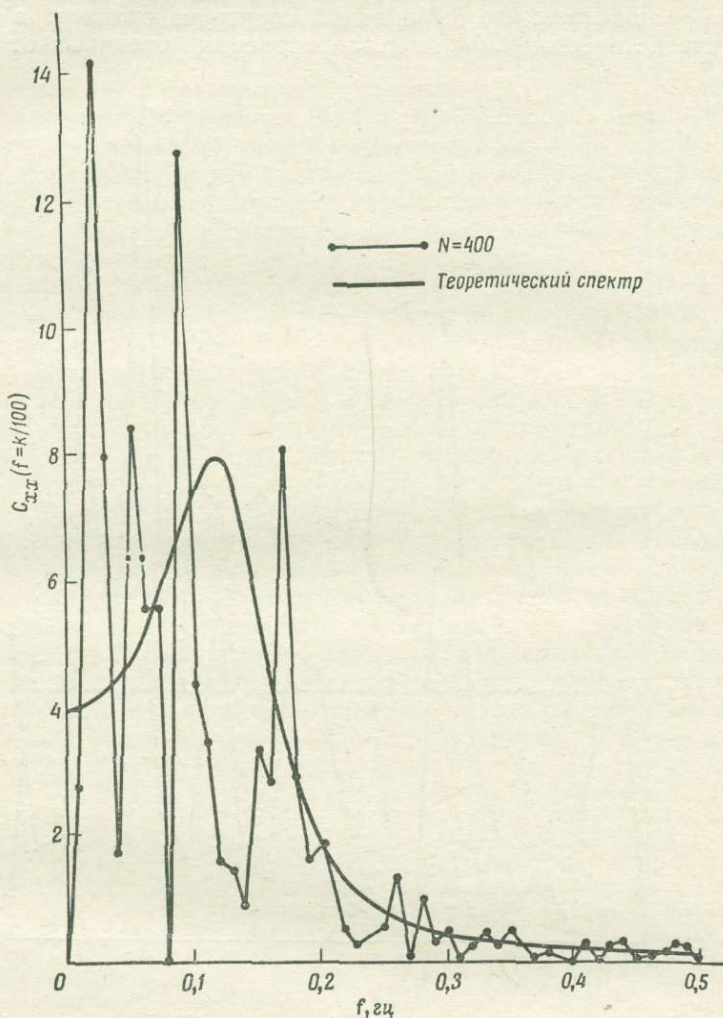


Рис. 6.2. Выборочный спектр для реализации процесса авторегрессии второго порядка.

Резюме. Для детерминированных сигналов спектр является пределом (в обычном математическом смысле) выборочного спектра $C_{xx}(f)$ при безграничном увеличении длины записи. Однако, как показывает пример с белым шумом, поведение функции

$C_{xx}(f)$ для временного ряда является настолько неустойчивым, что она становится бесполезной для оценивания. Основная причина, по которой анализ Фурье неприменим к временным рядам, заключается в том, что он основан на предположении, что амплитуды, частоты и фазы *фиксированы*. Для временных же рядов характерны случайные изменения амплитуд, частот и фаз. Поэтому тот вывод, что анализ Фурье для временных рядов следует видоизменить, учитывая их случайную природу, не является неожиданным.

6.1.3. Соотношение между выборочной спектральной плотностью и выборочной ковариационной функцией

Прежде чем дать более точное определение спектра стационарного случайного процесса, мы выведем фундаментальное соотношение, связывающее выборочный спектр и выборочную ковариационную функцию.

Из определения выборочного спектра (6.1.6) мы имеем

$$C_{xx}(f) = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi ft} dt \int_{-T/2}^{T/2} x(t') e^{+j2\pi ft'} dt'. \quad (6.1.8)$$

При замене переменных

$$u = t - t', \quad v = t'$$

в двойном интеграле (6.1.8) область интегрирования преобразуется так, как показано на рис. 6.3. При этом (6.1.8) переходит в

$$C_{xx}(f) = \int_0^T \left[\frac{1}{T} \int_{-T/2}^{(T/2)-u} x(v) x(v+u) dv \right] e^{-j2\pi fu} du + \\ + \int_{-T}^0 \left[\frac{1}{T} \int_{-(T/2)-u}^{T/2} x(v) x(v+u) dv \right] e^{-j2\pi fu} du.$$

Вводя функцию $c_{xx}(u)$, определенную равенством (5.3.5), мы получаем

$$C_{xx}(f) = \int_{-T}^T c_{xx}(u) e^{-j2\pi fu} du, \quad -\infty \leq f \leq \infty. \quad (6.1.9)$$

Следовательно, *выборочный спектр, или выборочная спектральная плотность, является преобразованием Фурье от выборочной ковариационной функции*. Обратное по отношению к (6.1.9) преобразование Фурье можно записать в виде

$$c_{xx}(u) = \int_{-\infty}^{\infty} C_{xx}(f) e^{j2\pi fu} df, \quad -T \leq u \leq T, \quad (6.1.10)$$

откуда при $u = 0$ получаем

$$c_{xx}(0) = s_T^2 = \int_{-\infty}^{\infty} C_{xx}(f) df. \quad (6.1.11)$$

Таким образом, выборочная спектральная плотность показывает, как дисперсия, или средняя мощность, сигнала $x(t)$ распределена по частотам.

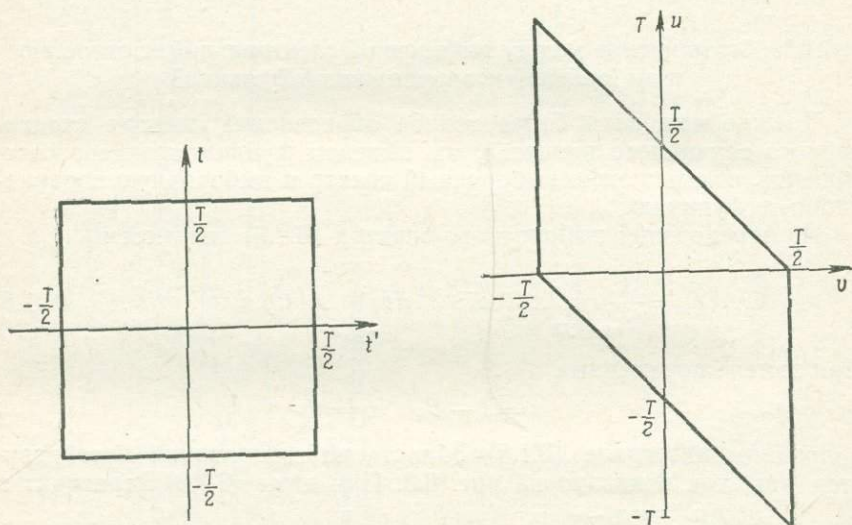


Рис. 6.3. Преобразование координат для выборочного спектра.

Для дискретного времени выборочный спектр равен

$$C_{xx}(f) = \Delta \sum_{k=-(N-1)}^{N-1} c_{xx}(k) e^{-j2\pi f k \Delta}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.1.12)$$

что соответствует формуле (6.1.9). Обратное преобразование (6.1.12) дает

$$c_{xx}(u) = \int_{-1/2\Delta}^{1/2\Delta} C_{xx}(f) e^{j2\pi f u} df, \quad -N\Delta \leq u \leq N\Delta, \quad (6.1.13)$$

что соответствует формуле (6.1.10).

Пары преобразований Фурье (6.1.9), (6.1.10) и (6.1.12), (6.1.13) являются математическими тождествами, которые верны независимо от того, является ли $x(t)$ детерминированным сигналом или реализацией случайного процесса. В следующем разделе дается интерпретация предельного значения $C_{xx}(f)$ для случая, когда $x(t)$ — реализация стационарного случайного процесса.

6.2. СПЕКТР

6.2.1. Определение спектра случайного процесса

Для описания изменчивости функции $C_{xx}(f)$, продемонстрированной в разд. 6.1.2, необходимо рассмотреть запись $x(t)$, $-T/2 \leq t \leq T/2$, как один из многих возможных временных рядов, которые могли бы быть наблюдаемы, т. е. как реализацию случайного процесса. Таким образом, изменчивость записи будет охарактеризована случайными величинами $X(t)$, $-T/2 \leq t \leq T/2$, как указывалось в гл. 5. При этом выборочная спектральная плотность $C_{xx}(f)$ в некоторой точке f рассматривается как реализация случайной величины $C_{XX}(f)$, точно так же, как $c_{xx}(u)$ считается реализацией случайной величины $c_{XX}(u)$ *). Получив распределение $C_{XX}(f)$ или ее моменты, можно объяснить неустойчивое поведение $C_{xx}(f)$, показанное на рис. 6.1 и 6.2.

Используя (6.1.9), получаем первый момент оценки, соответствующей выборочному спектру $C_{xx}(f)$:

$$E [C_{XX}(f)] = \int_{-T}^T E [c_{XX}(u)] e^{-j2\pi fu} du,$$

что можно с помощью (5.3.13) записать в виде

$$E [C_{XX}(f)] = \int_{-T}^T \gamma_{XX}(u) \left(1 - \frac{|u|}{T}\right) e^{-j2\pi fu} du. \quad (6.2.1)$$

Таким образом, (6.2.1) дает среднее распределение (по всем возможным временным рядам длины T) мощности по частотам. При увеличении длины записи T первый момент $E [C_{XX}(f)]$ стремится к

$$\Gamma_{XX}(f) = \lim_{T \rightarrow \infty} E [C_{XX}(f)] = \int_{-\infty}^{\infty} \gamma_{XX}(u) e^{-j2\pi fu} du. \quad (6.2.2)$$

Математические вопросы, связанные с этим предельным переходом, более полно обсуждаются в [1].

*) Случайная величина $C_{XX}(f)$ определяется равенством

$$C_{XX}(f) = \frac{1}{T} \left| \int_{-T/2}^{T/2} X(t) e^{-j2\pi ft} dt \right|^2 = \int_{-T}^T c_{XX}(u) e^{-j2\pi fu} du, \quad -\infty \leq f \leq \infty.$$

Авторы используют для $C_{XX}(f)$ термин *sample spectrum estimator*, который мы будем переводить как *оценка, соответствующая выборочному спектру*. — Прим. перев.

Функция $\Gamma_{XX}(f)$ называется *спектральной плотностью* *). Равенство (6.2.2) показывает, что спектральная плотность является преобразованием Фурье от ковариационной функции процесса $X(t)$. Пользуясь табл. 2.3, получаем обратное преобразование

$$\gamma_{XX}(u) = \int_{-\infty}^{\infty} \Gamma_{XX}(f) e^{j2\pi fu} df. \quad (6.2.3)$$

Положив $u = 0$ в (6.2.3), получаем

$$\gamma_{XX}(0) = \sigma_X^2 = \int_{-\infty}^{\infty} \Gamma_{XX}(f) df. \quad (6.2.4)$$

Следовательно, $\Gamma_{XX}(f)$ показывает, как дисперсия процесса $X(t)$ распределена по частотам аналогично тому, как (6.1.9) показывает, как распределена по частотам средняя мощность одной конкретной реализации длины T . В частности, вклад в дисперсию процесса $X(t)$, который вносят частоты в интервале от f до $f+df$, равен приблизительно $\Gamma_{XX}(f)df$. Отметим, что, согласно определению (6.1.6), $\Gamma_{XX}(f)$ является неотрицательной для всех f .

Для дискретного времени соотношения, соответствующие (6.2.1) — (6.2.3), имеют вид

$$E[C_{XX}(f)] = \Delta \sum_{k=-(N-1)}^{N-1} \gamma_{XX}(k) \left(1 - \frac{|k|}{N}\right) e^{-j2\pi fk\Delta}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.2.5)$$

$$\Gamma_{XX}(f) = \lim_{N \rightarrow \infty} E[C_{XX}(f)] = \Delta \sum_{k=-\infty}^{\infty} \gamma_{XX}(k) e^{-j2\pi kf\Delta}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.2.6)$$

и

$$\gamma_{XX}(k) = \int_{-1/(2\Delta)}^{1/(2\Delta)} \Gamma_{XX}(f) e^{j2\pi fk\Delta} df, \quad k = 0, \pm 1, \pm 2, \dots \quad (6.2.7)$$

Некоторые примеры. Для выяснения вопроса о том, какую информацию содержат спектры, на рис. 6.4 и 6.5 показаны теоретические спектры (спектральные плотности) процессов авторегрессии

*) В оригинале power spectrum. Наряду с термином «спектральная плотность» мы будем также использовать для функции $\Gamma_{XX}(f)$ (там, где это не приводит к неясностям) более короткие названия «спектр» или «спектр мощности». — Прим. перев.

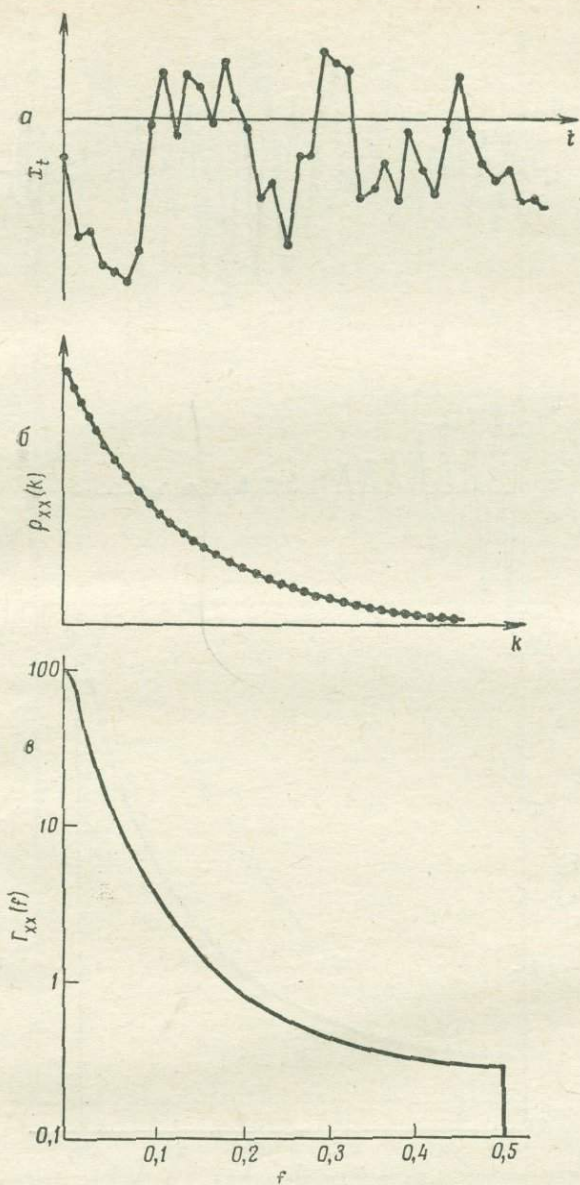


Рис. 6.4. а — реализация, б — автокорреляционная функция и в — спектр дискретного процесса авторегрессии первого порядка ($\alpha_1 = +0,9$).

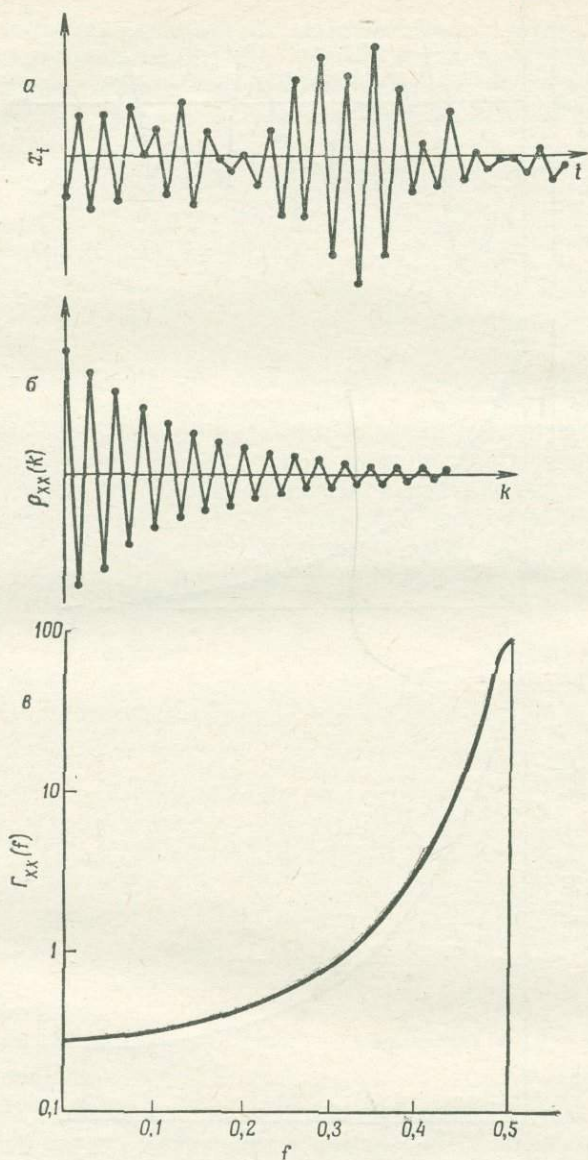


Рис. 6.5. *a* — реализация, *б* — автокорреляционная функция и *в* — спектр дискретного процесса авторегрессии первого порядка ($\alpha_1 = -0,9$).

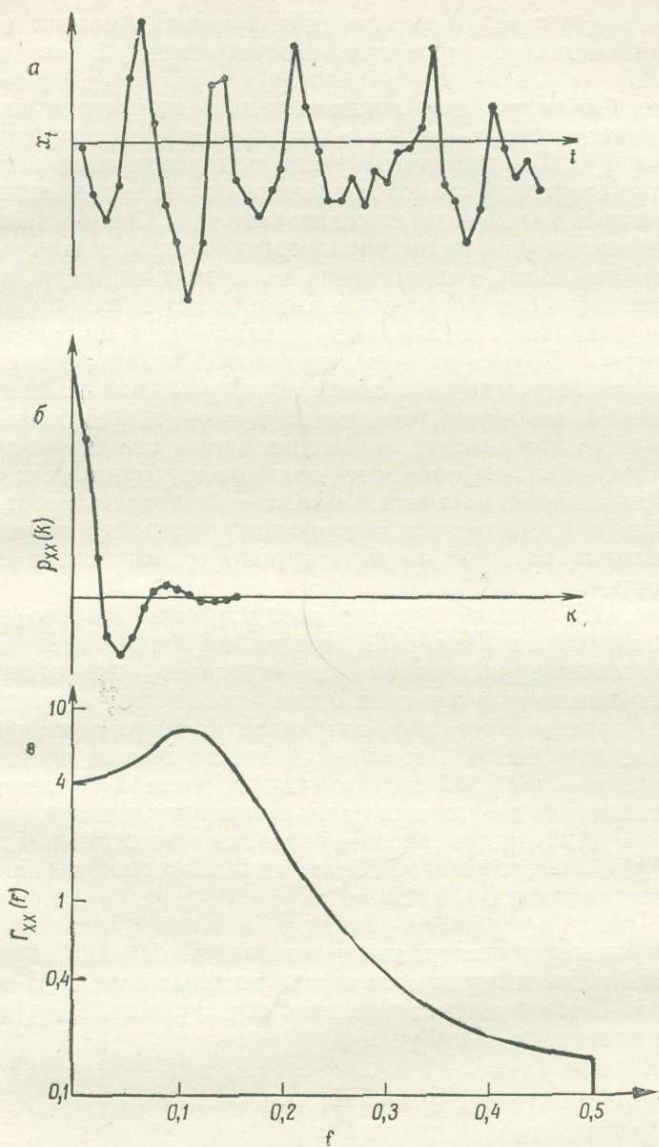


Рис. 6.6. а — реализация, б — автокорреляционная функция и в — спектр дискретного процесса авторегрессии второго порядка ($\alpha_1=1,0$; $\alpha_2=-0,5$).

первого порядка и их корреляционные функции. Аналитическое выражение для спектра процесса авторегрессии будет получено в разд. 6.2.5.

Из рис. 6.4 видно, что, когда параметр авторегрессии $\alpha_1 = 0,9$, ряд изменяется плавно, и это находит отражение в том, что корреляционная функция плавно затухает при увеличении запаздывания. Соответствующий спектр принимает большие значения на низких частотах и малые — на высоких частотах. Следовательно, для плавно изменяющихся рядов характерны спектры, у которых большая часть мощности сосредоточена на низких частотах. Заметим, что на рис. 6.4, 6.5 и 6.6 спектры изображены в *логарифмическом масштабе*, детальнее показывающем их в более широком диапазоне амплитуд. Другая причина, по которой спектр лучше изображать в логарифмическом масштабе, будет указана позднее.

На рис. 6.5 мы видим, что, когда $\alpha_1 = -0,9$, ряд очень быстро осциллирует, и это находит отражение в том, что корреляционная функция меняет знак. Соответствующий спектр принимает большие значения на высоких частотах и малые — на низких частотах. Следовательно, для быстро осциллирующих рядов характерны спектры, у которых большая часть мощности сосредоточена на высоких частотах.

На рис. 6.6 показан процесс авторегрессии второго порядка. Как указывалось в разд. 5.2.4, соответствующий временной ряд является квазипериодическим со «средним» периодом около 8 сек. Корреляционная функция отражает это периодическое поведение; она представляет собой затухающую синусоидальную волну с периодом 8 сек. Соответствующий этому случаю спектр имеет пик на частоте $f_0 = 0,125$ гц. Так как процесс $X(t)$ не является точно периодическим, его спектр не сосредоточен на единственной частоте $f_0 = 0,125$ гц, но рассеян по всем частотам в диапазоне $-0,5 \leq f \leq 0,5$ гц. Впрочем, большая часть мощности сосредоточена вблизи частоты $f_0 = 0,125$ гц.

Нормированная спектральная плотность. Иногда приходится сравнивать временные ряды, значения которых измерены в разных масштабах. В таких случаях полезно нормировать $\Gamma_{XX}(f)$, разделив ее на дисперсию σ_X^2 . Функция

$$\frac{\Gamma_{XX}(f)}{\sigma_X^2}$$

называется *нормированной спектральной плотностью* *). Из (6.2.2)

*) В оригинале spectral density function (спектральная плотность). В нашей литературе спектральной плотностью называют функцию $\Gamma_{XX}(f)$. Поэтому мы будем называть функцию $\Gamma_{XX}(f)\sigma_X^2$ нормированной спектральной плотностью или часто, ради краткости, *нормированным спектром*. — Прим. перев.

получаем, что
$$\frac{\Gamma_{XX}(f)}{\sigma_X^2} = \int_{-\infty}^{\infty} \rho_{XX}(u) e^{-j2\pi fu} du, \quad (6.2.8)$$

так что нормированная спектральная плотность является преобразованием Фурье от корреляционной функции.

Далее, нормированный спектр, будучи пределом неотрицательных функций, сам является неотрицательной функцией. Так как интеграл от нормированного спектра равен единице, то с математической точки зрения он обладает теми же свойствами (3.1.8), что и плотность вероятности. В разд. 6.3 будет показано, что аналогия между нормированным спектром и плотностью вероятности распространяется и на оценивание этих двух функций по записям конечной длины.

Использованный в этом разделе способ определения спектра не является единственно возможным. Другой способ, основанный на собственных значениях ковариационной матрицы случайного процесса, приводится в разд. 11.1.2.

Замечания относительно определений спектра, используемых в технических работах. В разд. 6.1.1 мы уже сделали несколько критических замечаний по поводу определения спектральной плотности в виде

$$\Gamma(f) = \lim_{T \rightarrow \infty} C_{xx}(f),$$

которое обычно приводится в учебниках по электротехнике (см., например, [2, 3]). Возражение против такого определения состоит в том, что если $x(t)$ — реализация стационарного случайного процесса, то соответствующая случайная величина $C_{XX}(f)$ не сходится ни в каком статистическом смысле к предельному значению.

Дальнейшая путаница проистекает из-за неправильного использования фундаментального равенства (6.1.9), доказанного выше. Из того, что выборочная ковариационная функция $c_{xx}(u)$ сходится при $T \rightarrow \infty$ во вполне определенном статистическом смысле к $\gamma_{XX}(u)$, делается неправильный вывод, что допустима перестановка интегрирования и перехода к пределу

$$\lim_{T \rightarrow \infty} C_{xx}(f) = \int_{-T}^T \lim_{T \rightarrow \infty} c_{xx}(u) e^{-j2\pi fu} du = \int_{-\infty}^{\infty} \gamma_{XX}(u) e^{-j2\pi fu} du = \Gamma_{XX}(f).$$

В разд. 5.3.3 было показано, что среднеквадратичная ошибка оценки ковариационной функции $c_{XX}(u)$ имеет порядок $1/T$, и поэтому ее распределение концентрируется все теснее около $\gamma_{XX}(u)$ при $T \rightarrow \infty$. Таким образом, $c_{XX}(u)$ является *состоятельной оценкой* $\gamma_{XX}(u)$. Другими словами, *средняя по времени* величина $c_{xx}(u)$ сходится к *средней по ансамблю* величине $\gamma_{XX}(u)$. Это

свойство обычно называют эргодическим. Для его выполнения требуется, чтобы $\gamma_{xx}(u)$ убывала достаточно быстро.

Однако из того, что эргодическое свойство имеет место для $c_{xx}(u)$, никоим образом не следует, что оно справедливо для его преобразования Фурье $C_{xx}(f)$. В самом деле, если имеется состоятельная оценка статистического параметра, то ее преобразование Фурье обычно не является состоятельной оценкой для преобразования Фурье этого параметра*). Иначе говоря, $C_{xx}(f)$ является примером выборочной функции, для которой эргодическое свойство не имеет места.

Интуиция подсказывает, что в такой ситуации интересно посмотреть, что происходит с функцией $c_{xx}(u)$ при фиксированном запаздывании u , когда длина записи T возрастает. В этом случае $c_{xx}(u)$ собирает в себе все больше и больше информации в виде произведений $x(t)x(t+u)$, и, следовательно, информация, содержащаяся в $c_{xx}(u)$ относительно $\gamma_{xx}(u)$, неограниченно возрастает при $T \rightarrow \infty$. Позднее мы увидим, что информация, содержащаяся в $C_{xx}(f)$ относительно $\Gamma_{xx}(f)$, рассеяна в полосе частот $f \pm 1/T$. При увеличении T полная информация, содержащаяся в $C_{xx}(f)$, распределяется по полосам частот, число которых увеличивается, а ширина стремится к нулю. Точный результат состоит в том, что при увеличении T можно оценивать среднюю мощность в полосе частот, ширина которой безгранично уменьшается; однако эффективность выборочной оценки мощности в этой сужающейся полосе не улучшается.

6.2.2. Интеграл от спектральной плотности (спектральная функция)

Случай, когда спектральной плотности не существует. Ранее спектральная плотность была определена с помощью предела

$$\lim_{T \rightarrow \infty} E [C_{xx}(f)] = \lim_{T \rightarrow \infty} \int_{-T}^T \gamma_{xx}(u) \left(1 - \frac{|u|}{T}\right) e^{-j2\pi fu} du,$$

при условии, что этот предел существует. Чтобы $\Gamma_{xx}(f)$ была конечной, достаточно выполнения неравенства

$$|\Gamma_{xx}(f)| = \left| \int_{-\infty}^{\infty} \gamma_{xx}(u) e^{-j2\pi fu} du \right| \leq \int_{-\infty}^{\infty} |\gamma_{xx}(u)| du \leq M, \quad (6.2.9)$$

где M — конечная константа. Следовательно, достаточное (но не необходимое) условие существования конечной спектральной плотности состоит в том, что $\gamma_{xx}(u)$ убывает достаточно быстро при $u \rightarrow \infty$, так что интеграл (6.2.9) сходится.

* Подразумевается, что как оценка, так и статистический параметр зависят от времени t , так что от них можно брать преобразование Фурье. — Прим. перев.

В качестве примера случайного процесса, для которого это условие не выполнено, рассмотрим процесс

$$X(t) = A \cos 2\pi f_0 t + B \sin 2\pi f_0 t = R \cos(2\pi f_0 t + \varphi), \quad (6.2.10)$$

где A и B — независимые случайные величины с нулевым средним значением и дисперсией σ^2 . Каждая реализация $x(t)$ является косинусоидальной волной $R \cos(2\pi f_0 t + \varphi)$, имеющей постоянную амплитуду, частоту и фазу. Но при переходе от одного члена ансамбля к другому амплитуда и фаза изменяются случайным образом, в то время как частота остается фиксированной. Из (6.2.10) получаем

$$E [X(t)] = E [A] \cos 2\pi f_0 t + E [B] \sin 2\pi f_0 t = 0.$$

Следовательно,

$$\begin{aligned} \gamma_{XX}(u) &= E [X(t) X(t+u)] = \\ &= E [(A \cos 2\pi f_0 t + B \sin 2\pi f_0 t)(A \cos 2\pi f_0(t+u) + \\ &\quad + B \sin 2\pi f_0(t+u))] = \\ &= \sigma^2 [\cos 2\pi f_0 t \cos 2\pi f_0(t+u) + \sin 2\pi f_0 t \sin 2\pi f_0(t+u)] = \\ &= \sigma^2 \cos 2\pi f_0 u. \end{aligned}$$

Функция $\gamma_{XX}(u)$ не стремится к нулю при $u \rightarrow \infty$, так что для нее интеграл (6.2.9) расходится. Однако можно определить спектральную плотность через δ -функции, используя (2.2.12):

$$\Gamma_{XX}(f) = \frac{\sigma^2}{2} [\delta(f - f_0) + \delta(f + f_0)].$$

Следовательно, спектральную плотность случайного процесса (6.2.10) можно считать равной двум δ -функциям, имеющим площадь $\sigma^2/2$ и сосредоточенным на частотах $f = \pm f_0$.

Интеграл от спектральной плотности (спектральная функция). Даже в том случае, когда спектральная плотность содержит δ -функции, имеет смысл говорить о дисперсии процесса, в котором оставлены только частоты, не превосходящие некоторой частоты f' . Эту дисперсию формально можно получить, интегрируя спектральную плотность. Так, интегрируя (6.2.2) от $f = -f'$ до $f = f'$, мы получаем *спектральную функцию* *).

$$I_{XX}(f') = \int_{-f'}^{f'} \Gamma_{XX}(f) df = \int_{-\infty}^{\infty} \gamma_{XX}(u) \frac{\sin 2\pi f' u}{\pi u} du, \quad 0 \leq f' \leq \infty. \quad (6.2.11)$$

*) Обычно в определении спектральной функции нижний предел интегрирования берут равным нулю. — Прим. перев.

Эта функция похожа на функцию распределения, так же как нормированный спектр похож на плотность вероятности. Таким образом, мы имеем

$$\begin{aligned} I(0) &= 0, \\ I(\infty) &= \sigma^2 \end{aligned}$$

и

$$I(f_1) \leq I(f_2)$$

при $f_1 \leq f_2$. Если спектральная плотность содержит δ -функцию на частоте $f = f_0$, т. е. $\gamma_{XX}(u)$ содержит компоненту $k \cos 2\pi f_0 u$, то спектральная функция имеет скачок величины k на частоте f_0 .

Для дискретного времени спектральная функция имеет вид

$$\begin{aligned} I_{XX}(f') &= \int_{-f'}^{f'} \Gamma_{XX}(f) df = \Delta \sum_{k=-\infty}^{\infty} \gamma_{XX}(k) \frac{\sin 2\pi k f' \Delta}{\pi k \Delta}, \\ 0 &\leq f' < \frac{1}{2\Delta}. \end{aligned} \quad (6.2.12)$$

6.2.3. Спектр белого шума

В разд. 5.2.1 чисто случайный процесс, или белый шум, $Z(t)$ был определен как процесс, имеющий ковариационную функцию $\gamma_{ZZ}(u) = \sigma_Z^2 \delta(u)$. Этот процесс имеет бесконечную дисперсию и поэтому не может быть случайным процессом в обычном смысле. Однако мы показали, что его можно рассматривать как предел при $\tau \rightarrow 0$ процесса Башелье—Винера $Y(t)$, имеющего ковариационную функцию

$$\gamma_{YY}(u) = \begin{cases} 0, & |u| > \tau, \\ \frac{\sigma_Z^2}{\tau} \left(1 - \frac{|u|}{\tau}\right), & |u| \leq \tau. \end{cases}$$

Отсюда, согласно определению (6.2.2), процесс Башелье—Винера имеет спектральную плотность

$$\Gamma_{YY}(f) = \int_{-\tau}^{\tau} \frac{\sigma_Z^2}{\tau} \left(1 - \frac{|u|}{\tau}\right) e^{-j2\pi f u} du = \sigma_Z^2 \left(\frac{\sin \pi f \tau}{\pi f \tau}\right)^2, \quad -\infty \leq f \leq \infty.$$

В пределе при $\tau \rightarrow 0$ функция $\Gamma_{YY}(f)$ стремится к константе для всех f :

$$\lim_{\tau \rightarrow 0} \Gamma_{YY}(f) = \Gamma_{ZZ}(f) = \sigma_Z^2. \quad (6.2.13)$$

Процесс $Z(t)$ называется *белым шумом* по аналогии с белым светом в оптике, содержащим все оптические частоты с приблизительно одинаковой интенсивностью. Строго говоря, белый шум

нельзя реализовать физически, так же как и единичный импульс, который можно рассматривать как математический аналог единичного импульса в технике.

Способы генерации белого шума. При определении белого шума для дискретного времени не возникает никаких трудностей, так как ковариационная функция дискретного белого шума Z_t равна

$$\gamma_{zz}(u) = \begin{cases} \sigma_z^2, & u = 0, \\ 0, & u = \pm\Delta, \pm 2\Delta, \pm 3\Delta, \dots \end{cases}$$

Используя (6.2.6), получаем

$$\Gamma_{zz}(f) = \sigma_z^2 \Delta, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.2.14)$$

так что все частоты в интервале $-1/2\Delta \leq f < 1/2\Delta$ несут одну и ту же мощность, или дисперсию.

Дискретный белый шум можно очень просто получить из непрерывного небелого шума. Предположим, например, что имеется источник непрерывного небелого шума, ковариационная функция которого равна нулю при $u > u_0$. Ясно, что если мы возьмем отсчеты процесса $X(t)$, отстоящие друг от друга на $\Delta > u_0$, то получим процесс Z_t с ковариационной функцией (6.2.14).

Частотная интерпретация этого метода генерации дискретного белого шума из непрерывного небелого шума состоит в следующем. Частота выборки $1/\Delta$ настолько мала, что простирается очень много наложений частот спектра $\Gamma_{xx}(f)$ (см. разд. 2.4.2). Поэтому спектр дискретного сигнала (отсчитываемого в дискретные моменты времени), равный сумме налагающихся участков $\Gamma_{xx}(f)$, будет становиться все более пологим, т. е. $\Gamma_{zz}(f)$ стремится к константе в интервале $-1/2\Delta \leq f \leq 1/2\Delta$. Этот процесс проиллюстрирован на рис. 2.11 для одного частного случая. Заметим, что, обсуждая вопросы, связанные с белым шумом, мы ничего не предполагали относительно плотности вероятности $Z(t)$. *Белый шум $Z(t)$ может иметь любую плотность вероятности.*

Строго говоря, белый шум нереализуем физически, но можно получить очень хорошее приближение к нему. Например, флуктуирующий ток в электронной лампе дает очень хорошее приближение, так как его спектр мощности по существу равен константе в интервале от 0 до 100 Мгц. Этот шум, называемый обычно *дробовым*, создается в результате случайной эмиссии электронов с катода лампы.

Другим физическим примером шума, являющегося приблизительно белым в широком диапазоне частот, служит *тепловой шум*. Этот шум представляет собой напряжение (или ток) в проводнике, обладающем сопротивлением R , вызванное тепловым движением электронов. Его спектр мощности почти постоянен в широком диапазоне частот и равен

$$\Gamma_{xx}(f) = 4RkT,$$

где T — абсолютная температура и k — постоянная Больцмана. Более детальное обсуждение дробового и теплового шумов можно найти в [2].

6.2.4. Спектр линейного процесса

Мы сейчас получим выражение для спектральной плотности выхода устойчивой линейной системы, на вход которой подается стационарный процесс. В том случае, когда на вход подается белый шум, выходной спектр является спектром стационарного линейного процесса.

Рассмотрим выходной процесс $X(t)$ устойчивой линейной системы с откликом на единичный импульс $h(u)$, когда входным процессом служит $Z(t)$. Из (5.2.8) ковариационная функция процесса $X(t)$ равна

$$\gamma_{XX}(u) = \int_0^{\infty} \int_0^{\infty} h(v) h(v') \gamma_{ZZ}(u + v - v') dv dv',$$

и, следовательно, из (6.2.2) спектральная плотность выхода равна

$$\begin{aligned} \Gamma_{XX}(f) &= \int_{-\infty}^{\infty} \gamma_{XX}(u) e^{-j2\pi fu} du = \\ &= \int_{-\infty}^{\infty} e^{-j2\pi fu} \int_0^{\infty} \int_0^{\infty} h(v) h(v') \gamma_{ZZ}(u + v - v') dv dv' du = \\ &= \int_0^{\infty} h(v) e^{j2\pi fv} dv \int_0^{\infty} h(v') e^{-j2\pi fv'} dv' \int_{-\infty}^{\infty} \gamma_{ZZ}(y) e^{-j2\pi fy} dy, \end{aligned}$$

где $y = u + v - v'$.

Отсюда

$$\Gamma_{XX}(f) = H(-f) H(f) \Gamma_{ZZ}(f) = |H(f)|^2 \Gamma_{ZZ}(f), \quad -\infty \leq f \leq \infty. \quad (6.2.15)$$

Это фундаментальное свойство утверждает, что *спектральная плотность выхода линейной системы получается из спектральной плотности входа с помощью умножения на квадрат модуля частотной характеристики системы.*

Если $Z(t)$ — белый шум со спектральной плотностью $\Gamma_{ZZ}(f) = \sigma_z^2$ и ковариационной функцией $\gamma_{ZZ}(u) = \sigma_z^2 \delta(u)$, то $X(t)$ является линейным процессом со спектральной плотностью

$$\Gamma_{XX}(f) = \sigma_z^2 |H(f)|^2, \quad -\infty \leq f \leq \infty. \quad (6.2.16)$$

Для дискретного времени соотношение, соответствующее (6.2.15), имеет вид

$$\Gamma_{XX}(f) = |H(f)|^2 \Gamma_{ZZ}(f), \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.2.17)$$

где $H(f) = \sum_{k=0}^{\infty} h_k e^{-j2\pi f k \Delta}$.

Если вход является чисто случайным процессом с дисперсией σ_Z^2 , то выход представляет собой линейный процесс со спектральной плотностью

$$\Gamma_{XX}(f) = \Delta \sigma_Z^2 |H(f)|^2, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.2.18)$$

Из (6.2.15) или (6.2.18) видно, что если есть источник белого шума и подходящий переменный аналоговый (или цифровой) фильтр, то можно получить случайный процесс с любым заданным спектром. В следующем разделе мы приведем некоторые примеры разнообразных спектров, которые можно получить с помощью линейной фильтрации белого шума.

6.2.5. Спектры процессов авторегрессии и скользящего среднего

Непрерывный процесс авторегрессии первого порядка. Рассмотрим непрерывный процесс авторегрессии первого порядка

$$T \frac{dX(t)}{dt} + (X(t) - \mu) = Z(t),$$

где $Z(t)$ — белый шум. Эта линейная система имеет отклик на единичный импульс

$$h(u) = \begin{cases} \frac{1}{T} e^{-u/T}, & 0 \leq u < \infty, \\ 0, & u < 0, \end{cases}$$

и частотную характеристику

$$H(f) = \frac{1}{1 + j2\pi f T}.$$

Отсюда, используя (6.2.16), получаем спектральную плотность процесса $X(t)$:

$$\Gamma_{XX}(f) = \frac{\sigma_Z^2}{1 + (2\pi f T)^2}, \quad -\infty \leq f \leq \infty. \quad (6.2.19)$$

График функции (6.2.19) изображен на рис. 2.3, а. Из него видно, что большая часть мощности, или дисперсии, сосредоточена на низких частотах.

Дискретный процесс авторегрессии первого порядка. Для дискретного времени процесс авторегрессии первого порядка имеет вид

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + Z_t.$$

В этом случае

$$h_k = \alpha_1^k, \quad k=0, 1, \dots, \infty,$$

и

$$H(f) = \frac{1}{1 - \alpha_1 e^{-j2\pi f \Delta}}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}.$$

Отсюда, используя (6.2.18), находим спектральную плотность процесса X_t :

$$\Gamma_{XX}(f) = \frac{\Delta \sigma_Z^2}{1 + \alpha_1^2 - 2\alpha_1 \cos 2\pi f \Delta}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.2.20)$$

Спектр (6.2.20) изображен на рис. 6.4 и 6.5 для случаев $\alpha_1 = +0,9$ и $\alpha_1 = -0,9$ соответственно, причем в обоих случаях $\Delta = 1$ и $\sigma_Z^2 = 1$. Как отмечалось в разд. 6.2.1, при положительном α_1 большая часть мощности спектра сосредоточена на низких частотах, а для отрицательных α_1 — на высоких частотах. Заметим из (6.2.20), что $\Gamma_{XX}(f)$ при $\alpha_1 > 0$ равна $\Gamma_{XX}(1/2\Delta - f)$ при $\alpha_1 < 0$.

Непрерывный процесс авторегрессии второго порядка. Рассмотрим непрерывный процесс авторегрессии второго порядка

$$a_2 \frac{d^2 X}{dt^2} + a_1 \frac{dX}{dt} + a_0 (X(t) - \mu) = Z(t).$$

В этом случае частотная характеристика равна

$$H(f) = \frac{1}{a_2 (j2\pi f)^2 + a_1 (j2\pi f) + a_0},$$

и, следовательно, спектральная плотность процесса равна

$$\Gamma_{XX}(f) = \frac{\sigma_Z^2}{(a_0 - a_2 4\pi^2 f^2)^2 + (2\pi f a_1)^2}. \quad (6.2.21)$$

Выражение (6.2.21) может давать как низкочастотные спектры (a_1 или a_2 велико), так и спектры с явно выраженным пиком (если характеристическое уравнение $a_2 p^2 + a_1 p + a_0 = 0$ имеет комплексные корни).

Дискретный процесс авторегрессии второго порядка. Дискретный процесс авторегрессии второго порядка (5.2.31), а именно

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + Z_t,$$

имеет частотную характеристику

$$H(f) = \frac{1}{1 - \alpha_1 e^{-j2\pi f\Delta} - \alpha_2 e^{-j4\pi f\Delta}}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta},$$

и, следовательно, спектральную плотность

$$\Gamma_{XX}(f) = \frac{\Delta\sigma_Z^2}{1 + \alpha_1^2 + \alpha_2^2 - 2\alpha_1(1 - \alpha_2)\cos 2\pi f\Delta - 2\alpha_2\cos 4\pi f\Delta}, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.2.22)$$

Для некоторых значений параметров α_1, α_2 выражение (6.2.22) представляет низкочастотный либо высокочастотный спектр, подобно дискретному процессу первого порядка. Но кроме таких спектров можно получить и спектры, имеющие пик, либо, наоборот, корытообразную впадину на некоторой частоте f_0 внутри интервала частот. Это происходит в случае, если $|\alpha_1(1 - \alpha_2)| < 4\alpha_2$. Частота f_0 , на которой получается пик либо впадина, определяется из выражения

$$\cos 2\pi f_0\Delta = -\frac{\alpha_1(1 - \alpha_2)}{4\alpha_2}.$$

Например, временной ряд, изображенный на рис. 6.6, получен с помощью процесса авторегрессии второго порядка с параметрами $\alpha_1 = 1, \alpha_2 = -0,5$. Спектр этого процесса имеет пик в точке $f_0 = 0,125/\Delta$ гц.

Четыре типа спектров, которые можно получить с помощью процесса авторегрессии второго порядка, перечислены на рис. 6.7. Интересная особенность, выявленная с помощью этого рисунка, заключается в том, что область $\alpha_1^2 + 4\alpha_2 < 0$ (в этой области корреляционная функция является затухающей синусоидой) частично перекрывается с областью $|\alpha_1(1 - \alpha_2)| \geq 4|\alpha_2|$, где спектр не имеет пиков внутри интервала частот (на рис. 6.7 последняя область заштрихована). Для высокочастотного спектра это не является неожиданным, так как даже процесс авторегрессии первого порядка при $\alpha_1 < 0$ имеет осциллирующую корреляционную функцию, хотя его спектр и не имеет внутренних пиков. Однако и для низкочастотного спектра корреляционная функция может осциллировать, и при этом не будет ярко выраженных внутренних пиков. Обычно считают, что осцилляция корреляционной функции сопровождается пиком в спектре, но этот пример показывает, что для этого амплитуда затухающих осцилляций корреляционной функции должна быть достаточно большой.

Общие процессы авторегрессии — скользящего среднего. Общий непрерывный процесс авторегрессии — скользящего среднего (5.2.21) имеет вид

$$a_m \frac{d^m X}{dt^m} + \dots + a_1 \frac{dX}{dt} + a_0 (X(t) - \mu) = b_1 \frac{d^l Z}{dt^l} + \dots + b_l \frac{dZ}{dt} + b_0 Z(t).$$

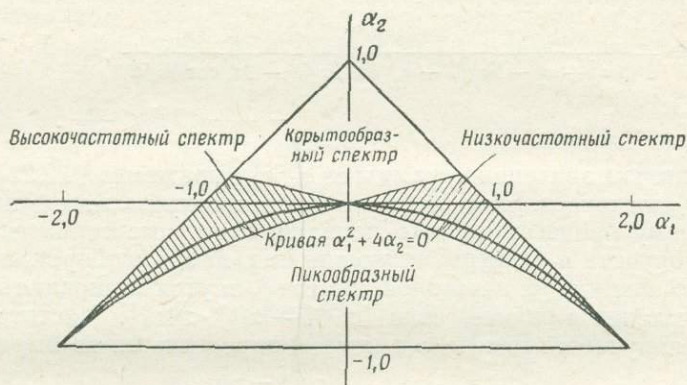


Рис. 6.7. Область устойчивости и классификация спектров для дискретных процессов авторегрессии второго порядка.

Его спектральная плотность равна

$$\Gamma_{XX}(f) = \sigma_Z^2 \left| \frac{b_0 + b_1 j 2\pi f + \dots + b_l (j 2\pi f)^l}{a_0 + a_1 j 2\pi f + \dots + a_m (j 2\pi f)^m} \right|^2, \quad -\infty \leq f \leq \infty. \quad (6.2.23)$$

Аналогично для дискретного времени процесс смешанного типа (5.2.50), а именно

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \dots + \alpha_m (X_{t-m} - \mu) + Z_t + \beta_1 Z_{t-1} + \dots + \beta_l Z_{t-l},$$

имеет спектральную плотность

$$\Gamma_{XX}(f) = \Delta \sigma_Z^2 \left| \frac{1 + \beta_1 e^{-j 2\pi f \Delta} + \dots + \beta_l e^{-j 2\pi f \Delta l}}{1 - \alpha_1 e^{-j 2\pi f \Delta} - \dots - \alpha_m e^{-j 2\pi f \Delta m}} \right|^2, \quad -\frac{1}{2\Delta} \leq f \leq \frac{1}{2\Delta}. \quad (6.2.24)$$

Из выражения (6.2.23) видно, что для того, чтобы $\Gamma_{XX}(f)$ была интегрируемой спектральной плотностью, соответствующей случайному процессу $X(t)$ с конечной дисперсией σ_X^2 , нужно, чтобы число l

удовлетворяло условию $l \leq m - 1$. Заметим, что в дискретном случае нет никаких ограничений на l .

Выражения (6.2.23) и (6.2.24) получены с помощью подстановки частотных характеристик (2.3.19) и (2.3.32) в (6.2.15) и (6.2.18) соответственно. В общем случае эти спектры могут иметь несколько пиков или впадин, если соответствующие характеристические уравнения имеют комплексные корни.

6.3. СПЕКТРАЛЬНЫЕ ОЦЕНКИ

6.3.1. Вероятностные свойства оценок, соответствующих выборочному спектру, для случая белого шума

Введение. Табл. 6.1 наводит на мысль о том, что оценка, соответствующая выборочному спектру,

$$C_{ZZ}(f) = \frac{\Delta}{N} \left[\left(\sum_{t=-n}^{n-1} Z_t \cos 2\pi f t \Delta \right)^2 + \left(\sum_{t=-n}^{n-1} Z_t \sin 2\pi f t \Delta \right)^2 \right],$$

$$-\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.3.1)$$

для чисто случайного процесса (дискретного белого шума) имеет дисперсию, не зависящую от числа наблюдений N . С другой стороны, среднее значение выборочного спектра по частоте близко к теоретическому значению спектра. Это указывает на то, что оценка, соответствующая спектру, не является состоятельной, т. е. ее распределение не стягивается к истинному значению спектра при увеличении объема выборки.

Чтобы убедиться, что это действительно так, рассмотрим случайные величины, соответствующие действительной и мнимой составляющим Фурье дискретного процесса Z_t , ($-n \leq t \leq n-1$). Они задаются равенствами

$$A(f) = \sum_{t=-n}^{n-1} Z_t \cos 2\pi f t \Delta,$$

$$B(f) = \sum_{t=-n}^{n-1} Z_t \sin 2\pi f t \Delta, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.3.2)$$

В таком случае оценку (6.3.1) можно записать в виде

$$C_{ZZ}(f) = \frac{\Delta}{N} [A^2(f) + B^2(f)], \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.3.3)$$

Исследовав свойства случайных величин $A(f)$ и $B(f)$, можно вывести и вероятностные свойства $C_{ZZ}(f)$. В этом разделе будет показано, что если Z_t — чисто случайный нормальный процесс с нулевым

средним значением и дисперсией σ_z^2 , то для гармонических частот, (частот f_k , кратных основной гармонике) $f_k = k/N\Delta$, справедливы следующие утверждения:

1) случайные величины

$$Y(f_k) = \frac{2C_{ZZ}(f_k)}{\Delta\sigma_z^2}, \quad k = \pm 1, \pm 2, \dots, \pm(n-1) \quad (6.3.4)$$

имеют χ^2 -распределение с двумя степенями свободы;

2) Если $f_k = 0$, или же $f_k = -1/2\Delta$, то случайные величины

$$Y(f_k) = \frac{C_{ZZ}(f_k)}{\Delta\sigma_z^2} \quad (6.3.5)$$

имеют χ^2 -распределение с одной степенью свободы;

3) случайные величины $Y(f_k)$ взаимно независимы для $k = 0, \pm 1, \pm 2, \dots, \pm(n-1), -n$.

Этими результатами мы воспользуемся в разд. 6.3.2 при выводе критерия для проверки гипотезы о том, что шум является белым. В разд. 6.3.3 дается краткое изложение более общих результатов, относящихся к вероятностным свойствам оценок, соответствующих выборочным спектрам. Эти результаты получены для произвольных частот и для процессов, не являющихся белым гауссовским шумом. Доказательства приведены в приложении П9.1.

χ^2 -свойство оценки, соответствующей выборочному спектру. Так как $E[Z_i] = 0$, то из (6.3.2) следует, что

$$E[A(f)] = 0 = E[B(f)].$$

Отсюда для гармоник $f_k = k/N\Delta$ получаем

$$\begin{aligned} \text{Var}[A(f_k)] &= E[A^2(f_k)] = \\ &= \sigma_z^2 \sum_{t=-n}^{n-1} \cos^2 2\pi f_k t \Delta = \begin{cases} \sigma_z^2 \frac{N}{2}, & k = \pm 1, \pm 2, \dots, \pm(n-1), \\ \sigma_z^2 N, & k = 0, -n. \end{cases} \end{aligned} \quad (6.3.6)$$

Аналогично находим

$$\text{Var}[B(f_k)] = \begin{cases} \sigma_z^2 \frac{N}{2}, & k = \pm 1, \pm 2, \dots, \pm(n-1), \\ 0, & k = 0, -n. \end{cases} \quad (6.3.7)$$

Далее при $k \neq l$ имеем

$$\text{Cov}[A(f_k), A(f_l)] = \sigma_z^2 \sum_{t=-n}^{n-1} \cos 2\pi f_k t \Delta \cos 2\pi f_l t \Delta = 0, \quad (6.3.8)$$

$$\text{Cov}[B(f_k), B(f_l)] = 0.$$

Кроме того, для любых k и l справедливо равенство

$$\text{Cov} [A(f_k), B(f_l)] = 0. \quad (6.3.9)$$

Так как $A(f_k)$ и $B(f_k)$ являются линейными комбинациями гауссовских величин, то они также имеют гауссовское распределение. Поэтому каждая из случайных величин

$$\frac{A^2(f_k)}{\text{Var} [A(f_k)]} = \frac{2A^2(f_k)}{N\sigma_Z^2}, \quad \frac{B^2(f_k)}{\text{Var} [B(f_k)]} = \frac{2B^2(f_k)}{N\sigma_Z^2}$$

имеет χ^2 -распределение с одной степенью свободы. Из (6.3.8) и (6.3.9) видно, что эти величины независимы, поскольку $A(f_k)$ и $B(f_k)$ имеют нормальное распределение. Поэтому их сумма

$$\frac{2}{N\sigma_Z^2} [A^2(f_k) + B^2(f_k)] = \frac{2C_{ZZ}(f_k)}{\Delta\sigma_Z^2} = Y(f_k)$$

имеет χ^2 -распределение с двумя степенями свободы.

При $k=0$ или $k=-n$ величина $B(f_k)$ тождественно равна нулю. Следовательно, случайная величина

$$Y(f_k) = \frac{A^2(f_k)}{\text{Var} [A(f_k)]} = \frac{C_{ZZ}(f_k)}{\Delta\sigma_Z^2}, \quad k=0, -n,$$

имеет χ^2 -распределение с одной степенью свободы. Из равенств (6.3.8) и (6.3.9) следует, что случайные величины $Y(f_k)$ для различных частот независимы, так как они получаются из независимых гауссовских величин $A(f_k)$, $B(f_k)$. Таким образом, утверждения (1), (2) и (3) доказаны.

Пользуясь этими результатами, можно объяснить флуктуирующее поведение выборочного спектра на рис. 6.1. В разд. 6.2.3 было показано, что спектр чисто случайного процесса равен константе

$$\Gamma_{ZZ}(f) = \sigma_Z^2 \Delta, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}.$$

Используя (3.3.6) и только что доказанные утверждения, получаем

$$E \left[\frac{2C_{ZZ}(f_k)}{\Delta\sigma_Z^2} \right] = 2,$$

т. е.

$$E [C_{ZZ}(f_k)] = \sigma_Z^2 \Delta = \Gamma_{ZZ}(f_k).$$

Следовательно, для гармонических частот оценка, соответствующая выборочному спектру, является *несмещенной* в случае, если шум белый. Это объясняет близость средних значений в табл. 6.1 к их теоретическим значениям.

Аналогично, используя (3.3.6), получаем

$$\text{Var} \left[\frac{2C_{ZZ}(f_k)}{\Delta\sigma_Z^2} \right] = 4,$$

т. е.

$$\text{Var} [C_{ZZ}(f_k)] = \sigma_Z^4 \Delta^2 = \Gamma_{ZZ}^2(f_k). \quad (6.3.10)$$

Равенства (6.3.10) показывают, что по крайней мере для гармонических частот дисперсия этой оценки равна константе, *независящей от объема выборки*. Это объясняет тот факт, что выборочные оценки дисперсии случайной величины $C_{ZZ}(f_k)$ не уменьшаются с увеличением объема выборки, как видно из табл. 6.1. Важно отметить, что даже для негауссовского процесса Z_t случайные величины $A(f)$ и $B(f)$ будут приближенно гауссовскими в силу центральной предельной теоремы. Поэтому величина $C_{ZZ}(f)$ будет иметь распределение, близкое к χ^2 -распределению с двумя степенями свободы, независимо от того, какое распределение у процесса Z_t .

Дисперсионный анализ. Важность полученных выше результатов легче оценить, если рассмотреть разложение полной суммы квадратов случайных величин Z_t . По теореме Парсевала (6.1.3) имеем

$$\sum_{t=-n}^{n-1} Z_t^2 = \sum_{k=-n}^{n-1} \frac{C_{ZZ}(f_k)}{\Delta}.$$

Используя то, что $C_{ZZ}(f_k) = C_{ZZ}(-f_k)$, получаем

$$\frac{1}{\sigma_Z^2} \sum_{t=-n}^{n-1} Z_t^2 = \frac{1}{\Delta\sigma_Z^2} \left[C_{ZZ}(0) + 2 \sum_{k=1}^{n-1} C_{ZZ}(f_k) + C_{ZZ}(f_n) \right]. \quad (6.3.11)$$

Так как Z_t/σ_Z — независимые нормальные величины с нулевыми средними значениями и единичными стандартными отклонениями, то стоящая в левой части равенства (6.3.11) случайная величина имеет χ^2 -распределение с N степенями свободы. Доказанные выше утверждения показывают в таком случае, что эта величина представляется в виде суммы двух χ^2 -величин с одной степенью свободы или $(n-1)$ χ^2 -величин с двумя степенями свободы. Таким образом, полное число степеней свободы раскладывается на следующие слагаемые:

$$N = 2n = 1 + 2(n-1) + 1.$$

Для нечетных N член с одной степенью свободы, соответствующий $k = -n$, исчезает из (6.3.11). Это разложение представляет собой частный случай метода, называемого в статистике *дисперсионным*

анализом. Если $E[Z_t] \neq 0$, то проведенный выше анализ справедлив, но разложение (6.3.11) в этом случае удобнее записывать в виде

$$\frac{1}{\sigma_Z^2} \sum_{t=-n}^{n-1} (Z_t - \bar{Z})^2 = \frac{1}{\Delta \sigma_Z^2} \left[2 \sum_{k=1}^{n-1} C_{ZZ}(f_k) + C_{ZZ}(f_n) \right], \quad (6.3.12)$$

где \bar{Z} — среднее арифметическое значение случайных величин Z_t .

6.3.2. Критерий для проверки гипотезы о том, что шум белый

Необходимость критерия. На практике часто возникают ситуации, когда требуется проверить гипотезу о том, что наблюдаемый временной ряд является реализацией белого шума. Пример такой ситуации приведен в разд. 5.3.5, где критерий для проверки того, что шум белый, был применен к случайным гауссовским числам, полученным с помощью вычислительной машины. Другим примером служит проверка подобранной модели, например процесса авторегрессии (5.2.39). Модель можно считать адекватной, если остаточные ошибки (между подобранной моделью и данными) образуют белый шум.

Приведенный в разд. 5.3.5 критерий для проверки того, что шум белый, полезен тогда, когда подозревают наличие «локальных корреляций», т. е. когда есть подозрение, что соседние точки временного ряда коррелированы. Иногда требуется обнаружить отклонения от белого шума, вызванные периодическими эффектами. Так, например, после подбора модели для экономического временного ряда, содержащего сезонные вариации, несоответствие модели могло бы выразиться в периодичности остаточных ошибок. В таком случае более подходящим является частотный критерий, основанный на выборочном спектре. Один такой критерий приведен ниже; его надо рассматривать как дополнение к критерию разд. 5.3.5, основанному на корреляционной функции.

Критерий. Равенство (6.2.14) показывает, что спектр дискретного белого шума имеет вид

$$\Gamma_{ZZ}(f) = \Delta \sigma_Z^2, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}.$$

Отсюда спектральная функция

$$I_{ZZ}(f) = \int_{-f}^f \Gamma_{ZZ}(g) dg = 2\Delta \sigma_Z^2 f, \quad 0 \leq f \leq \frac{1}{2\Delta},$$

линейно зависит от частоты.

Предположим, что выборочный спектр $C_{zz}(f)$ сосчитан для гармонических частот $f_k = k/N\Delta$, $k = 0, 1, \dots, N/2$. Рассмотрим тогда оценки $I(f_k)$ спектральной функции

$$I(f_k) = \frac{1}{N\Delta} \sum_{l=1}^k C_{zz}(f_l). \quad (6.3.13)$$

Заметим, что $C_{zz}(0) = 0$, если вычитается среднее значение. Так как $E[C_{zz}(f_k)] = \Gamma_{zz}(f_k) = 2\Delta\sigma_z^2$, то

$$E[I(f_k)] = 2\Delta\sigma_z^2 f_k = 2\Delta\sigma_z^2 \frac{k}{N\Delta}$$

и, следовательно, $I(f_k)$ является несмещенной оценкой $I_{zz}(f_k)$. На практике удобнее нормировать $I(f_k)$, разделив ее на σ_z^2 . В этом случае $I(1/2\Delta) = 1$. Поскольку σ_z^2 неизвестна в практических ситуациях, ее следует заменить на оценку S_z^2 , так что в окончательном виде оценка нормированной спектральной функции имеет вид $I(f_k)/S_z^2$. Таким образом, соответствующая выборочная оценка, сочитанная по временному ряду, равна

$$\frac{i(f_k)}{s_z^2} = \frac{1}{N\Delta s_z^2} \sum_{l=1}^k C_{zz}(f_l).$$

Если построить график этой выборочной оценки, беря в качестве аргументов точки $2\Delta f_k$, то точки графика должны лежать близко к отрезку, соединяющему точки $(0, 0)$ и $(1, 1)$. Так как $I(f_k)$ представляет собой сумму случайных величин с одинаковым распределением, то можно применить критерий Колмогорова — Смирнова [4], чтобы узнать, являются ли отклонения выборочной оценки нормированной спектральной функции от прямой линии значимыми (обычно этот критерий применяют для проверки значимости отклонений выборочной функции распределения от теоретической).

Два примера. В табл. 6.2 приведены значения $C_{zz}(f_k)$ для одной из выборок случайных нормальных чисел, использованных для вычислений в табл. 6.1. Здесь $N = 100$, $\Delta = 1$ и, следовательно, $f_k = 0,01; 0,02; \dots; 0,50$. На рис. 6.8 показан график $i(f_k)/s_z^2$ в зависимости от k для этого ряда. Из этого рисунка видно, что отклонения от прямой невелики. Чтобы получить точное заключение о величине этих отклонений, можно при больших N воспользоваться критерием значимости Колмогорова — Смирнова [4]. Он состоит в том, что надо построить полосу $\pm \lambda/(N/2 - 1)^{1/2}$ около теоретической прямой. Для уровней значимости 0,95 и 0,75 λ равно 1,36 и 1,02 соответственно. В нашем случае $N/2 = 50$; поэтому 95%-ные границы равны $\pm 1,36/\sqrt{49} = \pm 0,19$; 75%-ные границы равны $\pm 0,15$. Эти границы

Таблица 6.2

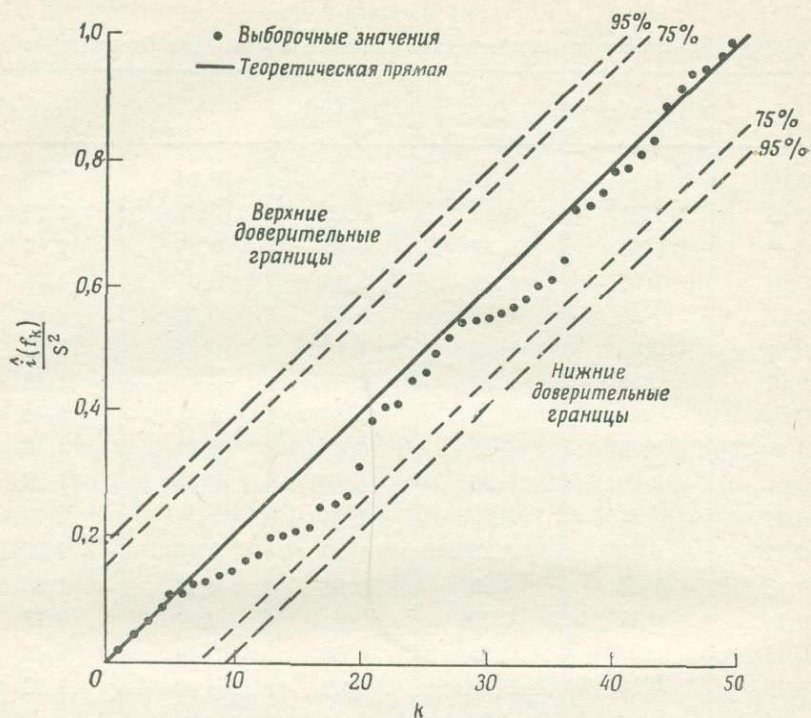
Выборочный спектр на гармонических частотах для выборки белого шума

f_k	$C_{zz}(f_k)$	f_k	$C_{zz}(f_k)$	f_k	$C_{zz}(f_k)$
0,01	1,13	0,17	1,91	0,34	1,75
0,02	1,41	0,18	0,15	0,35	0,25
0,03	0,74	0,19	0,85	0,36	1,84
0,04	1,08	0,20	2,49	0,37	3,98
0,05	1,28	0,21	3,89	0,38	0,22
0,06	0,06	0,22	1,13	0,39	1,52
0,07	0,85	0,23	0,53	0,40	1,48
0,08	0,23	0,24	1,86	0,41	0,44
0,09	0,71	0,25	0,47	0,42	1,16
0,10	0,79	0,26	1,87	0,43	1,20
0,11	0,51	0,27	1,35	0,44	2,73
0,12	0,46	0,28	1,29	0,45	1,66
0,13	1,38	0,29	0,06	0,46	1,34
0,14	0,11	0,30	0,24	0,47	0,17
0,15	0,37	0,31	0,56	0,48	1,43
0,16	0,50	0,32	0,68	0,49	1,03
		0,33	0,44		

показаны пунктиром на рис. 6.8, и мы видим, что значения $i(f_k)/s^2$ попадают целиком между ними. Поэтому нет никаких доводов против того, что выборка получена из белого шума. Интерпретация 75%-ных границ, например, заключается в том, что в среднем на каждом четвертом графике *максимальное отклонение* от теоретической прямой будет выходить за границы, даже если процесс на самом деле является белым шумом.

В табл. 6.3 показаны результаты вычислений для этого критерия, выполненных по ионосферным данным из табл. 2.1. Для этих данных $s^2=196,4$. Значения $C_{zz}(f_k)$ из табл. 6.3 можно получить, умножая вклады в среднеквадратичное значение, помещенные в табл. 2.2, на $N=12$.

Из рис. 6.9 видно, что выборочная оценка спектральной функции сильно отклоняется от прямой линии, поскольку $i(f_1)/s^2$ примерно в два раза больше соответствующего среднего значения для белого шума, а $i(f_2)/s^2$ почти в три раза больше. Доверительные границы, о которых мы говорили выше, здесь неприменимы, так как



Р и с. 6.8. Проверка того, что шум белый, использующая интеграл от периодограммы.

Таблица 6.3

Применение критерия проверки белого шума
к ионосферным данным

f_k	$C_{zz}(f_k)$	$\frac{1}{N S^2} \sum_{l=1}^k C_{zz}(f_l)$
0,083	753,6	0,32
0,166	1322,4	0,88
0,250	38,4	0,90
0,333	18,0	0,91
0,417	78,0	0,94
0,500	146,0	1,00

N слишком мало. Фактически, в этом случае и не требуется никакого критерия значимости, поскольку значения $i(f_k)/s^2$ так велики при $f_k = 0,083$ и $0,166$.

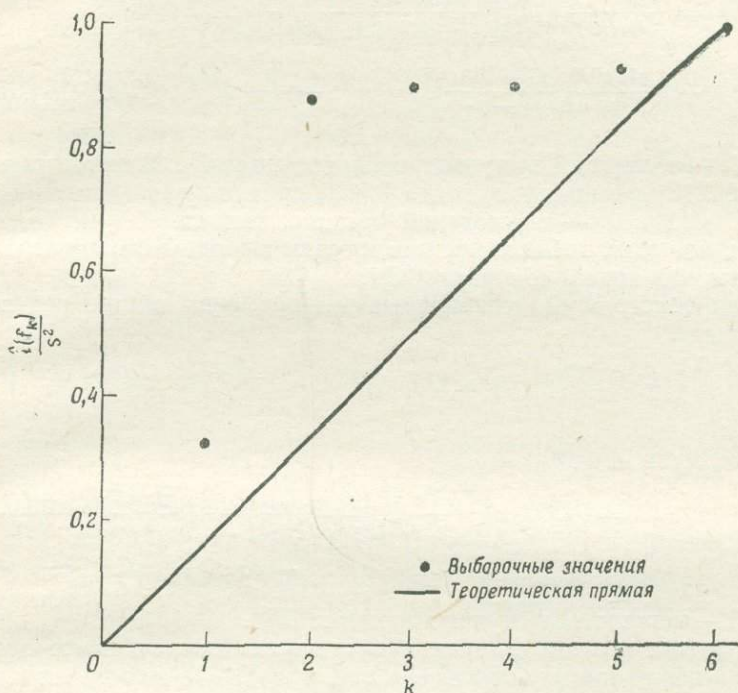


Рис. 6.9. Проверка ионосферных данных на случайность.

6.3.3. Общие результаты о выборочных спектрах для белого шума

В разд. 6.3.1 выведены выражения для среднего значения и ковариаций оценки, соответствующей выборочному спектру, на гармонических частотах $f_k = k/N\Delta$ в предположении, что Z_t — гауссовский процесс. В приложении П9.1 выведены более общие результаты, применимые для любых частот и для негауссовских процессов.

Моменты оценок, соответствующих выборочному спектру, для белого шума. Для дискретного времени эти более общие результаты имеют вид

$$E [C_{ZZ}(f)] = \Gamma_{ZZ}(f) = \sigma_z^2 \Delta, \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}, \quad (6.3.14)$$

и

$$\begin{aligned} \text{Cov} [C_{ZZ}(f_1), C_{ZZ}(f_2)] = & \\ = \frac{K_4 \Delta^2}{N} + \sigma_Z^4 \Delta^2 \left[\left(\frac{\sin \pi N \Delta (f_1 + f_2)}{N \sin \pi \Delta (f_1 + f_2)} \right)^2 + \right. & \\ \left. + \left(\frac{\sin \pi N \Delta (f_1 - f_2)}{N \sin \pi \Delta (f_1 - f_2)} \right)^2 \right], & \quad -\frac{1}{2\Delta} \leq f_1, f_2 < \frac{1}{2\Delta}, \quad (6.3.15) \end{aligned}$$

где K_4 — четвертый кумулянт распределения Z_t . Можно проверить, что (6.3.15) равно нулю, когда f_1 и f_2 кратны фундаментальной частоте $1/N\Delta$ и Z_t — гауссовский процесс, так что $K_4=0$. Таким образом, при этих предположениях оценки выборочного спектра независимы, как показано в разд. 6.3.1.

Для белого шума с непрерывным временем общие результаты имеют вид

$$E [C_{ZZ}(f)] = \Gamma_{ZZ}(f) = \sigma_Z^2, \quad -\infty \leq f \leq \infty, \quad (6.3.16)$$

и

$$\begin{aligned} \text{Cov} [C_{ZZ}(f_1), C_{ZZ}(f_2)] = & \\ = \frac{K_4}{T} + \sigma_Z^4 \left[\left(\frac{\sin \pi T (f_1 + f_2)}{\pi T (f_1 + f_2)} \right)^2 + \left(\frac{\sin \pi T (f_1 - f_2)}{\pi T (f_1 - f_2)} \right)^2 \right], & \\ -\infty \leq f_1, f_2 \leq \infty, & \quad (6.3.17) \end{aligned}$$

где K_4 — четвертый кумулянт процесса $Z(t)$.

Заметим, что ковариация спектральных оценок имеет порядок $1/T$ для негауссовских процессов, т. е. при $K_4 \neq 0$, в то время как для гауссовских процессов $K_4=0$ и ковариация имеет порядок $1/T^2$. В частном случае, когда f_1 и f_2 — значения, кратные $1/T$, ковариация равна нулю. Далее, дисперсия спектральных оценок без учета членов порядка $1/T$ и более высокого равна

$$\text{Var} [C_{ZZ}(f)] = \sigma_Z^4.$$

Это показывает вообще, что $C_{ZZ}(f)$ не является состоятельной оценкой $\Gamma_{ZZ}(f)$.

χ^2 -свойства оценок, соответствующих выборочному спектру, для случая белого шума. В разд. 6.3.1 было показано, что если Z_t является гауссовским белым шумом, то $2C_{ZZ}(f)/\Delta\sigma_Z^2$ имеет χ^2 -распределение с двумя степенями свободы для гармоник $f_k = k/N\Delta$. В приложении П9.1 этот результат обобщается следующим образом. Для гауссовского белого шума распределение величины $2C_{ZZ}(f)/\Delta\sigma_Z^2$ точно совпадает с χ^2 -распределением с двумя степенями свободы, в то время как для негауссовских процессов при больших N это со-

впадение распределений имеет приближенный характер. Для непрерывного времени результаты формулируются точно так же, за исключением того, что они относятся к $C_{zz}(f)/\sigma_z^2$.

6.3.4. Сглаживание спектральных оценок

Способ сглаживания Бартлетта. Один прием, который можно использовать для получения спектральных оценок, имеющих дисперсию, меньшую, чем у $C_{zz}(f)$, был предложен Бартлеттом [5]. Предположим, что вместо вычисления $C_{zz}(f)$ по реализации белого шума длины $N=400$, как это делалось в разд. 6.1.2, эта реализация разбивается на $k=8$ рядов длины $N/k=50$ и выборочный спектр $C_{zz}^{(i)}(f)$, $i=1, 2, \dots, 8$, вычисляется для каждого ряда длины 50. Среднее значение этих восьми выборочных спектров на частоте f равно

$$\bar{C}_{zz}(f) = \frac{1}{8} \sum_{i=1}^8 C_{zz}^{(i)}(f), \quad -\frac{1}{2\Delta} \leq f < \frac{1}{2\Delta}. \quad (6.3.18)$$

Оно называется *выборочной сглаженной спектральной оценкой* на частоте f .

На рис. 6.10 построены графики $\bar{C}_{zz}(f)$ и функции $C_{zz}(f)$, вычисленной по всем 400 членам, для частот $f=0; 0,02; \dots; 0,5$ гц. Отметим, что $\bar{C}_{zz}(f)$ меняется более плавно и проходит ближе к $\Gamma_{zz}(f)$. В табл. 6.4 показаны средние значения, дисперсии и среднеквадратичные ошибки $\bar{C}_{zz}(f)$ и $C_{zz}(f)$ при усреднении по частоте. Согласно (6.3.10), дисперсия каждой $C_{zz}^{(i)}(f)$ равна σ_z^4 . Так как Z_t — белый шум, то отдельные ряды разбиения независимы и, следовательно, дисперсия $\bar{C}_{zz}(f)$ равна $\sigma_z^4/8$. Отношение двух наблюдаемых дисперсий из табл. 6.4 ($0,139/0,826 = 1/5,94$) незначимо отличается от ожидаемого значения $1/8$. Следовательно, с помощью усреднения,

Таблица 6.4

Моменты несглаженной и сглаженной выборочных спектральных оценок (усреднение проводилось по частоте)

	Среднее значение	Дисперсия	Среднеквадратичная ошибка
$C_{zz}(f)$	0,95	0,826	0,828
$\bar{C}_{zz}(f)$	0,94	0,139	0,143

или сглаживания, величин, относящихся к отдельным частям разбиения исходного ряда, дисперсию спектральной оценки можно уменьшить в нужное число раз. В предельном случае можно было бы использовать разбиение исходного ряда на отдельные ряды из двух членов, и при этом дисперсия уменьшилась бы до $2\sigma_z^4/N$. Чтобы понять, почему не имеет смысла так поступать, необходимо

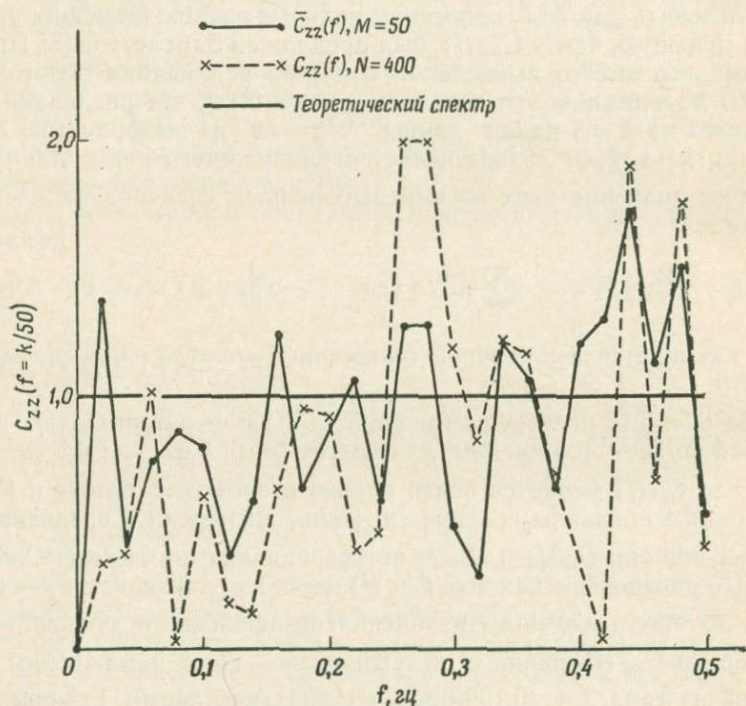


Рис. 6.10. Выборочный спектр и сглаженная выборочная оценка спектра для нормального белого шума.

внимательно рассмотреть процедуру сглаживания и вывести моменты сглаженных оценок.

Корреляционные и спектральные окна. Из (6.2.1) математическое ожидание оценки, соответствующей выборочному спектру, равно

$$E[C_{xx}(f)] = \int_{-T}^T \left(1 - \frac{|u|}{T}\right) \gamma_{xx}(u) e^{-j2\pi fu} du. \quad (6.3.19)$$

Оно представляет собой преобразование Фурье от произведения функции $\gamma_{xx}(u)$ и функции

$$w(u) = \begin{cases} 1 - \frac{|u|}{T}, & |u| \leq T, \\ 0, & |u| > T. \end{cases} \quad (6.3.20)$$

Отсюда, используя теорему о свертке (2.4.3), получаем

$$E[C_{xx}(f)] = \int_{-\infty}^{\infty} T \left(\frac{\sin \pi Tg}{\pi Tg} \right)^2 \Gamma_{xx}(f-g) dg, \quad (6.3.21)$$

поскольку преобразование Фурье функции $w(u)$ равно

$$W(f) = T \left(\frac{\sin \pi Tf}{\pi Tf} \right)^2. \quad (6.3.22)$$

Равенство (6.3.21) показывает, что математическое ожидание оценки $C_{xx}(f)$ соответствует как бы просматриванию теоретического спектра $\Gamma_{xx}(f)$ через спектральное окно $W(f)$. В терминологии гл. 2 $E[C_{xx}(f)]$ соответствует пропусканию теоретического спектра $\Gamma_{xx}(f)$ через фильтр с «откликом на единичный импульс» $W(f)$. Названия спектральное окно для $W(f)$ и корреляционное окно для $w(u)$ были введены Блэкманом и Тьюки [6].

Поскольку $W(f)$ в (6.3.22) при больших T ведет себя подобно δ -функции, из (6.3.21) и (2.2.5) следует, что

$$\lim_{T \rightarrow \infty} E[C_{xx}(f)] = \Gamma_{xx}(f),$$

т. е. $C_{xx}(f)$ — асимптотически несмещенная оценка $\Gamma_{xx}(f)$. Однако для записи конечной длины из (6.3.21) видно, что $C_{xx}(f)$ является смещенной оценкой $\Gamma_{xx}(f)$ со смещением

$$B(f) = E[C_{xx}(f)] - \Gamma_{xx}(f).$$

Для белого шума $\Gamma_{xx}(f) = \Delta\sigma_z^2$, и равенство (6.3.21) сводится к

$$E[C_{xx}(f)] = \Delta\sigma_z^2$$

для всех T . Следовательно, для белого шума оценка, соответствующая выборочному спектру, является несмещенной для всех T .

Спектральное окно $W(f)$, грубо говоря, действует при сглаживании как узкая щель, порядок ширины которой равен $1/T$, так что для больших T естественно считать $\Gamma_{xx}(f)$ приблизительно константой внутри этой щели. Поэтому (6.3.21) сводится к

$$E[C_{xx}(f)] \approx \Gamma_{xx}(f) \int_{-\infty}^{\infty} T \left(\frac{\sin \pi Tg}{\pi Tg} \right)^2 dg = \Gamma_{xx}(f).$$

Таким образом, для достаточно больших T смещение неоглаженного выборочного спектра будет малым.

Спектральное окно Бартлетта. Рассмотрим теперь математическое ожидание случайной оценки $\bar{C}_{XX}(f)$, используемой в способе сглаживания Бартлетта. При разбиении исходного ряда на k рядов, каждый из которых имеет длину M , из (6.1.9) получаем

$$C_{XX}^{(i)}(f) = \int_{-M}^M c_{XX}^{(i)}(u) e^{-j2\pi fu} du.$$

Отсюда сглаженная спектральная оценка равна

$$\bar{C}_{XX}(f) = \frac{1}{k} \sum_{i=1}^k C_{XX}^{(i)}(f) = \int_{-M}^M \bar{c}_{XX}(u) e^{-j2\pi fu} du, \quad (6.3.23)$$

где

$$\bar{c}_{XX}(u) = \frac{1}{k} \sum_{i=1}^k \left\{ \frac{1}{M} \int_{(i-1)M}^{iM-u} X(t) X(t+u) dt \right\}, \quad u \geq 0, \quad (6.3.24)$$

а для $u < 0$ эта функция определяется аналогично (5.3.9). Математическое ожидание $\bar{c}_{XX}(u)$ в таком случае равно

$$E[\bar{c}_{XX}(u)] = \gamma_{XX}(u) \left(1 - \frac{|u|}{M}\right)$$

и

$$\begin{aligned} E[\bar{C}_{XX}(f)] &= \int_{-M}^M \left(1 - \frac{|u|}{M}\right) \gamma_{XX}(u) e^{-j2\pi fu} du = \\ &= \int_{-\infty}^{\infty} \Gamma_{XX}(f-g) M \left(\frac{\sin \pi g M}{\pi g M}\right)^2 dg. \end{aligned} \quad (6.3.25)$$

Следовательно, разделение записи длины T на k частей длины $M = T/k$ каждая и построение сглаженной спектральной оценки (6.3.23) эквивалентно сглаживанию выборочного спектра с помощью окна

$$W(f) = M \left(\frac{\sin \pi f M}{\pi f M}\right)^2. \quad (6.3.26)$$

Во временной области это эквивалентно умножению ковариационной функции на корреляционное окно

$$w(u) = \begin{cases} 1 - \frac{|u|}{M}, & |u| \leq M, \\ 0, & |u| > M. \end{cases} \quad (6.3.27)$$

Окна (6.3.26) и (6.3.27) называются спектральным и корреляционным окнами Бартлетта. График спектрального окна Бартлетта

изображен на рис. 6.11. Видно, что он симметричен относительно начала координат и имеет нули в точках $f = \pm 1/M, \pm 2/M, \dots$. Таким образом, ширина окна (т. е. расстояние между первыми нулями с каждой стороны) равна $2/M$. Следовательно, выбирая длину M отрезка разбиения, можно регулировать ширину спектрального окна. Мы уже показали, что, выбирая M небольшим, можно сделать малой дисперсию спектральной оценки. А малые значения M , как мы видим, соответствуют большим значениям ширины спектрального

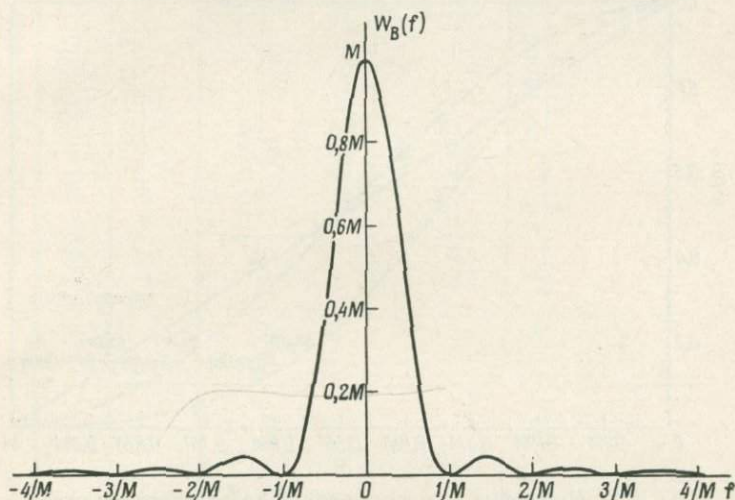


Рис. 6.11. Спектральное окно Бартлетта $W_B(f) = M (\sin \pi f M / \pi f M)^2$.

окна. Однако, если ширина окна велика, то происходит сглаживание на большом диапазоне частот, т. е. «отклик на единичный импульс» $W(f)$ очень широк, что может привести к большому смещению $B(f) = E[\bar{C}_{XX}(f)] - \Gamma_{XX}(f)$. Таким образом, как и для всех статистических оценок, нужно выбирать компромисс между дисперсией и смещением. В следующем разделе такое компромиссное решение изучается для более общего способа сглаживания выборочных спектров.

6.3.5. Спектральные окна и сглаженные спектральные оценки

Один общий класс сглаженных спектральных оценок. Описанный выше способ сглаживания Бартлетта показывает, что большую дисперсию оценки, соответствующей выборочному спектру, можно уменьшить, вводя корреляционное окно (6.3.27). Это наводит на

мысль о том, чтобы рассмотреть более общие сглаженные спектральные оценки вида

$$\bar{C}_{XX}(f) = \int_{-\infty}^{\infty} w(u) c_{XX}(u) e^{-j2\pi fu} du = \int_{-\infty}^{\infty} \bar{c}_{XX}(u) e^{-j2\pi fu} du, \quad (6.3.28)$$

у которых дисперсия будет меньше, чем у несглаженной оценки $C_{XX}(f)$.

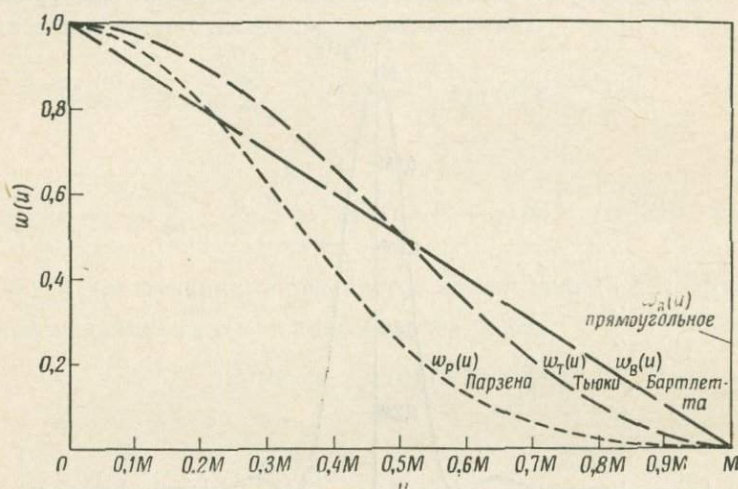


Рис. 6.12. Некоторые распространенные корреляционные окна.

Корреляционное окно $w(u)$ в (6.3.28) удовлетворяет условиям:

- 1) $w(0) = 1$,
 - 2) $w(u) = w(-u)$,
 - 3) $w(u) = 0, \quad |u| > T$.
- (6.3.29)

На практике условие (3) заменяют на

- 4) $w(u) = 0, \quad |u| \geq M, \quad M < T$,

так как при этом нужно будет вычислять ковариации лишь до запаздывания M . Примеры корреляционных окон, широко применяемых в спектральном анализе, приведены в табл. 6.5, а их графики построены на рис. 6.12. Преобразования Фурье этих корреляционных окон, т. е. спектральные окна $W(f)$, показаны на рис. 6.13.

Используя свойство свертки (2.4.3), равенство (6.3.28) можно записать в виде

$$\bar{C}_{XX}(f) = \int_{-\infty}^{\infty} W(g) C_{XX}(f - g) dg. \quad (6.3.30)$$

Корреляционные и спектральные окна

Название	Формула для корреляционного окна	Формула для спектрального окна
Прямоугольное	$w_R(u) = \begin{cases} 1, & u \leq M \\ 0, & u > M \end{cases}$	$W_R(f) = 2M \left(\frac{\sin 2\pi f M}{2\pi f M} \right), \quad -\infty \leq f \leq \infty$
Бартлетта	$w_B(u) = \begin{cases} 1 - \frac{ u }{M}, & u \leq M \\ 0, & u > M \end{cases}$	$W_B(f) = M \left(\frac{\sin \pi f M}{\pi f M} \right)^2, \quad -\infty \leq f \leq \infty$
Тьюки	$w_T(u) = \begin{cases} \frac{1}{2} \left(1 + \cos \frac{\pi u}{M} \right), & u \leq M \\ 0, & u > M \end{cases}$	$W_T(f) = M \left\{ \frac{\sin 2\pi f M}{2\pi f M} + \frac{1}{2} \frac{\sin 2\pi M \left(f + \frac{1}{2} M \right)}{2\pi M \left(f + \frac{1}{2} M \right)} + \frac{1}{2} \frac{\sin 2\pi M \left(f - \frac{1}{2} M \right)}{2\pi M \left(f - \frac{1}{2} M \right)} \right\} =$ $= M \left(\frac{\sin 2\pi f M}{2\pi f M} \right) \left(\frac{1}{1 - (2fM)^2} \right), \quad -\infty \leq f \leq \infty$
Парзена	$w_P(u) = \begin{cases} 1 - 6 \left(\frac{u}{M} \right)^2 + 6 \left(\frac{ u }{M} \right)^5, & u \leq \frac{M}{2} \\ 2 \left(1 - \frac{ u }{M} \right)^3, & \frac{M}{2} < u \leq M \\ 0, & u > M \end{cases}$	$W_P(f) = \frac{3}{4} M \left(\frac{\sin \pi f M / 2}{\pi f M / 2} \right)^4, \quad -\infty \leq f \leq \infty$

где $S_{xx}(f)$ определена в примечании переводчика на стр. 263 и

$$W(f) = \int_{-\infty}^{\infty} w(u) e^{-j2\pi fu} du. \quad (6.3.31)$$

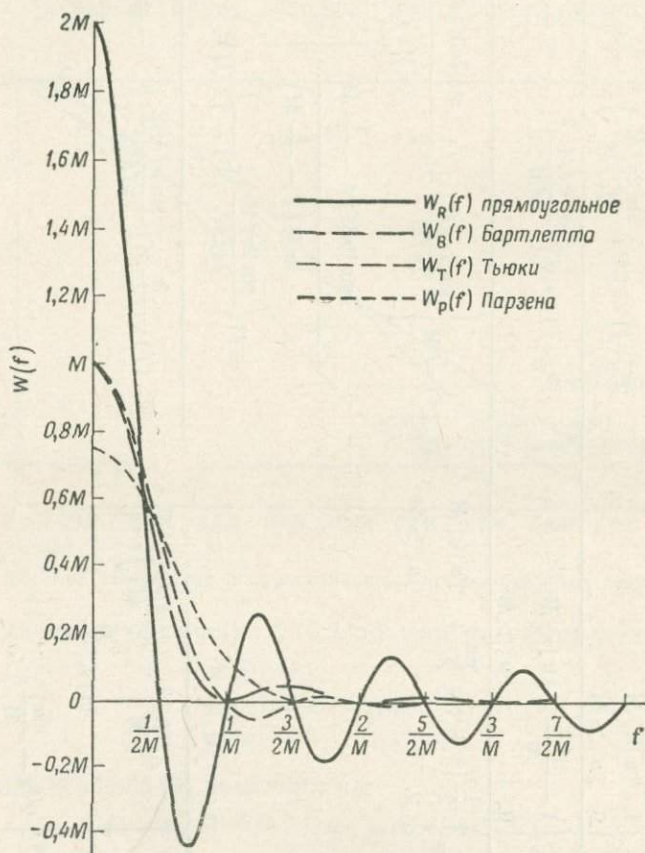


Рис. 6.13. Некоторые распространенные спектральные окна.

Обратное преобразование

$$w(u) = \int_{-\infty}^{\infty} W(f) e^{j2\pi fu} df. \quad (6.3.32)$$

дает возможность по спектральному окну $W(f)$ вычислить корреляционное окно $w(u)$. В соответствии со свойствами (6.3.29) спек-

ральное окно $W(f)$ удовлетворяет следующим условиям:

- 1) $\int_{-\infty}^{\infty} W(f) df = \omega(0) = 1,$
- 2) $W(f) = W(-f),$ (6.3.33)
- 3) $W(f)$ действует при сглаживании как щель ширины $\sim 2/M.$

Математическое ожидание сглаженной спектральной оценки. Беря математическое ожидание от обеих частей (6.3.30), получаем

$$E[\bar{C}_{XX}(f)] = \int_{-\infty}^{\infty} W(g) E[C_{XX}(f-g)] dg. \quad (6.3.34)$$

Однако, как показывает (6.3.21), для больших T

$$E[C_{XX}(g)] \approx \Gamma_{XX}(g),$$

следовательно,

$$E[\bar{C}_{XX}(f)] \approx \int_{-\infty}^{\infty} W(g) \Gamma_{XX}(f-g) dg = \bar{\Gamma}_{XX}(f). \quad (6.3.35)$$

Функцию $\bar{\Gamma}_{XX}(f)$ будем называть *средним сглаженным спектром*.

Теперь нам понадобится материал разд. 2.4.1. Поскольку спектральное окно $W(f)$ удовлетворяет условию (6.3.33) — (3), функция $\bar{\Gamma}_{XX}(f)$ будет выглядеть как несколько искаженная функция $\Gamma_{XX}(f)$. Этот эффект показан на рис. 2.10, где $\Gamma_{XX}(f)$ соответствует функции $|S_i(f)|$, $\bar{\Gamma}_{XX}(f)$ соответствует функции $|S_0(f)|$, а корреляционные окна $\omega(u)$ соответствуют временным окнам $w(t)$. Из рис. 2.10 видно, что чем меньше ширина корреляционного окна, тем сильнее отличается $\bar{\Gamma}_{XX}(f)$ от $\Gamma_{XX}(f)$. Следовательно, для того чтобы смещение

$$B(f) = E[\bar{C}_{XX}] - \Gamma_{XX}(f) = \bar{\Gamma}_{XX}(f) - \Gamma_{XX}(f)$$

было малым, нужно выбирать большое M . Это противоречит упоминавшемуся выше требованию выбора малого значения M для того, чтобы дисперсия $\text{Var}[\bar{C}_{XX}(f)]$ была небольшой. В разд. 4.2.3 было показано, что нужно выбирать компромиссное решение, учитывая и дисперсию, и смещение оценки. Те же самые рассуждения применимы и к оценкам спектра. Смещение можно сделать малым, лишь сужая $W(f)$, т. е. выбирая ее как можно ближе к δ -функции. С другой стороны, узкое спектральное окно $W(f)$ приводит

к большой дисперсии. Поэтому разумная процедура состоит в минимизации среднеквадратичной ошибки [7]:

$$\text{Var} [\bar{C}_{XX}(f)] + B^2(f).$$

Точная природа компромисса, который нужно сделать, будет зависеть от плавности изменения теоретического спектра $\Gamma_{XX}(f)$. Например, если $\Gamma_{XX}(f)$ очень плавно меняется, то дисперсию можно уменьшить с помощью широкого окна, не внося серьезного смещения. В частности, если $\Gamma_{XX}(f)$ плавно меняется в диапазоне $-1/M \leq (f-g) \leq 1/M$, то (6.3.36) приблизительно равно

$$E [\bar{C}_{XX}(f)] \approx \Gamma_{XX}(f) \int_{-\infty}^{\infty} W(g) dg = \Gamma_{XX}(f) \quad (6.3.36)$$

в силу (6.3.33) и (6.3.34). Следовательно, если теоретический спектр изменяется достаточно плавно, то получается фактически несмещенная оценка, хотя спектральное окно при этом делается широким для снижения дисперсии.

Приближенные выражения для смещения. Если нельзя считать, что теоретический спектр изменяется плавно по сравнению со спектральным окном, то можно, следуя Парзену [8], приблизительно подсчитать смещение, соответствующее данному спектральному окну. Используя (6.3.28) и (5.3.13), мы можем записать смещение для больших T также в виде

$$\begin{aligned} B(f) &= E \left[\int_{-\infty}^{\infty} w(u) c_{XX}(u) e^{-j2\pi fu} du \right] - \int_{-\infty}^{\infty} \gamma_{XX}(u) e^{-j2\pi fu} du \approx \\ &\approx \int_{-\infty}^{\infty} (w(u) - 1) \gamma_{XX}(u) e^{-j2\pi fu} du, \end{aligned} \quad (6.3.37)$$

Подставляя в эту формулу корреляционные окна $w(u)$ из гл. 6.5, получаем следующие выражения для смещений, соответствующих этим окнам:

$$\begin{aligned} B_B(f) &\approx -\frac{1}{M} \int_{-\infty}^{\infty} |u| \gamma_{XX}(u) e^{-j2\pi fu} du, \\ B_T(f) &\approx -\frac{\pi^2}{4M^2} \int_{-\infty}^{\infty} u^2 \gamma_{XX}(u) e^{-j2\pi fu} du + O\left(\frac{1}{M^4}\right) = \\ &= \frac{0,063}{M^2} \Gamma_{XX}^{(2)}(f) + O\left(\frac{1}{M^4}\right), \quad (6.3.38) \\ B_P(f) &\approx -\frac{6}{M^2} \int_{-\infty}^{\infty} u^2 \gamma_{XX}(u) e^{-j2\pi fu} du + O\left(\frac{1}{M^3}\right) = \\ &= \frac{0,152}{M^2} \Gamma_{XX}^{(2)}(f) + O\left(\frac{1}{M^3}\right). \end{aligned}$$

В приведенных выше выражениях $\Gamma_{xx}^{(2)}(f)$ — вторая производная спектра.

Эти формулы показывают следующее:

1. Если $\Gamma_{xx}^{(2)}(f)$ отрицательна (как, например, в окрестности пика), то смещение отрицательно, и поэтому в окрестностях пиков оценки будут обычно давать заниженные значения. Наоборот, если $\Gamma_{xx}^{(2)}(f)$ положительна (как, например, в окрестности впадины), то смещение положительно, и в этих точках оценки будут обычно давать завышенные значения.

2. Чем меньше ширина пика или впадины, тем больше $\Gamma_{xx}^{(2)}(f)$ и, следовательно, тем больше смещение.

3. Смещение $B_B(f)$ для окна Бартлетта имеет порядок $1/M$, и поэтому оно будет, вообще говоря, больше, чем смещения для окон Тьюки и Парзена, которые имеют порядок $1/M^2$.

4. Смещение уменьшается с увеличением M , т. е. с уменьшением ширины окна.

5. При одинаковом значении точки отсечения M , т. е. максимального запаздывания, на котором корреляционное окно отлично от нуля, окно Парзена дает большее смещение, чем окно Тьюки. Это происходит из-за того, что спектральное окно Парзена шире, чем спектральное окно Тьюки (см. рис. 6.13). Однако дисперсия оценки Парзена меньше, чем дисперсия оценки Тьюки при одном и том же значении M , как будет показано в разд. 6.4.1.

Формулы (6.3.38) полезны для качественного описания свойств смещения, однако для получения количественной картины нужно построить график среднего сглаженного спектра, как будет показано в разд. 7.1.

6.4. ДАЛЬНЕЙШИЕ СВОЙСТВА СГЛАЖЕННЫХ СПЕКТРАЛЬНЫХ ОЦЕНОК

Мы уже исследовали одно важное свойство спектральной оценки, а именно ее смещение. Другое важное свойство описывается ее дисперсией. В разд. 6.3.4 было получено приближенное выражение для дисперсии в частном случае белого шума при использовании окна Бартлетта. Теперь мы обобщим этот результат на случай произвольного процесса и произвольного окна. Зная дисперсию, можно на любой частоте построить доверительный интервал для истинного спектра. В этом разделе показано, что если две частоты стоят друг от друга достаточно далеко, то ковариация оценок на этих частотах почти равна нулю. Поэтому для таких частот доверительные интервалы можно строить независимо.

6.4.1. Ковариация сглаженных спектральных оценок

Вывод точного выражения для ковариации сглаженных оценок на двух частотах довольно сложен. Поэтому здесь мы дадим эвристический вывод результатов, а более подробное изложение будет приведено в приложении П9.1.

В способе, излагаемом здесь, мы воспользуемся тем фактом (см. (5.2.6)), что любой случайный процесс $X(t)$ со спектром $\Gamma_{XX}(f)$ можно представить в виде белого шума $Z(t)$, пропущенного через линейный фильтр. Воспользовавшись этим фактом и формулами разд. 6.3.3 для ковариаций оценок, соответствующих выборочному спектру, в случае белого шума, мы сможем вывести выражения для аналогичных ковариаций, но в случае произвольного случайного процесса. Затем уже несложно получить выражения для ковариаций сглаженных спектральных оценок.

Ковариация оценок, соответствующих выборочному спектру. Рассмотрим случайный процесс $X(t)$ со спектром $\Gamma_{XX}(f)$, получаемый из белого шума $Z(t)$ по формуле

$$X(t) = \int_0^{\infty} h(u) Z(t-u) du, \quad -\infty \leq t \leq \infty. \quad (6.4.1)$$

Согласно (6.2.16), спектр этого процесса можно записать в виде

$$\Gamma_{XX}(f) = \sigma_Z^2 |H(f)|^2, \quad -\infty \leq f \leq \infty. \quad (6.4.2)$$

Для конечного отрезка процесса $X(t)$ (6.4.1) можно приближенно записать следующим образом:

$$X(t) = \int_0^{\infty} h(u) Z(t-u) du \approx X_T(t), \quad -T/2 \leq t \leq T/2, \quad (6.4.3)$$

где

$$X_T(t) = \int_0^{\infty} h(u) Z_T(t-u) du, \quad -T/2 \leq t \leq T/2. \quad (6.4.4)$$

В (6.4.4) $Z_T(t)$ обозначает конечный отрезок процесса $Z(t)$. На интервале $-T/2 \leq t \leq T/2$ два процесса $X(t)$ и $X_T(t)$ будут идентичны, за исключением некоторого участка вблизи начала интервала, при условии, что отклик на единичный импульс $h(u)$ убывает до нуля за время, малое по сравнению с T . Мы предположим, что этим «начальным эффектом» можно пренебречь.

В таком случае оценку, соответствующую выборочному спектру

$$C_{XX}(f) = \frac{1}{T} \left| \int_{-T/2}^{T/2} X(t) e^{-j2\pi ft} dt \right|^2,$$

можно приближенно записать в виде

$$C_{XX}(f) \approx \frac{1}{T} \left| \int_{-T/2}^{T/2} X_T(t) e^{-j2\pi f t} dt \right|^2 = \\ = \frac{1}{T} \left| \int_{-T/2}^{T/2} \int_0^{\infty} h(u) Z_T(t-u) du e^{-j2\pi f t} dt \right|^2 = |H(f)|^2 C_{ZZ}(f). \quad (6.4.5)$$

Таким образом, оценка, соответствующая выборочному спектру, для процесса $X(t)$ приближенно равна соответствующей оценке для белого шума, умноженной на квадрат модуля частотной характеристики фильтра. Поскольку $2C_{ZZ}(f)/\sigma_Z^2$ распределена приближенно как χ^2 с двумя степенями свободы при любых f , то из (6.4.5) следует, что величина

$$\frac{2C_{XX}(f)}{\sigma_Z^2 |H(f)|^2} = \frac{2C_{XX}(f)}{\Gamma_{XX}(f)} \quad (6.4.6)$$

также распределена приближенно как χ^2 с двумя степенями свободы. Теперь можно использовать результаты разд. 6.3.3 для спектральных оценок белого шума. Так как $E[C_{ZZ}(f)] = \sigma_Z^2$, то из (6.3.16) получаем

$$E[C_{XX}(f)] \approx |H(f)|^2 \sigma_Z^2 = \Gamma_{XX}(f). \quad (6.4.7)$$

Аналогично, так как

$$\text{Cov}[C_{XX}(f_1), C_{XX}(f_2)] = \text{Cov}[|H(f_1)|^2 C_{ZZ}(f_1), |H(f_2)|^2 C_{ZZ}(f_2)] = \\ = |H(f_1)|^2 |H(f_2)|^2 \text{Cov}[C_{ZZ}(f_1), C_{ZZ}(f_2)],$$

то из (6.3.17) следует, что

$$\text{Cov}[C_{XX}(f_1), C_{XX}(f_2)] \approx \\ \approx |H(f_1)|^2 |H(f_2)|^2 \sigma_Z^4 \left[\left(\frac{\sin \pi T (f_1 + f_2)}{\pi T (f_1 + f_2)} \right)^2 + \left(\frac{\sin \pi T (f_1 - f_2)}{\pi T (f_1 - f_2)} \right)^2 \right], \quad (6.4.8)$$

где мы пренебрегли членом, содержащим K_4 . Так как $\Gamma_{XX}(f) = \sigma_Z^2 |H(f)|^2$, то на двух разных частотах f_1 и f_2 ковариация оценок, соответствующих выборочному спектру, для линейного процесса равна

$$\text{Cov}[C_{XX}(f_1), C_{XX}(f_2)] \approx \Gamma_{XX}(f_1) \Gamma_{XX}(f_2) \left[\left(\frac{\sin \pi T (f_1 + f_2)}{\pi T (f_1 + f_2)} \right)^2 + \right. \\ \left. + \left(\frac{\sin \pi T (f_1 - f_2)}{\pi T (f_1 - f_2)} \right)^2 \right]. \quad (6.4.9)$$

Формула (6.4.9) показывает, что для любого гауссовского случайного процесса $X(t)$

$$\text{Cov}[C_{XX}(f_1), C_{XX}(f_2)] \approx O\left(\frac{1}{T^2}\right), \quad f_1 \neq f_2,$$

$$\text{Var}[C_{XX}(f)] \approx \Gamma_{XX}^2(f).$$

Таким образом, мы получили обобщение результатов разд. 6.3.3, которые были получены только для белого шума. Заметим, что для больших T выражение в квадратных скобках в (6.4.9) ведет себя подобно δ -функции с множителем $1/T$. Кроме того, ковариация в точности равна нулю, когда частоты $(f_1 + f_2)$ и $(f_1 - f_2)$ кратны величине $1/T$.

Ковариация сглаженных спектральных оценок. Из (6.3.30) сглаженную спектральную оценку $\bar{C}_{XX}(f)$ для процесса $X(t)$ можно записать в виде

$$\bar{C}_{XX}(f) = \int_{-\infty}^{\infty} C_{XX}(g) W(f-g) dg,$$

а, следовательно, ковариация $\bar{C}_{XX}(f_1)$ и $\bar{C}_{XX}(f_2)$ равна

$$\text{Cov}[\bar{C}_{XX}(f_1), \bar{C}_{XX}(f_2)] =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(f_1-g) W(f_2-h) \text{Cov}[C_{XX}(g), C_{XX}(h)] dg dh.$$

Заменяя $\text{Cov}[C_{XX}(g), C_{XX}(h)]$ на (6.4.9) и интегрируя по h , получаем

$$\text{Cov}[\bar{C}_{XX}(f_1), \bar{C}_{XX}(f_2)] \approx$$

$$\approx \frac{1}{T} \int_{-\infty}^{\infty} \Gamma_{XX}^2(g) W(f_1-g) [W(f_2+g) + W(f_2-g)] dg, \quad (6.4.10)$$

при условии, что T настолько велико, что члены $\sin^2 \pi f T / (\pi f T)^2$ ведут себя как δ -функции. Равенство (6.4.10) является окончательным результатом, но можно еще вывести полезное приближение, предположив, что $\Gamma_{XX}(f)$ изменяется плавно на ширине спектрального окна $W(f)$. При этом предположении (6.4.10) переходит в

$$\text{Cov}[\bar{C}_{XX}(f_1), \bar{C}_{XX}(f_2)] \approx$$

$$\approx \frac{\Gamma_{XX}^2(f)}{T} \int_{-\infty}^{\infty} W(f_1-g) [W(f_2+g) + W(f_2-g)] dg, \quad (6.4.11)$$

где $f_1 \leq f \leq f_2$.

Равенство (6.4.11) показывает, что ковариация сглаженных спектральных оценок пропорциональна площади перекрытия спектральных окон с центрами в f_1 и f_2 . Следовательно, если спектральные окна почти не перекрываются, ковариация будет очень малой. Некоторые численные значения для ковариаций сглаженных спектральных оценок при использовании различных окон будут даны в разд. 7.2.

Дисперсия сглаженных спектральных оценок. Если $f_1=f_2=f$, то (6.4.10) сводится к

$$\text{Var} [\bar{C}_{XX}(f)] \approx \frac{\Gamma_{XX}^2(f)}{T} \int_{-\infty}^{\infty} W^2(g) dg, \quad (6.4.12)$$

где мы пренебрегли членом $\int_{-\infty}^{\infty} W(g)W(g+2f)dg$, малым по сравнению с $\int_{-\infty}^{\infty} W^2(g) dg$. Воспользовавшись теоремой Парсеваля, равенство (6.4.12) можно переписать в эквивалентном виде

$$\text{Var} [\bar{C}_{XX}(f)] \approx \frac{\Gamma_{XX}^2(f)}{T} \int_{-\infty}^{\infty} w^2(u) du = \Gamma_{XX}^2(f) \frac{I}{T}. \quad (6.4.13)$$

Например, для окна Бартлетта $w_B(u)$ из табл. 6.5 имеем

$$I = \int_{-M}^M \left(1 - \frac{|u|}{M}\right)^2 du = \frac{2}{3} M$$

и, следовательно,

$$\text{Var} [\bar{C}_{XX}(f)] \approx \frac{\Gamma_{XX}^2(f)}{T} \left(\frac{2}{3} M\right).$$

Это показывает, что дисперсию сглаженной спектральной оценки можно уменьшить, выбрав точку отсечения M корреляционного окна малой. Но, как указывалось в разд. 6.3.5, при уменьшении M увеличивается смещение, искажающее теоретический спектр, так как спектральное окно при этом расширяется. В таком случае, как показывает формула (6.4.10), спектральные оценки на соседних частотах будут сильнее коррелированы из-за более полного перекрытия спектральных окон. Поэтому точный выбор M является очень важным вопросом. Этот вопрос обсуждается в гл. 7.

Заметим, что поскольку $\text{Var} [C_{XX}(f)] \approx \Gamma_{XX}^2(f)$, то величина

$$\frac{\text{Var} [\bar{C}_{XX}(f)]}{\Gamma_{XX}^2(f)} = \frac{1}{T} \int_{-\infty}^{\infty} w^2(u) du = I/T \quad (6.4.14)$$

равна относительному уменьшению дисперсии, вызванному сглаживанием, т. е. использованием сглаженной спектральной оценки вместо оценки, соответствующей выборочному спектру. Значения отношений (6.4.14), соответствующих спектральным окнам из табл. 6.5, приведены в третьем столбце табл. 6.6. Видно, что все они имеют вид $c(M/T)$, где c — некоторая константа, зависящая от окна.

Таблица 6.6

Свойства спектральных окон

Название окна	Формула окна	Отношение дисперсий I/T	Число степеней свободы	Нормированная ширина полосы частот b_1
Прямоугольное	$2M \frac{\sin 2\pi fM}{2\pi fM}$	$2 \frac{M}{T}$	$\frac{T}{M}$	0,5
Бартлетта	$M \left(\frac{\sin \pi fM}{\pi fM} \right)^2$	$0,667 \frac{M}{T}$	$3 \frac{T}{M}$	1,5
Тьюки	$M \frac{\sin 2\pi fM}{2\pi fM} \times \frac{1}{1 - (2fM)^2}$	$0,75 \frac{M}{T}$	$2,667 \frac{T}{M}$	1,333
Парзена	$\frac{3}{4} M \left(\frac{\sin (\pi fM/2)}{\pi fM/2} \right)^4$	$0,539 \frac{M}{T}$	$3,71 \frac{T}{M}$	1,86

Предположим, например, что точка отсечения M равна $0,1T$. Тогда для окна Бартлетта $I/T = 2/3 (0,1) = 0,067$. Следовательно, беря точку отсечения на расстоянии 10% длины записи, мы снизим дисперсию сглаженной спектральной оценки до 6,7% от дисперсии оценки, соответствующей выборочному спектру. Соответствующие величины для окон Тьюки и Парзена равны 7,5% и 5,4% соответственно. Следовательно, при фиксированном M из трех рассматриваемых окон наименьшую дисперсию дает окно Парзена. Это объясняется тем, что, как видно из рис. 6.13, окно Парзена является более широким и плоским, чем два остальных. В результате оно приводит к большим смещениям. Поэтому сравнения окон, сделанные только с учетом дисперсии, могут ввести в заблуждение, как мы увидим позднее.

6.4.2. χ^2 -приближение к распределению сглаженных спектральных оценок

В разд. 6.3.5 было показано, что оценка, соответствующая выборочному спектру $S_{xx}(f)$, такова, что величина $2S_{xx}(f)\Gamma_{xx}(f)$ имеет приблизительно χ^2 -распределение с двумя степенями свободы. В этом разделе мы покажем, что соответствующий результат

для сглаженной спектральной оценки состоит в том, что величина $\sqrt{\bar{C}_{XX}(f)}/\Gamma_{XX}(f)$ распределена приближенно как χ^2 с ν степенями свободы, где $\nu > 2$. Это означает, что сглаженные спектральные оценки будут иметь гораздо больше степеней свободы, чем оценка, соответствующая выборочному спектру, что приводит к уменьшению их дисперсии.

Оценка $C_{XX}(f)$ есть преобразование Фурье оценки ковариационной функции $c_{XX}(u)$, причем $c_{XX}(u) = 0$ вне интервала $-T \leq u \leq T$. Если внутри интервала $-T \leq u \leq T$ функция $c_{XX}(u)$ представляется некоторой периодической функцией $c_{XX}^p(u)$, такой, что $c_{XX}^p(u) = c_{XX}^p(u+2T)$, то функция $c_{XX}^p(u)$ представляется в виде ряда Фурье

$$c_{XX}^p(u) = \sum_{i=-\infty}^{\infty} C_{XX}^p\left(\frac{i}{2T}\right) e^{j(2\pi i u/2T)}.$$

Поскольку корреляционное окно $w(u) = 0$ при $|u| \geq M$, функции $\bar{c}_{XX}(u) = c_{XX}(u)w(u)$ и $\bar{c}_{XX}^p(u) = c_{XX}^p(u)w(u)$ совпадают при всех u , так что сглаженная спектральная оценка имеет два эквивалентных представления

$$\bar{C}_{XX}(f) = \int_{-\infty}^{\infty} W(f-g) C_{XX}(g) dg$$

и

$$\bar{C}_{XX}(f) = \sum_{l=-\infty}^{\infty} W\left(f - \frac{l}{2T}\right) C_{XX}^p\left(\frac{l}{2T}\right).$$

Но

$$\frac{C_{XX}(l/2T)}{2T} = C_{XX}^p\left(\frac{l}{2T}\right),$$

и, следовательно,

$$\bar{C}_{XX}(f) = \frac{1}{2T} \sum_{l=-\infty}^{\infty} C_{XX}\left(\frac{l}{2T}\right) W\left(f - \frac{l}{2T}\right).$$

Таким образом, сглаженная спектральная оценка является взвешенной суммой случайных величин $C_{XX}(l/2T)$ на субгармонических частотах $l/2T$. Эти случайные величины распределены как χ^2 с двумя степенями свободы. Следовательно, пользуясь результатами разд. 3.3.5, распределение величины $\bar{C}_{XX}(f)$ можно приблизить с помощью распределения величины $a\chi_{\nu}^2$, где a — константа, и χ_{ν}^2 — случайная величина, имеющая χ^2 -распределение с ν степенями

свободы. Из (3.3.14) и (3.3.15) можно вычислить константы a и v :

$$v \approx \frac{2 \{E [\bar{C}_{XX}(f)]\}^2}{\text{Var} [\bar{C}_{XX}(f)]}, \quad (6.4.15)$$

$$a \approx \frac{E[\bar{C}_{XX}(f)]}{v}. \quad (6.4.16)$$

Предполагая, что истинный спектр изменяется плавно по сравнению со спектральным окном, получаем из (6.3.36)

$$E [\bar{C}_{XX}(f)] \approx \Gamma_{XX}(f)$$

и из (6.4.13)

$$\text{Var} [\bar{C}_{XX}(f)] \approx \frac{\Gamma_{XX}^2(f)}{T} \int_{-\infty}^{\infty} w^2(u) du.$$

Поэтому, подставляя эти выражения в (6.4.15) и (6.4.16), имеем

$$v = \frac{2T}{\int_{-\infty}^{\infty} w^2(u) du} = \frac{2T}{I}, \quad (6.4.17)$$

$$a = \frac{\Gamma_{XX}(f)}{v}. \quad (6.4.18)$$

Следовательно, случайная величина $v\bar{C}_{XX}(f)/\Gamma_{XX}(f)$ имеет χ^2 -распределение с v степенями свободы, где v задается равенством (6.4.17). Таким образом, число степеней свободы сглаженной спектральной оценки зависит от окна $w(u)$.

В столбце 4 табл. 6.6 приведены степени свободы, соответствующие спектральным окнам, указанным в столбце 2. Например, если используется окно Бартлетта с точкой отсечения M на расстоянии одной десятой длины записи (т. е. $M/T=0,1$), то число степеней свободы оценки равно $3/0,1=30$. Чем больше число степеней свободы, тем надежнее оценка в том смысле, что ее дисперсия меньше. Однако, как указывалось выше, должен выбираться некоторый компромисс между числом степеней свободы и смещением.

Из табл. 6.6 видно, что широкое окно, такое, как окно Парзена $W_P(f)$, дает меньшую дисперсию и, следовательно, большее число степеней свободы, чем более узкое окно, такое, как окно Бартлетта $W_B(f)$. Это находится в согласии со сделанным выше замечанием о том, что чем шире окно, тем меньше дисперсия.

6.4.3. Доверительные границы для спектра

Поскольку $\sqrt{v} \bar{C}_{XX}(f) / \Gamma_{XX}(f)$ имеет χ^2 -распределение с v степенями свободы, где v задается равенством (6.4.17), то отсюда получаем

$$\Pr \left\{ x_v \left(\frac{\alpha}{2} \right) < \frac{\sqrt{v} \bar{C}_{XX}(f)}{\Gamma_{XX}(f)} \leq x_v \left(1 - \frac{\alpha}{2} \right) \right\} = 1 - \alpha, \quad (6.4.19)$$

где $x_v(\alpha/2)$ определяется из равенства $\Pr \{ \chi_v^2 \leq x_v(\alpha/2) \} = \alpha/2$. Отсюда, используя точно такие же рассуждения, что и в разд. 3.3.2, получаем, что интервал

$$\frac{\sqrt{v} \bar{C}_{XX}(f)}{x_v [1 - (\alpha/2)]}, \quad \frac{\sqrt{v} \bar{C}_{XX}(f)}{x_v (\alpha/2)} \quad (6.4.20)$$

является $100(1 - \alpha) \%$ -ным доверительным интервалом для $\Gamma_{XX}(f)$. Для заданного отношения T/M значение v , соответствующее данному спектральному окну, можно взять из столбца 4 табл. 6.6. После этого доверительный интервал можно вычислить по (6.4.20), взяв с рис. 3.10 множители $v/x_v(\alpha/2)$ и $v/x_v[1 - (\alpha/2)]$. Например, выборочная сглаженная спектральная оценка, приведенная на рис. 6.10, была получена с помощью окна Бартлетта при $M/T = 0,125$. Поэтому из табл. 6.5 находим $v = 3/0,125 = 24$. На частоте $f = 0,1$ гц $\bar{C}_{ZZ}(f) = 0,804$, и, пользуясь рис. 3.10, находим 95%-ные доверительные границы для $\Gamma_{ZZ}(f)$:

$$0,61 \cdot 0,804 = 0,49; \quad 1,94 \cdot 0,804 = 1,56.$$

Аналогично 95%-ные доверительные границы для $\Gamma_{ZZ}(f)$, полученные с помощью несглаженного выборочного спектра на той же частоте $f = 0,1$ гц, равны

$$0,27 \cdot 0,622 = 0,169; \quad 39,5 \cdot 0,622 = 24,6.$$

Эти границы значительно шире, так как при этом выборочной оценке соответствует меньшее число степеней свободы.

Заметим, что равенство (6.4.19) дает доверительный интервал для $\Gamma_{XX}(f)$ лишь на одной конкретной частоте f . Если задать доверительные интервалы на q частотах, на которых оценки независимы, то уровень доверия будет $(1 - \alpha)^q$, что обычно значительно меньше, чем $1 - \alpha$. Отметим еще, что дисперсия будет полно характеризовать свойства оценки лишь в том случае, когда мало смещение, как отмечалось в разд. 6.3.5. Поэтому построенные выше доверительные интервалы будут иметь значение лишь тогда, когда спектральное окно достаточно узкое, так что нет заметного смещения.

Доверительные интервалы в логарифмическом масштабе. В разд. 7.1.2 будет показано, что выборочные спектральные оценки

нужно строить в логарифмическом масштабе, так чтобы изменчивость спектра могла быть выражена удобным образом. Логарифмический масштаб является также разумным с технической точки зрения, так как обычно важны *относительные* изменения мощности. Со статистической точки зрения также важно строить спектры в логарифмическом масштабе, так как при этом построение доверительного интервала для спектра сводится к откладыванию около выборочной спектральной оценки одного и того же интервала для всех частот. Таким образом, из (6.4.20) доверительный интервал для $\lg \Gamma_{xx}(f)$ равен

$$\lg \bar{C}_{xx}(f) + \lg \frac{v}{x_y [1 - (a/2)]}, \quad \lg \bar{C}_{xx}(f) + \lg \frac{v}{x_y (a/2)}. \quad (6.4.21)$$

Поэтому при построении выборочной оценки спектра доверительный интервал для всех частот можно указать одним вертикальным отрезком.

Рассмотрим, например, выборочную сглаженную спектральную оценку $\bar{C}_{xx}(f)$ на рис. 6.10, для которой $v=24$. Из рис. 3.10 и (6.4.21) находим, что 95%-ные доверительные интервалы для $\lg \Gamma_{xx}(f)$ равны

$$\lg \bar{C}_{xx}(f) + \lg 0,61; \quad \lg \bar{C}_{xx}(f) + \lg 1,94.$$

Для $\bar{C}_{xx}(f)$, построенной на логарифмической бумаге, 95%-ный доверительный интервал можно было бы получить, просто построив точки (0,61; 1,0; 1,94), взятые с рис. 3.10, в виде вертикального отрезка в логарифмическом масштабе. Этот способ мы проиллюстрируем в разд. 7.2 и в других местах книги.

6.4.4. Ширина полосы частот спектрального окна

В разд. 6.4.1 было показано, что полезную характеристику спектрального окна дает величина $I = \int_{-\infty}^{\infty} w^2(u) du$, так как I/T есть мера уменьшения дисперсии оценки, обусловленного сглаживанием с помощью спектрального окна. Следовательно, для получения небольшой дисперсии нужно выбрать $w(u)$ так, чтобы I было мало. Для заданного окна этого можно достичь, уменьшив M . Полезной характеристикой окна является также его ширина. В следующих разделах будет показано, что для получения хорошей оценки пика спектра «ширина» спектрального окна должна быть того же порядка, что и ширина пика. Поскольку спектральное окно отлично от нуля для большинства частот f в диапазоне $-\infty \leq f \leq \infty$, необходимо определить точнее понятие «ширины» спектрального окна.

Один способ определения ширины, или *ширины полосы частот*, спектрального окна, который используют статистики [9], состоит

в следующем. Рассматривают «полосовое» спектральное окно

$$W(f) = \frac{1}{h}, \quad -\frac{h}{2} \leq f \leq \frac{h}{2}.$$

Это спектральное окно представляет собой прямоугольник в частотной области, ширина которого равна h ; таким образом, ширина полосы частот этого окна $b = h$. Из (6.4.13) получаем дисперсию сглаженной спектральной оценки, использующей это спектральное окно:

$$\text{Var} [\bar{C}_{XX}(f)] \approx \frac{\Gamma_{XX}^2(f)}{T} \cdot \frac{1}{h} = \frac{\Gamma_{XX}^2(f)}{Tb}.$$

Для оценки, использующей спектральное окно отличное от прямоугольного, естественно определить ширину полосы частот окна как ширину такого прямоугольного окна, которое дает ту же самую дисперсию, т. е.

$$\text{Var} [\bar{C}_{XX}(f)] \approx \frac{\Gamma_{XX}^2(f)}{T} \cdot \frac{1}{b} = \frac{\Gamma_{XX}^2(f)}{T} \int_{-\infty}^{\infty} w^2(u) du. \quad (6.4.22)$$

Отсюда ширина полосы частот равна

$$b = \frac{1}{I} = \frac{1}{\int_{-\infty}^{\infty} w^2(u) du} = \frac{1}{\int_{-\infty}^{\infty} W^2(f) df}. \quad (6.4.23)$$

Например, для прямоугольного корреляционного окна $w_R(u)$ и корреляционного окна Бартлетта $w_B(u)$ из табл. 6.5 значения ширины полосы частот равны $1/2M$ и $1/3M$ соответственно.

Иногда удобнее пользоваться *нормированной шириной полосы частот* b_1 , соответствующей значению $M=1$. Величина b_1 определяется из соотношения

$$b = \frac{b_1}{M} = \frac{1}{\int_{-\infty}^{\infty} w^2(u) du}. \quad (6.4.24)$$

Например, значения нормированной ширины полосы частот для прямоугольного корреляционного окна и окна Бартлетта равны $1/2$ и $3/2$ соответственно.

В пятом столбце табл. 6.6 приведены значения нормированной ширины для окон из табл. 6.5. Мы видим, что окно Парзена w_P имеет нормированную ширину полосы частот примерно в 1,4 раза больше, чем окно Тьюки w_T .

Инженеры узнают в выражении (6.4.23) определение ширины полосы частот шума, пропущенного через фильтр. Точное определение ширины полосы частот не очень существенно. Например, некоторые авторы [10] используют в качестве такого определения расстояние между точками, в которых мощность убывает до половины своего максимального значения. Мы предпочли определение (6.4.23) из-за

того, что оно использует всю форму спектрального окна и поэтому с помощью этого определения легче отличать окна по их форме, чем по определению, основанному на точках половины максимума. Из (6.4.23) видно, что дисперсия спектральной оценки обратно пропорциональна ширине полосы частот спектрального окна. Действительно, из (6.4.22) и (6.4.23) получаем

$$(\text{Дисперсия}) \times (\text{Ширина полосы частот}) = \text{Константа}. \quad (6.4.25)$$

Следовательно, небольшая дисперсия соответствует широкой полосе частот и большая дисперсия — узкой полосе частот. Кроме того, равенство (6.4.17) показывает, что число степеней свободы ν сглаженной оценки равно

$$\nu = \frac{2T}{T} = 2 \left(\frac{T}{M} \right) b_1. \quad (6.4.26)$$

Следовательно, из того, что полоса частот широкая, вытекает, что число степеней свободы сглаженной оценки велико, а дисперсия мала. Обратно, из того, что ширина полосы частот невелика, следует, что число степеней свободы мало и, следовательно, дисперсия велика. Поскольку в разд. 6.3.5 было показано, что смещение уменьшается при увеличении M , то отсюда следует, что малому смещению соответствует и малое значение ширины полосы частот.

В следующей главе введенные здесь понятия применяются к практическим задачам оценивания спектра наблюдаемых временных рядов.

ЛИТЕРАТУРА

1. Doob J. L., Stochastic Processes, John Wiley, New York, 1953. (Русский перевод: Дуб Дж., Вероятностные процессы, М., ИЛ, 1956.)
2. Laping J. H., Battin R. H., Random Processes in Automatic Control, McGraw-Hill, New York, 1956. (Русский перевод: Лэнинг Дж. Х., Бэттин Р. Г., Случайные процессы в задачах автоматического управления, М., ИЛ, 1958.)
3. James H. M., Nichols N. B., Phillips R. S., Theory of Servomechanisms, McGraw-Hill, New York, 1947. (Русский перевод: Джеймс, Никольс, Филлипс, Теория следящих систем, М., ИЛ, 1953.)
4. Hald A., Statistical Theory with Engineering Applications, John Wiley, New York, 1952. (Русский перевод: Хальд А., Математическая статистика с техническими приложениями, М., ИЛ, 1956.)
5. Bartlett M. S., An Introduction to Stochastic Processes with Special Reference to Methods and Applications, Cambridge Univ. Press, Cambridge, 1953. (Русский перевод: Бартлетт М., Введение в теорию случайных процессов, М., ИЛ, 1958.)
6. Blackman R. B., Tukey J. W., The Measurement of Power Spectra from the Point of View of Communications Engineering, Dover, New York, 1958.
7. Grenander U., Rosenblatt M., Statistical Analysis of Stationary Time Series, John Wiley, New York, 1957.
8. Parzen E., Technometrics, 3, 167 (1961).
9. Jenkins G. M., Technometrics, 3, 133 (1961).
10. Priestley M. B., Technometrics, 4, 551 (1962).

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА К ВЫПУСКУ 1

- 1*. Яглом А. М., Введение в теорию стационарных случайных функций, Усп. мат. наук, 7, вып. 5 (51), 3—162 (1952).
- 2*. Рао С. Р., Линейные статистические методы и их приложения, М., изд-во «Наука», 1968.
- 3*. Линник Ю. В., Метод наименьших квадратов и основы теории обработки наблюдений, ГИФМЛ, М., 1962.
- 4*. Гельфанд И. М., Шиллов Г. Е., Обобщенные функции и действия над ними, М., ГИФМЛ, 1958.
- 5*. Бейтман Г., Эрдейи А., Таблицы интегральных преобразований, I, изд-во «Наука», М., 1969.
- 6*. Большев Л. Н., Смирнов Н. В., Таблицы математической статистики, М., изд-во «Наука», 1968.
- 7*. Шеффе Г., Дисперсионный анализ, М., ГИФМЛ, 1963.
- 8*. Джалаларидзе К. О., Об оценке параметров спектра гауссовского стационарного процесса с рациональной спектральной плотностью, Теория вероятностей и ее применения, XV, № 3, 548—554 (1970).
- 9*. Ибрагимов И. А., Об оценке методом максимального правдоподобия параметров спектральной плотности стационарного процесса, Теория вероятностей и ее применения, XII, № 1, 128—134 (1967).
- 10*. Хеннан Э., Анализ временных рядов, М., изд-во «Наука», 1964.
- 11*. Мак-Кракен Д., Дорн У., Численные методы и программирование на ФОРТРАНе, М., изд-во «Мир», 1969.

УКАЗАТЕЛЬ

- Автоковариационная функция 19, 194
Автокорреляционная функция 19, 182, 185
Амплитуда 38, 60
Ансамбль 179
- Башелье — Винера процесс 198
Белый шум с ограниченной полосой частот 186
Боде графики 60—61
- Вариационное исчисление 249
Вероятность 81
— плотность 84
Вероятные области 152
Весовая функция 27, 53—55; см. Функция отклика на единичный импульс
Временное окно 68
— — ширина 70
Временные ряды 175
Выборочная оценка ковариаций 220
— — корреляционной функции 224
— — максимального правдоподобия 150
— — среднего правдоподобия 155
— — спектральная плотность 257, 261
Выборочное распределение 102, 115, 117
— дисперсии 104
Выборочный спектр 257
— — белого шума 258
— — свойства оценок 287
- Гармоники 36
Гаусс К. Ф. 116
Гаусса теорема 135
Гистограмма 86
- Данные акселерометра 87, 89
— ионосферные 37
— о сроке службы ламп 149
— о токе коллектора 83
— о транзисторах 78, 121, 147, 160
— радиолокатора 177
— «скорость—время» 135
— турбогенератора 17, 19—20
Двойники 72
Дельта-функция 47
Диаграмма разброса 87
Дискретизация сигнала 70
Дисперсионный анализ 282
Дисперсия 92, 96
— нелинейной функции 100
Доверительные интервалы 119—124, 133, 137
— области 141, 171
- Единичного скачка функция 50
- Запаздывание 185
- Ковариационная функция выборочная 213
— — линейного процесса 194
— — свойства оценок 214
— — эргodicность 220
Ковариация 95, 99
Комплексное сопряжение 40
Комплексные амплитуды 39
Корреляционная функция 182
— — оценивание 222
— — процесса авторегрессии 200
— — свойства 192
Корреляционное окно 290, 294
Коэффициент корреляции 97
— усиления 60

Критерий минимума среднеквадратичной ошибки 189
— значимости 116, 131

Линейная теория наименьших квадратов 165

Линейные системы 52
— уравнения в конечных разностях 65

Линейный процесс 195
— — ковариационная функция
— — моменты 251
— — оценивание параметров 230
— — пример 196
— — спектр 274

Маргинальное правдоподобие 160
— распределение вероятностей 89

Математическое ожидание 91, 96

Матрица ковариаций 90, 113
— — оценок 142, 168

Метод выборочных распределений 117
— максимального правдоподобия 230

— наименьших квадратов 211

— правдоподобия 115

Моменты выборочные 103
— линейных функций от случайных величин 94, 113
— многомерные 94

— случайного процесса 182

— случайных величин 91

— старшие 94

Мощность средняя 39

Найквистова частота 257

Наложение частот 70

Нестационарность среднего значения 187

— — — и дисперсии 188

Нормальное распределение 159

Нормальные уравнения 135

Обобщенные функции 46

Обобщенный принцип наименьших квадратов 166

Оператор сдвига 67

Ортогональность оценок 143

Оценивание остаточной дисперсии 170

Оценка 118—120

— выбор 118

— максимального правдоподобия 119, 126

— свойства 124—126

Парсеваля теорема 38, 76

Передаточная функция 66

Период 34

Периодограмма 257; см. Выборочный спектр

Плотность вероятности 85

— — — нормальная 85

— — — двумерная 89

— — — нормированная 94

Полоса частот ограниченная 41

Постоянная времени 26, 53

Принцип наименьших квадратов 134
— правдоподобия 147

Проверка того, что шум белый 283

Программа MULTICOR 253

Процессы авторегрессии 200, 218

— — выборочные оценки 236

— — корреляционные функции 204

— — определение порядка 240

— — оценивание параметров 230

— — спектры 275

— скользящего среднего — авторегрессии 24, 205

— — — оценивание параметров 247

— — — спектры 278

— — — конечного порядка 199

— — — оценивание параметров 243

Разностные уравнения 65

Распределение вероятностей

— — безусловное 89; см. Маргинальное

— — биномиальное 81

— — многомерное 90

— — нормальное (гауссовское) 19, 93

— — Пуассона 93

— — χ^2 104, 111

Свертка 52, 64, 76

Сглаженные спектральные оценки 289

— — — дисперсия 303

— — — ковариация 302

— — — распределение 304

— — — свойства 299

Синтез следящих систем 189

Случайная величина 81

Случайное блуждание 188

— поле 178

Случайный процесс 179

— — строго стационарный 185

— — телеграфный сигнал 209

Событие 81

- Спектр 263
 - белого шума 272
 - доверительные границы 307
 - линейного процесса 274
 - мощности 21
 - процессов авторегрессии и скользящего среднего 275
 - случайного процесса 263
 - средний сглаженный 297
- Спектральная плотность 264
 - — нормированная 268
 - функция 270
- Спектральное окно 68, 290
 - — Бартлетта 292
 - — Парзена 294
 - — прямоугольное 294
 - — свойства 304
 - — Тьюки 294
 - — ширина полосы частот 308
- Спектральные оценки 279
 - — свойства 299
 - — сглаживание 289
- Среднее значение 91, 96
 - — нелинейной функции 99
- Среднеквадратичная ошибка 125
- Стандартное отклонение 93
- Стационарность 182
- Степени свободы 106, 110
- Стьюдента t -распределение 108, 132

- Теория наименьших квадратов 115, 134
 - проверки гипотез 131

- Уиттекера интерполяционная формула 71
- Усиление 59
- Условная плотность вероятности 181
- Устойчивость 59, 63, 66

- Фаза 38
- Фазовый угол 60
- Фильтр 61
 - цифровой 23
- Фишера количество информации 131, 152
 - F -распределение 110
- Функция отклика квадратичная 154
 - — на единичный импульс 54
 - — — скачок 55
 - — неквадратичная 154
 - — приближение 155
 - — свойства 147
 - правдоподобия 146
 - распределения 83
 - — двумерная 87
- Фурье анализ 33
 - интегралы 41
 - комплексные ряды 39
 - коэффициенты 34
 - линейчатый спектр 39
 - преобразование 42, 44—52
 - — свойства 74—76

- Характеристическое уравнение системы 63

- Центральная предельная теорема 85, 138

- Частота 36
 - основная 36
- Частная корреляционная функция 242
- Частотное распределение 79
- Частотные характеристики 57, 60

- Ширина полосы частот 61
- Шум белый 194, 229, 283
 - — выборочный спектр 258
 - — спектр 272
 - — способы генерации 273

- Эргодичность 270

- z -преобразования 66; см. Оператор сдвига

ОГЛАВЛЕНИЕ

Предисловие к русскому изданию	5
Предисловие	9
Обозначения	12
Глава 1. ЦЕЛИ И СРЕДСТВА АНАЛИЗА ВРЕМЕННЫХ РЯДОВ	15
1.1. Временные ряды и случайные процессы	15
1.2. Описание временных рядов во временной и частотной областях	17
1.3. Цели анализа временных рядов	24
1.4. Круг вопросов, рассмотренных в данной книге	31
Литература	31
Глава 2. АНАЛИЗ ФУРЬЕ	33
2.1. Введение	33
2.2. Преобразования Фурье и их свойства	44
2.3. Линейные системы и свертки	52
2.4. Применения в анализе временных рядов	63
Литература	73
Приложение П2.1. Операторные свойства преобразований Фурье	74
Глава 3. ТЕОРИЯ ВЕРОЯТНОСТЕЙ	78
3.1. Частотные распределения и распределения вероятностей	78
3.2. Моменты случайных величин	91
3.3. Выборочные распределения	101
Литература	112
Приложение П3.1. Моменты линейных функций от случайных величин	113
Глава 4. ВВЕДЕНИЕ В ТЕОРИЮ СТАТИСТИЧЕСКИХ ВЫВОДОВ	115
4.1. Историческое развитие теории статистических выводов	116
4.2. Применение метода выборочных распределений к статистическим выводам	117
4.3. Оценивание с помощью наименьших квадратов	134
4.4. Выводы, основанные на функции правдоподобия	146
4.5. Резюме	162
Литература	164
Приложение П4.1. Линейная теория наименьших квадратов	165
Литература	174
Глава 5. ВВЕДЕНИЕ В АНАЛИЗ ВРЕМЕННЫХ РЯДОВ	175
5.1. Стационарные и нестационарные случайные процессы	175
5.2. Корреляционная и ковариационная функции	192

5.3. Оценивание ковариационных функций	210
5.4. Оценивание параметров линейного процесса	230
Литература	248
Приложение П5.1. Вариационное исчисление	249
Приложение П5.2. Моменты линейного процесса	251
Приложение П5.3. Логическая схема программы вычисления ковариаций	253
Глава 6. СПЕКТР	255
6.1. Выборочный спектр	255
6.2. Спектр	263
6.3. Спектральные оценки	279
6.4. Дальнейшие свойства сглаженных спектральных оценок	299
Литература	310
ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА К ВЫПУСКУ 1	311
Указатель	312

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., дом 2, издательство «Мир».

Г. Дженкинс, Д. Ваттс

СПЕКТРАЛЬНЫЙ АНАЛИЗ И ЕГО ПРИЛОЖЕНИЯ

Том I

Редактор Э. А. Медушевская
Художник А. Д. Смеляков
Художественный редактор В. М. Варлашин
Технический редактор Е. Н. Лебедева

Сдано в набор 26/I 1971 г.
Подписано к печати 18/VI 1971 г.
Бумага кн.-журн. 60×90¹/₁₆=10 бум. л.
20 печ. л. Уч.-изд. л. 18,26.
Изд. № 27/5888.

Цена 1 р. 06 к. Зак. 1210.

ИЗДАТЕЛЬСТВО «МИР»

Москва, 1-й Рижский пер., 2

Главполиграфпром Комитета по печати при Совете Министров СССР. Отпечатано в ордена Трудового Красного Знамени Ленинградской типографии № 2 имени Евгении Соколовой, Измайловский пр., 29.
Набрано в Ленинградской типографии № 8, Прачечный пер., д. 6.

ГОТОВИТСЯ К ИЗДАНИЮ В 1972 ГОДУ

Бонди Г., Гипотезы и мифы в физической теории,
Англия, 1967 (с дополн. 1970 г.), перев. с англ.

Бурное развитие физики и астрофизики с особой остротой поставило вопрос о роли теории. Как связана теория и наблюдения? Как возникают новые теоретические идеи? Как они проявляются и уточняются?

Видный физик-теоретик Г. Бонди рассматривает эти интересные и важные вопросы на примере частной и общей теории относительности, а также космологии. Он прослеживает, как последовательное развитие идей классической ньютоновской механики неизбежно приводит к теории относительности, как дальнейшее развитие общей теории относительности решающим образом зависит от последующего развития экспериментальных исследований, в частности изучения гравитационных волн.

Русский перевод книги дополнен переводом популярной статьи автора о гравитационных волнах и проблемах их обнаружения.

Расчитана на широкие круги читателей, интересующихся наукой, прежде всего школьников старших классов и преподавателей средней школы.

Заказы направляйте в местные книготорги

ГОТОВИТСЯ К ИЗДАНИЮ В 1972 ГОДУ

Чепмен С., Линдзен Р., Атмосферные приливы, Голландия, 1970, перев. с англ.

Данная монография является фактически первой в мировой литературе по этому вопросу. В гл. I рассматривается история вопроса и вводятся основные понятия. Во второй главе рассматриваются данные многочисленных наблюдений и их обработки, позволившие выделить у поверхности Земли солнечные суточные и лунные приливы. В третьей главе излагаются вопросы количественной теории атмосферных приливов, указаны основные ограничения этой теории и возможные пути ее дальнейшего развития. Последнее делает книгу особенно ценной. Весь материал изложен очень четко и логично.

Книга представляет большой интерес для метеорологов и особенно для специалистов по аэрономии и физике верхней атмосферы.

Заказы направляйте в местные книготорги

ГОТОВИТСЯ К ИЗДАНИЮ В 1972 ГОДУ

Р. Дикке, Гравитация и Вселенная, США, 1970, перевод. с англ.

Содержание книги составляют три лекции Роберта Дикке, прочитанные в Американском философском обществе. Первая лекция посвящена обзору знаменитого эксперимента по проверке эквивалентности тяготеющей и инертной масс.

Во второй лекции «Сплюсненность Солнца и тяготение» автор провел подробный анализ этой сложной проблемы, исходя как из теории Ньютона, так и из теории Эйнштейна, а также из сформулированной им самим скалярно-тензорной теории тяготения. В третьей лекции Дикке рассматривает проблемы космологии.

Книга, написанная признанным и оригинальным исследователем, будет желанным приобретением для любого специалиста в области астрономии, физики, космологии и геофизики.

Заказы направляйте в местные книготорги

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

Только предварительный заказ на книги в магазинах, торгующих научно-технической литературой, обеспечит своевременное получение интересующих вас книг.

90