

А. Н. ЖОВИНСКИЙ, В. Н. ЖОВИНСКИЙ



ИНЖЕНЕРНЫЙ ЭКСПРЕСС-АНАЛИЗ СЛУЧАЙНЫХ ПРОЦЕССОВ



• ЭНЕРГИЯ •

ОГЛАВЛЕНИЕ

Предисловие	3
Глава первая. Методические особенности экспресс-анализа случайных процессов	6
1-1. Объект экспресс-анализа	6
1-2. О глазомерной классификации случайных процессов	13
Глава вторая. Методы экспресс-анализа распределения вероятностей	15
2-1. Распределение вероятностей как характеристика свойств случайного процесса	15
2-2. Оценки вероятностей	16
2-3. Оценки функций распределения вероятностей	19
2-4. Статистическая проверка согласия экспериментального распределения с предполагаемым	25
2-5. Графическая проверка согласия с предполагаемым распределением	34
2-6. Выявление вида закона распределения по оценкам плотностей и моментов	36
Глава третья. Определение средних значений	37
3-1. Оценка среднего значения по равноотстоящим отсчетам	37
3-2. Определение среднего значения с помощью размаховых оценок	41
3-3. Оценивание средних значений с помощью порядковых статистик	44
3-4. Рекуррентное усреднение	49
3-5. Текущее усреднение	51
Глава четвертая. Определение среднеквадратических отклонений	54
4-1. Оценка среднеквадратического отклонения по равноотстоящим отсчетам	54
4-2. Определение среднеквадратического отклонения по размаху случайного процесса	55
4-3. Определение среднеквадратического отклонения по размаху выборки	58
Глава пятая. Определение корреляционных и спектральных характеристик случайных процессов	62
5-1. Оценивание корреляционных процессов	62
5-2. Оценка интервала корреляции	74
5-3. Определение спектральных характеристик случайных процессов	79
Глава шестая. Анализ стационарности случайных процессов	84
6-1. Необходимость контроля стационарности	84
6-2. Контроль постоянства среднего значения	85
6-3. Контроль постоянства дисперсии	95
6-4. Проверка стационарности распределений	99
Приложение	102
Список литературы	111

**БИБЛИОТЕКА
ПО РАДИО
ЭЛЕКТРОНИКЕ**

Выпуск 61

**А. Н. ЖОВИНСКИЙ
В. Н. ЖОВИНСКИЙ**

ИНЖЕНЕРНЫЙ ЭКСПРЕСС-АНАЛИЗ СЛУЧАЙНЫХ ПРОЦЕССОВ



МОСКВА «ЭНЕРГИЯ» 1979

ПРЕДИСЛОВИЕ

Вооружить инженеров простыми методами получения информации о свойствах и статистических характеристиках случайных процессов, реализуемыми вручную с помощью самых доступных инструментов — вот задача, которую поставили перед собой авторы.

Эта задача может показаться неблагодарной. Действительно, стоит ли в наше время при развитых системах обработки информации, в арсенале которых имеются ЭВМ и специализированные приборы, заниматься изучением методов ручной обработки случайных процессов? Ответ на этот вопрос по глубокому убеждению авторов может быть только утвердительным. Да, стоит! И вот почему. Методы теории и практики случайных процессов находят все более широкое применение, выходя далеко за рамки таких традиционных областей, как кибернетика, радиотехника, физика и авиация.

Очень часто исследователи самых разных специальностей проводят работы в удалении от своих лабораторий и вычислительных центров. Решения по поводу функционирования тех или иных устройств, условий проводимого эксперимента или прерывания работы приходится принимать «на месте», не дожидаясь обработки всех материалов. С другой стороны, измерение характеристик случайных процессов, основанное на методах математической статистики, имеет одну парадоксальную особенность: чем точнее требуется обработать случайный процесс, тем большими знаниями о нем должен заранее располагать исследователь. Поэтому перед планированием обработки с помощью средств вычислительной техники или специальной аппаратуры необходимо располагать некоторым объемом априорных сведений, которые обычно называют моделью процесса. При этом

окончательная обработка может быть организована точнее и дешевле.

Объем априорных сведений о процессе может быть весьма различным: может быть известен только диапазон изменения мгновенных значений процесса, что позволяет выбрать лишь масштаб для регистрации реализации процесса. Более информативной моделью может служить процесс, о котором априори известно, что он стационарен и эргодичен, известна его спектральная плотность и вид плотности распределения мгновенных значений, а неизвестны лишь некоторые параметры распределения.

При обработке всегда предполагается, что исследователь принимает некоторую первичную модель. Дальнейшие экспериментальные исследования увеличивают объем знаний о случайном процессе и это увеличивает ее информативность.

Практическое использование описываемых методов экспресс-анализа вполне доступно для читателя, знакомого с обычным вузовским курсом теории вероятностей. Под термином «экспресс-анализ» здесь понимается получение статистических характеристик случайных процессов с определенной точностью и достоверностью при минимальных затратах труда за небольшое время.

Предлагаемая читателю книга состоит из шести глав. Гл. 1 знакомит читателя с объектом экспресс-анализа, т. е. с видами обрабатываемых случайных процессов. Гл. 2 посвящена получению оценок одномерных распределений и проверке справедливости предположений об их свойствах. Эта глава готовит читателя к вопросам оценивания таких параметров распределения, как среднее значение и дисперсия, которые изложены в последующих двух главах. В гл. 5 описаны методы не слишком трудоемкого определения интервалов корреляции, корреляционных функций и спектральных плотностей (спектров) случайных процессов. На базе понятий, рассмотренных в предыдущих главах, в гл. 6 описаны приемы выявления нестационарности процессов, основанные на сравнении статистических характеристик различных участков реализации случайных процессов. Свойство стационарности чрезвычайно важно при формировании модели процесса и полезно при обработке, так как позволяет использовать широкий арсенал методов анализа стационарных процессов.

Объектом описанных в книге методов экспресс-анализа является чаще всего реализация процесса, записанная, например, на бумаге с известными масштабами по обеим осям координат. В качестве инструментов для экспресс-анализа могут быть использованы циркуль, логарифмическая линейка, микрокалькуляторы.

Таким образом, роль экспресс-анализа обрабатываемых процессов сводится к «разведке» их свойств, подобно тому, как при прокладке самой совершенной транспортной магистрали по ней первым пешком проходит изыскатель.

Авторы старались уделить основное внимание практической направленности описываемых методов и по возможности снабдить читателя вспомогательными расчетными графиками и номограммами.

Главы 1 и 4 написаны В. Н. Жовинским, гл. 2 и 6 — А. Н. Жовинским, гл. 3 и 5 — совместно.

Авторы выражают признательность д-ру техн. наук Г. Я. Мирскому, любезно предоставившему авторам некоторые ранее не опубликованные материалы, вошедшие в гл. 5 книги, благодарят инженеров И. М. Коломенского за подготовку и проведение отдельных расчетов на ЭВМ, Е. С. Загадскую и Е. А. Крылова за помощь, оказанную в практической проверке и выборке рекомендаций по использованию новых методов инженерного анализа случайных процессов.

Авторы будут признательны читателям за критические замечания по содержанию книги, которые следует направлять по адресу: 113114, Москва, Шлюзовая наб., 10, издательство «Энергия».

Авторы

ГЛАВА ПЕРВАЯ

МЕТОДИЧЕСКИЕ ОСОБЕННОСТИ ЭКСПРЕСС-АНАЛИЗА СЛУЧАЙНЫХ ПРОЦЕССОВ

1-1. ОБЪЕКТ ЭКСПРЕСС-АНАЛИЗА

«Neu miser, quia frequenter impeditus ero deinceps!»¹ — так думали древние и они оказались правы. Случайные процессы (помехи, случайные сигналы, шумы) в настоящее время являются предметом внимательного изучения широкого круга исследователей и объектом инженерного анализа, о котором идет речь в данной книге. Большинство задач оперативного анализа связано с изучением одномерных случайных процессов — временных случайных функций. Эти функции могут быть представлены либо в непрерывном виде (например, как реализации — осциллограммы), либо в дискретном — последовательным набором чисел, представляющих собой отсчеты реализации случайного процесса, либо, наконец, набором случайных чисел, принадлежащих некоторой общей совокупности, статистические характеристики которой подлежат выяснению.

При постановке задачи исследования статистических характеристик случайных процессов необходимо иметь представление об их классификации. Это связано с тем, что многие методы получения статистических характеристик, используемые для исследования одних групп процессов, оказываются непригодными для других групп. Так, например, как подчеркивается в излагаемом материале, наиболее сильно аппарат экспресс-анализа разработан для стационарных гауссовых (или нормальных) процессов. В отдельных случаях методы экспресс-анализа процессов этой группы удается применить для анализа некоторых видов нестационарных процессов. Необ-

¹ Горе мне, ибо впредь я буду часто встречать помехи (лат.).

ходимые рекомендации помещены в соответствующих разделах книги.

Предполагается, что случайный процесс записывается на каком-либо регистрирующем устройстве. В качестве таких устройств широко используются различного рода самописцы, шлейфовые осциллографы и другие приборы. Следует отметить, что даже на этапе регистрации необходимо иметь хотя бы грубое представление о вероятном диапазоне изменения сигнала и ширине его спектра. Эти исходные данные позволяют записать сигнал без искажений и в удобном для обработки масштабе. Наконец, запись произведена, и в руки инженера-исследователя попадает реализация случайного процесса. С этого момента можно начинать анализ, первой операцией которого является визуальное изучение и классификация сигнала.

Как уже отмечалось, одним из важных свойств процессов является характеристика их стационарности.

Напомним, что *стационарным процессом* называется такой случайный процесс, статистические характеристики которого не зависят от начала отсчета времени. Процессы, для которых корреляционная функция зависит только от разности временных аргументов, а моменты первых двух порядков (математическое ожидание и дисперсия) не зависят от времени, называют стационарными в широком смысле.

Следует указать, что существуют процессы, для которых статистические характеристики, полученные усреднением по достаточно большому интервалу времени, могут служить оценками статистических характеристик, полученных усреднением по достаточно большой совокупности реализаций. Такие процессы называют эргодическими.

Нестационарным процессом называют такой случайный процесс, у которого хотя бы одна статистическая характеристика зависит от времени. На рис. 1-1 представлены реализации различных случайных процессов, из которых процессы на рис. 1-1,а и б являются стационарными, а на рис. 1-1,в—г — нестационарными. При рассмотрении этих реализаций бросается в глаза их различие, связанное с тем, что в сигнале б существует некоторая средняя частота и, наоборот, в сигнале а такую частоту указать не представляется возможным. Очевидно, что спектр сигнала а богаче различными частотными

составляющими, чем спектр сигнала *б*. Если средняя частота просматривается уже на осциллограмме, то можно утверждать, что спектральная плотность такого сигнала (здесь сигнал *б*) имеет максимум на этой частоте и нельзя этого утверждать относительно сигнала *а*. Сигналы вида *б*, характеризующиеся наличием в спек-

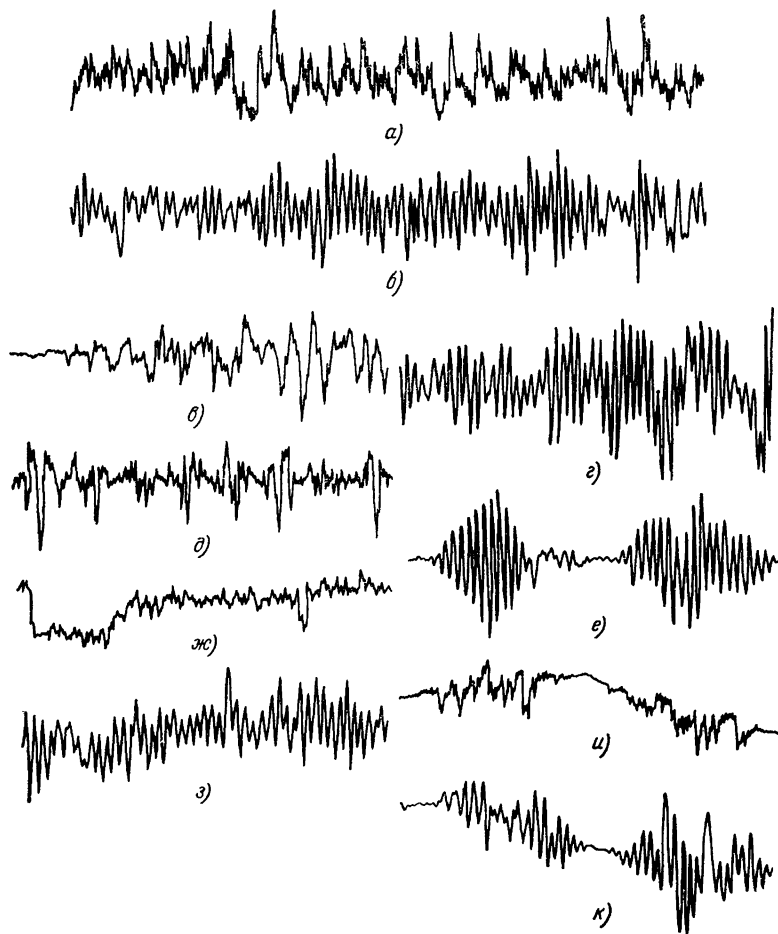


Рис. 1-1. Виды случайных процессов.

а — широкополосный стационарный; *б* — узкополосный стационарный; *в* — широкополосный нестационарный по дисперсии; *г* — узкополосный нестационарный по дисперсии; *д* — широкополосный периодически нестационарный по дисперсии; *е* — узкополосный периодически нестационарный по дисперсии; *ж* — широкополосный нестационарный по математическому ожиданию; *з* — узкополосный нестационарный по математическому ожиданию; *и* — широкополосный нестационарный по дисперсии; *к* — узкополосный нестационарный по математическому ожиданию и периодически нестационарный по дисперсии.

тре средней частоты, называют узкополосными. Такие сигналы могут порождаться в радиотехнических системах, звенья которой настроены на определенную частоту, а на их входе действует широкополосный сигнал. В системах автоматики встречаются чаще колебательные звенья, для которых подъем частотной характери-

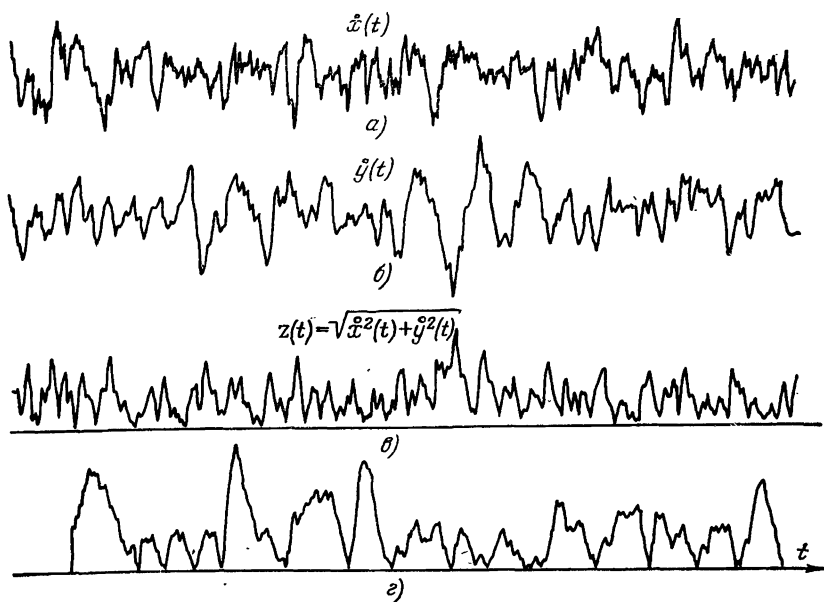


Рис. 1-2. Стационарные процессы с различными видами амплитудного распределения.

а и б — нормальные случайные процессы; в — релеевский случайный процесс; г — односторонний нормальный случайный процесс.

стики в области резонансной частоты не столь велик, как в избирательных радиотехнических цепях. Если на вход такой системы действует белый шум, то спектральная плотность будет иметь подъем в области резонансной частоты, и эта частота так же может быть обнаружена на осциллограмме, как и для узкополосных нормальных сигналов. За оценку этой частоты может быть принята величина, обратная среднему значению удвоенного интервала времени между нулями процесса [28], причем интервалы между нулями суть расстояния по оси времени между пересечениями центрированным (с нулевым средним значением) случайным процессом нулевого уровня с любым знаком производной.

Другим признаком классификации является вид одномерной плотности распределения вероятностей мгновенных значений случайного процесса. На практике очень часто встречаются нормальные или гауссовы процессы, для которых указанная одномерная плотность представляет собой нормальное распределение. Эти процессы наиболее хорошо изучены. Примеры реализаций таких процессов показаны на рис. 1-2, а и б. Процессы с другими видами распределений представлены на рис. 1-2, в и г.

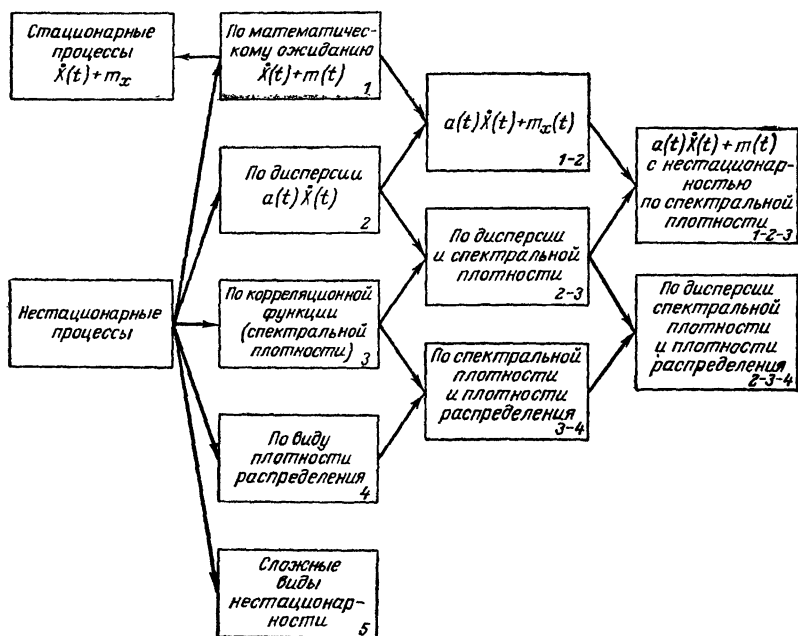


Рис. 1-3. Классификация случайных процессов.

Так, процесс, образованный из двух нормальных процессов в виде $Z(t) = \sqrt{\dot{X}^2(t) + \dot{Y}^2(t)}$ (рис. 1-2, в), имеет релеевское распределение, а процесс на рис. 1-2, г, представляющий собой результат двухполупериодного выпрямления нормального процесса $\dot{X}(t)$, т. е. $|\dot{X}(t)|$, — одно-стороннее нормальное распределение.

Стационарные процессы можно считать частным видом нестационарных процессов, примерная схема классификации которых показана на рис. 1-3.

Большой класс случайных процессов можно представить в виде суммы

$$Y(t) = a(t) \dot{X}(t) + m_x(t), \quad (1-1)$$

где $\dot{X}(t)$ — стационарная случайная функция времени, $a(t)$ и $m_x(t)$ — неслучайные функции времени, причем $m_x(t)$ является средним значением $Y(t)$.

Процессы со слагаемым $m_x(t)$ в виде регулярной (неслучайной) функции и при $a(t)=1$ относятся к процессам нестационарным по среднему значению (группа 1).

Если $\dot{X}(t)$ рассматривается как помеха, а $m_x(t)$ — как полезный сигнал, то такие помехи иногда называют аддитивными (т. е. суммируемыми с сигналом). Примером может служить выходной сигнал усилителя, когда полезный сигнал складывается с внутренним шумом усилителя (рис. 1-1, ж — широкополосная помеха, рис. 1-1, з — узкополосная помеха).

Процесс, нестационарный по дисперсии (группа 2), описывается формулой (1-1) при $m_x(t)=\text{const}$ [в том числе $m_x(t)=0$], если $a(t)$ — неслучайная (часто медленно меняющаяся) функция времени. Такие случайные процессы иногда называют мультипликативными. Пример широкополосной мультипликативной помехи показан на рис. 1-1 (сигнал в). С сигналами такого вида часто сталкиваются при изучении процессов модуляции и при исследованиях так называемого углового шума радиолокационных приемников [26]. Такой сигнал с узкополосной помехой показан на рис. 1-1 (сигнал г). Интересный случай представляют собой случайные процессы с периодической нестационарностью, т. е. когда функция $a(t)$ периодична. На рис. 1-1 показаны реализации мультипликативных помех с гармонической нестационарностью (сигнал д — широкополосный процесс, сигнал е — узкополосный процесс). Случайный сигнал с периодической нестационарностью встречается при изучении электрокардиограмм, например систологические шумы являются случайными процессами с периодичностью, соответствующей частоте сердечных сокращений [36].

Очевидно, что по характеру записанных процессов можно иногда судить (когда позволяет длительность реализации) о характере нестационарности по диспер-

сии. Появление случайной составляющей и почти полное ее пропадание через равные промежутки времени позволяют говорить о периодической нестационарности, а вид огибающей иногда позволяет сделать более определенные выводы о ее виде (гармоническая, меандровая и т. д.).

Процессы, нестационарные по спектральной плотности (корреляционной функции), изменяют свои частотные свойства во времени (группа 3). Для таких процессов корреляционная функция является функцией двух переменных, т. е. $R(t_1, t_2)$.

Случайные сигналы, нестационарные по спектральной плотности, возникают, например, в нестационарных линейных системах управления.

Процессы могут быть *нестационарными по одномерной плотности распределения* (группа 4), когда последняя изменяется во времени. Такие процессы встречаются в нестационарных нелинейных системах, когда нелинейные характеристики отдельных блоков зависят от времени.

Процессы со сложными видами нестационарности (группа 5) трудно охарактеризовать общими свойствами и лучше иллюстрировать примером. Таким примером могут служить шумы человека-оператора, возникающие при выполнении им операции слежения за меткой входного случайного сигнала по индикатору путем совмещения ее с ответной меткой [20]. Интенсивность этих шумов $\sigma_x(t)$ (среднеквадратическое отклонение) является сложной функцией времени:

$$\sigma_x(t) = \varphi[a|X(t)| + b|X'(t)|], \quad (1-2)$$

где a, b — постоянные коэффициенты; $|X(t)|, |X'(t)|$ — соответственно модули выходного сигнала и его производной.

Кроме рассмотренных видов нестационарных процессов, на практике встречаются комбинированные процессы вида 1—2, 1—3, 2—3, 3—4 и 1—2—3. Реализации, соответствующие виду 1—2 для широкополосных и узкополосных процессов, показаны на рис. 1-1, и, к соответственно (при периодической нестационарности). Процессы группы 1—2 являются характерными для измерителей угловой скорости линии визирования, используемых в системах самонаведения [26]. Процессы группы 1—2—3 являются типичными для выходного сигнала

коррелометров, в которых каждое измерение корреляционной функции $R_x(\tau)$ дает реализацию случайного процесса с математическим ожиданием $m_x(t) = R_x(t)$ и аддитивной случайной составляющей $n_R(\tau)$ с изменяющейся интенсивностью [с увеличением аргумента τ дисперсия $\sigma_R^2(\tau)$ увеличивается до значения $\sigma_R^2(\tau_{\max}) = \sigma_x^2/T$, где T — интервал наблюдения]. При этом изменяется спектральная плотность процесса $n_R(\tau)$, а одномерная плотность распределения нормализуется (рис. 1-4, а, б) при больших значениях τ .

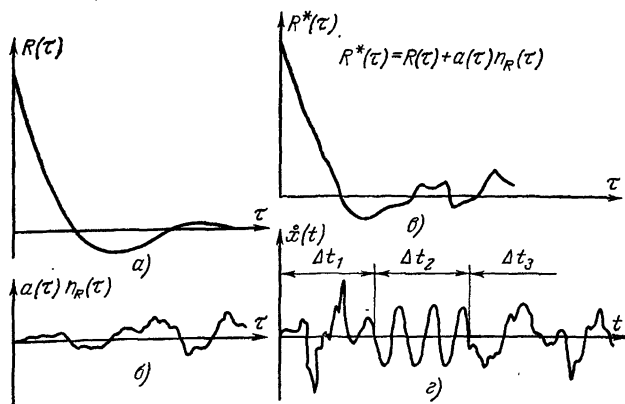


Рис. 1-4. Оценка корреляционной функции как пример нестационарного случайного процесса по дисперсии, математическому ожиданию и плотности распределения:

а — корреляционная функция; б — аддитивная измерительная помеха; в — оценка корреляционной функции; г — процесс с нестационарной спектральной плотностью.

Процессы группы 3—4 (пример реализации показан на рис. 1-4, г) встречаются в энцефалографии (тэта ритмы) [34]. На рис. 1-4 четко видны три интервала: Δt_1 , Δt_2 , Δt_3 , где наблюдаемый процесс изменяет свои свойства как по виду плотности распределения, так и по виду корреляционной функции.

1-2. О ГЛАЗОМЕРНОЙ КЛАССИФИКАЦИИ СЛУЧАЙНЫХ ПРОЦЕССОВ

Целью глазомерной классификации является попытка извлечения информации для принятия просейшей первичной модели при непосредственном наблюдении реализаций случайных процессов, например, на осциллографе или их визуального изучения после регистрации.

Примеры таких реализаций были рассмотрены в § 1-1. Первый этап анализа состоит в том, чтобы вынести суждение о стационарности процесса или в противном случае определить по возможности вид нестационарности в рамках обсуждающейся классификации.

Решение этой задачи порой невозможно, так как до сих пор отсутствует единая методика, пригодная для анализа нестационарных процессов. По мнению Д. Бендата и А. Пирсола [1], это объясняется тем, что суждение о нестационарности является просто негативным суждением, которое констатирует отсутствие свойств стационарности, а не определяет характер нестационарности процесса.

Основная цель выявления нестационарности направлена на определение свойств самого случайного процесса, поведение которого объективно отражает изучаемое явление или поведение исследуемой системы (объекта). Поэтому при определении вида нестационарности процессов необходимо учитывать всю априорную информацию об источнике сигнала или о породившей его системе. Чем больший объем исходных сведений принят во внимание, тем надежнее классификация процессов.

Рассмотренные в § 1-1 виды нестационарности случайных процессов условно можно разделить на быструю и медленную. Эта условность связана с временем наблюдения. Если нестационарность в пределах реализации определяется визуально, то ее можно называть «быстрой». Обнаружение медленной нестационарности представляет собой более трудную и тонкую задачу. Эта задача сводится, по крайней мере, к выявлению той или тех статистических характеристик процесса, которые изменяются во времени. Некоторые методы обнаружения медленной нестационарности будут рассмотрены в гл. 6, а в настоящем параграфе внимание читателя обращается на характерные особенности процессов с быстрой нестационарностью, поддающейся грубой глазомерной оценке.

Рассмотрим классификационные группы сигналов согласно рис. 1-3.

Нестационарные процессы группы 1 (1—2) характеризуются тем, что среднее значение сигнала изменяется во времени. При рассмотрении рис. 1-1 (ж — к) видно, что средняя линия, относительно которой наблюдаются флуктуации сигнала, не сохраняет параллельности некоторой другой линии, выбранной в качестве отсчетной (например, линия края осциллограммы). Эта средняя линия на отдель-

ных участках имеет различный наклон, что соответствует изменению среднего значения.

Группа 2 (1—2) характеризуется изменением интенсивности сигнала или его размаха. Так, для сигналов ϵ и z на рис. 1-1 характерно нарастание интенсивности со временем. На записи сигнала ϵ видно, что случайный сигнал дважды возникает и затухает. Если бы длительность реализации была больше и интервалы между отрезками времени существования случайного сигнала и длительности его «вырождения» сохранялись постоянными, то можно было бы говорить о периодической нестационарности. Менее заметно изменение интенсивности сигнала δ (рис. 1-1) и ярко выражено в сигналах u и k .

Группа 3. Увидеть характерные признаки изменения корреляционной функции можно лишь в редких случаях, один из которых показан на рис. 1-4,2. На интервале Δt_2 корреляционная функция близка к косинусоиду, если разделить процесс на локально-стационарные интервалы. Определить на глаз вид корреляционных функций на других участках не представляется возможным.

Изложенное позволяет иногда определить нестационарность по группам 4 и 1—2—3.

Очевидно, что возможности глазомерной классификации весьма ограничены, но в то же время, если других источников информации о сигнале нет, «живое общение» с объектом исследования позволит сделать выбор первичной малоинформативной модели случайного процесса.

ГЛАВА ВТОРАЯ

МЕТОДЫ ЭКСПРЕСС-АНАЛИЗА РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

2-1. РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ КАК ХАРАКТЕРИСТИКА СВОЙСТВ СЛУЧАЙНОГО ПРОЦЕССА

Большие трудности получения и наглядности использования многомерных распределений обусловили широкое распространение на практике одномерных распределений вероятностей. Поэтому основное внимание будет сосредоточено только на одномерных распределениях.

При обработке экспериментальных данных вместо функции распределения $W(x)$ случайного процесса $X(t)$ и плотности вероятностей $w(x)$ используют оценку распределения $W^*(x)$ и оценку плотности $w^*(x)$, дающих с той или иной погрешностью информацию о вероятностной структуре процесса. При получении оценок $W^*(x)$ и $w^*(x)$ непрерывное множество значений x неизбежно приходится заменять конечным (и притом не слишком

большим) набором значений x_i , $i=0, 1, \dots, L$. Тогда эти оценки формируются как совокупности вероятностей

$$\begin{aligned} W^*(x) &= P^*[X(t) \leq x_i] \text{ и } w^*(x) = \\ &= \frac{1}{x_i - x_{i-1}} P^*[x_{i-1} < X(t) \leq x_i] \end{aligned} \quad (2-1)$$

и их свойства зависят от оценок вероятностей попадания процесса в интервалы $(-\infty, x_i]$ или $(x_{i-1}, x_i]$. Поэтому ниже описываются способы оценивания вероятностей, а только потом способы формирования оценок $W^*(x)$ и $w^*(x)$.

Существуют три основные задачи, которые решаются с помощью оценок распределений случайных процессов.

Первая из этих задач — определение вероятностей, с которыми процесс может быть выше или ниже заданных уровней. К такой форме сводятся многие задачи помехоустойчивости, надежности технических систем и ряд других.

Вторая задача связана с экспериментальной проверкой гипотетического предположения о виде распределения процесса. Если предположение после статистического анализа подтверждается, то свойства предполагаемого распределения (например, нормальности) могут быть использованы в дальнейшем. Отклонения от ожидаемого распределения дают исследователю пищу для размышлений о правильности модели или условий эксперимента.

Третья практически важная задача возникает тогда, когда о природе исследуемого процесса известно мало и его математическая модель строится после экспериментального изучения. В этом случае на первом этапе изучаются оценки распределений, а на втором делаются попытки их аппроксимации.

О некоторых способах решения указанных задач, доступных по трудоемкости ручному анализу, будет рассказано ниже.

2-2. ОЦЕНКИ ВЕРОЯТНОСТЕЙ

Оценки вероятностей для реализаций случайных процессов могут быть вычислены двумя способами. Первый способ основан на понятии частоты некоторого события. Согласно этому способу оценку формируют по процессу, представленному независимыми отсчетами.

Это значит, что значения процесса выбираются через интервал времени, превосходящий интервал корреляции (§ 5-2).

Другой способ, менее трудоемкий, пригоден для осциллограмм непрерывных процессов. В основу способа положен расчет относительного времени пребывания процесса в заданных пределах. В этом случае разбиение записи процесса на большое число равных интервалов не производится, а обрабатываются длительности участков записи, расположенных в некоторых пределах по оси ординат.

Оценка по дискретным отсчетам. Предположим, что имеется N независимых отсчетов x_i и из них n отсчетов связано с некоторым событием A , представляющим интерес для исследователя. Например, событие A состоит в выполнении неравенства $x_i > U$, где U — заданный уровень. Тогда частота выполнения этого неравенства, равная отношению n/N , может служить оценкой вероятности $P_A = P\{x_i > U\}$. Известно [4], что среднее значение $m[n/N] = P_A$, поэтому говорят, что частота является несмещенной оценкой. Кроме того, оценка $P^* = n/N$ обладает еще несколькими важными свойствами.

Оценка P^* — состоятельна, т. е. с ростом N она становится все более точной оценкой P_A . Как всякая случайная величина, оценка P^* характеризуется распределением вероятностей $W(P^*)$. С ростом N , если истинная вероятность не слишком близка к 0 или 1, распределение $W(P^*)$ сходится к нормальному. В этом состоит свойство асимптотической нормальности оценки.

Асимптотическое распределение $W(P^*)$ описывается средним значением $m_{P^*} = m[n/N] = P_A$ и дисперсией $\sigma_{P^*}^2 = P_A(1 - P_A)/N$. Поэтому относительная погрешность оценки P^* выражается отношением σ_{P^*}/m_{P^*} , равным:

$$\epsilon_{P^*} = \sqrt{(1 - P_A)/N}. \quad (2-2)$$

Таким образом, на точность оценки влияют объем выборки N и истинное значение вероятности P_A . Чем меньше вероятность, тем труднее ее оценивать. При определении вероятностей, близких к единице, погрешность ϵ_{P^*} стремится к нулю. Но в этом случае оценка P^* указывает лишь близость P_A к единице. Как правило, исследователю этой информации недостаточно, и возникает вопрос о том, насколько близка оцениваемая вероятность к 1. Такой вопрос эквивалентен требованию оценить малую вероятность $1 - P_A$ [32]. Это следует иметь в виду при использовании оценок больших вероятностей.

Качество оценок можно также определять с помощью доверительного интервала, задаваемого границами, внутрь которых попадает истинное значение оценки с доверительной вероятностью P_d . Так, если оценивать точность оценки P^* парой чисел $\pm \sigma_{P^*}$ то при нормальном распределении $W(P^*)$ это соответствует довери-

тельной вероятности $P_d=0,67$. При малых объемах выборки и при значениях P_d , близких к 0 или 1, распределение $W(P^*)$ отличается от нормального, а доверительные интервалы становятся несимметричными относительно величины P^* . Относительные значения ϵ_P нижней и верхней границ доверительного интервала приведены на рис. 2-1 для точного распределения $W(P^*)$ [25]. График на рис. 2-1 построен для $P_d=0,95$ и объема выборки от 20 до 1000. Аналогичные границы доверительного интервала при малых вероятностях (от 0,01 до 0,2) приведены на рис. 2-2 для $P_d=0,8$, так как назначение

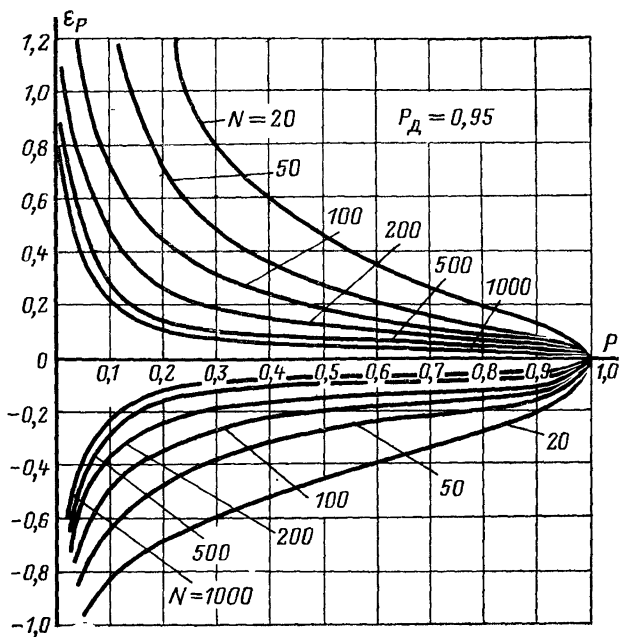


Рис. 2-1. Зависимость границ доверительного интервала относительной погрешности оценки вероятности от ее истинного значения для разных объемов выборки ($P_d=0,95$).

доверительной вероятности $P_d=0,95$ для малых значений P_d приводит к слишком большой неопределенности. Графики могут быть использованы как для расчета погрешности вычисленной оценки, так и для планирования объема выборки при заданных требованиях к погрешности оценок вероятности. В первом случае по полученному значению P^* , которое откладывают по оси P для реального N по графикам на рис. 2-1 и 2-2, определяется погрешность ϵ_P , а во втором по ожидаемому значению P и требуемой погрешности ϵ_P определяют объем выборки N , обеспечивающий нужную точность оценки, как показано на графиках.

Оценка вероятности по относительному времени пребывания. Пусть, как и в предыдущем случае, предстоит найти вероятность

события A , состоящего в том, что значения процесса $X(t)$ принадлежат некоторой области по оси ординат, например области $x \geq U$. Если по осциллограмме реализации $x(t)$ заметить участки длительностью Δt_i , для которых $x(t) \geq U$, то оценкой вероятности P_A может служить отношение

$$P^* = \sum_i \Delta t_i / T, \quad (2-3)$$

где T — длительность реализации.

Точность такого метода оценивания вероятности задается относительной погрешностью ϵ_{P^*} , равной:

$$\epsilon_{P^*} = \sqrt{\frac{2(1-P_A)\tau_{кx}}{TP_A}}, \quad (2-4)$$

где $\tau_{кx}$ — интервал корреляции процесса $x(t)$.

В этой формуле правильнее было бы использовать τ_k процесса $u(t)=1$ при $x(t) < U$ и $u(t)=0$ при $x(t) \geq U$, но это сложно.

Поэтому оценка погрешности оказывается завышенной. В каждом конкретном случае степень завышения различна. Так, для нормального процесса $x(t)$ с нормированной корреляционной функцией (КФ) вида $\rho(\tau) = e^{-\alpha|\tau|}$ для $P_A=0,5$ завышение составляет 20%, а для $P_A=0,98$ или 0,02 погрешность завышается в три раза [32].

2-3. ОЦЕНКИ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Оценки функций распределений $W^*(x)$ и плотностей вероятностей $w^*(x)$ строятся как комбинации вероятностей попадания в интервалы $(-\infty, x_i]$ или $(x_{i-1}, x_i]$. Поскольку при уменьшении интервала уменьшается и вероятность попадания в него, нельзя брать интервалы очень малыми, так как, как было показано в § 2-2, точность оценок вероятностей окажется невысокой.

Как правило, упомянутые комбинации вероятностей строятся так, чтобы получить ступенчатые аппроксимации искомых функций в виде

$$W^*(x) = W^*_i(x) \text{ при } x < x_i, \quad i=0, 1, \dots, L \quad (2-5)$$

и

$$w^*(x) = w^*_i(x) \text{ при } x_{i-1} < x \leq x_i, \quad i=0, 1, \dots, L, \quad (2-6)$$

2*

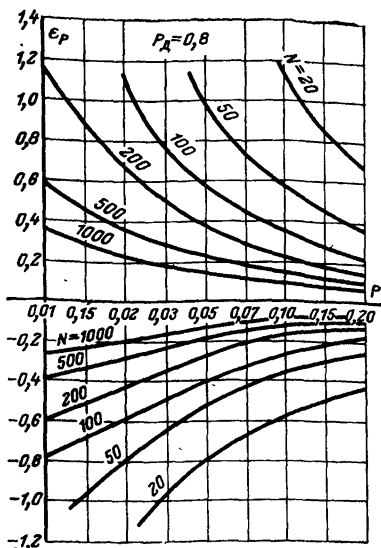


Рис. 2-2. Границы доверительного интервала относительной погрешности оценки малых вероятностей ($P_d=0,8$).

причем

$$\left. \begin{aligned} W^*_i &= \tilde{n}_i/N, \\ w^*_i(x) &= \tilde{n}_i/Nh, \end{aligned} \right\} \quad (2-7)$$

где \tilde{n}_i — число членов выборки, попавших в интервал $(x_{i-1}, x_i]$; \tilde{n}_i — число членов выборки, меньших x_i :

$$\tilde{n}_i = \sum_{l=0}^{l=i} \tilde{n}_l;$$

$h = x_i - x_{i-1}$ — ширина каждого из L обычно равных между собой разрядов.

Погрешность смещения g для плотности $w^*(x)$ может быть рассчитана по формуле [1]

$$g = \frac{h^2}{24} w''(x), \quad (2-8)$$

где $w''(x)$ — вторая производная плотности $w(x)$.

Эта формула соответствует приписыванию значения $w^*_i(x)$ середине интервала $(x_{i-1}, x_i]$. Как показано в [31], погрешность примерно в 10 раз увеличится, если $w^*_i(x)$ приписывать крайним точкам интервала. При построении эмпирических распределений $W^*(x)$ в силу свойств распределений $W(x)$ корректнее значения $W^*_i(x)$ приписывать абсциссам x_i , а график строить не ступенчатой линией, а кусочно-линейной, соединяя прямыми точки с координатами (x_i, W^*_i) .

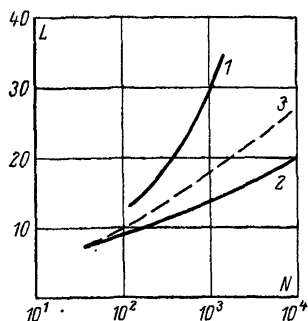
На практике при $N < 10^4$ более существенны статистические погрешности, а погрешности смещения не превышают единиц процентов и ими можно пренебречь [32].

Статистические погрешности зависят от трех параметров, с которыми оперируют при построении оценок $W^*(x)$ и $w^*(x)$. Этими параметрами являются уже упомянутые объем выборки N , ширина разряда h и количество разрядов L .

Величина N должна выбираться по возможности большей. Так, например, при $N=200$ для средней по оси ординат части распределений относительные статистические погрешности составляют 20—40%. Для значений $w^*(x)$ вблизи моды точность выше, а на «хвостах» распределений относительные погрешности могут составлять сотни процентов.

Рис. 2-3. Зависимости числа разрядов гистограммы от объема выборки.

1 — согласно оптимизации критерия χ^2 ; 2 — кривая, рассчитанная по формуле $L=k \lg N$; 3 — промежуточная рекомендуемая зависимость.



По выбору числа разрядов L в литературе даются различные рекомендации. Так, в [1] и [4] имеются ссылки на различные исследования, устанавливающие оптимальное значение L в зависимости от объема выборки N с точки зрения использования оценок распределений во второй из указанных в § 2-1 задач. В этих работах за основу принято наилучшее использование критерия χ^2 для проверки согласия распределений (§ 2-4). Обобщенная зависимость $L(N)$ по данным этих исследований приведена в виде кривой 1 на рис. 2-3. Согласно другим рекомендациям [5] зависимость $L(N)$ может быть представлена в форме $L(N)=k \lg N$, где постоянная k лежит в диапазоне от 3 до 5. Такая зависимость представлена для $k=4,6$ в виде кривой 2 на рис. 2-3. По опыту авторов может быть рекомендована промежуточная кривая 3.

Ширина разряда h равна размаху выборки, определяемому разностью между наибольшим x_{\max} и наименьшим x_{\min} значениями отсчетов, деленному на L . После расчета величины h рассчитывают границы по оси абсцисс $x_i=x_{\min}+ih$. Построение оценок состоит в сортировке массива отсчетов x по разрядам. Сортировка заключается в прибавлении единицы к величине $W^*_i(x)$, если $x_{i-1} < x \leq x_i$. Значения эмпирического распределения могут быть получены суммированием согласно формуле (2-5).

Относительные статистические погрешности $\varepsilon_w(x)$ и $\varepsilon_w(x)$ определяются формулами, аналогичными формуле (2-2) для некоррелированных отсчетов [32],

$$\varepsilon_w(x) = \sqrt{\frac{1}{N} \frac{1 - W(x)}{W(x)}}; \quad (2-9)$$

$$\varepsilon_w(x) = \sqrt{\frac{1}{N} \frac{1 - w(x)h}{w(x)h}}. \quad (2-10)$$

Если погрешность вычисляется до расчета оценок, то в качестве $W(x)$ и $w(x)$ подставляют ожидаемые значения. Если оценки уже получены, то значения $\varepsilon_w(x)$ и $\varepsilon_w(x)$ вычисляются при абсциссах x , интересующих расчетчика. Формулы (2-9) и (2-10) соответствуют среднему квадратическому отклонению для нормального распределения случайной составляющей погрешности, поэтому они характеризуют доверительный интервал для $P_d=0,67$. Другой доверительный интервал может быть получен умножением величины ε на коэффициент Z_Φ из табл. П1.

Несколько забегаая вперед, отметим, что группированные данные, по которым строится оценка $w^*(x)$, удобны для вычисления моментов. Оценки среднего и дисперсии выражаются в этом случае формулами

$$m_x^* = \frac{1}{N} \sum_{i=1}^L \tilde{n}_i \xi_i$$

и

$$D_x^* = \frac{1}{N} \sum_{i=1}^L \tilde{n}_i \xi_i^2 - m_x^{*2} - \frac{1}{12} h^2,$$

где $\xi_i = x_{\min} + h(i-0,5)$.

Метод относительного времени пребывания. Согласно этому методу оценки $W^*(x)$ и $w^*(x)$ находятся по величинам $P_i^*[x < x_i]$ и $P_i^*[x_{i-1} < x \leq x_i]$, которые измеряются относительным временем пребывания процесса $x(t)$ в заданных пределах. Выбор числа разрядов и их границ выполняется так же, как и в методе дискретных отсчетов. Благодаря этому ошибками смещения можно пренебречь.

Сами же вероятности оцениваются отношением T_i/T , где $T_i = \sum_j \Delta t_{ij}$, Δt_{ij} — длительность j -го участка, на котором процесс $x(t)$ пребывает в области i -го разряда, как показано на рис. 2-4.

Исследования случайных погрешностей оценок распределений показывают, что они определяются длиной исследуемой реализации, истинным значением оцениваемых в каждом разряде вероятностей и КФ процесса

$x(t)$. В общем случае анализ оказывается достаточно сложным. Поэтому в литературе приводятся частные случаи, которыми можно пользоваться точно, если выполняются предположения, сделанные при их выводе, или для ориентировочных расчетов, если в справедливости сделанных предположений нет уверенности.

Для нормальных случайных процессов с произвольной КФ относительные случайные погрешности для оценок распределений и плотностей приведены соответственно на рис. 2-5 и рис. 2-6. По оси абсцисс на графиках

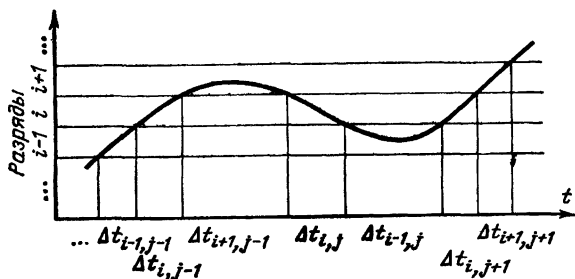


Рис. 2-4. Подготовка осциллограммы при оценке распределения методом относительного пребывания в пределах i -го разряда.

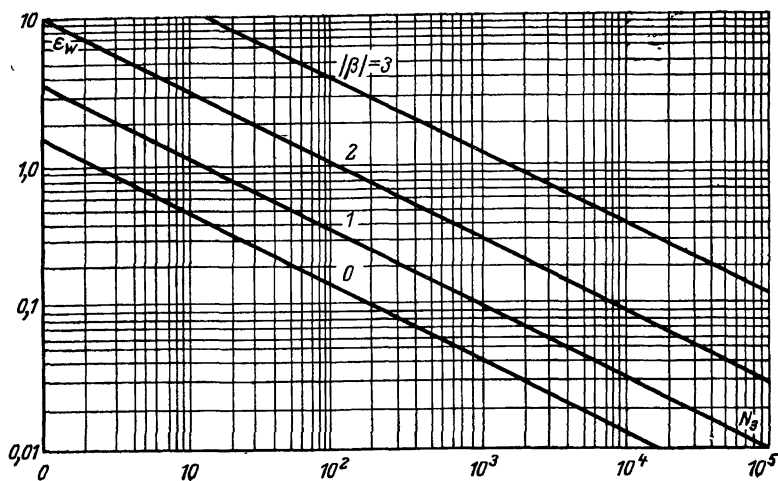


Рис. 2-5. Зависимость относительных случайных погрешностей оценки распределения вероятностей от нормированной продолжительности анализа при различных нормированных уровнях нормальных процессов ($\beta = x/\sigma$).

отложено эквивалентное число независимых отсчетов $N_0 = T/\tau_{\text{кх}}$. При конкретизации вида КФ оценки погрешностей оказываются в 1,5—2 раза меньше [10]. Но, как правило, КФ заранее не известна. Поэтому и удобно применять менее точный, но более универсальный метод, описанный в [32].

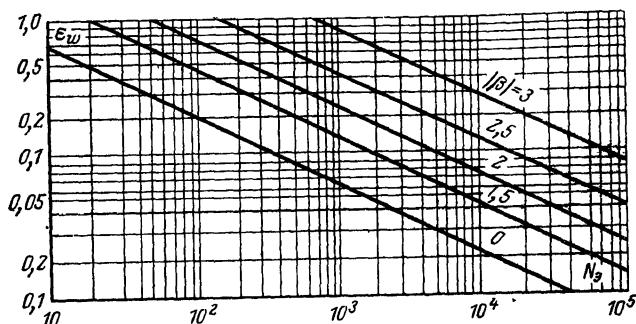


Рис. 2-6. Зависимость относительных случайных погрешностей оценки плотности вероятностей от нормированной продолжительности анализа при различных нормированных уровнях анализа нормальных процессов.

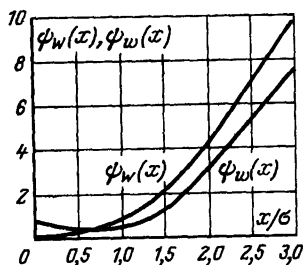


Рис. 2-7. Вспомогательные функции $\psi_w(x/\sigma)$ и $\psi_w(x/\sigma)$ для расчета погрешностей оценки распределений узкополосных нормальных процессов.

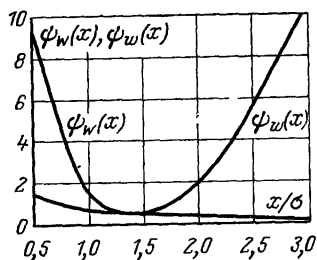


Рис. 2-8. Вспомогательные функции $\psi_w(x/\sigma)$ и $\psi_w(x/\sigma)$ для расчета погрешностей оценки распределений релейских процессов.

Погрешности для узкополосных нормальных процессов могут быть рассчитаны по формуле [10]

$$\varepsilon = \frac{1}{\sqrt{\alpha T}} \psi(\beta), \quad (2-11)$$

где α — параметр КФ вида $\rho(\tau) = e^{-\alpha|\tau|} \cos \omega_0 \tau$, а функции $\psi(\beta)$ для оценок распределений $W^*(x)$ и плотностей

$w^*(x)$ заданы на рис. 2-7 с соответствующими индексами при $\beta=x/\sigma$. По формуле (2-11) можно рассчитать погрешности для широкополосного релеевского процесса с КФ $\rho(\tau)=e^{-\alpha|\tau|}$, если вспомогательную функцию $\psi(\beta)$ отсчитать по графикам, приведенным на рис. 2-8.

2-4. СТАТИСТИЧЕСКАЯ ПРОВЕРКА СОГЛАСИЯ ЭКСПЕРИМЕНТАЛЬНОГО РАСПРЕДЕЛЕНИЯ С ПРЕДПОЛАГАЕМЫМ

В практике контроля режимов технических устройств и систем, а также при планировании обработки экспериментальных материалов большую пользу приносит знание распределения процесса. Распределение может быть известно из физических соображений или из математических свойств процесса. Однако всегда возникает необходимость проверить, правильны ли предположения о распределении по результатам эксперимента. Проверку выполняют с помощью критериев согласия, которые позволяют ответить на вопрос о том, с какой достоверной вероятностью P_d отклонение эмпирического распределения от предполагаемого может объясняться лишь случайным разбросом и с какой вероятностью предположение о распределении должно быть отвергнуто. В последнем случае исходное предположение, называемое нулевой гипотезой, считают несправедливым, так как отличия распределений носят неслучайный характер.

При использовании критериев согласия по экспериментальным данным рассчитывают некоторую величину, называемую статистикой критерия. Для каждого из критериев формируется правило расчета такой статистики. Статистика критерия является мерой различия между гипотетическим распределением $W(x)$ и оценкой $W^*(x)$.

Применение критериев согласия основано на том, что распределение статистики критерия для нулевой гипотезы известно. Тогда, сравнивая эмпирическое значение статистики критерия с табличным, можно сделать вывод о значимости различия функций $W(x)$ и $W^*(x)$. Строго говоря, критерии согласия позволяют установить, следует ли отвергнуть предположение о виде распределения, но не дают оснований утверждать, что гипотетическое распределение справедливо.

Ниже приведено несколько критериев согласия. Каждый из них имеет свои особенности, и поэтому, если есть возможность, желательно использовать различные критерии и сравнить выводы. Совпадение выводов подтверждает их справедливость. Если выводы различных критериев противоречат друг другу, то объем выборки надо увеличить.

Критерий χ^2 Пирсона. Расчеты по этому критерию начинаются с сортировки данных по разрядам аналогично тому, как это делалось при оценке плотности $w^*(x)$. Статистика критерия X^2 равна:

$$X^2 = \sum_{i=1}^L \frac{\tilde{n}_i - NP_i}{NP_i}, \quad (2-12)$$

где L — число разрядов; n_i — число отсчетов в каждом разряде; $P_i = W(x_i) - W(x_{i-1})$ — теоретическая вероятность попадания отсчета в разряд для гипотетического распределения $W(x)$.

Для нулевой гипотезы, состоящей в том, что испытуемая выборка принадлежит процессу с распределением $W(x)$, величина X^2 приблизительно подчиняется распределению χ^2 [4] практически для любого вида $W(x)$.

Вычисленная статистика X^2 сравнивается с критическим значением $Z_{\chi^2}(\nu, P_d)$. Доверительная вероятность P_d

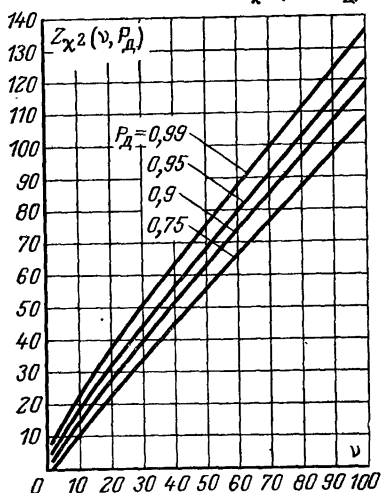


Рис. 2-9. Критические значения критерия χ^2 .

есть вероятность принять гипотезу о предполагаемом распределении, если она верна. Доверительной вероятностью задаются при расчетах в зависимости от практической важности проверки согласия и обрабатываемого объема выборки. Обычно принимают P_d равным от 0,8 до 0,999. Подробнее о выборе P_d сказано в гл. 6.

Второй аргумент ν критического значения называется числом степеней свободы и равен $\nu = L - q - 1$, где q — число неизвестных параметров гипотетического распре-

деления, которые предварительно приходится находить для того, чтобы вычислить вероятности P_i . Так, например, предстоит проверить гипотезу о нормальности распределения $W^*(x)$. Если среднее и дисперсия гипотетического распределения известны, как бывает при контроле генераторов шума, то $q=0$ и $v=L-1$. Если оба параметра заранее не известны, то их придется оценивать, и тогда $q=2$ и $v=L-3$.

По величинам v и P_d на рис. 2-9, построенном согласно [2], можно найти критическое значение $Z_{\chi^2}(v, P_d)$. Если оказывается, что $X^2 \leq Z_{\chi^2}(v, P_d)$, то нулевая гипотеза принимается. Отвергается гипотеза при $X^2 > Z_{\chi^2}(v, P_d)$.

В экспресс-анализе удобен вариант критерия χ^2 , называемый критерием Романовского [5]. Этот вариант не требует критических значений $Z_{\chi^2}(v, P_d)$, так как в нем используется нормализованная и нормированная статистика, равная $|X^2 - v|/\sqrt{2v}$. Если эта величина меньше 3, то нулевая гипотеза принимается. Доверительная вероятность соответствует в этом случае критическим значениям нормального распределения (см. табл. П1).

Достоинство критерия χ^2 состоит в том, что он может применяться для проверки различных распределений и даже значений отдельных вероятностей. К недостаткам критерия относятся необходимость большого объема выборки и относительно большая трудоемкость вычислений, к тому же результат, получаемый по критерию χ^2 , зависит от способа разбиения на разряды.

При выборе числа разрядов L можно использовать изложенные выше соображения (§ 2-3). Иногда [49] рекомендуется брать наибольшее число L , не превышающее $N/5$. Зависимость L от объема выборки определяется тем, сколько отсчетов попадает в каждый из разрядов. Часто в литературе подчеркивается, что число отсчетов в разрядах не должно быть меньше 5. Однако и меньшие количества не приводят к серьезным ошибкам, особенно при больших значениях v [4, 24].

Пример [38]. Пусть по записи процесса $x(t)$ снято $N=130$ отсчетов, вычислены величины $m_x^*=5,15$ и $\sigma_x^*=0,76$, а также произведено разбиение на семь разрядов для построения гистограммы. Результаты разбиения приведены в табл. 2-1.

Требуется с доверительной вероятностью, равной 0,95, проверить гипотезу о нормальности распределения $W(x)$.

Для расчетов по формуле (2-16) надо вычислить теоретические средние значения NP_i , где вероятности P_i равны соответственно $P_1=W(3,6)-W(3,0)$, $P_2=W(4,2)-W(3,6)$ и т. д., причем теоретическое распределение рассчитывается по таблицам нормального распределения с эмпирическими параметрами m^*_x и σ^*_x . Таким обра-

Таблица 2-1

Группирование данных для построения гистограммы

Значения	Число отсчетов
3,0—3,6	2
3,6—4,2	8
4,2—4,8	35
4,8—5,4	43
5,4—6,0	22
6,0—6,6	15
6,6—7,2	5
Всего	130

зом, $P_1=0,0207$; $P_2=0,0849$; $P_3=0,2172$; $P_4=0,3065$; $P_5=0,2393$; $P_6=0,1033$ и $P_7=0,0281$. Умножим вероятности P_i на объем выборки $NP_1=2,69$; $NP_2=11,04$; ...; $NP_7=3,65$. Теперь может быть рассчитана статистика критерия

$$X^2 = \frac{(2 - 2,69)^2}{2,69} + \frac{(8 - 11,04)^2}{11,04} + \dots + \frac{(5 - 3,65)^2}{3,65} = 6,22.$$

Число степеней свободы ν равно $7 - 3 = 4$. По графикам на рис. 2-9 находим критическое значение $Z_{\chi^2}(\nu; P_d) = Z_{\chi^2}(4; 0,95)$. Поскольку $X^2 < Z_{\chi^2}(4; 0,95)$, то нулевая гипотеза о нормальности распределения $W(x)$ принимается.

Критерий Колмогорова. В этом критерии согласие распределений проверяется по наибольшему расхождению между ними. Исходная информация об эмпирическом распределении $W^*_N(x)$ задается вариационным рядом, представляющим собой выборку $x_1, x_2, \dots, x_i, \dots, x_N$, упорядоченную так, что $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(j)} \leq \dots \leq x_N$. Целое число j , означающее номер порядковой статистики $x_{(j)}$ в вариационном ряду, называется рангом отсчета x_i . Распределение $W^*_N(x)$ оказывается ступенчатой функцией со ступеньками, равными $1/N$ в точ-

ках x_i . По полученным данным находится статистика критерия, имеющая вид:

$$H = \max(H^+, H^-); \quad (2-13)$$

$$\left. \begin{aligned} H^+ &= \max_{1 \leq j \leq N} \left[\frac{j}{N} - W(x_j) \right]; \\ H^- &= \max_{1 \leq j \leq N} \left[W(x_j) - \frac{j-1}{N} \right], \end{aligned} \right\} \quad (2-14)$$

где $W(x_j)$ — предполагаемое распределение, значения которого вычисляются или берутся из таблиц.

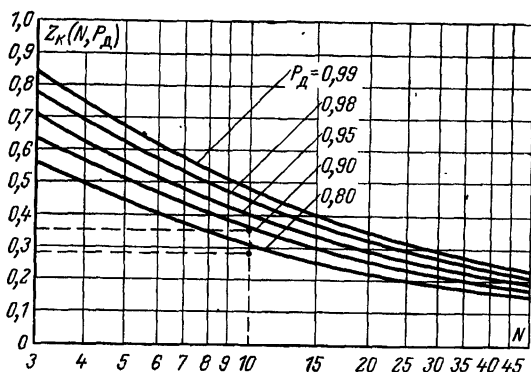


Рис. 2-10. Критические значения критерия Колмогорова.

Статистика H сравнивается с критическим значением распределения Колмогорова $Z_K(N, P_d)$ для заданного объема выборки и доверительной вероятности P_d . Критические значения при $N \leq 50$ представлены на рис. 2-10. При больших N критические значения $Z_K(P_d)$ для статистики вида \sqrt{NH} приведены в табл. П1. Если при выбранной вероятности P_d статистика меньше соответствующего критического значения, то гипотеза о предполагаемом распределении принимается. В противном случае расхождение нельзя признать случайным.

Существенного облегчения расчетов можно достичь при графическом использовании критерия. При этом построение вариационного ряда необязательно, так как функция $W_N^*(x)$ может быть построена прямо по точкам x_i , отложенным по оси абсцисс. Наносить гипотетическое распределение на график можно по любым табличным значениям. Статистика H является наибольшим

отклонением $|W^*_N(x) - W(x)|$ на графике, а строгое определение N по формулам (2-13) и (2-14) необходимо лишь при расчете без наглядного представления распределений, чтобы избежать ошибок вычисления $W^*_N(x)$ в точках разрыва.

Любые практические данные имеют конечную точность, например два десятичных знака, а отсчеты x_i могут иметь совпадающие значения. Следовательно, можно считать, что мы всегда имеем дело с группировкой данных в некоторые разряды. Поэтому в литературе встречается применение критерия Колмогорова к группированным данным, где отсчеты трактуются как верхние границы разрядов $x_k + h/2$ (x_k — середины разрядов). Разумеется, выводы по группированным данным из-за потери части информации тем менее точны, чем меньше число разрядов. Количественное выражение этой зависимости в литературе, известной авторам, не приводится.

Достоинство критерия Колмогорова состоит в меньшей трудоемкости расчетов и возможности строить доверительные границы для оценок распределений $W^*(x)$. По сравнению с критерием χ^2 критерий Колмогорова позволяет точнее выявить различия распределений. Более высокая «чувствительность» сказывается в том, что, если для принятия решения о распределении по критерию χ^2 требуется N отсчетов, для решения по критерию Колмогорова, обладающего тем же «качеством», нужно всего $N^{4/5}$ отсчетов [24]. Это значит, что вместе $N=50$ можно использовать 23 отсчета, а вместе $N=1000$ — всего 250. Кроме того, критерий может быть применен при столь малых объемах выборки, при которых критерий χ^2 вообще неприменим.

Пример. Некоторая система контролируется по равномерности распределения фазового сдвига $\varphi(t)$ между случайным узкополосным и опорным колебаниями. По записи процесса $\varphi(t)$ на выходе фазометра сняты 10 отсчетов в градусах, которые помещены в первом столбце табл. 2-2. Теоретические значения $W(\varphi)$ для равномерного распределения в диапазоне от 0 до 360°, соответствующие отсчетам, приведены во втором столбце таблицы

В следующих столбцах даны рассчитанные по полученной выборке ранги j и статистики H_j^+ и H_j^- . По ряду значений H_j^+ и H_j^- легко найти статистику критерия H , равную 0,36. Сопоставление с рис. 2-10 вынуждает отвергнуть гипотезу о равномерности распределения при $P_d \leq 0,9$. После регулировки системы проведена аналогичная серия измерений, результаты которых даны во второй част.

Экспериментальные данные примера по проверке согласия распределений по критериям Колмогорова и χ^2

Эксперимент № 1					Эксперимент № 2				
φ	W	j	H_j^+	H_j^-	φ	W	j	H_j^+	H_j^-
194	0,54	9	0,36	-0,26	166	0,46	4	-0,06	0,16
47	0,13	2	0,07	0,03	248	0,69	6	-0,09	0,19
180	0,5	8	0,3	-0,2	50	0,14	1	-0,04	0,14
245	0,68	10	0,32	-0,22	202	0,56	5	-0,06	0,16
40	0,11	1	0,01	0,11	292	0,81	8	-0,01	0,11
166	0,46	7	0,24	0,14	252	0,7	7	0	0,1
155	0,43	5	0,07	-0,03	324	0,9	10	0,1	0
97	0,27	3	0,03	-0,07	140	0,39	2	-0,19	0,29
133	0,37	4	0,03	-0,07	317	0,88	9	0,02	0,08
162	0,45	6	0,15	-0,05	162	0,45	3	-0,15	0,25

табл. 2-2. Получено, что $H=0,29$. Данные эксперимента № 2 не противоречат нулевой гипотезе.

Критерий ω^2_N Крамера — Мизеса — Смирнова. В этом критерии используется среднеквадратичное расхождение между эмпирическим и гипотетическим распределениями. В качестве исходного материала, как и в критерии Колмогорова, применяется вариационный ряд. Статистика критерия вычисляется по формуле [43]

$$\omega^2_N = \frac{1}{12N} + \sum_{j=1}^N \left[W(x_j) - \frac{2j-1}{2N} \right]^2. \quad (2-15)$$

Статистика ω^2_N сравнивается с критическим значением $Z_\omega(P_d)$, вычисленным для распределения $a(Z)$ при нулевой гипотезе. Значения $Z_\omega(P_d)$ приведены в табл. П1. При $\omega^2_N \leq Z_\omega(P_d)$ гипотеза принимается, и наоборот, по аналогии с описанными критериями.

По своим свойствам критерий ω^2_N близок критерию Колмогорова, но при его использовании несколько увеличивается сложность расчетов. Однако этот критерий применим при малом числе отсчетов. Считается [24], что удовлетворительные результаты могут быть получены при N , равном нескольким отсчетам.

Проиллюстрируем применение критерия ω^2_N на данных из предыдущего примера. Для эксперимента № 1 статистика ω^2_N равна:

$$\omega^2_N = 1/(12 \times 10) + [0,54 - (2 \times 9 - 1)/(2 \times 10)]^2 + \\ + [0,13 - (2 \times 2 - 1)/(2 \times 10)]^2 + \dots + [0,37 - (2 \times 4 - 1)/(2 \times 10)]^2 + \\ + [0,45 - (2 \times 6 - 1)/(2 \times 10)]^2 = 0,0083 + 0,0961 + \\ + 0,0004 + \dots + 0,0004 + 0,01 = 0,29.$$

Согласно табл. П1 критические значения равны: $Z_{\omega}(0,8) = 0,24$ и $Z_{\omega}(0,9) = 0,35$. Следовательно, при $P_d \approx 0,8$ гипотеза о равномерном распределении отвергается. Расчет для эксперимента № 2 дает $\omega^2_N = 0,18$, что позволяет принять нулевую гипотезу. Таким образом, выводы, полученные с помощью критериев Колмогорова и ω^2_N , совпадают.

Критерий нормальности Шапиро — Уилка. Для проверки на нормальность пригодны все приведенные выше критерии. Однако относительно недавно (1965 г.) был предложен эффективный критерий нормальности, пригодный при небольших объемах выборки.

Рассмотрим порядок использования этого критерия. По выборке формируется вариационный ряд. Далее вычисляется сумма b взвешенных с коэффициентами a_{n-i+1} разностей между наибольшими и наименьшими значениями выборки, начиная с самых крайних, т. е.

$$b = a_n(x_n - x_1) + a_{n-1}(x_{n-1} - x_2) + \dots \\ \dots + a_{n-k+1}(x_{n-k+1} - x_k) = \sum_{i=1}^k a_{n-i+1}(x_{n-i+1} - x_i), \quad (2-16)$$

где $k=N/2$ при четном N и $k=(N-1)/2$ при нечетном. При нечетном N средний член вариационного ряда x_{k+1} не используется. Коэффициенты a_{n-i+1} приведены в табл. П2 [49]. Статистика критерия имеет вид:

$$V = b^2 / S_x, \quad (2-17)$$

где

$$S_x = \sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2$$

— величина, пропорциональная выборочной дисперсии (§ 4-1).

Для получения выводов о справедливости нормального закона необходимо задаться доверительной вероятностью P_d и сравнить статистику V с критическими значениями $Z_V(N, P_d)$, приведенными в виде графика на рис. 2-11, построенного по материалам [49]. В противо-

положность предыдущим случаям гипотеза принимается при $V \geq Z_V(N, P_d)$ и отвергается при $V < Z_V(N, P_d)$.

Критерий применим для проверки согласия эмпирических распределений с законами, приводимыми к нормальным, например с логарифмически нормальным; известна также модификация, позволяющая применять критерий к экспоненциальному распределению [49].

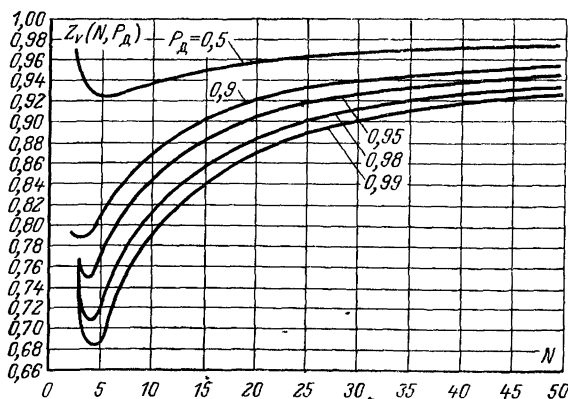


Рис. 2-11. Критические значения критерия Шапиро — Уилка.

Пример. Пусть имеется 10 независимых отсчетов процесса $x(t)$: $-0,6; 0,33; 0,55; 0,12; -1,29; -0,5; -1,05; 1,95; 0,07; 1,83$. Вариационный ряд, сформированный из отсчетов, имеет вид: $x_1 = -1,29$, $x_2 = -1,05$; $x_3 = -0,6$; $x_4 = -0,5$; $x_5 = 0,07$; $x_6 = 0,12$; $x_7 = 0,33$; $x_8 = 0,55$; $x_9 = 1,83$; $x_{10} = 1,95$. Величины N и k равны соответственно 10 и 5. Найдя суммы $\sum_{i=1}^N x_i = 1,41$ и $\sum_{i=1}^N x_i^2 = 10,95$, вычислим:

$$S_x = \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 / N = 10,95 - 1,41^2 / 10 = 10,75.$$

По табл. П2 выбираем коэффициенты a_{10}, a_9, \dots, a_6 для $N=10$ и рассчитываем величину b по формуле (2-16): $b = 0,57[1,95 - (-1,29)] + 0,33[1,83 - (-1,05)] + 0,21[0,55 - (-0,6)] + 0,14[0,33 - (-0,5)] + 0,04[0,12 - (-0,07)] = 2,71$.

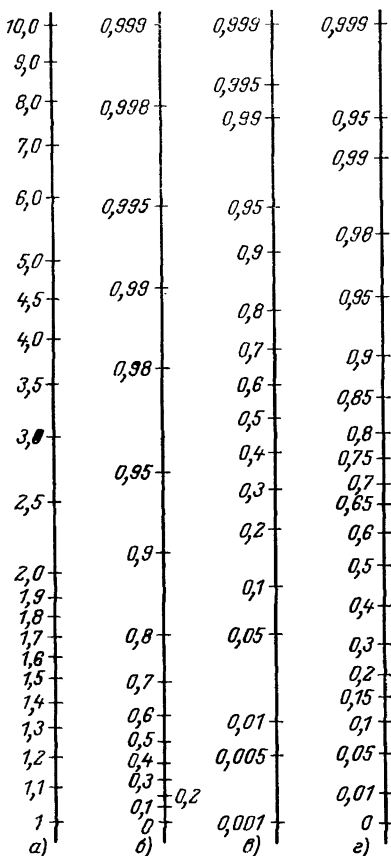
Тогда статистика V критерия равна:

$$V = \frac{b^2}{S_x} = \frac{2,71^2}{10,75} = 0,69.$$

Сопоставление V с критическими значениями $Z_V(N, P_d)$ на рис. 2-11 позволяет принять нулевую гипотезу с доверительной вероятностью $P_d > 0,99$.

2-5. ГРАФИЧЕСКАЯ ПРОВЕРКА СОГЛАСИЯ С ПРЕДПОЛАГАЕМЫМ РАСПРЕДЕЛЕНИЕМ

В основе графической проверки лежит такой способ представления эмпирического распределения, чтобы функция $W^*(x)$ в определенной координатной сетке была близка к прямой линии, если она соответствует гипотетическому распределению. Такая сетка задается нелинейной осью ординат. Вид нелинейности зависит от типа проверяемого распределения. Чтобы построить вероятную шкалу для требуемого распределения $W(x)$,



надо откладывать по оси ординат квантили x , а надписывать значение $W(x)$. Наиболее просто построение сетки, или, как ее называют, вероятностной бумаги, для табулированных распределений.

В некоторых случаях удобно делать нелинейной и ось абсцисс. Например, если по оси ординат отложить нормальную шкалу, а по оси абсцисс — функцию $\lg x$, то будет образована сетка для логарифмически - нормального распределения. Иногда удобно производить дополнительное нелинейное преобразование и по оси ординат. Так строится, например, сетка для распределения Вейбулла используемого при контроле надежности [51]. Несколько широко используются шкалы приведено на рис. 2-12, а—г.

Рис. 2-12. Наиболее распространенные шкалы.

а — логарифмическая; б — экспоненциальная; в — нормальная; г — пелевская.

Если проверяется согласие эмпирического распределения с полностью известным, то вероятно

стную шкалу построить легко. Если же известен лишь тип гипотетического распределения, а параметры не известны, то сетку можно построить для распределений, форма которых устойчива при изменении параметров. К таким распределениям относятся однопараметрические, например релеевское и экспоненциальное, а также двухпараметрические, нормальное или лапласово, коэффициенты асимметрии и эксцесса которых (§ 2-6) не зависят от других параметров.

Как правило, функции $W^*(x)$, даже выбранные из известного распределения, отклоняются от прямой на участках, удаленных от средней части. И наоборот, самые различные распределения, построенные на какой-либо вероятностной бумаге, мало отклоняются от прямой вблизи моды распределения. Отклонение от прямой, имеющее систематический характер, подсказывает исследователю, какое распределение стоило бы попытаться подобрать.

Графическая проверка согласия представляет собой очень приближенный, но удобный способ контроля оценок распределений. Отметим также, что по вероятностной бумаге легко определять квантили распределений, а по ним находить их параметры. Для любых распределений по графикам сразу определяется медиана: $W(\text{Me}_x) = 0,5$. У однопараметрических распределений параметры m_x и σ_x пропорциональны медиане. Для нормального распределения, например, σ_x задается как квантиль $W(x = -\sigma_x) = 0,1587$ или $W(x = \sigma_x) = 0,8413$.

Пример. На рис. 2-13 изображена релеевская сетка с равномерной шкалой x в относительных единицах. На сетке соединены точки эмпирического распределения некоторого эксперимента. По графику видно, что в пределах от 5 до 80—85% вероятности распределение можно считать близким к релеевскому. Легко определяются и параметры распределения на основе соотношений для теоретического распределения Релея. Медиана обозначена на графике $x_3 = 2$. Другие характеристики равны: $m_x = x_{54,4\%} = x_4 = 2,2$; $\sigma_x = x_{19,3\%} = x_1 = 1,14$; мода соответствует $x_{39,4\%} = x_2 = 1,68$. В то же время, если бы требовалось более точное соответствие подобранного распределения на больших вероятностях, следовало бы искать другое распределение.

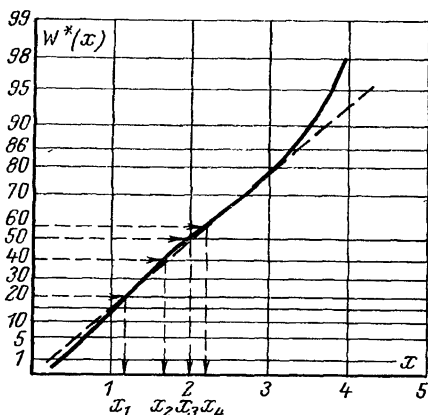


Рис. 2-13. Пример использования релеевской вероятностной шкалы.

2-6. ВЫЯВЛЕНИЕ ВИДА ЗАКОНА РАСПРЕДЕЛЕНИЯ ПО ОЦЕНКАМ ПЛОТНОСТЕЙ И МОМЕНТОВ

В экспресс-анализе бывает удобно использовать характерные точки оценок $W^*(x)$ и $w^*(x)$ для идентификации распределений. Если учесть априорную информацию о распределениях, то эти точки, как и оценки моментов, могут подсказать исследователю вид оцениваемого закона распределения.

Для идентификации функций $w^*(x)$ используют наибольшее значение — моду распределений и ординату моды, величины m_x^* и σ_x^* , определенным образом связанные с параметрами распределений, а также три коэффициента, описывающие рассеяние и форму распределения. Первый из них — уже упоминавшийся коэффициент вариации κ_x , равный

$$\kappa_x = \sigma_x / m_x, \quad (2-18)$$

характеризует размытость распределения относительно среднего, второй — коэффициент асимметрии K_a , равный

$$K_a = m [(x - m_x)^3] / \sigma_x^3, \quad (2-19)$$

характеризует «кособокость» плотности $w(x)$. Для симметричных распределений $K_a = 0$. Если левый склон $w(x)$ круче правого, то $K_a > 0$, и наоборот. Приплюснутость вершины функции $w(x)$ описывается коэффициентом эксцесса K_{Ξ} :

$$K_{\Xi} = m [(x - m_x)^4] / \sigma_x^4 - 3. \quad (2-20)$$

Мерой эксцесса считается K_{Ξ} нормального распределения, равный нулю. У распределений с более острой вершиной $K_{\Xi} > 0$, например у лапласова распределения. При более тупой вершине, например у равномерного распределения, $K_{\Xi} < 0$.

В табл. ПЗ приведены характеристики некоторых широко используемых на практике распределений. Приведенное в таблице гамма-распределение часто бывает удобно для аппроксимации различных односторонних распределений, так как может принимать самые различ-

ные формы в зависимости от параметров. Для аппроксимации экспериментально полученной плотности $w^*(x)$ вычисляют m_x и σ_x и рассчитывают параметры распределения по формулам $\alpha_T = \sigma_x^2/m_x^2 - 1$, $\beta_T = \sigma_x^2/m_x$.

Рассмотрим пример практического применения соотношений из табл. ПЗ. Пусть при небольшом количестве разрядов оценена плотность $w(x)$ и она имеет экспоненциально спадающий характер. Возникает вопрос: надо ли пересчитать гистограмму на большее количество разрядов, чтобы выявить начальный возрастающий участок плотности, или этого делать не стоит? Ответ на этот вопрос дает коэффициент вариации. Если он близок к единице, то пересчет не даст ничего нового. Если же v_x существенно меньше единицы, то информация о начальном участке потеряна и имеет смысл повторить расчет $w^*(x)$ при меньшей ширине разряда h .

ГЛАВА ТРЕТЬЯ

ОПРЕДЕЛЕНИЕ СРЕДНИХ ЗНАЧЕНИЙ

3.1. ОЦЕНКА СРЕДНЕГО ЗНАЧЕНИЯ ПО РАВНООТСТОЯЩИМ ОТСЧЕТАМ

Среднее значение или математическое ожидание является одной из важнейших характеристик случайных процессов. Во многих случаях практики измерений величина среднего значения несет в себе информацию об измеряемом параметре, которая часто оказывается искаженной случайными факторами. К этой задаче близко примыкает задача фильтрации полезного сигнала, подверженного действию шумов.

При определении среднего значения случайного процесса могут возникнуть две задачи. Первая задача — определить среднее значение по реализации случайного процесса конечной длительности, которую можно выбирать, планируя эксперимент с целью получения оценки среднего с заданной точностью. Вторая задача может быть сформулирована так: имеется реализация случайного процесса конечной и неизменной длительности, необходимо получить оценку среднего, определить погрешность оценки и, если можно, минимизировать ее рациональным выбором количества отсчетов. При определении среднего значения часто пользуются формулой среднего арифметического для равноотстоящих на ин-

тервал T_0 отсчетов процесса $X(t)$

$$m_x^* = \frac{1}{N} \sum_{i=0}^{N-1} x_i(iT_0), \quad (3-1)$$

где $T_0 = T/(N-1)$.

Оценка среднего в форме (3-1) является состоятельной, эффективной и несмещенной, при этом дисперсия оценки среднего (3-1) для независимых отсчетов убывает с ростом N в соответствии с зависимостью

$$D[m_x] = D_x/N. \quad (3-2)$$

В практике оперативного анализа обычно приходится иметь дело с небольшим объемом выборки, поэтому вопросы точности и достоверности получения оценок m_x^* играют существенную роль при трактовке полученных результатов.

Случайная величина оценки распределена нормально ($N > 6$) со средним m_x и $\sigma = \sqrt{D[m_x]}$. Поэтому, если получена оценка σ_x^* , ширину доверительного интервала можно выразить через разность $\Delta_m = m_x^* - m_x$ и найти доверительную вероятность P_d :

$$P_d[m_x^* - \varepsilon_{1m} \sigma_x^* \leq m_x \leq m_x^* + \varepsilon_{1m} \sigma_x^*] = 2 \cdot \int_0^{\varepsilon_{1m}} w_\varepsilon(\varepsilon) d\varepsilon,$$

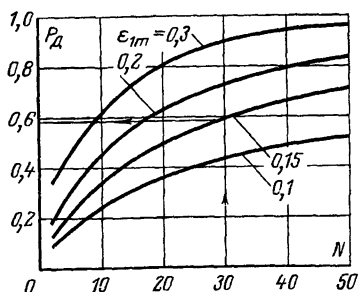


Рис. 3-1. Зависимость доверительной вероятности от объема выборки и относительной ошибки при определении математического ожидания как среднего арифметического.

где случайная величина ε_{1m} , определяемая равенством $\varepsilon_{1m} = \Delta_m / \sigma_x^*$ и характеризующая относительную погрешность, подчинена закону распределения $w_\varepsilon(\varepsilon)$, называемому «распределением Стьюдента» [1].

На рис. 3-1 показаны зависимости доверительной вероятности P_d от ε_{1m} и объема выборки N , рассчитанные для нормальных случайных чисел x_i по формулам, приведенным в [29].

Как указывалось, формула (3-2) соответствует независимым отсчетам, т. е. при $T_0 > \tau^*_k$ (интервал между отсчетами больше оценки интервала корреляции). Для решения первой из сформулированных в начале параграфа задач достаточно выбрать интервал T так, чтобы из соображений точности обеспечить требуемое число независимых отсчетов, т. е. $T \geq T_0 N > \tau^*_k N$.

Однако такая возможность существует не всегда: иногда интервал времени T не может быть увеличен и тогда необходимо решать вторую задачу. Известно [32], что дисперсия оценки среднего для коррелированных отсчетов для процессов с нормированной корреляционной функцией $\rho_x(iT_0)$ равна:

$$D[m_x] = \frac{D_x}{N} \left[1 + 2 \sum_{i=1}^{N-1} \left(1 - \frac{i}{N} \right) \rho_x(iT_0) \right]. \quad (3-3)$$

Из этой формулы следует, что при бесконечном делении интервала T , т. е. при искусственном увеличении объема выборки N , величина дисперсии $D[m_x]$ стремится к значению D_x . Таким образом, искусственное увеличение объема выборки в пределах интервала T приводит к увеличению дисперсии. Аналогичный результат получается, если в пределе рассматривать только один отсчет. Тогда из (3-2) следует, что $D[m_x] = D_x$. Из этих рассуждений следует, что дисперсия оценки среднего, полученного по отсчетам выборки ограниченной длительности, имеет минимум при определенном объеме выборки $N_{\text{опт}}$. Другими словами, существует оптимальное число $N_{\text{опт}}$, которое минимизирует дисперсию оценки среднего. Это положение было подмечено и сформулировано в виде теоремы С. Я. Виленкиным [6].

С целью оптимизации объема выборки по минимуму дисперсии оценки для коррелированных отсчетов при оценке среднего авторами были выполнены расчеты на ЦВМ в соответствии с (3-3) для некоторых видов корреляционных функций.

Графики зависимости $D[m_x]/D_x = \eta$ от N и αT показаны на рис. 3-2 для $\rho_x(\tau) = e^{-\alpha|\tau|}$. Из приведенных кривых видно, что величина η довольно быстро уменьшается до минимума, а затем справа от минимума медленно возрастает, всегда находясь выше кривой для некоррелированных отсчетов. Из показанных кривых получено уравнение для определения оптимального количества отсчетов при ограничен-

ном времени наблюдения T , минимизирующих статистическую погрешность оценки среднего для $\rho_x(\tau) = e^{-\alpha|\tau|} N_{\text{опт}} = 1 + E[1,6\alpha T + 0,06(\alpha T^2)]$, где $E[\cdot]$ — целочисленная функция. Аналогичные расчеты были выполнены для $\rho_x(\tau) = \frac{\sin \omega_b \tau}{\omega_b \tau}$, а основные результаты которых сведены в табл. 3-1.

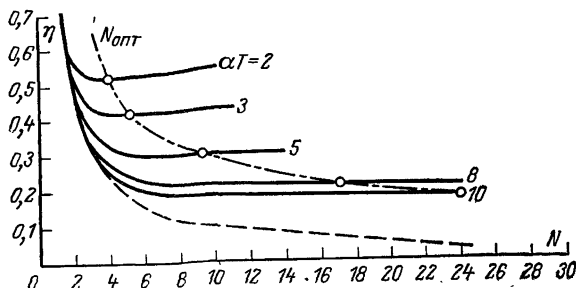


Рис. 3-2. Зависимость относительной погрешности оценки среднего от числа корреляционных отсчетов для $\rho_x(\tau) = e^{-\alpha|\tau|} N_{\text{опт}} = 1 + E[1,6\alpha T + 0,06(\alpha T^2)]$; штриховая линия соответствует некоррелированным отсчетам $\eta = 1/N$.

Для рассматриваемого случая оптимальное число отсчетов, минимизирующее статистическую погрешность оценки среднего значения, для $\omega_b T \geq 10$ равно:

$$N_{\text{опт}} = 1 + E[0,25\omega_b T] = 1 + E\left\{\frac{\pi}{2} T f_b\right\}.$$

Таблица 3-1

Зависимость главного минимума относительной статистической погрешности оценки среднего от $\omega_b T$ и N для процессов с НКФ $\rho_x(\tau) = \sin \omega_b T / \omega_b T$

$\omega_b T$	N	Относительная погрешность $D_{\text{опт}} m_x / D_\lambda$	Относительная погрешность для некоррелированных отсчетов
10	3	0,215	0,333
20	6	0,115	0,167
30	8	0,090	0,125
50	13	0,050	0,077
80	18	0,035	0,050

Последний результат особенно важен при анализе процессов в трудновоспроизводимых условиях, т. е. в тех случаях, когда эксперимент трудно или дорого повторить или невозможно спланировать. Если спектр процесса можно принять близким к прямоуголь-

ному в полосе от 0 до f_v при длительности реализации \tilde{T} , то легко определить оптимальное число отсчетов для оценки среднего значения. Пусть $T=10$ с, а $f_v=1$ Гц, тогда $N_{\text{опт}}=1+E[0,5\pi\cdot 10]=16$, т. е. среднее целесообразно определять по шестнадцати равноотстоящим отсчетам.

Для процессов с экспоненциально-косинусными корреляционными функциями можно принять следующую эмпирическую зависимость оптимального числа коррелированных отсчетов от $T/\tau_{\text{к}}^*$, минимизирующих статистическую погрешность оценки среднего (при длительности реализации T и оценке интервала корреляции $\tau_{\text{к}}^*$):

$$N_{\text{опт}}=1+2E\left[\frac{T}{\tau_{\text{к}}^*}\right]=1+2N_{\text{э}}.$$

Эту зависимость целесообразно использовать при относительно малых значениях $N_{\text{э}}\leq 4$, т. е. при относительно коротких реализациях. В тех случаях, когда корреляционные функции процессов знакопеременны, определение среднего по коррелированным $N_{\text{опт}}$ отсчетам может обладать меньшими статистическими погрешностями, чем оценка по $N_{\text{э}}$ некоррелированным отсчетам.

3.2. ОПРЕДЕЛЕНИЕ СРЕДНЕГО ЗНАЧЕНИЯ С ПОМОЩЬЮ РАЗМАХОВЫХ ОЦЕНОК

Размаховые оценки обычно связывают с действиями над наибольшей максимальной x_1 и наименьшей минимальной x_2 зарегистрированными величинами выбросов в заданной реализации случайного процесса. Достаточно просто оценка среднего значения может быть получена для нормального случайного процесса, реализация которого показана, например, на рис. 3-3.

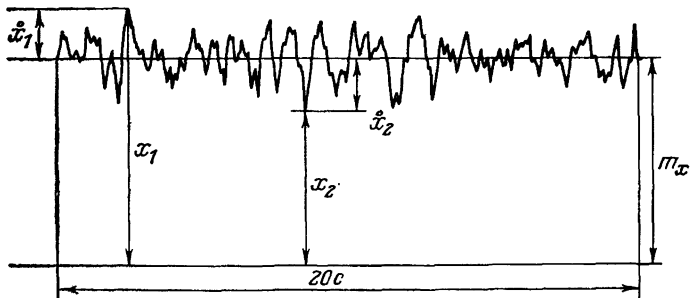


Рис. 3-3. Определение математического ожидания как полусуммы наибольшего и наименьшего значений нормального случайного процесса в данной реализации.

Иногда значения x_1 и x_2 называют соответственно максимантом и минимантом реализации процесса. Тогда можно получить оценку m_x^* среднего значения в виде

$$m_x^* = 0,5(x_1 + x_2) = m_x + 0,5(\dot{x}_1 - \dot{x}_2) = m_x + \Delta, \quad (3-4)$$

т. е. в качестве оценки среднего значения можно принимать полусумму максиманта и миниманта реализации случайного процесса. На рис. 3-3 $m_x = 43$, а $m_x^* = 0,5(53 + 32) = 42,5$. Величина Δ представляет собой случайную погрешность оценки среднего. Описанная процедура является очень простой и пригодна для определения среднего для процессов с симметричным распределением.

Для широкого использования размаховых оценок необходимо дать точностные характеристики метода. Точность полученных оценок будет рассмотрена на примере нормальных процессов.

С целью установления правил выбора максиманта и миниманта будем считать, что каждый из них превосходит некоторый уровень x_0 , в β раз больший значения среднеквадратического отклонения процесса σ_x , т. е. $x_0 = \beta\sigma_x$. Таким образом, ассортимент экстремумов процесса ограничен снизу значением $\beta\sigma_x$. Такое ограничение аналогично пропусканию случайного процесса через нелинейный преобразователь с характеристикой

$$\psi = \begin{cases} 0 & \text{при } x(t) < x_0; \\ x(t) & \text{при } x(t) \geq x_0. \end{cases} \quad (3-5)$$

Тогда дисперсия D_x каждой из величин x_1, x_2 может быть записана в виде [29]

$$\begin{aligned} \frac{D_{x_1}}{D_{x_2}} &= \frac{1}{2} \{(\beta^2 + 1) [1 - \Phi(\beta)]\} - \beta\varphi(\beta) - \\ &- \left\{ \frac{1}{2} \beta [1 - \Phi(\beta)] - \varphi(\beta) \right\}^2 = \gamma(\beta), \end{aligned} \quad (3-6)$$

где

$$\varphi(\beta) = \frac{1}{\sqrt{2\pi}} e^{-\beta^2/2}; \quad \Phi(\beta) = 2 \int_0^\beta \varphi(\beta) d\beta. \quad (3-7)$$

Дисперсия погрешности Δ с учетом того, что минимант и максимант реализации случайного процесса разделены интервалом времени, как правило, большим

интервала корреляции, и считаются независимыми, может быть определена в виде

$$D_{\Delta} = D [0,5 (\dot{x}_1 - \dot{x}_2)] = 0,25 (D_{x_1} + D_{x_2}) = 0,5 D_x, \quad (3-8)$$

так как $D_{x_1} = D_{x_2}$.

Распределение величины Δ будет нормальным с параметрами $m_{\Delta} = 0$ и $\sigma_{\Delta} = \sigma_x \gamma(\beta) / \sqrt{2}$, функция $\gamma(\beta)$ равна левой части (3-6).

Погрешность получения оценки m_x^* можно характеризовать относительной безразмерной величиной $\varepsilon_{2m} = \Delta / \sigma_x$, тогда доверительная вероятность, т. е. вероятность $P_d(\varepsilon_{2m}, \beta)$ того, что среднее значение m_x находится в интервале $m_x^* \pm \varepsilon_{2m} \sigma_x$, соответствующая неравенству

$$P_d(\varepsilon_{2m}, \beta) = P_d[m_x^* - \varepsilon_{2m} \sigma_x \leq m_x \leq m_x^* + \varepsilon_{2m} \sigma_x], \quad (3-9)$$

может быть определена интегрированием (это табулированная функция):

$$P_d(\varepsilon_{2m}, \beta) = \frac{2}{\sqrt{2\pi}} \int_0^{\varepsilon_{2m}} e^{-\varepsilon_{2m}^2 / 2} d\varepsilon_{2m} = 2\Phi_0'(\varepsilon_{2m}). \quad (3-10)$$

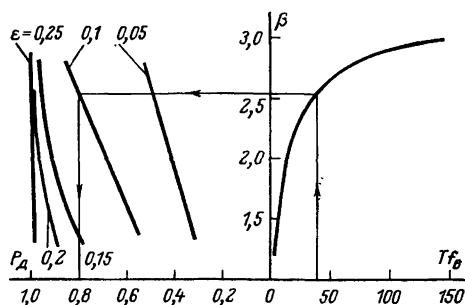


Рис. 3-4. Номограмма для определения доверительной вероятности при получении оценки математического ожидания как среднего нескольких полусумм ранжированных наибольших и наименьших значений нормального случайного процесса.

Результаты расчетов по формуле (3-10) с учетом соотношений (3-7) и (3-8) показаны на рис. 3-4 в левой части графика, из которого видно, что с увеличением β при равных величинах относительной погрешности доверительная вероятность возрастает.

Если необходимо определять среднее с меньшими погрешностями, то в качестве оценки можно использо-

вать среднее арифметическое от n полусумм экстремальных пар $(x_{1i} + x_{2i})$.

Использование экстремальных пар состоит в том, что находят среднее значение полусумм, убывающих по порядку экстремумов: первый по величине максимум — первый по величине минимум, второй максимум — второй минимум и т. д. Очевидно, что погрешность может быть уменьшена вдвое при $n=4$, т. е. уменьшится доверительный интервал при сохранении $P_d(\varepsilon_{2m}, \beta)$.

Выбор значения β для получения оценки математического ожидания (и среднеквадратического отклонения — § 4-3) основан на теории выбросов нормальных случайных процессов и подробно рассмотрен в [18, 19]. Зависимость параметра β от времени наблюдения T (длительность осциллограммы) и верхней граничной частоты спектра сигнала f_v нанесена на правую часть графика в виде соответствующей кривой (рис. 3-4). Таким образом, кривые рис. 3-4 образуют номограмму.

Правило пользования номограммой состоит в следующем. Если ориентировочно известна граничная частота спектра сигнала, то, находя произведение Tf_v , проводят вертикальную линию на правом графике до пересечения с кривой, соответствующей ошибке. Затем, опуская перпендикуляр из точки пересечения вспомогательной линии с соответствующей кривой, находят доверительную вероятность. Пример построения на номограмме выполнен для следующего случая: $Tf_v=40$, $\beta=2,54$ при $\varepsilon_{2m}=0,1$, $P_d(0,1; 2,54)=0,8$.

Пример использования описанного метода, показанный на рис. 3-3, по первой полусумме экстремальных значений $0,5(x_1+x_2)$ дает оценку $m^*_x=42,5$, в то время как заранее известное значение $m_x=43$ (в условных единицах — миллиметры в натуральном масштабе осциллограммы).

Следует отметить, что при практическом определении точности среднеквадратического отклонения σ_x неизвестно, поэтому в неравенстве (3-9) вместо σ_x можно использовать оценку σ^*_x , полученную любым (в частности, размаховым) методом. При этом доверительные вероятности несколько уменьшаются, и описанный метод определения точности становится приближенным. Аналогичным образом может быть разработан метод получения оценок среднего значения для других симметричных видов распределений.

3-3. ОЦЕНИВАНИЕ СРЕДНИХ ЗНАЧЕНИЙ С ПОМОЩЬЮ ПОРЯДКОВЫХ СТАТИСТИК

При решении ряда практических задач приходится оценивать среднее значение совокупности некоррелированных случайных чисел, при этом оценку среднего значения выборки объемом N можно получить со значи-

тельно меньшей трудоемкостью, чем при вычислении среднего арифметического, если использовать размаховые оценки [3]. Для этого напомним, что размещенные в порядке возрастания случайные числа x_i образуют вариационный ряд, в котором после перенумерации членов ряда соблюдается неравенство $x_{(i)} \leq x_{(i+k)}$, $k > 0$.

При такой нумерации член $x_{(i)}$ получил название i -й порядковой статистики. Крайние члены ряда $x_{(1)}$ и $x_{(N)}$ — экстремумы выборки — иногда называют соответственно минимантом и максимантом выборки, в отличие от миниманта и максиманта процесса. Для процесса эти величины выбирались из континуума значений (т. е. из всей непрерывной совокупности). Величина разности $\Delta_1 = x_{(N)} - x_{(1)}$ характеризует размах выборки, а разности

Таблица 3-2

Пример определения оценки среднего значения по выборочной медиане

Номер выборки	Порядковый номер статистики				
	1	2	3	4	5
1	0	3	5	6	9
2	0	1	6	6	9
3	1	4	5	7	9
4	2	3	4	7	8

вида $\Delta_i = x_{(N+1-i)} - x_{(i)}$ именуют подразмахами. Статистика $x_{(h)}$ при нечетном числе $N = 2k - 1$ и величина $Me^* = 0,5[x_{(h)} + x_{(h+1)}]$ при четном $N = 2k$ носят название выборочной медианы и, как видно в последнем случае, является полусуммой центральных порядковых статистик или средних членов вариационного ряда, т. е. медиана является границей, выше и ниже которой лежит ровно по половине объема выборки.

Если генеральная совокупность имеет симметричное распределение, то выборочная медиана может служить оценкой среднего значения. При наличии нескольких выборок среднее арифметическое выборочных медиан также может быть принято за оценку среднего значения.

Пример. Пусть имеется несколько выборок случайных чисел, являющихся результатом моделирования некоторой задачи — пример в табл. 3-2. Каждую выборку представим в виде вариационного

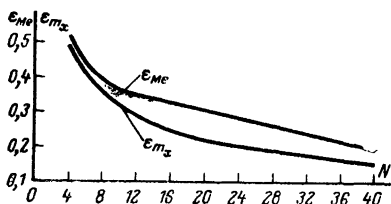


Рис. 3-5. Сравнение зависимостей относительных статистических погрешностей оценок математического ожидания от объема выборки по медиане и среднему арифметическому для нормальной генеральной совокупности.

Для нормального распределения генеральной совокупности дисперсия медианы может быть определена как

$$[24] D_{Me} = \frac{\pi}{2N} \sigma_x^2.$$

Если в качестве нормированной статистической погрешности рассмотреть величину $\epsilon_{Me} = \sigma_{Me} / \sigma_x$, то соответственно $\epsilon_{Me} = \sqrt{\pi / 2N}$. Интересно отметить, что аналогично отношение для нормированной погрешности оценки математического ожидания как среднего значения, вытекающего из формулы (3-2), составляет $\epsilon_{m_x} = \sqrt{1/N}$.

Сопоставление ϵ_{Me} и ϵ_{m_x} показывает, что нормированная статистическая погрешность определения медианы по одному размаху всегда оказывается несколько большей, чем при определении среднего значения. Однако незначительное уменьшение точности «окупается» существенным сокращением трудоемкости обработки. На рис. 3-5 показаны зависимости ϵ_{Me} и ϵ_{m_x} от N .

Следует также иметь в виду, что если оценка среднего получается как среднее значение выборочных медиан, полученных по каждой из q групп выборок, то погрешность ϵ_{Me} уменьшится до $\epsilon_{Meq} = \sqrt{\pi / (2qN)}$.

Для ряда законов распределения относительные погрешности ϵ_{Me} и параметр $\eta_m = \epsilon_{m_x} / \epsilon_{Me}$ сведены в табл. 3-3.

Доверительная вероятность может быть найдена при учете того, что выборочная медиана (при нормальном

ряда и найдем оценки средних значений. Выборочные медианы составляют соответственно 5; 6; 5; 4, а их среднее арифметическое равно 5. В то же время оценка среднего значения по среднему арифметическому всех шестнадцати чисел равна 4,7.

Из этого примера видна близость полученных оценок. Однако ответ на вопрос о точности получаемой оценки через выборочную медиану требует определения дисперсии этой оценки.

распределении генеральной совокупности) представляет собой нормальную случайную величину со средним значением, равным m_x и $\sigma_{Me} = \sqrt{D_{Me}}$ среднеквадратичного отклонения СКО. При расчете можно использовать

Таблица 3-3

Относительные статистические погрешности определения выборочной медианы и их сравнение с погрешностями оценки математического ожидания по среднему арифметическому

Вид распределения	N = 3		N = ∞
	σ_{Me}^*	η_m	η_m
Двустороннее экспоненциальное (Лапласа)	0,565	1,022	1,41
Нормальное	0,688	0,842	0,796
Симпсона	0,771	0,823	0,814
Равномерное	0,905	0,745	0,575

значение σ_{Me}^* , полученное из опыта через σ_x^* , тогда равенство

$$P_d [Me^* - B_m \sigma_{Me}^* \leq m_x \leq Me^* + B_m \sigma_{Me}^*] = 2\Phi_0(B_m), \quad (3-11)$$

где

$$\Phi_0(B_m) = \frac{1}{\sqrt{2\pi}} \int_0^{B_m} e^{-t^2/2} dt,$$

будет приближенным. Величина B_m является по существу критическим значением $Z_\Phi(P_d)$ нормального распределения. Ее значения приведены в табл. П1.

Другой путь определения математического ожидания выборочной совокупности состоит в вычислении так называемых оптимальных линейных оценок μ_x^* :

$$\mu_x^* = \sum_{i=1}^n a_i x_{(i)}. \quad (3-12)$$

Оптимальность линейных оценок понимается в том смысле, что дисперсия этих оценок меньше дисперсии оценки математического ожидания по среднему арифметическому значению, причем они сравнительно просто вычисляются. Эти оценки являются своего рода компромиссом между определением математического ожидания методом максимального правдоподобия [3] и вычислением простого среднего арифметического. В выражении (3-12) каждой порядковой

статистике (члену вариационного ряда) $x_{(i)}$ соответствует постоянный весовой коэффициент a_i . Величины a_i зависят от вида распределения генеральной совокупности и обладают тем свойством, что

$$\sum_{i=1}^n a_i = 1.$$

Вторым свойством оптимальных весов a_i является их симметричность для симметричных распределений: $a_i = a_{n+1-i}$. Интересно отметить, что оптимальные линейные оценки дают возможность определять математическое ожидание для любых вариантов усеченной выборки, пропадания отдельных членов вариационного ряда [3]. Там же показано, что при определении математического ожидания для выборки из нормальной генеральной совокупности оптимальная линейная оценка совпадает с оценкой по среднему арифметическому, которая была рассмотрена в § 3-1. Для некоторых других распределений коэффициенты a_i и отношение $\sigma[\mu^*]/\sigma_x$ приведены в табл. П4.

Коэффициентами, приведенными в табл. П4, удобно пользоваться для оценивания среднего и определения точности оценок для приведенных видов распределений.

Укажем еще на один прием определения математического ожидания, аналогичный тому, который обсуждался в § 3-2. В качестве оценки математического ожидания можно принять полусумму максиманта и миниманта вариационного ряда.

Если величину σ_x^* получить пропорционально размаху выборки (как будет показано ниже):

$$\sigma_x^* = \frac{1}{2\beta(N)} [x_{(N)} - x_{(1)}], \quad (3-13)$$

где $\beta(N)$ — коэффициент, зависящий от объема выборки N , то формула для определения $D_1[m_x]$ может быть принята для практических расчетов в виде [3]

$$D_1[m_x] = \frac{\pi [x_{(N)} - x_{(1)}]^2}{2N^2\beta^2(N)} \left[e^{\frac{x_{(1)}^2\beta^2(N)}{[x_{(N)} - x_{(1)}]^2}} + e^{\frac{x_{(N)}^2\beta^2(N)}{[x_{(N)} - x_{(1)}]^2}} \right]. \quad (3-14)$$

Пример. Пусть экстремумы [вариационного ряда из десяти членов ($N = 10$)] соответственно равны $x_{(N)} = 6$, $x_{(1)} = -4$, а $\frac{1}{2\beta(N)} = 0,325$ (см. рис. 4-4); $m_x^* = 0,5(6 - 4) = 1$; $\sigma_x^* = 0,325 [6 - (-4)] = 3,25$, тогда расчеты по (3-14) дают:

$$\sigma_1[m_x^*] = \sqrt{D_1[m_x^*]} = 2,4.$$

Учитывая, что оценка m_x^* представляет собой нормально распределенную случайную величину со средним значением m_x и среднеквадратическим отклонением $\sigma_1[m_x]$, можно найти доверительную

вероятность P_d , назначив доверительный интервал, например $\pm 1,5\sigma^*_{1}[m_x]$, тогда

$$P_d\{m^*_x - 1,5\sigma^*_{1}[m_x] \leq m_x \leq m^*_x + 1,5\sigma^*_{1}[m_x]\} = 2\Phi_0(1,5);$$

$$P_d[-2,6 \leq m_x \leq 4,4] = 0,87.$$

3-4. РЕКУРРЕНТНОЕ УСРЕДНЕНИЕ

В некоторых случаях эксперимент может длиться достаточно долго, и к моменту получения исследователем следующего отсчета могут быть рассчитаны оценки по результатам прошлых измерений. В таких случаях можно использовать методы рекуррентного усреднения [30, 39].

Рекуррентное усреднение основано на представлении оценки начального момента порядка k в виде

$$m^*_{k,n} = \frac{n-1}{n} m^*_{k,n-1} + \frac{1}{n} x^k_n. \quad (3-15)$$

Формула (3-15) для среднего по объему выборки n связывает оценку со средним, вычисленным по объему выборки $n-1$ и n -м отсчетом, взвешенными с соответствующими множителями.

Рекуррентный способ дает возможность экономии объема выборки прямо в ходе эксперимента за счет контроля точности оценки. При расчете рекуррентной оценки возникает вопрос: когда следует остановить процесс оценивания? В общем виде ответ почти очевиден — расчет надо прекратить тогда, когда разница между $m^*_{k,n}$ и m_k станет устойчиво меньше, чем заранее выбранная погрешность оценки m^*_k . Одно из правил остановки рекуррентного оценивания основано на текущем рекуррентном контроле старших моментов [39]. Так, при оценивании среднего $m^*_{1,n}$ параллельно может вычисляться оценка второго момента $m^*_{2,n}$, по которой рассчитывается текущее среднее квадратическое отклонение:

$$\sigma^*_n[m^*] = \sqrt{\frac{1}{n} (m^*_{2,n} - m^{*2}_{1,n})}. \quad (3-16)$$

Если значение величины $\sigma^*_n[m^*]$ стало меньше допустимого уровня, то эксперимент можно прекратить. Другое правило состоит в том, что эксперимент прекращают, когда требуемый десятичный знак оценки перестает меняться от одного значения n к другому [30]. Еще один принцип остановки процесса рекуррентного усреднения основан на контроле разности $\Delta_n = m^*_n - m^*_{n-1}$

[48] и применении к последовательности Δ_n теории серий успехов [47]. Под успехом понимают превышение величиной $|\Delta_n|$ заданного уровня. Однако используемые в [48] критерии малоприспособлены в экспресс-анализе из-за необходимости обрабатывать большие объемы выборки. Поэтому представляется целесообразным использовать теорию серий успехов в иной форме.

Предположим, что имеется стационарная последовательность независимых случайных отсчетов x_i . Для каж-

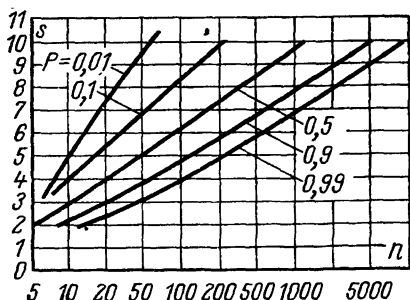


Рис. 3-6. Зависимость наибольшей длины серии «успехов» s , появляющейся с вероятностью P при n испытаниях.

видно, что контроль появления серии длиной 5—8 при n , лежащем в диапазоне 15—500, может явиться способом проверки, соответствует ли уровень U медиане последовательности.

Для применения теории серий необходимо из соображений требуемой точности выбрать уровень δ , характеризующий абсолютное допустимое отклонение Δ_n . Наблюдая за величиной $|\Delta_n|$, рано или поздно можно зарегистрировать серию заданной длины s . Поэтому, останавливая процедуру оценивания при s , равном 5—8, можно утверждать, что медиана разброса $|\Delta_n|$ стала меньше уровня δ с вероятностью, определяемой рис. 3-6. Рассматривая свойства последовательности Δ_n , можно показать, что $\Delta_{n+1}\Delta_n=0$, и найти дисперсию $D[\Delta_n]$, которая оказывается равной $2D_x/n^2$. Таким образом, величины Δ_n независимы и их дисперсия убывает по закону n^2 . Остановка оценивания может произойти через $n=n_1+n_2$ шагов, где n_1 определяется отношением σ_x/δ , а n_2 — графиками на рис. 3-6.

Каждого x_i проверяется событие $x_i < U$, понимаемое как успешный исход испытания. Для $U = Me_x$, т. е. уровня, равного медиане, известно распределение наименьшего числа испытаний n , при котором с вероятностью P наибольшая длина серии будет не менее s [37]. На рис. 3-6 приведено семейство кривых равной вероятности, соответствующих указанным распределениям.

По приведенным кривым

Пример. Рассмотрим процесс установления оценки среднего случайных чисел, равномерно распределенных на интервале $[0,1]$ (в этом случае $m_x=0,5$ и $\sigma_x \approx 0,3$). Зададимся $\delta=0,02$ и $s=8$. На рис. 3-7 показана последовательность $|\Delta_n|$. Из графика видно, что $n_1=12$, $n_2=s=8$. Оценивание остановлено после 20 расчетов m_n^* .

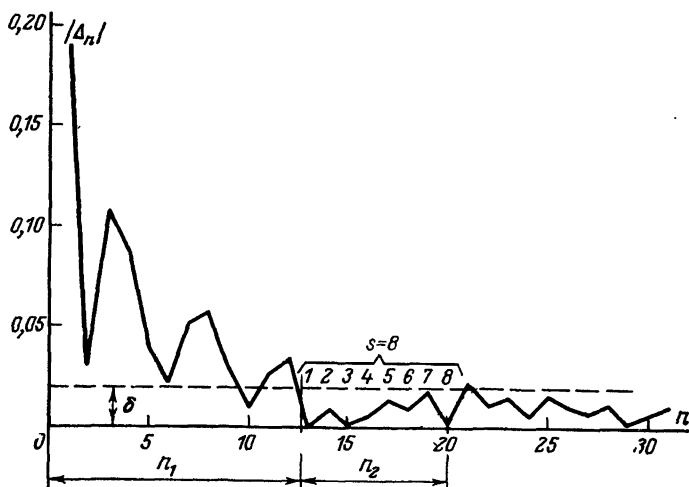


Рис. 3-7. Процесс установления отклонения $|\Delta_n|$ последовательной оценки среднего на n -м шаге от оценки среднего на $n-1$ -м шаге.

Описанный здесь метод усреднения полезен, когда проведение каждого последующего эксперимента занимает много времени и, кроме того, трудоемко или дорого. В этом случае назначать заранее объем выборки нежелательно. Лучше использовать последовательную оценку и остановить эксперимент немедленно после достижения требуемой точности.

3.5. ТЕКУЩЕЕ УСРЕДНЕНИЕ

Достаточно большой класс задач экспресс-анализа состоит в определении среднего значения за интервал времени, меньший, чем полное возможное время наблюдения T . В этом случае исходный процесс $X(t)$ можно представить в виде

$$X(t) = m(t) + \dot{X}(t), \quad (3-17)$$

где $m(t)$ — более медленная, чем $\dot{X}(t)$, составляющая, полученная текущим усреднением.

Следует заметить, что функция $m(t)$ не может быть определена до тех пор, пока не оговорены условия текущего усреднения, т. е. вид частотной характеристики алгоритма сглаживания и его параметры, а также спектр процесса $X(t)$.

Рассмотрим два крайних случая представления спектра $S_X(f)$ процесса $X(t)$. Пусть, как это показано на рис. 3-8,а, спектр имеет два выраженных пика. Один из

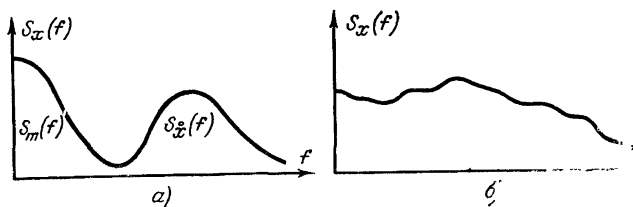


Рис. 3-8. Примеры спектров случайных процессов.

а — с явно выраженными пиками медленной и быстрой составляющих; б — без выраженных пиков.

них принадлежит медленной составляющей $S_m(f)$, а второй $S_X(f)$ — быстрой компоненте процесса $X(t)$.

В такой ситуации текущее среднее определено достаточно четко. Во многих задачах оно описывается как фильтрация «сигнала» $m(t)$ из смеси с «шумом» $\dot{X}(t)$. При этом $m(t)$ может быть любой. Если $m(t)$ — случайная функция, то она может быть и стационарной, и нестационарной. При текущем усреднении $m(t)$ ведет себя как нестационарная функция, если интервал усреднения T_T достаточно мал. Такие процессы рассматриваются как локально-нестационарные, даже если они и стационарны на длительном интервале T .

Другой возможный вид спектра показан на рис. 3-8,б. Здесь не видно выраженных пиков спектра и поэтому не ясно, что же считать средним (или «сигналом»). В таких ситуациях определение текущего среднего зависит от конкретных требований и нормативов того вида техники, к которому относится задача. Так, например, при исследовании шумов в телефонных каналах связи усредняют процесс на интервале, равном одной минуте и одному часу. Свойства среднeminутных и среднечасовых процессов нормированы и должны отвечать определенным требованиям [23].

В экспресс-анализе удобны два алгоритма текущего усреднения. Первый из них заключается в вычислении традиционной оценки среднего по $2n+1$ точкам и приписывании среднего точке с номером j :

$$m_j = \frac{1}{2n+1} \sum_{i=-n}^n x_j. \quad (3-18)$$

Последовательность m_j представляет собой отсчеты текущего среднего. В непрерывной форме алгоритм подобен интегрированию на интервале $(t-T_T, t+T_T)$.

Если скорости изменения процессов $\dot{X}(t)$ и $m(t)$ значительно отличаются, то сглаживание может выполняться вручную проведением на глаз средней линии суммарного процесса [при симметричном распределении $W(x)$]. Анализ погрешности в этом случае требует достаточно большой информации о процессе [40].

Ручное сглаживание тем точнее, чем меньше интервал корреляции процесса $x(t)$ по сравнению с T_T , чем меньше дисперсия исходного процесса и чем линейнее функция $m(t)$ на интервале усреднения T_T . Проводя гладкую линию $m(t)$, исследователь по существу сглаживает кусочно-линейную зависимость по участкам длительностью T_T .

В другом алгоритме, более удобном для спектров, показанных на рис. 3-8,б, используют усреднение отсчетов с весом, экспоненциально убывающим по мере удаления от точки, в которой вычисляется среднее. Алгоритм в дискретной форме задается разностным рекуррентным уравнением

$$m_j = \alpha x_j + (1-\alpha) m_{j-1}, \quad (3-19)$$

которое является уравнением цифрового фильтра первого порядка [9] с коэффициентом сглаживания α . Величину α можно оптимизировать, если известен вид функции $m(t)$. Чаще же просто задаются частотой среза ω_c сглаживающего фильтра, тогда [1]

$$\alpha = 1 - e^{-\omega_c T_0}. \quad (3-20)$$

где T_0 — шаг дискретизации процесса $x(t)$, с которым сняты отсчеты по осциллограмме.

Высокие статистические показатели этого метода усреднения подтверждены специальным анализом [27]. Кроме того, этот вид сглаживания предпочтителен для

отфильтровывания медленных составляющих при оценивании корреляционных функций локально-нестационарных процессов [21].

Таким образом, при сильно зашумленных записях можно усреднять вручную. Когда усреднять на глаз трудно, используют текущее сглаживание, причем экспоненциальное сглаживание имеет преимущества перед другими способами, например текущим сглаживанием с равным весом.

ГЛАВА ЧЕТВЕРТАЯ

ОПРЕДЕЛЕНИЕ СРЕДНЕКВАДРАТИЧЕСКИХ ОТКЛОНЕНИЙ

4-1. ОЦЕНКА СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ ПО РАВНООТСТОЯЩИМ ОТСЧЕТАМ

Среднеквадратическое отклонение (СКО) случайного процесса широко используется для характеристики точности работы различных систем. Размерность СКО совпадает с размерностью случайного процесса. СКО (σ_x) равно квадратному корню из дисперсии D_x . Известно [29], что для некоррелированных равноотстоящих отсчетов x_i сигнала (или случайных величин) оценка дисперсии D_x находится в соответствии с алгоритмом:

$$D_x^* = \frac{\sum_{i=1}^N (x_i - m_x^*)^2}{N-1}, \quad (4-1)$$

или

$$D_x^* = \left[\frac{\sum_{i=1}^N x_i^2}{N} - (m_x^*)^2 \right] \frac{N}{N-1}.$$

При больших N последний множитель в правом алгоритме мало отличается от единицы. Оценка m_x^* в (4-1) вычисляется, например, как среднее арифметическое по формуле (3-1).

Относительной мерой точности определения статистической дисперсии принято считать отношение СКО дисперсии к ее величине. Для нормального распределе-

ния величин [29] и некоррелированных отсчетов это отношение равно:

$$\frac{\sigma [D^*_x]}{D_x} = \sqrt{\frac{2}{N-1}}. \quad (4-2)$$

Формула (4-2) справедлива для оценки дисперсии, в которой использована оценка математического ожидания (3-1).

Ширина доверительного интервала и доверительной вероятности может быть определена [29] как

$$P_d [(1 - \varepsilon_{1\sigma}) \sigma^*_x \leq \sigma_x \leq (1 + \varepsilon_{1\sigma}) \sigma^*_x] = \int_{\frac{1}{1+\varepsilon_{1\sigma}}}^{\frac{1}{1-\varepsilon_{1\sigma}}} \omega(v) dv, \quad (4-3)$$

где $v = \sigma^*_x / \sigma_x$ — случайная величина с плотностью распределения $\omega(v)$, часто называемого распределением χ^2 .

Свойства плотности χ^2 достаточно хорошо изучены [42, 45]. Поэтому здесь приведем окончательные результаты расчетов, объединенные в графики, показанные на рис. 4-1.

Правило пользования этими графиками аналогично правилу определения доверительных интервалов по рис. 3-1. Например, при $N=20$ и $\varepsilon_{1\sigma}=0,2$ доверительная вероятность $P_d=0,785$ (все действия указаны стрелками).

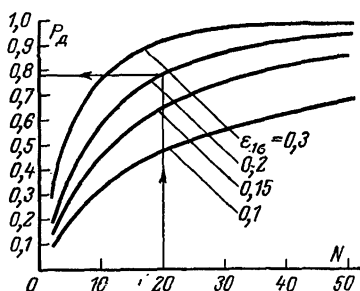


Рис. 4-1. Зависимость доверительных вероятностей от относительной статистической погрешности и объема выборки при определении СКО прямым методом.

4-2. ОПРЕДЕЛЕНИЕ СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ ПО РАЗМАХУ СЛУЧАЙНОГО ПРОЦЕССА

Описанный в § 4-1 метод определения СКО, точность которого достаточно хорошо изучена, является в значительной мере трудоемким. В задачах оперативного анализа очень удобным является метод получения оценки СКО по размаху случайного процесса, практически не требующий вычислительных операций.

Размахом реализации случайного процесса Δ_r длительностью T называется разность между максимантом и

минимантом реализации случайного сигнала $x(t)$. Иначе говоря, размах есть разность между главными экстремумами сигнала на интервале $[0, T]$. Обращаясь к рис. 3-2 и соотношению (3-4), найдем разность и преобразуем ее к виду

$$\Delta_{\sigma} = x_1 - x_2 = \dot{x}_1 - \dot{x}_2 = 2x_0 = 2 \frac{x_0}{\sigma_x^*} \sigma_x^* = 2\beta \sigma_x^*, \quad (4-4)$$

где x_0 — некоторый заданный условный уровень.

Из 4-4 следует, что оценку СКО можно получить как отношение

$$\sigma_x^* = \Delta \sigma / 2\beta_n. \quad (4-5)^*$$

Простое получение оценки СКО делением размаха сигнала на 6 (для нормальных процессов) стало при-

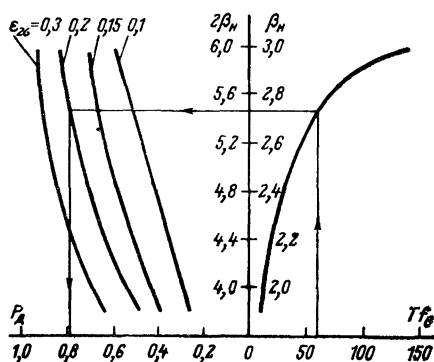


Рис. 4-2. Номограмма для определения коэффициента β_n и доверительной вероятности при получении оценки СКО по размаху нормального случайного процесса в зависимости от параметра Tf_v .

вычным для многих инженеров-исследователей. Однако деление размаха на 6 оказывается не всегда правильным даже для нормальных процессов. Оценка точности и доверительной вероятности приведена в [18, 19]. Результаты расчетов по определению относительного уровня в зависимости от интервала наблюдения T , верхней граничной частоты процесса показаны на правых частях рис. 3-3 и 4-2. На рис. 4-2 приведена номограмма для определения СКО. Слева приведены графики для полу-

* Здесь индекс «н» параметра β означает принадлежность к нормальному распределению в отличие от индекса «р», означающего принадлежность к релеевскому распределению.

чения значений доверительной вероятности P_d в зависимости от относительной погрешности $\varepsilon_{2\sigma} = \Delta\beta_n/\beta_n = \Delta\sigma_x^*/\sigma_x$:

$$P_d = P[(1 - \varepsilon_{2\sigma})\sigma_x^* \leq \sigma_x \leq (1 + \varepsilon_{2\sigma})\sigma_x^*].$$

Из рис. 4-2 следует, что при фиксированном Tf_v стремление к уменьшению $\varepsilon_{2\sigma}$ приведет к уменьшению P_d .

Рассмотренный метод относился к нормальным случайным процессам, но его можно распространить и на процессы с релеевским распределением. Напомним, что плотность релеевского распределения имеет вид:

$$w_p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}; \quad m_x = \frac{\sigma \sqrt{\pi}}{\sqrt{2}} \approx 1,25\sigma;$$

$$D_x \approx \frac{4 - \pi}{2} \approx 0,43\sigma^2. \quad (4-6)$$

Среднеквадратическое отклонение составляет:

$$\sigma_x = 0,655\sigma. \quad (4-7)$$

Если ввести, как и раньше, параметр $\beta_p = x_0/\sigma$, то оценку σ^* можно получить делением максимально наблюдаемого значения процесса x_1 на β_p .

Из теории выбросов случайных релеевских процессов с шириной спектра Δf известно [46], что среднее число максимумов, появляющихся в единицу времени и превосходящих достаточно высокий относительный уровень β_p , равно:

$$\mu = \Delta f \sqrt{\frac{\pi}{6}} \beta_p \exp\left(-\frac{\beta_p^2}{2}\right).$$

Тогда вероятность того, что на интервале $[0, T]$ найдется хотя бы один относительный максимум, превосходящий β_p , может быть представлена так

$$P(\beta_p, T\Delta f) = 1 - \exp\left[-T\Delta f \sqrt{\frac{\pi}{6}} \beta_p \exp\left(-\frac{\beta_p^2}{2}\right)\right].$$

Дифференцируя последнее выражение по β_p , представляющее собой функцию распределения этого параметра, можно найти связь между $T\Delta f$ и β_p для моды распределения:

$$T\Delta f = \sqrt{\frac{6}{\pi}} \frac{1}{\beta_p} \exp(\beta_p^2/2). \quad (4-8)$$

Соотношение (4-8) позволяет определить наиболее выгодное значение β_p , соответствующее максимальной вероятности обнаружения хотя бы одного максимума, превосходящего β_p . Зависимость β от Tf_v показана в правой

части рис. 4-3, в левой части которого построены кривые доверительных вероятностей, которые, как можно показать, определяются зависимостью

$$P_d [(1 - \varepsilon_{3\sigma}) \sigma^* \leq \sigma \leq (1 + \varepsilon_{3\sigma}) \sigma^*] = \\ = \exp [-(1 + \varepsilon_{3\sigma}) e^{-\beta_p^2} \varepsilon_{3\sigma}] - \exp [-(1 - \varepsilon_{3\sigma}) e^{\beta_p^2} \varepsilon_{3\sigma}]. \quad (4-9)$$

Так же как и раньше, кривые на рис. 4-3 можно использовать как номограмму для определения того значения β_p , на которое нужно разделить величину наибольшего из максимумов, обнаруженных на реализации

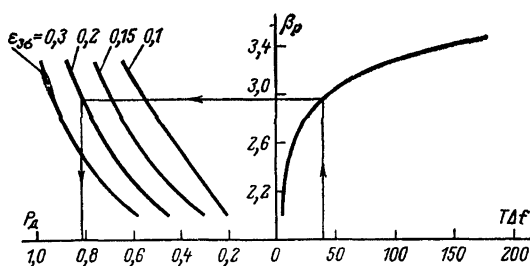


Рис. 4-3. Номограмма для определения коэффициента β_p и доверительной вероятности при получении оценки СКО по размаху релеевского случайного процесса в зависимости от параметра $T\Delta f$.

релеевского случайного процесса с шириной спектра Δf для получения оценки σ . Правило пользования номограммой указано стрелками на рис. 4-3.

Таким образом, на осциллограмме длительностью T отыскивается максимант x_1 и легко находится оценка параметра распределения $\sigma^* = x_1/\beta_p$, а затем определяют оценки m_x^* и σ_x^* (4-6) и (4-7).

4.3. ОПРЕДЕЛЕНИЕ СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ ПО РАЗМАХУ ВЫБОРКИ

Размаховые методы находят широкое практическое применение при обработке дискретных статистических массивов. Естественно, что размах выборки является случайной величиной, среднее значение и дисперсия которой зависят от объема выборки.

Как и раньше (4-5), оценку среднеквадратического отклонения σ_x^* можно получить по формуле

$$\sigma_x^* = \frac{x_{(N)} - x_{(1)}}{2\beta(N)} = \frac{\Delta_{1,2}}{2\beta(N)}, \quad (4-10)$$

где $x_{(N)} - x_{(1)}$ — разность экстремумов вариационного ряда (см. § 3-3).

Как отмечалось, для случайных процессов коэффициент 2β зависит от объема выборки N (а для случайных процессов — от величины Tf_B).

Значение $2\beta(N)$ приведено в виде кривой на рис. 4-4 на основе данных, приведенных в [3], и построена кри-

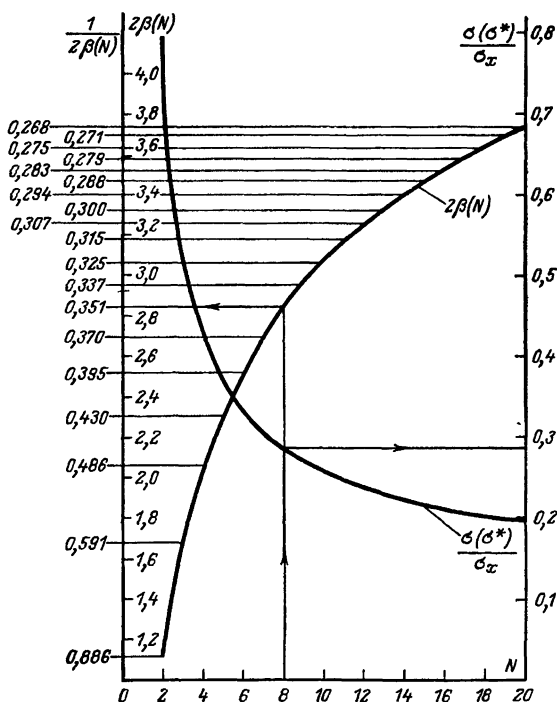


Рис. 4-4. Зависимость коэффициента 2β и относительного СКО оценки при определении СКО по размаху выборки объемом N .

вая $\sigma[\sigma_x^*]$. Интересно отметить, что при $N \geq 12$ точность оценки среднего арифметического значения σ_{xq}^* , полученного по выборкам q групп объема n каждая ($nq=N$) оказывается выше, чем точность одиночной оценки σ_x^* , полученной по размаху одной Δ_{1i} группы объемом N .

Было обнаружено [3], что погрешность получаемой оценки СКО наименьшая при $n=8$. Таким образом, можно рекомендовать следующее правило обработки результатов: планировать эксперимент так, чтобы общее число $N=8q$, а оценку СКО получать при $2\beta(8)=2,85$ как среднее значение, т. е.

$$\sigma_x^* = 0,351 \frac{\sum_{i=1}^q \Delta_{1i}}{q}. \quad (4-11)$$

В этом случае в соответствии с результатами, отображенными на рис. 4-4,

$$\frac{\sigma[\sigma_x^*]}{\sigma_x} = \frac{0,288}{\sqrt{q}}. \quad (4-12)$$

Еще один способ получения размаховой оценки СКО состоит в суммировании нескольких подразмахов (§ 3-3) и умножении полученной суммы на коэффициент $B(N, l)$, зависящий от объема выборки N и от числа складываемых подразмахов l , т. е.

$$\sigma_x^* = B(N, l) \sum_{i=1}^l \Delta_i, \quad (4-13)$$

где i принимает некоторые заданные значения $\leq l$.

Таблица значений коэффициентов $B(N, l)$ и сравнение точностей различных оценок СКО при разном объеме выборки приведена в табл. П5.

При весьма малых объемах выборки, т. е. при $N \leq 5$, удовлетворительные результаты получаются при использовании оптимальных линейных оценок, согласно которым

$$\sigma_{\text{хонт}}^* = \sum_{i=1}^N b_i x_{(i)}, \quad (4-14)$$

где b_i — постоянные весовые коэффициенты; $x_{(i)}$ — i -я порядковая статистика.

Значения b_i обладают тем свойством, что $\sum_{i=1}^N b_i = 0$.

Значения b_i при разных объемах выборки $N \leq 5$ для некоторых симметричных распределений можно найти в табл. П6 (таблица заимствована из [3]), там же приведены СКО, относительные погрешности этой оценки.

Таблица 4-1

**Исходные данные для расчетного примера
определения оценок СКО для нормальной
генеральной совокупности**

Номер порядковой статистики	Выборка № 1	Вариацион- ный ряд № 1	Выборка № 2	Вариацион- ный ряд № 2
1	+1,2	-6,0	+5,8	-4,2
2	+2,0	-4,8	-0,3	-0,3
3	+4,8	-4,7	-4,2	+0,5
4	+0,7	-4,0	+3,5	+3,5
5	+1,1	-2,4	+0,5	+5,8
6	-0,3	-1,2		
7	+4,0	-0,3		
8	-4,0	+0,7		
9	-6,0	+1,1		
10	+2,0	+1,2		
11	-4,7	+1,2		
12	-2,4	+2,0		
13	-4,8	+2,0		
14	-1,2	+3,0		
15	+1,2	+4,0		
16	+3,0	+4,8		

Пример получения оценок σ_x^* по всем упомянутым формулам приведен в табл. 4-1, 4-2.

Таблица 4-2

**Результаты расчета СКО по различным формулам
для примера из табл. 4-1**

Номер форму- лы	Номер выбор- ки	Параметры формулы	σ_x^*	$\sigma [\sigma^*]$	Примечание
(4-10)	1	$1/2\beta = 0,283$	3,06	0,65	По полной выборке
(4-11)	1	$q = 2$ $1/2\beta = 0,351$	3,12	0,635	По двум группам объемом $0,5N = 8$
(4-13)	1	$B = 0,131$ $i = 1, 2, 4$	3,36	0,625	По полной выборке
(4-14)	2	$b_1 = -0,372 = b_5$ $b_2 = -0,135 = b_4$ $b_3 = 0$	4,23	1,540	По частичной выбор- ке из общей сово- купности
(4-1)	1	$\sigma^* = \sqrt{\tilde{D}}$	3,31	0,61	По полной выборке

На основании изложенного можно рекомендовать следующее применение описанных методов при соответствующей модели случайного процесса:

если исследователь располагает реализацией случайного процесса, то оценку СКО целесообразно получать по размаху и подразмахам реализации;

если имеется выборка достаточного объема, то применение размаховых оценок типов (4-10) и (4-11) приведет к удовлетворительным результатам по точности, так же как и использование соотношений (4-1);

если объем выборки мал ($N \leq 5$), то лучшие результаты дают оптимальные линейные оценки (4-14).

ГЛАВА ПЯТАЯ

ОПРЕДЕЛЕНИЕ КОРРЕЛЯЦИОННЫХ И СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК СЛУЧАЙНЫХ ПРОЦЕССОВ

5-1. ОЦЕНИВАНИЕ КОРРЕЛЯЦИОННЫХ ФУНКЦИЙ

Для стационарных случайных процессов корреляционная функция (КФ) определяется выражением

$$R_x(\tau) = m[X(t)X(t+\tau)] \quad (5-1)$$

и дает сведения о среднем значении, дисперсии процесса и о степени статистической связи между отсчетами, разделенными интервалом времени τ . Однако среднее и дисперсия могут быть оценены более простыми средствами, описанными в предыдущих главах. Поэтому часто используется другой вид корреляционной функции, называемой иногда ковариационной и отличающейся от (5-1) тем, что действия производят с центрированными процессами.

Нормированная корреляционная функция (НКФ) вида $\rho_x(\tau) = R_x(\tau)/R_x(0)$, где $R_x(0) = D_x$, дает сведения только о корреляции сдвинутых на интервал τ отсчетов.

Рассмотрим методы приближенного оценивания НКФ и КФ.

Метод определения НКФ через оценку условного среднего значения [32]. Запишем теорему упреждения [19] для центрированного стационарного случайного процесса в виде, связывающем прогнозируемое на время

τ значение процесса $\dot{X}(t+\tau)$ через его значение в момент t и НКФ $\rho_x(\tau)$:

$$\dot{X}(t+\tau) = \dot{X}(t)\rho_x(\tau) + \dot{Y}(t)\sqrt{1-\rho_x^2(\tau)}, \quad (5-2)$$

где $\dot{Y}(t)$ — случайный независимый от $\dot{X}(t)$ процесс, среднеквадратическое отклонение которого $\sigma_y = \sigma_x$ (близкий к белому шуму).

Предположим, что в момент времени t значение $\dot{X}(t) = U = \text{const}$, тогда

$$\dot{X}(t+\tau) = U\rho_x(\tau) + \dot{Y}(t)\sqrt{1-\rho_x^2(\tau)}. \quad (5-3)$$

Из последнего соотношения следует, что условное математическое ожидание

$$m[\dot{X}(t+\tau) | \dot{X}(t) = U] = U\rho_x(\tau),$$

так как $m[\dot{Y}(t)] = 0$, откуда

$$\rho_x(\tau) = \frac{1}{U} m[\dot{X}(t+\tau) | \dot{X}(t) = U]. \quad (5-4)$$

О выборе уровня U будет рассказано ниже. Соотношение (5-2) положено в основу для определения НКФ при экспресс-анализе.

Получение НКФ эргодического стационарного случайного процесса осуществляется следующим образом. Реализация процесса регистрируется на осциллограмме в течение времени T . В начале осциллограммы, обычно в левой ее части, откладывается относительно нулевого уровня величина U по оси ординат (для процессов с симметричным распределением можно отложить два значения $\pm U$). Отмечаются точки пересечения линии, проведенной параллельно оси времени (линии уровня U) с реализацией процесса $x(t)$. Для копирования процесса нужно приготовить полосу вспомогательной бумаги (пергамента или кальки). Желательно, чтобы ширина полосы бумаги была не меньше ширины осциллограммы. Длина полосы вспомогательной бумаги должна соответствовать требуемому максимальному значению аргумента корреляционной функции $\tau_{\text{макс}}$ в том же масштабе k_t , мм/с, что и осциллограмма. Можно принять $\tau_{\text{макс}} = \tau_{\text{к}}^*$ (§ 5-2). В левой части вспомогательной полосы следует провести две взаимно перпендикулярные линии, одна из которых направлена вправо (ось времени, нулевая линия процесса), а вторая проведена вверх и вниз из выбранного начала. На оси ординат откладывается вверх в том же масштабе, что и на осциллограмме, значение U . Далее вспомогательная бумага накладывается на осциллограмму так, чтобы их оси времени совпали. Перемещая вспомогательную бумагу вдоль оси времени, добиваются совпадения отмеченной на ее оси ординат точки U с первой точкой пересечения, проведенной на осциллограмме линии уровня U с реализацией $x(t)$.

После совмещения указанных точек отрезок реализации процесса длительностью τ^*_0 с осциллограммы копируется на вспомогательную бумагу. Затем на эту же бумагу наносятся путем наложения друг на друга участки реализации процесса длительностью τ^*_0 , каждый из которых имеет, в соответствии с изложенным, начальные координаты $(0, U)$ на вспомогательной бумаге.

После нанесения всех отрезков реализации случайного процесса на вспомогательную бумагу интервал τ_k , а по длине бумаги $l = k_1 \tau^*_0$ делят на восемь — десять равных интервалов Δt для монотонных КФ, а для осциллирующих КФ — на четыре — шесть интервалов на период КФ. Более подробная методика выбора интервалов Δt из условий заданной ошибки аппроксимации описана в [19, 32].

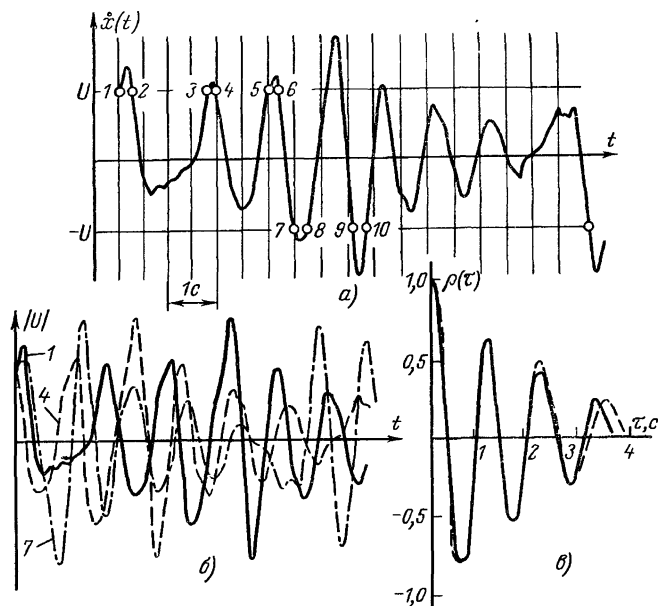


Рис. 5-1. Получение оценки нормированной корреляционной функции методом условного среднего.

а — реализация случайного процесса; *б* — отрезки реализаций, нанесенные на вспомогательную бумагу; *в* — сравнение оценки и нормированной корреляционной функции процесса [здесь сплошная линия для $\rho(\tau) = e^{-0,45\tau} \cos 6,28\tau$, штриховая — результат ручной обработки при $N=10$].

В сечениях, соответствующих значениям аргумента $i\Delta t$ ($i=0, 1, 2, \dots, n$), проводят усреднение ординат наложенных «кусков» реализации случайного процесса [по формуле (3-1) для получения оценки математического ожидания как среднего значения]. Полученные точки, соответствующие оценкам средних значений, соединяют плавной кривой, которая и представляет собой оценку нормированной корреляционной функции $\rho^*_x(\tau)$ в масштабе, равном U ($\rho^*_x(0) = U$ вместо единицы). На рис. 5-1 показан пример получения корреляционной функции описанным методом. На рис. 5-1, *в*

представлена реализация центрированного нормального случайного процесса, проведены уровни $\pm U$ и отмечены с 1-й по 10-ю точки пересечения процессом $x(t)$ уровней $\pm U$. На рис. 5-1,б показан отрезок вспомогательной бумаги с наложенными участками 1, 4, 7-й реализаций $x(t)$ (остальные участки на рисунке не показаны, чтобы его излишне не загромождать). Для точки пересечения 7-й реализации с уровнем $-U$ процесс копировался с обратным знаком, т. е. строился $-x(t)$. Результат усреднения всех десяти отрезков реализации иллюстрируется рис. 5-1,в, где сплошной линией проведена теоретическая корреляционная функция, с которой формировался процесс, а пунктиром отмечена полученная оценка $\rho_x^*(\tau)$. Видно достаточно хорошее совпадение полученного результата.

О связи погрешностей определения НКФ с выбором уровня U . Можно показать [32], что дисперсия оценки НКФ равна:

$$D[\rho_x^*(\tau)] = \frac{1 - \rho_x^2(\tau)}{U^2 N} D^*_{xx}. \quad (5-5)$$

Введем в рассмотрение относительную статистическую погрешность $\varepsilon_p = \sigma_p^*(\tau) / \rho_x(\tau)$ и параметр $\beta_p^* = U / \sigma_p^*(\tau)$, тогда

$$\varepsilon_p^2 = \frac{D[\rho_x^*(\tau)]}{U^2 \rho_x^2(\tau)} = \frac{1}{(\beta_p^*)^2} \frac{[1 - \rho_x^2(\tau)]}{N \rho_x^2(\tau)}. \quad (5-6)$$

При фиксированной длительности T реализации (осциллограммы) число ее участков, которые могут быть использованы при обработке, ограничено и не может превзойти среднее число выбросов случайного процесса на интервале T за уровень U .

Для процессов с симметричными распределениями (в частности, для нормальных процессов) число обрабатываемых участков реализации может быть увеличено вдвое, если использовать в качестве опорного уровня $\pm U$ с соответствующим учетом знаков обрабатываемых участков.

В соответствии с теорией выбросов число пересечений нормальным процессом уровня U с заданным знаком производной составляет:

$$\lambda(U) = \frac{\omega_1}{2\pi} e^{-\frac{U^2}{2\sigma_x^2}}, \quad (5-7)$$

где $\omega_1 = \sqrt{-\rho''_{xx}(0)}$, $\rho''_{xx}(0)$ — значение второй производной НКФ при $\tau=0$.

Если учесть, что при обработке принимаются в расчет участки пересечения процессом уровней $\pm U$ как с положительной, так и с отрицательной производной, величину $\lambda(U)$ следует увеличить в 4 раза. Тогда, учтя введенное выше обозначение β_p^* , получаем из (5-7) приближенное равенство

$$\lambda_{\text{полн}} \approx \frac{2\omega_1}{\pi} e^{-\frac{(\beta_p^*)^2}{2}}. \quad (5-8)$$

При этом число обрабатываемых участков N реализации случайного процесса может быть выражено формулой

$$N = \lambda_{\text{полн}} (T - \tau_k^*) = \frac{2\omega_1}{\pi} e^{-\frac{(\beta_p^*)^2}{2}} (T - \tau_k^*). \quad (5-9)$$

Подставляя результат (5-7) в (5-4), получаем, пренебрегая τ_k^* :

$$\varepsilon_p^2 \approx \frac{\pi}{2\omega_1} \frac{\exp [(\beta_p^*)^2/2]}{(\beta_p^*)^2} \frac{[1 - \rho_x^2(\tau)]}{\rho_x^2(\tau)}, \quad (5-10)$$

где $\omega_1 = \sqrt{-\rho''_x(0)}$, $\rho''_x(0)$ — значение второй производной НКФ при $\tau=0$.

Ошибка ε_p минимальна при $\beta_p^* = \sqrt{2}$. Таким образом, уровень U находится после определения σ_x^* , например, по размаху процесса, как указано в разделе 4-2, тогда

$$U = \sigma_x^* \sqrt{2}. \quad (5-11)$$

Характеристика ε_p не всегда удобна для практического использования. Для оценки дисперсии $\rho_x(\tau)$ удобно пользоваться соотношением (5-5).

Определение точности получения оценки корреляционной функции. Более предпочтительной по сравнению с ε_p для оценки точности является величина среднее квадратического отклонения оценки НКФ. С учетом (5-11) получим, извлекая квадратный корень из правой и левой частей равенства (5-5), выражение

$$\sigma_p^* = \sqrt{\frac{1 - \rho_x^2(\tau)}{2N}} \quad (5-12)$$

Эта величина по существу нормирована по отношению к $\rho(0) = 1$.

В табл. 5-1 приведена зависимость $\sigma_p^* [N, \rho(\tau)]$ по формуле (5-12). Из таблицы видно, что, например, для $\rho_x(\tau) = 0,5$ увеличение N в 10 раз (с 5 до 50) приводит к уменьшению σ_p в 3 раза (с 0,276 до 0,087).

Таблица 5-1

Зависимость $\sigma_p [N, \rho(\tau)]$ оценки нормированной корреляционной функции от числа N усредняемых отрезков реализации и уровня $\rho(\tau)$

$\rho(\tau)$	N					
	5	10	20	30	40	50
0,1	0,3200	0,2200	0,1580	0,1280	0,1120	0,0995
0,2	0,3000	0,2100	0,1550	0,1250	0,1100	0,0980
0,3	0,3050	0,2130	0,1520	0,1240	0,1080	0,0960
0,4	0,2900	0,2000	0,1480	0,1200	0,1000	0,0930
0,5	0,2760	0,1930	0,1380	0,1120	0,0980	0,0870
0,6	0,2500	0,1800	0,1270	0,1030	0,0900	0,0800
0,7	0,2290	0,1600	0,1140	0,0930	0,0808	0,0720
0,8	0,1900	0,1350	0,0950	0,0775	0,0675	0,0600
0,9	0,1380	0,0967	0,0690	0,0560	0,0488	0,0438
1	0	0	0	0	0	0

Доверительная вероятность попадания оценки НКФ в доверительный интервал $\rho_x(\tau) \pm 2Z_\Phi \sigma_p^*$ определяется равенством

$$P_d[\rho_x^*(\tau) - Z_\Phi \sigma_p^*(\tau) \leq \rho_x(\tau) < \rho_x^*(\tau) + Z_\Phi \sigma_p^*(\tau)] = 2\Phi(Z_\Phi), \quad (5-13)$$

где значения Z_Φ приведены в табл. П1.

Эта зависимость справедлива в силу того, что каждая оценка $\rho_x^*(\tau)$ представляет собой случайную, нормальную распределенную величину со средним значением $\rho_x(\tau)$ и со среднеквадратическим отклонением σ_p^* . Нормальность распределения оценки следует из того, что $\rho_x^*(\tau)$ получается суммированием достаточно большого числа случайных значений $\hat{x}(t)$. На рис. 5-2 приведен пример $\rho_x^*(\tau)$ с построенными доверительными интервалами $\pm \sigma_p^*(\tau)$ и $\pm 1,5\sigma_p^*(\tau)$, которым соответствуют доверительные вероятности 0,68 и 0,87 в соответствии с (5-11).

Можно показать на основании формул (5-7) и (5-9), что обрабатываемое число отрезков реализаций равно:

$$N = E \left[1,28 \left(\frac{T}{\tau_{\kappa}^*} - 1 \right) \right] + 1, \quad (5-14)$$

где $E[\cdot]$ — как и раньше, целочисленная функция; τ_{κ}^* — оценка интервала корреляции.

Соотношение (5-12) при заданной относительной погрешности и доверительной вероятности, определяемой по графику на рис. 3-1, может быть положено в основу планирования эксперимента по определению НКФ.

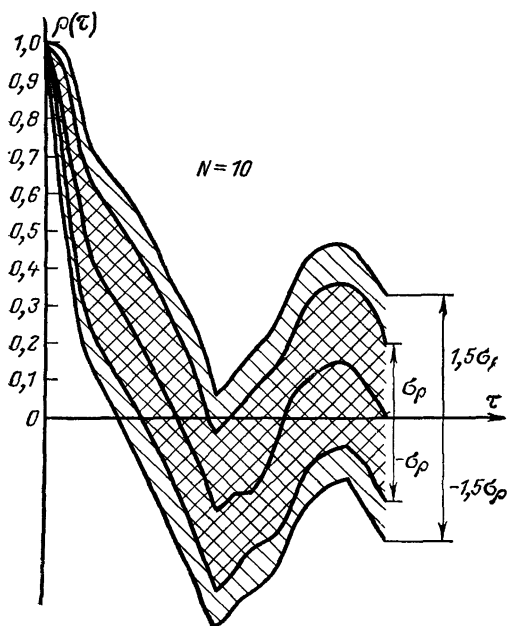


Рис. 5-2. Доверительные интервалы оценки нормированной корреляционной функции для $N=10$ и конкретного вида $\rho(\tau)$.

Для $P_d=0,68$ — двухсторонняя штриховка, для $P_d=0,87$ — односторонняя.

Определение оценки корреляционной функции через условное среднее знаковой функции. Этот метод оказывается более простым, чем описанный выше. Для эргодического нормального стационарного центрированного

процесса из формулы (5-3) можно записать:

$$\left. \begin{aligned} m[\dot{X}_\tau] &= m[\dot{X}(t+\tau) | \dot{X}(t) = U] = U\rho_x(\tau); \\ \sigma[\dot{X}_\tau] &= \sigma_x \sqrt{1 - \rho_x^2(\tau)}. \end{aligned} \right\} \quad (5-15)$$

Можно показать [33], что для центрированных нормальных процессов справедливо соотношение

$$m[\text{sign } \dot{X}_\tau] = 2\Phi_0 \left[\frac{\beta\rho_x(\tau)}{\sqrt{1 - \rho_x^2(\tau)}} \right] = 2\Phi_0(A). \quad (5-16)$$

Если в результате соответствующей обработки результатов эксперимента получена при известном β оценка математического ожидания в левой части равенства (5-16), то вычисление $\rho_x(\tau)$ не представляет серьезных трудностей. Для этого по таблицам функции Φ_0 достаточно найти значение аргумента A , а затем определить НКФ как

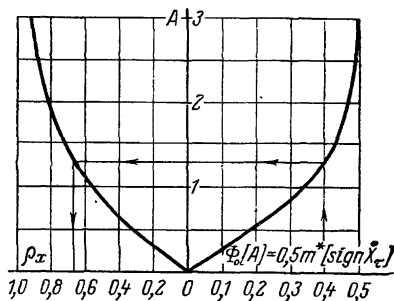


Рис. 5-3. Номограмма для определения нормированной корреляционной функции через условное среднее знаковой функции.

$$\rho_x^*(\tau) = \frac{A}{\sqrt{\beta^2 + A^2}}. \quad (5-17)$$

Будем считать, что величина $\beta = \sqrt{2}$ и выбрана, как и раньше, исходя из минимизации статистической погрешности, тогда при экспресс-анализе в качестве уровня U можно принимать величину $U = \sigma_x^* \sqrt{2}$, получаемую, например, через размаховую оценку СКО.

На рис. 5-3 показана номограмма для определения НКФ по условному среднему значению знаковой функции. Необходимые линии построения обозначены на рис. 5-3 стрелками. С допустимой для практики погрешностью (не превосходящей 6%) можно рекомендовать аппроксимацию выражения (5-17) в виде

$$\rho_x^*(\tau) = a_p \{m^*[\text{sign } \dot{X}_\tau]\},$$

где $a_p = 0,867$ при $0,05 \leq m^*[\text{sign } \dot{X}_\tau] \leq 0,996$, что соответствует $0,045 \leq \rho_x^*(\tau) \leq 0,9$. Этот результат, приведен-

ный в [33], позволяет получать оценку НКФ путем простого умножения на постоянный коэффициент оценки $m^* [\text{sign } \dot{X}_\tau]$.

Обработка осциллограмм описанным методом во многом аналогична рассмотренной выше.

По оценке σ_x^* выбирается уровень $U = \sigma_x^* \sqrt{2}$ и проводится на этом уровне горизонтальная линия параллельно оси абсцисс (или времени) осциллограммы. Точки пересечения уровня U с реализацией процесса $\dot{x}(t)$ являются опорными для получения отрезков реализации случайного процесса. Отличие заключается в том, что на удалении τ по оси времени от опорной точки усредняются не отсчеты процесса $\dot{x}_i(t + \tau)$, а лишь его знаки, т. е. $\text{sign } \dot{X}_{\tau i}$, до получения оценки $m^* [\text{sign } \dot{X}_\tau]$ (как среднего арифметического). Это обстоятельство позволяет не пользоваться вспомогательной бумагой, а делать разметку непосредственно на осциллограмме. Затем по найденной оценке $0,5m^* [\text{sign } \dot{X}_\tau]$ с помощью номограммы (рис. 5-3) через A или по приведенной выше аппроксимирующей формуле определяют оценку НКФ. Эта процедура повторяется для различных τ до получения оценки $\rho_x^*(\tau)$ в возможном диапазоне аргумента.

Следует отметить, что, несмотря на точное (по определению) значение $\rho_x(0) = 1$, в области малых τ возможны погрешности, так как в этой области наиболее ярко проявляется наличие всего двух уровней квантования, т. е. если $\dot{x}(t) = +U$, то при малых τ очень низка вероятность того, что $\dot{x}(t + \tau)$ станет меньше нуля, а для определения весьма малых вероятностей с высокой точностью требуются очень большие времена наблюдений.

Описанный метод является самым простым методом получения оценки НКФ непосредственно по осциллограмме.

Пример сравнения оценок НКФ, полученных при обработке реализации случайного процесса с помощью ЦВМ путем суммирования произведений по 4000 отсчетов, показан на рис. 5-4 (кривая 1). Кривая 2 на рис. 5-4 иллюстрирует применение описанного метода, причем оценка каждой ординаты нормированной корреляционной функции получена усреднением массива из 70 знаковых функций

задержанных отсчетов процесса. Интересно отметить, что вся процедура подобной обработки вручную занимает менее двух часов.

Ранговая корреляционная функция (РКФ). Для экспресс-анализа может быть предложен еще один способ получения оценки НКФ, основанный на оценке коэффициента ранговой корреляции Спирмена R_S [24]. При его вычислении используются ранги, и поэтому в расчетах

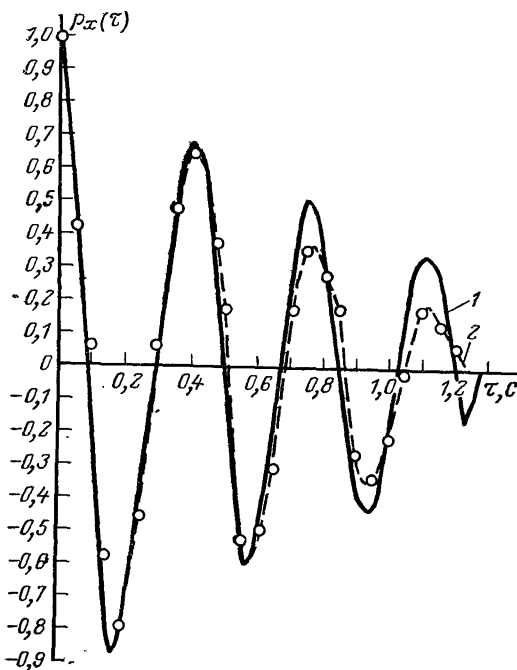


Рис. 5-4. Сравнение оценок нормированных корреляционных функций, полученных на ЦВМ по массиву из 4000 отсчетов (1) и методом условного среднего знаковой функции (2).

участвуют целые положительные числа, что упрощает вычисления. Значение R_S определяется по формуле

$$R_S = 1 - \frac{6 \sum d^2}{n(n-1)(n+1)}, \quad (5-18)$$

где n — объем каждой из выборок; d — разность рангов пар элементов выборок.

При построении РКФ анализируемые выборки формируются, как и при расчете обычной НКФ, путем сдвига отсчетов процесса $x(i\Delta t)$.

Например, для расчета $R_S(k\Delta t)$ первая выборка представляет собой последовательность $x(0), x(\Delta t), x(2\Delta t), \dots, x(n\Delta t)$, а вторая — последовательность $x(k\Delta t), x[(k+1)\Delta t], x[(k+2)\Delta t], \dots, x[(k+n)\Delta t]$. Таким образом, если R_S рассчитывается по n элементам, а максимальная задержка, для которой рассчитывается $R_S(k)$, равна k_m , то всего в расчете, как и для НКФ, используется $N=n+k_m$ отсчетов.

При больших задержках $k\Delta t$, т. е. когда исследуемые выборки уже независимы, функции $R_S(k)$ и $\rho(k)$ имеют близкие свойства.

Это нормальные последовательности, у которых $m[R_S] = m[\rho] = 0$, $D[\rho] \approx \frac{1}{n}$ и $D[R_S] \approx \frac{1}{n-1}$ [24]. Дисперсия коэффициента R_S несколько больше, так как зна-

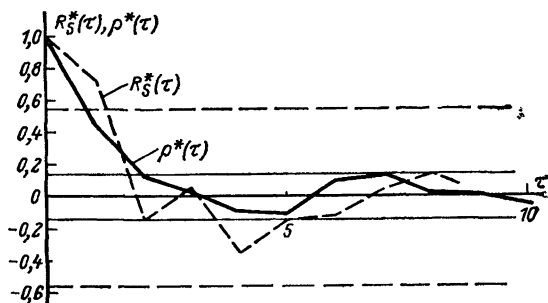


Рис. 5-5. Сопоставление НКФ, полученной на ЦВМ по 400 отсчетам, с РКФ, рассчитанной по 15 отсчетам вручную.

чения отсчетов заменены рангами и часть информации при этом потеряна. При малых задержках $k\Delta t$, когда корреляция между отсчетами существует, для нормальных процессов $x(t)$ имеется простая связь между коэффициентами корреляции [4]:

$$\rho = 2 \sin \left(\frac{\pi}{6} R_S \right). \quad (5-19)$$

При $n > 10$ оценка коэффициента ρ может строиться для нормальных процессов по оценке коэффициента ранговой корреляции Спирмена. Пересчет по формуле (5-19) несложен, но и он может не производиться, так как значения величин ρ и R_S отличаются не более чем на 5%.

Для процессов, отличных от нормального, формул перехода от R_S к ρ нет. Но тем не менее и в этом слу-

чае РКФ отражает степень статистической связи между отсчетами процесса и может быть полезна для оценивания интервала корреляции. В качестве примера на рис. 5-5 приведена НКФ осциллограммы фединга на участке тропосферной линии связи, рассчитанная на ЦВМ по объему выборки, равному 400 отсчетам, и РКФ, вычисленная по $n=15$ вручную [17]. На графике видно, что РКФ сохраняет основные тенденции поведения НКФ, несмотря на релеевское, а не гауссово, распределение выборки. Оценка РКФ на рис. 5-5 имеет большую дисперсию, так как она вычислялась по выборке меньшего объема.

Таблица 5-2

Порядок расчета суммы квадратов разностей рангов

Номер отсчета	x_i	$x_{1i},$ $k=0$	$r_0(i)$	$x_{2i},$ $k=3$	$r_3(i)$	$d_i=r_{0i}-r_{3i}$	d_i^2
1	-8	-8	3	7	10	-7	49
2	-6	-6	4	11	13	-9	81
3	2	2	7	10	12	-5	25
4	7	7	10	-10	3	7	49
5	11	11	13	-12	2	11	121
6	10	10	12	-3	5	7	49
7	-10	-10	2	2	7	-5	25
8	-12	-12	1	5	9	-8	64
9	-3	-3	5	-1	6	-1	1
10	2	2	8	8	11	-3	9
11	5	5	9	14	14	-5	25
12	-1	-1	6	17	15	-9	81
13	8	8	11	-15	1	10	100
14	14	14	14	-5	4	10	100
15	17	17	15	-3	8	7	49
16	-15			$\sum_{i=1}^{15} d_i^2$			828
17	-5						
18	-3						

В заключение рассмотрим на примере процедуру вычисления коэффициента ранговой корреляции. Пусть имеются отсчеты x_i осциллограммы $x(t)$. Найдем коэффициент $R^*_s(3)$ при $n=15$. Первая выборка $\{x_{1i}\}$ формируется при сдвиге $k=0$, а вторая $\{x_{2i}\}$ — при $k=3$. Обозначим ранги первой выборки $r_0(i)$, а ранги второй выборки $r_3(i)$. Сформированные выборки, ранги и порядок расчета разностей, требуемых для формулы (5-18), приведены в табл. 5-2. В результате расчетов получаем $R_s(3) = -0,47$.

Среднее квадратическое отклонение $R_s(k)$ при больших значениях аргумента равно $\sigma[R_s] = 1/\sqrt{n-1} = 0,071$.

5.2. ОЦЕНКА ИНТЕРВАЛА КОРРЕЛЯЦИИ

Понятие интервала корреляции (ИК) случайного процесса $x(t)$ характеризует интервал времени, разделяющий отсчеты процесса x_i , которые можно считать некоррелированными. В литературе известно несколько определений ИК τ_k , основанных на использовании НКФ $\rho(\tau)$ [28, 32, 41]. Эти определения имеют теоретико-вероятностное происхождение и весьма полезны в теоретических исследованиях. Однако в экспериментальной практике имеют дело не с НКФ, а с их оценками $\rho^*(\tau)$, отличающимися по своим свойствам от $\rho(\tau)$. Для оценки ИК наиболее важное отличие заключается в том, что если для эргодических процессов $\rho(\tau) \rightarrow 0$ при $\tau \rightarrow \infty$, то для $\rho^*(\tau)$ это несправедливо.

Оценка $\rho^*(\tau)$ представляет собой нестационарный случайный процесс, который при малом $\rho(\tau)$ можно считать стационарным с распределением, близким к нормальному, со средним $m_p = 0$, дисперсией $D_p \approx 1/n$ и корреляционной функцией $R_p(t_2 - t_1)$, зависящей от $\rho(\tau)$ [13].

Указанное отличие порождает некорректности, если оценку τ_k^* строить согласно теоретическим определениям, применяя их по отношению к оценке $\rho^*(\tau)$ вместо $\rho(\tau)$.

Приведем два примера таких некорректностей. Если использовать определение вида $\tau_k^* = \int_0^{T_1} \rho^*(\tau) d\tau$ вместо интеграла с бесконечным верхним пределом к нормальному процессу с НКФ вида $\rho(\tau) = e^{-\alpha|\tau|}$, дисперсия $D[\tau_k^*]$ оценки τ_k^* равна [17]:

$$D[\tau_k^*] = \frac{4}{\alpha^2 T} (\operatorname{ch} \alpha T_1 - 1), \quad (5-20)$$

где T — длительность реализации.

Из формулы следует, что если увеличивать интервал интегрирования, то дисперсия оценки ИК бесконечно возрастает, что делает такую оценку бесполезной. Другой пример некорректности дает использование определения, согласно которому τ_k^* задается выражением

$$\rho^*(\tau_k^*) = \rho_0, \quad (5-21)$$

где уровень ρ_0 выбирается заранее. Часто уровень ρ_0 рекомендуют выбирать достаточно малым (порядка 5—

10%). Противоречие такой рекомендации с практикой оценивания ИК состоит в том, что если выбирать ρ_0 без учета объема выборки, по которому оценивается НКФ, то очень большой разброс оценки $\rho^*(\tau)$ при больших τ может не позволить статистически достоверно оценить τ^*_k [16]. Уровень ρ_0 должен задаваться, исходя из конкретной задачи. Например, в технике разнесенного приема в системе связи практически некоррелированными считают сигналы, коэффициент корреляции которых не превосходит 0,36 [14], а при формировании нормальных случайных последовательностей в технике моделирования на ЦВМ значительным может считаться коэффициент корреляции порядка 0,01.

Чтобы избежать противоречий, связанных с применением упомянутых определений ИК, целесообразно оценку τ^*_k строить на основе теории проверки гипотез. Под оценкой τ^*_k будем понимать наименьшее значение поддержки, начиная с которого с заданной доверительной вероятностью P_d принимается нулевая гипотеза о независимости отсчетов, сдвинутых на $\tau \geq \tau_k$. Для проверки независимости должен быть выбран критерий $\psi(\tau)$. Принятие нулевой гипотезы определяется выполнением неравенства $|\psi(\tau)| \leq \psi_0$, где ψ_0 — уровень, выбираемый с учетом конкретной задачи. Такой подход напоминает оценку τ^*_k по (5-21), но в отличие от него, во-первых, допускает применение других, например ранговых, критериев независимости. Во-вторых, такой подход вынуждает увязывать выбранный уровень ψ_0 (критическое значение критерия) с объемом выборки N и доверительной вероятностью P_d . Эта увязка может быть выполнена по графикам, показанным на рис. 5-6 для нормального процесса $x(t)$, если в качестве $\psi(\tau)$ взять НКФ $\rho(\tau)$ или РКФ $R_S(\tau)$ (их критические значения очень близки и для экспресс-анализа могут считаться совпадающими). Выбрав по оси ординат критический уровень ψ_0 и кри-

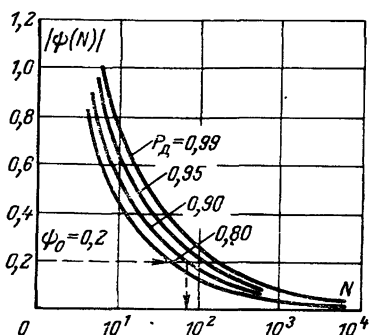


Рис. 5-6. Зависимость критических значений $\psi(N)$ от объема выборки для выборочного коэффициента корреляции и коэффициента ранговой корреляции Спирмена.

вую, соответствующую желаемой вероятности P_d , можно получить необходимый объем выборки N . Связь объема выборки N с длительностью реализации производится, как и при планировании расчета оценки НКФ, по верхней частоте процесса [13].

Использование непараметрических критериев независимости, таких как, например, РКФ $R_S(\tau)$, позволяет найти интервал «независимости», а не только ИК. Это целесообразно для распределений $x(t)$, отличных от нормального.

В качестве оценки ИК может быть предложена оценка среднего значения интервала между пересечениями случайным процессом нулевого уровня с заданным знаком производной [41]. Известно [28], что удвоенное среднее значение интервала между нулями стационарного нормального случайного процесса составляет:

$$\tau_0 = \frac{2\pi}{\omega_1} = \frac{2\pi}{\sqrt{-\rho''(0)}}. \quad (5-22)$$

Если принять это значение τ_0 в качестве величины интервала корреляции и найти значение $\rho(\tau_0)$, то можно убедиться в том, что значения $\dot{X}(t)$ и $\dot{X}(t+\tau)$ будут достаточно слабо коррелированы. Сравнительные значения $\tau_{м.к}^*$ и τ_0 для различных видов нормированных корреляционных функций $\rho(\tau)$ и значения $\rho(\tau_0)$ приведены в табл. 5-3, из которой следует, что с учетом сделанных примечаний для наиболее часто встречающихся в практических задачах видов корреляционных функций значения $\rho(\tau_0) \leq 0,05$. Во всех рассмотренных примерах (кроме шестого) при выборе длины реализации T , кратной τ_0 , возникает некоторая избыточность информации. Эта избыточность по существу и будет «платой» за априорную неинформированность исследователя [т. е. за незнание вида $\rho(\tau)$ и, следовательно, $\tau_{м.к}$]. Практически получение оценки τ_0^* сводится к определению удвоенного среднего интервала между нулями реализации процесса (Δt_i):

$$\tau_0 = 2 \frac{\sum_{i=1}^N \Delta t_i}{N}. \quad (5-23)$$

* Здесь по аналогии с [32] принято, что $\rho(\tau_{м.к}) \leq 0,05$.

Таблица 5-3

Сравнение максимального интервала корреляции $\tau_{\text{МК}}$ и удвоенного среднего значения интервала между нулями случайного процесса

$\rho(\tau)$	$\tau_{\text{МК}}$	$\tau_0 = \frac{2\pi}{\omega_0}$	$\rho(\tau_0)$	Примечание
$e^{-\gamma^2 \tau^2}$	$\frac{\sqrt{3}}{\gamma}$	$\frac{2\pi}{\gamma \sqrt{2}}$	$\ll 0,05$	—
$e^{-\gamma^2 \tau^2} \cos \omega_0 \tau$	$\frac{\sqrt{3}}{\gamma}$	$\frac{2\pi}{\sqrt{2\gamma^2 + \omega_0^2}}$	$< 0,05$	при $\frac{\omega_0}{\gamma} < 3,35$
$(1 + \alpha \tau) e^{-\alpha \tau}$	$\frac{4,5}{\alpha}$	$\frac{2\pi}{\alpha}$	0,014	—
$e^{-\alpha \tau} \left(\cos \omega_0 \tau + \frac{\alpha}{\omega_0} \sin \omega_0 \tau \right)$	$\frac{3}{\alpha}$	$\frac{2\pi}{\sqrt{\alpha^2 + \omega_0^2}}$	$< 0,05$	при $\frac{\omega_0}{\alpha} < 2,45$
$e^{-\alpha \tau} \left(1 + \alpha \tau + \frac{1}{3} \alpha^2 \tau^2 \right)$	$\frac{5,8}{\alpha}$	$\frac{2\pi \sqrt{3}}{\alpha}$	$\ll 0,05$	—
$\frac{\sin \omega_0 \tau}{\omega_0 \tau}$	$\frac{20}{\omega_0}$	$\frac{2\pi}{\omega_0}$	0	—

На рис. 5-7 показаны примеры реализаций случайных процессов с двумя видами корреляционных функций для реализации a $\rho_1(\tau) = e^{-0,7\tau} \cos 5,6\pi\tau$, для реализации b $\rho_1(\tau) = \frac{\sin 10\pi\tau}{10\pi\tau}$. На рис. 5-7, b

и $г$ показаны оценки корреляционных функций, полученные на ЦВМ для объема выборки $N=2000$ и путем оперативного анализа оценки τ_0^* . Точность получаемой оценки τ_0^* можно найти, определив $\sigma^*[\tau_0]$, как указано в гл. 4. Следует заметить, что, несмотря на то что плотность распределения случайной величины τ_0 не является нормальной, полученная оценка среднего значения $m[\tau_0^*]$ имеет уже при объеме выборки $n \geq 10$ распределение, близкое к нормальному [28]. Поэтому доверительный интервал определяется аналогично (3-11).

Принимая во внимание рассмотренные методы корреляционного анализа, можно рекомендовать их использование следующим образом:

оценка ИК через среднее значение интервала между нулями процесса является самой простой и оперативной

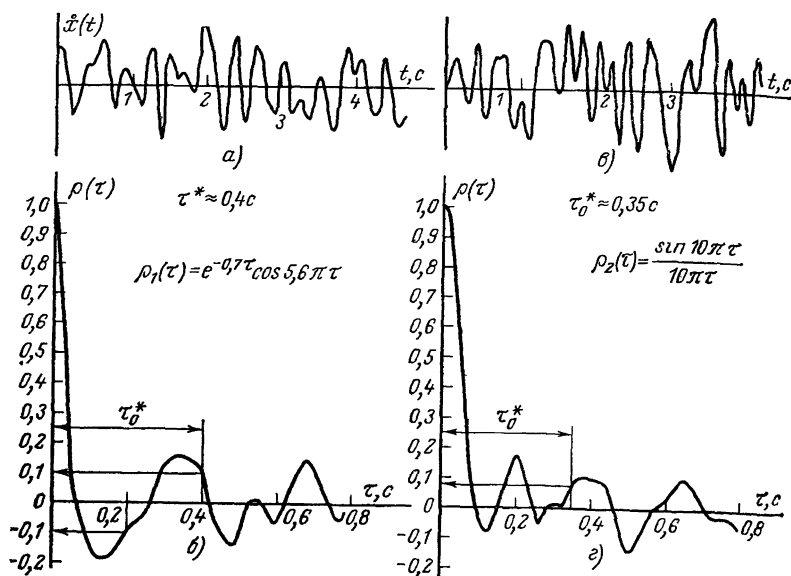


Рис. 5-7. Примеры определения интервала τ_0 по реализации нормального случайного процесса.

a — реализация случайного процесса с $\rho_1(\tau) = e^{-0,7\tau} \cos 5,6\pi\tau$; $б$ — оценка корреляционной функции $\rho_1^*(\tau)$; $в$ — реализация случайного процесса с $\rho_2(\tau) = \frac{\sin 10\pi\tau}{10\pi\tau}$; $г$ — оценка корреляционной функции $\rho_2^*(\tau)$.

при соответствующей модели процесса. Кроме самостоятельного значения, она может быть использована для планирования эксперимента по определению шага аргумента при оценивании КФ;

оценка НКФ по условному среднему целесообразна при наличии реализации процесса и соответствующей его модели. Это же относится и к оценке знаковой НКФ по условному среднему. Получение последней менее трудоемко, но связано с несколько большими погрешностями;

ранговая КФ легко определяется для стационарных процессов, отличных от нормального, являясь функцией, определяющей степень зависимости отсчетов, разделенных заданным интервалом времени (для таких процессов другие описанные методы непригодны);

ранговая КФ удобна в качестве функции связи и для нормальных процессов, особенно если последние заданы набором отсчетов;

при получении оценок НКФ интервал корреляции может быть определен непосредственно по кривой оценки НКФ.

5-3. ОПРЕДЕЛЕНИЕ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК СЛУЧАЙНЫХ ПРОЦЕССОВ

Определение спектральной плотности как преобразование Фурье от корреляционной функции. Если предварительно получена оценка НКФ по интервалу T_1 , то возможно применение преобразования Фурье от оценки корреляционной функции с учетом заранее выбранного «спектрального окна» $v(\tau)$ [32] и замены функции $\cos 2\pi f\tau$ на $\text{sign}[\cos(2\pi f\tau)]$.

Допуская при переходе относительную методическую погрешность, не превосходящую 30%, можно определять оценку нормированной спектральной плотности (СП) согласно выражению

$$\tilde{S}_x^*(f) \approx \frac{\pi}{2} \int_0^{T_1} v(\tau) \rho_x^*(\tau) \text{sign}[\cos(2\pi f\tau)] d\tau. \quad (5-24)$$

В этом случае первые два сомножителя в подынтегральном выражении (5-24) перемножаются аналитически, а последний учитывается присвоением знака указанному произведению. Интегрирование достаточно

просто осуществляется графически измерением соответствующих площадей под кривой подынтегральной функции для требуемых частот на интервале $[0, T_1]$.

Вид спектральных окон достаточно разнообразен, но в качестве одного из простейших может быть рекомендовано так называемое окно Барлетта $v(\tau) = |\tau|/T_1$.

Определение нормированной спектральной плотности с помощью метода знаковых функций. Другая возможность в получении оценки нормированной СП стационарного нормального процесса непосредственно по его реализации, представленной осциллограммой, основана на получении оценки вида [32]

$$\begin{aligned} \tilde{S}_{\text{зн},x}^*(f, T_0) = \frac{1}{T_0} \left\{ \left[\int_0^{T_0} \text{sign}[x(t)] \text{sign}[\cos 2\pi ft] dt \right]^2 + \right. \\ \left. + \left[\int_0^{T_0} \text{sign}[x(t)] \text{sign}[\sin 2\pi ft] dt \right]^2 \right\}. \end{aligned} \quad (5-25)$$

При получении этой оценки существует несколько особенностей. Реализация достаточно большой длительности T разбивается на достаточно большое число отрезков (так, что $T = NT_0$), для которых определяются оценки вида (5-25) по описанной ниже «технологии» а затем находят среднее арифметическое полученных кривых, чем в определенной степени преодолевается не состоятельность однократного оценивания, так как $\lim \tilde{S}_{\text{зн},x}(f, T_0) \neq \tilde{S}_{\text{зн},x}(f)$. Для получения оценок вида (5-25) реализацию центрированного процесса записывают в виде осциллограммы на соответствующей скорости, желательно с отметками времени. Общая длительность реализации выбирается равной $T = (8 \div 10) T_0 = (8 \div 10) \tau_0^*$. Об оценке τ_0^* говорилось выше. Затем заготавливается вспомогательная бумага (желательно полупрозрачная), на которой наносится (в том же масштабе времени, что и реализация) знаковая функция $\text{sign}[\sin 2\pi ft]$, представляющая собой симметричные относительно нулевой линии прямоугольные импульсы. На осциллограмме проводится вертикальная линия начала отсчета.

С этой и с нулевой линиями совмещается вспомогательная бумага, размещаемая вдоль осциллограммы. Далее операция интегрирования по (5-25) на интер-

вале T_0 заменяется суммированием числа совпадающих знаков отсчетов реализации $\dot{x}(t)$, попадающих на пересечения с отметками времени и функцией $\text{sign}[\sin 2\pi ft]$ и вычитанием числа несовпадающих отсчетов этих функций на интервале T_0 . Полученное число возводится в квадрат. Аналогично получают первое слагаемое с той лишь разницей, что вспомогательная бумага сдвигается

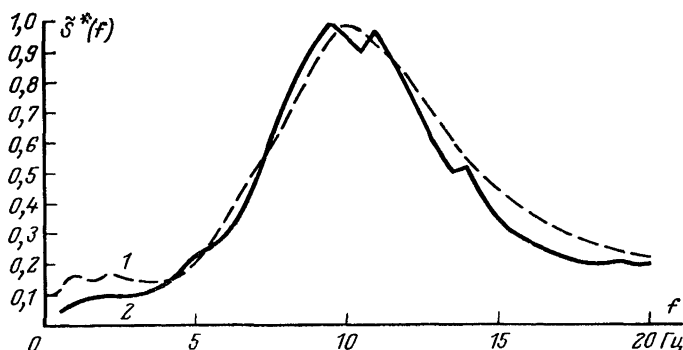


Рис. 5-8. Сравнение оценки спектральной плотности, полученной на ЦВМ через оценку корреляционной функции, определенной по множительному алгоритму, с оценкой спектральной плотности, полученной вручную знаковым методом.

1 — обработка на ЦВМ по преобразованию Фурье от КФ $N=4000$; 2 — обработка вручную по знаковому методу $N=2$, $n=100$.

на четверть периода относительно линии начала отсчета, чем обеспечивается получение функции $\text{sign}[\cos 2\pi ft]$. После суммирования квадратов слагаемых по (5-25) общий результат делится на общее число отсчетов процесса на интервале T_0 , определяемое как $n=T_0/t_0$ (t_0 — интервал между отметками времени). При ручной обработке величина $n=100$ оказывается достаточной, а процедура вполне приемлема по трудоемкости. Однако стремление к увеличению числа ординат оценки $\tilde{S}_{\text{зн}, x}(f)$ значительно увеличивает трудоемкость.

На рис. 5-8 показаны примеры сравнения оценки $\tilde{S}_{\text{зн}, x}(f)$ ($n=100$, $N=2$), полученной описанным методом, и оценки $S_x^*(f)$, вычисленной по традиционному алгоритму через КФ процесса на ЦВМ. Различия в величинах оценок в правой части от главного экстремума объ-

ясняется методическими погрешностями в области высоких частот, как отмечается в [32]. Описанный метод позволяет получать приемлемые по точности результаты, однако является достаточно трудоемким.

Определение нормированной спектральной плотности случайного процесса по распределению нулей. Для некоторого вида случайных процессов, представимых в виде

$$X(t) = A(t) \cos \varphi(t) = A(t) \cos 2\pi \int_{t_0}^t \omega(t) dt,$$

где $A(t)$, $\varphi(t)$, $\omega(t)$ — соответственно огибающая, фаза и частота случайного процесса, справедливо следующее соотношение [12] при независимости $A(t)$ и $\omega(t)$, связывающее $S_x(f)$ и плотность распределения $w_f(f)$ ее частотных составляющих:

$$S_x(f) = \sigma_x^2 w_f(f). \quad (5-26)$$

Функция $w_f(f)$ может быть получена пересчетом оценки плотности распределения $w_\tau(\tau)$ удвоенных интервалов $\tau = 2\tau_x$ между нулями центрированного процесса $\dot{X}(t)$. Пересчет легко осуществляется по формуле нелинейного преобразования ($f = 1/2\tau_x$), т. е.

$$w_f(f) = \frac{1}{f^2} w_\tau\left(\frac{1}{f}\right). \quad (5-27)$$

Таким образом, построив гистограмму $w_\tau^*(\tau)$ и преобразовав ее к виду $w_f^*(f)$, по (5-27) можно с точностью до масштабов получить оценку СП [22].

Чтобы избежать погрешностей, связанных с пересчетом масштабов, удобно операции перехода от $w_\tau(\tau)$ к $w_f(f)$ делать с нормированными кривыми. При этом для плотности $w_\tau(\tau)$ справедливо условие

$$\int_0^\infty w_\tau(\tau) d\tau = 1,$$

а нормировка кривой $w_f(f)$ выполняется из условия

$$w_f(f)|_{\text{макс}} = 1.$$

Оценка дисперсии $(\sigma_x^*)^2$ для изучаемого процесса может быть определена описанными в предыдущих разделах способами.

Допущения, принятые при получении (5-26), могут быть распространены на стационарный эргодический центрированный нормальный процесс $\dot{X}(t)$, пропущенный через нелинейный преобразователь, выходной сигнал которого $\xi(t)$ соответствует знаковой функции от входного процесса [28].

Проблема исследования погрешностей описанного метода еще не решена. Однако некоторые результаты его практического применения получены и иллюстрируются рис. 5-9, где сплошными линиями показаны теоретические спектральные плотности $S_x(f)$ двух процессов

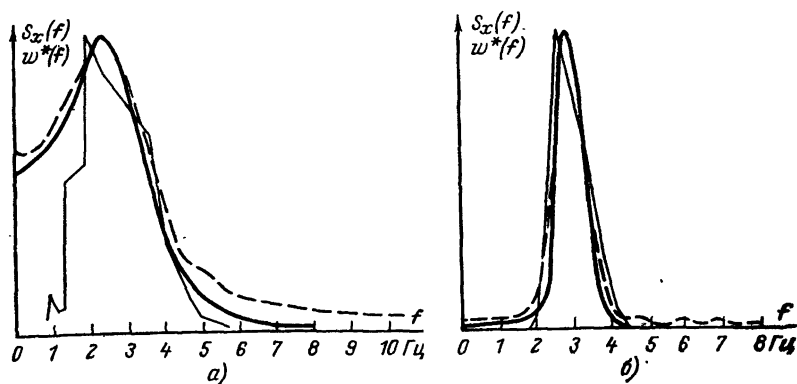


Рис. 5-9. Сравнение теоретических спектральных плотностей и построенных через гистограммы распределения нулей.

a — $f_0/a=5$; $б$ — $f_0/a=30$.

— теоретическая спектральная плотность; — — — — — оценка спектральной плотности, полученная через знаковую корреляционную функцию; — оценка спектральной плотности, полученная через распределение нулей процесса.

с экспоненциально-косинусными корреляционными функциями $\rho_x(\tau) = e^{-\alpha\tau} \cos 2\pi f_0 \tau$. Пунктирными линиями на этом рисунке показана оценка СП $S(t)$, нормированная по максимуму.

Процесс на левом рисунке является менее узкополосным, чем второй процесс. Для первого процесса $f_0/a=5$, а для второго $f_0/a=30$. Анализ погрешностей машинного эксперимента позволяет рекомендовать для получения оценки СП через распределения нулей процесса соотношение $f_0/a \geq 15$.

АНАЛИЗ СТАЦИОНАРНОСТИ СЛУЧАЙНЫХ ПРОЦЕССОВ**6-1. НЕОБХОДИМОСТЬ КОНТРОЛЯ СТАЦИОНАРНОСТИ**

В практике экспресс-анализа случайных процессов задача проверки стационарности изучаемого процесса занимает важное место. Суждение о стационарности позволяет исследователю уточнить модель процесса или выявить причины нарушения стабильности хода эксперимента.

Условно нестационарность можно рассматривать как «быструю» и «медленную». Если нестационарность заметна на глаз по осциллограмме процесса, то она может быть отнесена к категории «быстрой» (рис. 1-1, *е, ж, и, к*). В некоторых случаях изменения характеристик по конкретной записи процесса заметить не удастся. Чтобы выявить такие формы нестационарности в экспресс-анализе, могут применяться методы математической статистики, относящиеся к проверке статистических гипотез [2, 4, 8, 24]. Эти методы принципиально не отличаются от методов проверки согласия распределений (§ 2-4). Однако при проверке стационарности совокупность этих методов расширяется в двух направлениях. Во-первых, формулировки нулевой гипотезы становятся разнообразнее, при этом может проверяться равенство какого-либо параметра конкретному значению или равенство параметров в разных выборках. Может, например, предполагаться идентичность распределений без конкретизации их параметров и даже вида распределений и многое другое. Во-вторых, при различных исходных сведениях о проверяемых процессах используются различные статистические критерии.

Методика проверки стационарности, как и проверки согласия распределений, состоит в вычислении статистики критерия и сравнении ее с критическим значением для нулевой гипотезы. Для правильной трактовки результатов проверки стационарности, как и статистических гипотез вообще, надо помнить, что термины «отвергнуть» или «принять» гипотезу условны. В действительности из-за стохастичности исследуемых процессов всегда имеется вероятность ошибиться. С вероятностью $\alpha = 1 - P_d$ можно отвергнуть справедливую гипотезу (ошибка первого рода) или с вероятностью β принять гипотезу, когда верна альтернатива (ошибка второго

рода). Из трех величин α , β и n (объем одной из выборок) свободно выбранными могут быть только две. При фиксированном n , улучшая критерий за счет увеличения P_d , мы неизбежно его ухудшаем за счет роста вероятности β . Если увеличивать n при фиксированной величине P_d , то с вероятностью, стремящейся к единице, можно отвергнуть любую гипотезу. Оценить вероятность β удастся не всегда. Поэтому практически имеет смысл руководствоваться следующим правилом. Если в данной задаче опаснее отвергнуть нулевую гипотезу, когда она верна, то вероятность P_d следует выбирать равной 0,9 и выше. Если же важнее не принимать неверную альтернативу, то надо задавать P_d в диапазоне 0,7—0,9. При этом в экспресс-анализе допустимо вероятность β не контролировать.

6-2. КОНТРОЛЬ ПОСТОЯНСТВА СРЕДНЕГО ЗНАЧЕНИЯ

Методы сравнения средних значений, или критерии сдвига, рассмотрим, начиная с предположения о нормальности исследуемого процесса.

t -критерий Стьюдента. Схема применения критерия основана на оценке среднего значения m^*_1 и m^*_2 двух сопоставляемых участков реализации и формировании меры их различия $\Delta_m = f(m^*_1, m^*_2)$. В частном случае исследователя может интересовать отличие оценки среднего от некоторого значения m_0 . Тогда вместо оценки m^*_2 используется величина m_0 . Нулевая гипотеза состоит в том, что $m^*_1 = m^*_2$ или $m^*_1 = m_0$, а альтернативы могут быть разные: утверждение, что $m^*_1 \neq m^*_2$ ($m^*_1 \neq m_0$), или неравенство $m^*_1 \geq m^*_2$ (дрейф среднего «вверх» или «вниз»). Для нормального процесса распределение величины Δ_m известно и поэтому может быть задано критическое значение $Z(v, P')$, определяющее допустимое отклонение средних. Критическое значение выбирается либо согласно t -распределению Стьюдента с v степенями свободы $Z(v, P') = Z_t(v, P')$, либо согласно нормальному закону $Z(v, P') = Z_\Phi(P')$. Вид распределения и параметр P' , зависящий от доверительной вероятности P_d , определяются формулировкой нулевой гипотезы и альтернативы.

Перед сравнением с $Z(v, P')$ величина Δ_m нормируется параметром σ , характеризующим разброс и зависящим от объема выборок и оценок σ^*_{x1} и σ^*_{x2} или па-

параметров σ_{x1} и σ_{x2} , если они известны. Таким образом, проверка гипотезы о среднем состоит в проверке неравенства

$$\frac{\Delta_m}{\sigma} > Z(v, P'). \quad (6-1)$$

Возможные практические ситуации и все необходимые данные для расчета отражены в табл. 6-1. Критические значения $Z_\Phi(P')$ могут быть взяты из табл. П1, а $Z_t(v, P')$ определены по графикам на рис. 6-1. В строках с 1-й по 6-ю в табл. 6-1 N означает число отсчетов в изучаемой выборке, а в строках с 7-й по 12-ю N

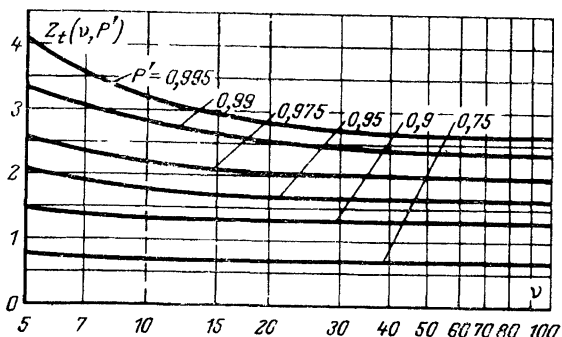


Рис. 6-1. Критические значения t -критерия.

относится к объему каждой из выборок. В экспресс-анализе, как правило, нет препятствий к использованию равных объемов выборок. Сведения по критериям для $N_1 \neq N_2$ можно найти в [42, 50]. Строки 8 и 11 таблицы содержат формулы для случая, когда процесс стационарен по дисперсии. Строки 9 и 12 — для случая априорной нестационарности дисперсии процесса. Для проверки альтернативы $m^*_1 < m^*_2$ можно воспользоваться строками с 9-й по 12-ю, переставив индексы у параметров.

Превышение нормированным расхождением Δ_m/σ критического уровня $Z(v, P')$ означает, что отклонения средних больше, чем допустимо при случайном разбросе. При этом считается, что факт нестационарности по среднему статистически установлен.

t -критерий обладает рядом практически удобных особенностей. Во-первых, распределение Стьюдента от-

Справочная таблица для применения t -критерия

Номер варианта	Альтерна- тива	Известно ли m_0	Известно ли σ_{x1} и σ_{x2}	Δm	$Z(\nu, P')$	P'	ν	σ
1	$m^*_1 \neq m_0$	Да	Да	$ m^*_1 - m_0 $	$Z_\Phi(P')$	$(1 + P_D)/2$	—	σ_{x1}/\sqrt{N}
2	$m^*_1 \neq m_0$	Да	Нет	$ m^*_1 - m_0 $	$Z_t(\nu, P')$	$(1 + P_D)/2$	$N-1$	σ^*_{x1}/\sqrt{N}
3	$m^*_1 > m_0$	Да	Да	$m^*_1 - m_0$	$Z_\Phi(P')$	P_D	—	σ_{x1}/\sqrt{N}
4	$m^*_1 > m_0$	Да	Нет	$m^*_1 - m_0$	$Z_t(\nu, P')$	P_D	$N-1$	σ^*_{x1}/\sqrt{N}
5	$m^*_1 < m_0$	Да	Да	$m_0 - m^*_1$	$Z_\Phi(P')$	P_D	—	σ_{x1}/\sqrt{N}
6	$m^*_1 < m_0$	Да	Нет	$m_0 - m^*_1$	$Z_t(\nu, P')$	P_D	$N-1$	σ^*_{x1}/\sqrt{N}
7	$m^*_1 \neq m^*_2$	Нет	Да	$ m^*_1 - m^*_2 $	$Z_\Phi(P') \sqrt{\frac{N}{2}}$	$(1 + F_D)/2$	—	$((\sigma^2_{x1} + \sigma^2_{x2})/N)^{1/2}$
8	$m^*_1 \neq m^*_2$	Нет	Нет, но $\sigma_{x1} \approx \sigma_{x2}$	$ m^*_1 - m^*_2 $	$Z_t(\nu, P')$	$(1 + P_D)/2$	$2(N-1)$	$((\sigma^*_{x1} + \sigma^*_{x2})/N)^{1/2}$

Номер варианта	Альтерна- тива	Известно ли m_0	Известно ли σ_{x1} и σ_{x2}	Δm	$Z(\nu, P')$	P'	ν	σ
9	$m_1^* \neq m_2^*$	Нет	Нет, но $\sigma_{x1} \neq \sigma_{x2}$	$ m_1^* - m_2^* $	$Z_t(\nu, P')$	$(1 + P_D)/2$	$(N+1) \times$ $\times \left(1 + \frac{2k}{k^2+1}\right) - 2$ $k = \frac{\sigma_{x1}^{*2}}{\sigma_{x2}^{*2}}$	$((\sigma_{x1}^{*2} + \sigma_{x2}^{*2})/N)^{1/2}$
10	$m_1^* > m_2^*$	Нет	Да	$m_1^* - m_2^*$	$Z_\Phi(P')$	P'_D	—	$((\sigma_{x1}^{*2} + \sigma_{x2}^{*2})/N)^{1/2}$
11	$m_1^* > m_2^*$	Нет	Нет, но $\sigma_{x1} \approx \sigma_{x2}$	$m_1^* - m_2^*$	$Z_t(\nu, P')$	P_D	$2(N-1)$	$((\sigma_{x1}^{*2} + \sigma_{x2}^{*2})/N)^{1/2}$
12	$m_1^* > m_2^*$	Нет	Нет, но $\sigma_{x1} \neq \sigma_{x2}$	$m_1^* - m_2^*$	$Z_t(\nu, P')$	P_D	$(N+1) \times$ $\times \left(1 + \frac{2k}{k^2+1}\right) - 2$ $k = \frac{\sigma_{x1}^{*2}}{\sigma_{x2}^{*2}}$	$((\sigma_{x1}^{*2} + \sigma_{x2}^{*2})/N)^{1/2}$

носителем быстро нормализуется. Поэтому для $N > 30$ можно принять $Z_t(v, P') \approx Z_\Phi(P')$ с погрешностью не более 10%. Кроме того, отклонение от нормальности изучаемого процесса не приводит к большим ошибкам. Недостатком t -критерия является чувствительность к зависимости между отсчетами.

Пример. Рассмотрим результаты контроля напряжения источника питания с номинальным напряжением 20 В. При измерении цифровым вольтметром получено 10 независимых отсчетов в вольтах: 19,72; 20,10; 19,82; 19,88; 19,66; 20,04; 19,92; 19,96; 20,04; 19,66. Нулевая гипотеза состоит в том, что дрейфа напряжения нет, а разброс определяется лишь случайными причинами. Теоретическое среднее $m_0 = 20$ В в данном случае известно, а дисперсия неизвестна. Альтернатива соответствует неравенству $m^*_{-1} \neq m_0$. Эти сведения позволяют выбрать для расчета строку 2 в табл. 6-1. Произведем необходимый расчет: $m^*_1 = 19,88$ В,

$$\Delta_m = |m^*_1 - m_0| = 0,12, \quad \sigma^{*2}_{x1} = \frac{1}{N-1} \sum_{i=1}^N (x_i - m^*_1)^2 =$$

$$= 0,026, \quad \sigma = \sqrt{\sigma^{*2}_{x1}/N} = 0,051.$$

Статистика критерия равна $\Delta_m/\sigma = 2,36$. Проведя на рис. 6-1 вертикальную линию от $N=10$ до пересечения с $Z_t = 2,36$, можно сделать вывод, что последовательность отсчетов стационарна при $P' \leq 0,975$. Эта величина соответствует доверительной вероятности $P_d = 2P' - 1 = 0,95$. Для более высоких значений P_d источник питания пришлось бы считать нестабильным.

Критерий Вилкоксона. Предыдущий критерий основывался на предположении о нормальности. Часто приходится иметь дело с другими распределениями, а иногда распределение обрабатываемого процесса просто неизвестно. В этих случаях вместо критериев, основанных на изучении параметров известного распределения, могут применяться критерии, использование которых не требует знания распределений — непараметрические критерии. К этому классу критериев относятся критерий Вилкоксона и его модификация — критерий Манна — Уитни. Нулевая гипотеза в критерии Вилкоксона состоит в предположении, что две сравниваемые выборки характеризуются одним и тем же распределением, причем безразлично каким. Если отсчеты каждой из выборок относятся к разным временным интервалам одного процесса, то нулевая гипотеза означает предположение стационарности. Известно, что критерий Вилкоксона почти не уступает t -критерию для нормальных распределений.

Вероятности ошибок при других распределениях могут для критерия Вилкоксона оказаться значительно меньше [4].

Статистика критерия Вилкоксона W представляет собой сумму рангов r_i , которые имеет одна из выборок в общем вариационном ряду (§ 2-4), построенном для обеих сравниваемых выборок. Например, если вариационный ряд для выборок X и Y имеет вид последовательности с рангами r_i , равными номерам позиций в общем вариационном ряду

$$\begin{array}{cccccccccccccccc} x_1 & x_2 & y_1 & y_2 & y_3 & x_3 & y_5 & y_6 & \dots & \dots & \dots & y_n & x_n \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \dots & \dots & \dots & N-1 & N, \end{array}$$

где $N=2n$, n — объем каждой из выборок (предполагая, что объемы выборок одинаковы), то статистика W равна сумме рангов всех иксов:

$$W = \sum_{i=1}^N r_i = 1 + 2 + 6 + \dots + N. \quad (6-2)$$

Критерий Вилкоксона справедлив для непрерывных распределений, где совпадений отсчетов быть не может. Но из-за приближенного характера регистрации процессов и считывания данных возможны совпадения значений x_i и y_i . Если совпадают элементы разных выборок x_i и y_i , то им приписывается одинаковый ранг, равный среднему арифметическому рангов, которые они бы имели, занимая те же позиции по отношению к другим отсчетам, но несколько различаясь между собой.

Например, участок вариационного ряда имеет вид:

Отсчеты	x_6	y_8	x_{10}	x_{11}	y_3	$y_9 \dots$
Значения	6,3	6,7	6,8	9,1	9,1	9,2...
Ранги	7	8	9	10,5	10,5	12...

Совпавшим значениям отсчетов x_{11} и y_3 , которые имели бы ранги 10 и 11, приписывается средний ранг 10,5, равный для обоих отсчетов. Если совпали элементы одной из выборок x_i и x_j , то им можно присвоить ранги, например, по порядку их появления (при вычислении W ранги окажутся теми же слагаемыми той же суммы).

Распределение статистики W симметрично относительно среднего $m_W = n(2n+1)/2$. Гипотеза стационарности принимается, если относительное отклонение $\varepsilon_W = |W - m_W|/m_W$ меньше критического значения $Z_W(n, P_d)$, представленного на рис. 6-2.

Критерий Вилкоксона удобен для применения, начиная с $n \geq 5$, не требует знания распределения процесса

$x(t)$, а расчеты облегчаются тем, что в них используются целые числа (ранги). По критерию Вилкоксона может быть выявлена не только нестационарность по среднему, но и по форме распределения.

Пример. Рассмотрим два эксперимента, в которых отсчеты процесса имитируются с помощью таблицы равномерно распределенных случайных чисел [49]. В эксперименте № 1 сопоставляются выборки X_1 и X_2 объемом $n=8$, в эксперименте № 2 — выборка X_1 и выборка X_3 , каждый элемент которой после извлечения из таблицы искус-

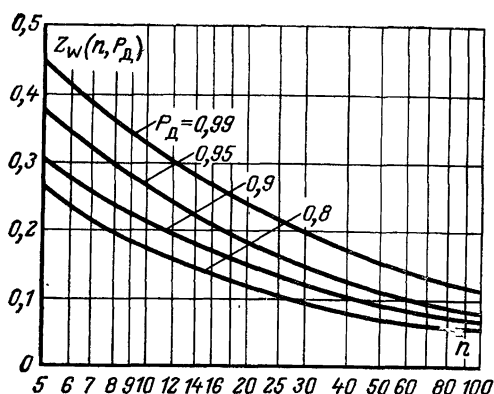


Рис. 6-2. Критические значения критерия Вилкоксона.

ственно увеличен на 0,12. Таким образом, выборка X_3 имеет известное смещение по среднему. Для каждой пары выборок образован общий вариационный ряд, проставлены ранги $r^{(j)}_i$ для i -й выборки в паре с j -й, вычислены статистики W и относительные величины e_w . Результаты расчетов сведены в табл. 6-2, где значения отсчетов умножены на 100. В эксперименте № 2 имеются совпадения.

Нулевая гипотеза правильно принимается в первом случае. Результат эксперимента № 2 при сопоставлении с графиками на рис. 6-2 показывает, что гипотеза может быть отвергнута лишь при доверительной вероятности, несколько меньше чем 0,8, а при больших значениях P_d нулевую гипотезу пришлось бы ошибочно принять. Поэтому эксперимент № 2 оказывается хорошей иллюстрацией того, как малый объем выборки может привести к пониженной чувствительности критерия и к ошибочному решению.

Непараметрический критерий сдвига [37]. Пусть имеется по N независимых отсчетов с первого и второго участков записи x_1, x_2, \dots, x_N и y_1, y_2, \dots, y_N . В последовательности x_i ищется наибольшее значение y_{\max} и подсчитывается число s отсчетов x_i , превышающих y_{\max} . Если это число s превышает критическое значение

Экспериментальные данные для примера применения критерия Вилкоксона

Номер отсчета	Эксперимент № 1				Эксперимент № 2			
	$r_1^{(2)}$	x_1	x_2	$r_2^{(1)}$	$r_1^{(3)}$	x_1	x_2	$r_3^{(1)}$
1	5	22	76	14	4	22	60	11
2	11	51	65	13	9	51	92	13,5
3	3	14	20	4	1	14	66	12
4	7	37	09	3	5	37	20	2
5	16	92	80	15	13,5	92	112	16
6	12	58	34	6	10	58	46	8
7	9	45	46	10	7	45	100	15
8	8	43	02	1	6	43	21	3
	71	—	—	65	55,5	—	—	80,5
ϵ_W	0,044				0,184			

Z_s для выбранной доверительной вероятности P_d , то делается вывод о том, что в реализации присутствует сдвиг среднего вверх. Если этот вид нестационарности не обнаружен, следует проверить число значений y_i , превышающих x_{\max} , что характеризует сдвиг среднего вниз. Критические значения Z_s равны [38]: для $N < 15$ при $P_d = 0,95$ $Z_s = 4$, при $P_d = 0,99$ $Z_s = 6$, для $N > 15$ при $P_d = 0,95$ $Z_s = 5$, при $P_d = 0,99$ $Z_s = 7$.

Критерий серий Вальда — Вольфовица [50]. Согласно этому критерию последовательность отсчетов x_i заменяется последовательностью знаков «+» или «—» в зависимости от выполнения неравенства $x_i > Me^*_x$ или $x_i < Me^*_x$, где Me^*_x — оценка медианы. Следующие подряд одинаковые знаки образуют группу, называемую серией. Количество серий q — величина случайная. При справедливости нулевой гипотезы, состоящей в предположении стационарности, количество серий группируется около своего среднего значения m_q :

$$m_q = \frac{2n_1 n_2}{n_1 + n_2} + 1, \quad (6-3)$$

где n_1 — число плюсов; n_2 — число минусов; $n_1 + n_2 = N$ — полное число отсчетов с отклонением σ_q :

$$\sigma_q = \frac{2n_1 n_2 [2n_1 n_2 - (n_1 + n_2)]}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}. \quad (6-4)$$

Если число серий q выйдет за границы критических значений $Z_q(P_d)$ для выбранной вероятности P_d , то нулевая гипотеза отвергается. Критические значения для больших объемов N определяются нормальным распределением величины q , которая может быть представлена в нормированном виде: $q_{\text{норм}} = (|q - m_q| - 0,5) / \sigma_q$.

Более удобно критические значения считывать с графиков, приведенных на рис. 6-3 (даже при относительно больших N). На графике даны абсолютные отклонения $Z_q(N, P_d) - N$, причем доверительная вероятность P_d определяется разностью верхней P_v и нижней P_n вероятностей, отмеченных на графике.

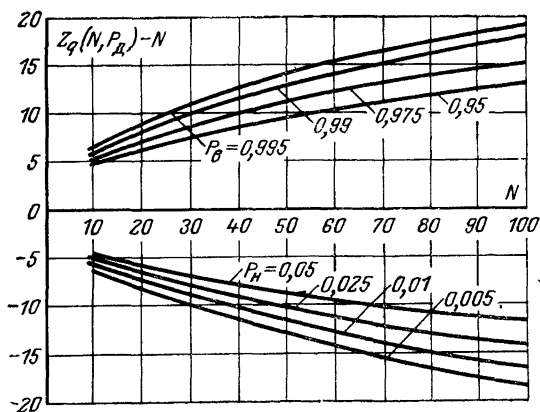


Рис. 6-3. Критические значения критерия серий.

Критерий можно использовать также для сравнения средних двух выборок. Если объемы выборок равны и в сумме составляют величину N , то величины $Z_q(N, P_d)$ могут быть также получены с помощью графиков рис. 6-3. При наличии многих выборок допустимо сравнивать последовательность различных статистических параметров, вычисленных по каждой выборке, например m^*_{xi} , σ^*_{xi} , p^*_i и т. п. Критерий серии прост для расчетов, но не относится к числу очень эффективных.

Пример. Данные, отраженные в таблице 6-3 [50], представляют собой усредненные за каждые из десяти интервалов времени показания расходомера. Проверяется гипотеза стационарности по среднему расходу продукта при $P_d = 0,90$. Медиана средних равна $(38,7 + 38,9) / 2 = 38,8$. В третьем столбце таблицы 6-3 отмечены знаки «+» при превышении медианы и «—» — в противном случае.

Число серий q равно 5 при $N=10$. По рис. 6-3 легко убедиться, что гипотеза стационарности должна быть принята.

Т а б л и ц а 6-3

**Экспериментальные данные для примеров
по применению критериев серий и инверсий**

Номер периода	Среднее за период	Превышение медианы	Число инверсий
1	36,5	—	1
2	43,0	+	7
3	44,5	+	7
4	36,9	+	4
5	38,1	—	2
6	32,6	—	0
7	38,7	—	1
8	41,7	+	2
9	41,1	+	1
10	36,8	—	0
Всего		5 серий	25 инверсий

Критерий инверсий [50]. Критерий строится на понятии инверсии, которая заключается в том, что в последовательности чисел за большим следует меньшее. Если отсчеты отражают стационарный процесс, то число инверсий I независимо от распределения $W(x)$ обладает известным распределением. Так, величина γ , определяемая формулой

$$\gamma = 1 - \frac{4I}{N(N-1)}, \quad (6-5)$$

имеет среднее значение, равное нулю, и дисперсию, равную [24]:

$$D_{\gamma} = \frac{2(2N-5)}{9N(N-1)}. \quad (6-6)$$

Величина γ , уже начиная с $N \geq 10$, имеет распределение, близкое к нормальному. Поэтому для проверки нулевой гипотезы, состоящей в предположении о том, что дрейфа среднего значения нет, можно использовать неравенство

$$|\gamma / \sqrt{D_{\gamma}}| < Z_{\Phi} \left(\frac{1 + P_{\alpha}}{2} \right), \quad (6-7)$$

где критическое значение нормального распределения выбирается из табл. П1.

Критерий инверсий является одним из наиболее эффективных непараметрических критериев, применим как к последовательности отсчетов процесса, так и к последовательностям его характеристик, но требует несколько большего объема вычислений, чем критерий серий.

Пример. В четвертой строке таблицы 6-3 указаны количества инверсий для каждого среднего. Всего инверсий 25. Теоретическое значение дисперсии D_{γ} (6-6) равно 0,037, а статистика критерия γ (6-5) равна $-0,11$.

Проверка неравенства $|-0,11/\sqrt{0,037}| = 0,575 < 1,96$, как и в прошлом случае, позволяет признать справедливость нулевой гипотезы.

6-3. КОНТРОЛЬ ПОСТОЯНСТВА ДИСПЕРСИИ

Для проверки стационарности по дисперсии применяются различные критерии масштаба в зависимости от априорной информации. Начнем рассмотрение критериев с допущения о нормальности процесса.

F-критерий Фишера—Снедекора. Если проверяется нулевая гипотеза о равенстве эмпирической дисперсии σ_1^{*2} , оцененной на некотором участке записи, предполагаемому значению σ_0^2 , то вычисляется статистика $F = \sigma_1^{*2} / \sigma_0^2$. Если сравниваются дисперсии двух участков реализации, то отношение F имеет вид $\sigma_1^{*2} / \sigma_2^{*2}$, где $\sigma_1^{*2} > \sigma_2^{*2}$. В первом слу-

Таблица 6-4

Справочная таблица по применению F-критерия

Номер варианта	Гипотеза	Альтернатива	Критерий
1	$F = 1$	$F \neq 1$	$Z_{\chi^2} [\nu, (1 + P_d)/2] / \nu > F > Z_{\chi^2} [\nu, (1 - P_d)/2] / \nu$
2	$F = 1$	$F > 1$	$Z_{\chi^2} [\nu, 1 - P_d] / \nu > F$
3	$F = 1$	$F < 1$	$F > Z_{\chi^2} [\nu, P_d] / \nu$
4	$F = k$	$F \neq k$	$kZ_F [\nu, (1 + P_d)/2] > F > k/Z_F [\nu, (1 + P_d)/2]$
5	$F = k$	$F > k$	$kZ_F [\nu, P] > F$

чае при нулевой гипотезе имеют дело с критическими значениями распределения χ^2 , которые могут быть взяты с графиков на рис. 2-9. Во втором случае используется распределение отношения выборочных дисперсий Фишера — Снедекора [8], критические значения которого указаны на рис. 6-4 (построен по таблицам из [2]). В обоих распределениях используются в качестве параметра число степеней свободы $\nu = N - 1$, где N — число независимых отсчетов на каждом из участков реализации. По-прежнему рассматривается случай равных объемов для обеих выборок. В качестве вероятности на

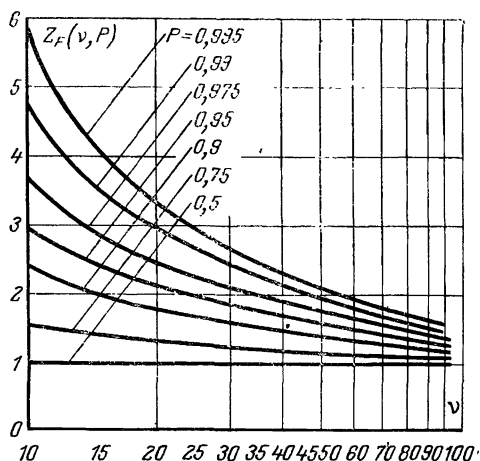


Рис. 6-4. Критические значения F -критерия.

рис. 6-4 используется некоторая функция P_d . Вид функции определяется формулировкой критерия. Возможные формулировки нулевой гипотезы и альтернативы, а также неравенств, при выполнении которых принимается нулевая гипотеза, даны в табл. 6-4. Если неравенства выполнены, то принимается гипотеза стационарности по дисперсии.

Пример. Пусть необходимо проконтролировать усилитель случайного сигнала по соотношению выходной и входной дисперсий. Номинальный коэффициент усиления $K_{yc} = 110$. Для контроля записаны процессы на выходе и входе усилителя, по которым снято по 75 отсчетов и рассчитаны величины $\sigma_{вых} = 18$ мВ и $\sigma_{вх} = 0,12$ мВ. Пусть допустимо считать, что в данном случае $K_{yc} = \sigma_{вых} / \sigma_{вх} = 150$. Предстоит проверить, значимо ли отклонение полученного значения K_{yc} от номинального или оно объясняется случайными факторами при

$P_d=0,95$. Согласно нулевой гипотезе среднее значение K_{yc} равно 110, а альтернатива состоит в том, что K_{yc} отличается от номинального (допускаются отклонения в любую сторону). Указанным условиям соответствует 4-я строка табл. 6-2.

Число степеней свободы равно $\nu = 75 - 1 = 74$, вероятность $(1 + P_d)/2 = (1 + 0,95)/2 = 0,975$. Критическое значение $Z_\Phi(74; 0,975) = 1,5$ (рис. 6-4). Статистика $F = \sigma_{\text{вых}}^2 / \sigma_{\text{вх}}^2 = 1,21 \times 10^4$. Граница в правом неравенстве равна $k/Z_\Phi(74; 0,975) = 110^2 / 1,5 = 0,806 \times 10^4$. Граница в левом неравенстве равна $kZ_\Phi(74; 0,975) = 110^2 \times 1,5 = 1,84 \times 10^4$. Левое неравенство не выполняется. Таким образом, коэффициент усиления значительно отклонился от номинального значения.

Критерий Кокрена. Этот критерий относится к многовыборочным критериям. Он применяется для выяснения вопроса о том, являются ли несколько оценок дисперсий σ_i^{*2} принадлежащими одному и тому же нормальному распределению. Если по нескольким участкам реализации вычислены оценки дисперсий σ_i^{*2} и вычисленная статистика G , равная:

$$G = \frac{\sigma_{\text{макс}}^{*2}}{K} \cdot \sum_{i=1}^K \sigma_i^{*2}, \quad (6-8)$$

превышает критическое значение $Z_G(\nu, K, P_d)$, где $\nu = N - 1$, N — объем каждой выборки, K — число участков (выборок), то предположение о стационарности следует отвергнуть. Зависимости критического значения

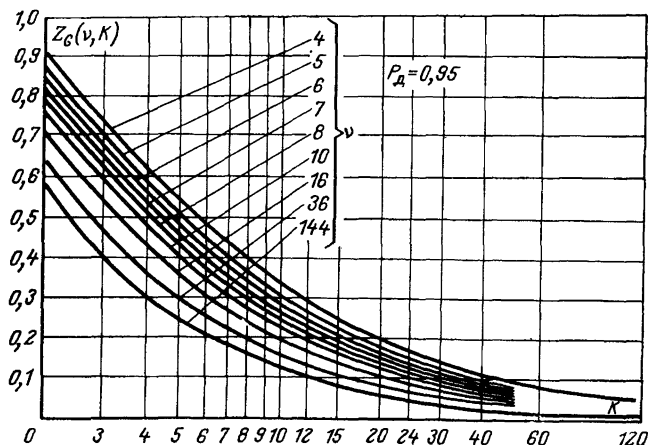


Рис. 6-5. Критические значения критерия Кокрена.

для распределения Кокрена от величин v и K при $P_d = 0,95$ построены на рис. 6-5 [2].

Пример. Рассчитана серия оценок дисперсии на 10 участках некоторого процесса: 2,7; 2,1; 3,8; 4,1; 2,0; 3,1; 3,5; 1,9; 2,8; 3,6. Каждая оценка вычислялась по 30 отсчетам. Таким образом, $v = N - 1 =$

$$= 29; \sum_{i=1}^{10} \sigma_i^2 = 29,6; \sigma_{\text{макс}}^2 = 4,1 \text{ и } G = 4,1/29,6 = 0,14. \text{ Согласно}$$

рис. 6-5, критическое значение $Z_G(29, 10, 0,95)$ равно 0,17. Следовательно, $Z_G > G$ и процесс можно считать стационарным по дисперсии.

Два приведенных критерия справедливы при предположении нормальности. В отличие от t -критерия для сравнения средних критерии для сравнения дисперсий очень чувствительны к нормальности исходных распределений, и их применение может приводить к серьезным ошибкам, если предположение о нормальности несправедливо.

Критерий Сиджела — Тьюки [50]. Этот критерий является непараметрическим ранговым критерием и может употребляться для двухвыборочного сравнения дисперсий. По существу процедура использования критерия состоит из двух частей. Первая ее часть — процедура Сиджела — позволяет свести критерий масштаба к критерию сдвига [7], а затем применять известные критерии для проверки равенства средних. Процедура Сиджела заключается в перераспределении вариационного ряда, в который вошли как отсчеты x_i первой выборки, так и отсчеты y_i второй выборки. Когда вариационный ряд построен в виде $v_1 \leq v_2 \leq \dots \leq v_{2N-1} \leq v_{2N}$, где $v_i = x_i$ или $v_i = y_i$, то ранги переназначают следующим образом. Ранг 1 приписывается наименьшему члену v_1 , ранг 2 — наибольшему v_{2N} , ранг 3 — наименьшему оставшемуся члену v_2 , ранг 4 — наибольшему оставшемуся члену v_{2N-1} и так далее попеременным переходом из одного края вариационного ряда к другому и обратно.

Пример. Пусть имеется две выборки — x_i : 12, 8, 23, 61 и y_i : 7, 31, 18, 25. Общий вариационный ряд, расположенный в порядке возрастания рангов, выглядит так: 7, 8, 12, 18, 23, 25, 31, 61. Тогда ряд, построенный с помощью процедуры Сиджела, имеет вид: 7, 61, 8, 31, 12, 25, 18, 23.

Второй частью критерия является использование критерия Вилкоксона (§ 6-2) к перестроенному вариационному ряду.

Критерий рассеяния [37]. Этот критерий аналогичен критерию сдвига из § 6-2 и очень прост для расчетов. Используются отсчеты двух выборок x_i и y_i , где $i=1, 2, \dots, n$. По второй выборке находятся наименьший отсчет y_{\min} и наибольший y_{\max} . Затем следует подсчитать число иксов s , удовлетворяющих неравенству $y_{\max} < x_i < y_{\min}$. Нулевая гипотеза, состоящая в предположении равенства дисперсий для первого и второго участков реализации, принимается, если $s > Z_s$. Критические значения Z_s приведены в табл. 6-5. Объемы каждой вы-

Т а б л и ц а 6-5

Критические значения критерия рассеяния

$P_{\text{д}}=0,95$		$P_{\text{д}}=0,99$	
Диапазон n	Z_s	Диапазон n	Z_s
Не более 6	5	Не более 10	7
От 7 до 25	6	От 11 до 20	8
От 25 до 50	7	От 21 до 50	9

борки $n=N/2$ по-прежнему целесообразно выбирать равными для обеих выборок.

В заключение этого раздела отметим, что кроме перечисленных здесь критериев могут быть применены описанные в § 6-2 **критерии серий и инверсий**. Для этих критериев определяется стационарность ряда оценок σ_i^* или σ_i^{*2} по результатам вычисления на нескольких интервалах.

6-4. ПРОВЕРКА СТАЦИОНАРНОСТИ РАСПРЕДЕЛЕНИЙ

Для проверки стационарности распределений может быть применена уже описанная логическая схема: из реализации извлекаются две или более выборки, состоящие из отсчетов x_i, y_i, \dots , снятых на первом, втором и так далее участках записи. По полученным отсчетам рассчитывается статистика критерия, которая сравнивается с критическим значением при выбранной доверительной вероятности. Удобно, что для решения рассматриваемой задачи не приходится вводить новые критерии.

Для проверки согласия распределений разных выборок применимы критерии согласия эмпирических распределений с теоретическими. Поэтому ниже будут

рассмотрены особенности критериев, описанных в § 2-4, связанные с использованием для разных выборок.

Критерий χ^2 . Пусть имеется n_x отсчетов x_i с первого участка записи, n_y отсчетов y_i со второго участка и т. д. Общее количество отсчетов равно N . Нулевая гипотеза состоит в том, что все l выборок принадлежат одной и той же генеральной совокупности, а отклонения в распределениях для каждого из участков объясняются лишь случайным характером процесса. Использование критерия χ^2 в такой постановке состоит в следующем [4]. Вся выборка разбивается на L разрядов, как это делается при построении гистограмм. Границы разрядов возьмем одинаковыми для всех l участков. После сортировки по разрядам количества отсчетов, попавшие в разряды, можно описать в форме таблицы

$$\begin{array}{c|c} \begin{array}{cccc} n_{x1} & n_{x2} & \dots & n_{xL} \end{array} & n_x \\ \begin{array}{cccc} n_{y1} & n_{y2} & \dots & n_{yL} \end{array} & n_y \\ \hline \begin{array}{cccc} \dots & \dots & \dots & \dots \end{array} & \dots \\ \hline \begin{array}{cccc} n_1 & n_2 & \dots & n_L \end{array} & N \end{array}$$

По таблице оцениваются вероятности принадлежности отсчетов каждому из участков $P^*_x = n_x/N$, $P^*_y = n_y/N$, и вычисляется статистика

$$X^2 = \sum_{j=1}^L \left[\frac{(n_{xj} - n_j P^*_x)^2}{n_j P^*_x} + \frac{(n_{yj} - n_j P^*_y)^2}{n_j P^*_y} + \dots \right], \quad (6-9)$$

которая сравнивается с критическим значением $Z_{\chi^2}(\nu, P)$ (см. рис. 2-9) для числа степеней свободы ν . В случае двух выборок с равным числом отсчетов $l=2$

$$n_x = n_y = N/2; \quad P^*_x = P^*_y = 1/2;$$

$$X^2 = 2 \sum_{j=1}^L \frac{1}{n_j} [(n_{xj} - n_j/2)^2 + (n_{yj} - n_j/2)^2] \quad (6-10)$$

и $\nu = L - 1$.

Критерий Смирнова [2]. Этот критерий основан на той же мере расхождения, что и критерий Колмогорова, и применяется для сравнения распределений двух выборок $x_1, x_2, \dots, x_s, \dots, x_m$ и $y_1, y_2, \dots, y_r, \dots, y_n$. Статистика критерия, являющаяся наибольшим значением модуля отклонения эмпирических распределений $W^*_m(x)$

и $W_n^*(y)$, для выборок произвольного объема может быть записана в виде

$$D = \max(D^+, D^-); \quad (6-11)$$

$$D^+ = \max_{1 \leq r \leq n} \left(\frac{r}{n} - W_m^*(y_r) \right) = \max_{1 \leq s \leq m} \left(W_n^*(x_s) - \frac{s-1}{m} \right);$$

$$D^- = \max_{1 \leq r \leq n} \left(W_m^*(y_r) - \frac{r-1}{n} \right) = \max_{1 \leq s \leq m} \left(\frac{s}{m} - W_n^*(x_s) \right).$$

Критические значения $Z_c(n, P_d)$ статистики D для равных объемов обеих выборок приведены на рис. 6-6. При $n > 40$ удобно пользоваться статистикой $0,707 \sqrt{nD}$. В этом случае критические значения $Z_c(P_d)$ могут быть заменены критическими значениями $Z_K(P_d)$ из табл. П1. Нулевая гипотеза о совпадении распределений принимается, если соответствующая статистика меньше критического значения для выбранной доверительной вероятности. Все замечания об упрощении расчетов, указанные в § 2-4 для критерия Колмогорова, остаются в силе и для критерия Смирнова.

Критерий Смирнова позволяет при увеличении объема выборок выявить сколь угодно малые отклонения распределений, но если известно, что отклонения вызваны сдвигом среднего или дисперсии, то пользоваться критерием Смирнова нецелесообразно из-за большой трудоемкости [4].

Пример, приводимый ниже, построен на выборках эксперимента № 2 в примере для критерия Вилкоксона. Результаты промежуточных расчетов сведены в табл. 6-6. Согласно таблице статистика D равна 0,5. Следовательно, неравенство $0,707 \sqrt{8} \times 0,5 = 1 < Z_c(8, P_d)$ не выполняется (см. рис. 6-6) даже для $P_d = 0,99$. Таким образом, в данном примере нулевая гипотеза отвергается с доверительной вероятностью $P_d > 0,99$.

Критерий ω_N^2 . Аналогично предыдущему случаю критерий Крамера — Мизеса (§ 2-4) может быть применен для

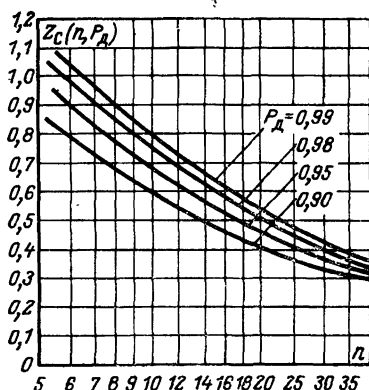


Рис. 6-6. Критические значения критерия Смирнова.

**Данные и результаты промежуточных расчетов
при использовании критерия Смирнова**

Номер отсчета	x_i	y_i	x'_i	y'_i	$W^*(y')$	$\frac{s-1}{m}$	$\frac{r-1}{n}$	D_i^+	D_i^-
1	22	60	14	20	1/8	1/8	0	0	1/8
2	51	92	22	21	2/8	1/8	1/8	1/8	0
3	14	66	37	46	3/8	5/8	2/8	-2/8	3/8
4	37	20	43	60	4/8	7/8	3/8	-3/8	4/8
5	92	112	45	66	5/8	7/8	4/8	-2/8	3/8
6	58	46	51	92	6/8	1	5/8	-2/8	3/8
7	45	100	58	100	7/8	1	6/8	-1/8	2/8
8	43	21	92	112	1	1	7/8	0	1/8

выявления нестационарности по виду эмпирических распределений. Асимптотическое распределение критерия ω_N^2 для двухвыборочного случая совпадает с распределением для одновыборочного случая [4]. Следовательно, могут быть использованы прежние критические значения Z_ω (табл. П1).

ПРИЛОЖЕНИЕ

СПРАВОЧНЫЕ ТАБЛИЦЫ

Таблица П1

**Критические значения нормального распределения
и распределений статистик критериев Колмогорова и ω_N^2**

P	$\alpha = 1 - P$	Z_Φ	Z_K	Z_ω
0,8	0,2	1,28	1,07	0,24
0,9	0,1	1,64	1,22	0,35
0,95	0,05	1,96	1,36	0,46
0,99	0,01	2,58	1,63	0,74
0,995	0,005	2,81	1,73	0,87

Коэффициенты a_{n-i+1} , используемые при проверке на нормальность с помощью критерия V , для N от 3 до 50, тысячные доли

N																				
$n-i+1$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
1	707	687	665	643	623	605	589	574	560	547	536	525	515	506	497	489	481	473	464	
2		168	241	281	303	316	324	329	331	332	332	332	330	329	327	325	323	321	318	
3				087	140	174	198	214	226	235	241	246	249	252	254	255	256	256	258	
4						056	095	122	143	158	170	180	188	194	199	203	206	208	212	
5								040	069	092	110	124	135	145	152	159	164	169	174	
6										030	054	073	088	100	111	120	127	133	140	
7												024	043	059	072	084	093	101	109	
8														020	036	050	061	071	080	
9																016	030	042	053	
10																		014	026	
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				
20																				
21																				
22																				
23																				
24																				
25																				


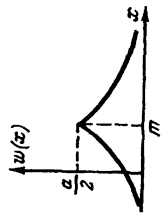
a_{n-i+1}	N																		
	22	23	24	25	26	27	28	29	30	32	34	36	38	40	42	44'	46	48	50
1	459	454	449	445	441	437	433	429	425	419	413	407	401	396	392	387	383	379	375
2	316	313	310	307	304	302	299	297	294	290	285	281	277	273	270	267	263	260	257
3	257	256	255	254	253	252	251	250	249	246	244	241	239	237	234	232	230	228	226
4	213	214	214	215	215	215	215	215	215	214	213	212	211	210	208	207	206	205	203
5	176	179	181	182	184	185	186	186	187	188	188	188	188	188	187	187	186	185	185
6	144	148	151	154	156	158	160	162	163	165	167	168	169	169	169	169	169	169	169
7	115	120	124	128	132	135	137	139	141	145	147	150	151	153	153	154	155	155	155
8	8	088	094	100	105	109	113	116	119	122	126	130	136	138	139	140	141	142	143
9	062	070	076	082	088	092	096	100	104	109	114	118	121	124	126	128	129	131	132
10	037	046	054	061	067	073	078	082	086	093	100	104	107	111	114	116	118	120	121
11	012	023	032	040	048	054	060	065	070	078	084	090	095	099	102	105	107	109	111
12			011	020	028	036	042	048	054	063	071	077	082	087	091	094	097	100	102
13					009	018	025	032	038	048	057	064	071	076	080	084	088	091	093
14							008	016	023	034	044	052	059	065	070	074	078	082	085
15									008	021	031	040	048	055	060	065	069	073	076
16										007	019	028	037	044	051	056	060	065	068
17											006	017	026	034	041	047	052	057	061
18												006	016	024	032	038	044	049	053
19													005	015	023	030	036	041	046
20														005	014	021	028	033	039
21															004	013	020	026	031
22																004	012	018	024
23																	004	011	017
24																		004	010
25																			003

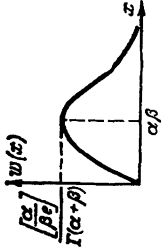
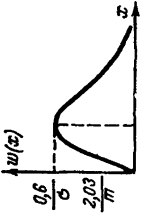
Примечание. Для $N > 30$ коэффициенты a_{n-i+1} даны для четных N . Для нечетных значений N коэффициенты получают линейной интерполяцией.

Таблица ПЗ

Параметры некоторых непрерывных распределений

Закон распределения	Формула плотности	График	Среднее значение m_x	СКО σ_x	Коэффициент вариации v_x	Коэффициент асимметрии K_a	Коэффициент эксцесса K_3
Равномерный	$w(x) = \begin{cases} 0 & -\infty < x < a \\ \frac{1}{b-a} & a < x < b \\ 0 & b < x < \infty \end{cases}$		$\frac{a+b}{2}$	$0,29 \times (b-a)$	0,145 при $a=0$ или $b=0$	0	-1,2
Нормальный	$w(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$		m	σ	σ/m	0	0
Односторонний нормальный	$w(x) = \begin{cases} 0 & x < 0 \\ \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} & x > 0 \end{cases}$		$0,8\sigma$	$0,6\sigma$	0,75	1	0,85

Закон распределения	Формула плотности	График	Среднее значение m_x	СКО σ_x	Коэффициент вариации v_x	Коэффициент асимметрии K_3	Коэффициент эксцесса K_3
Экспоненциальный	$w(x) = \begin{cases} 0 & x < 0 \\ ae^{-ax} & x > 0 \end{cases}$		$1/a$	$1/a$	1	2	6
Лапласа	$w(x) = \frac{a}{2} e^{-a x-m }$		m	$\frac{1,41}{a}$	$\frac{a}{1,41} m$	2,1	3

Закон распределения	Формула плотности	График	Среднее значение m_x	СКО σ_x	Коэффициент вариации v_x	Коэффициент асимметрии K_a	Коэффициент эксцесса K_3
Релея	$w(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & x > 0 \end{cases}$		1,25 σ	0,655 σ	0,52	0,63	-0,3
Гамма-распределение	$w(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)} x^{\alpha} e^{-x/\beta} & x > 0 \end{cases}$		($\alpha + 1$) β	$\sqrt{\alpha + 1} \beta$	$\frac{1}{\sqrt{\alpha + 1}}$	$\frac{2}{\sqrt{\alpha + 1}}$	$\frac{6}{\alpha + 1}$

Значение весовых коэффициентов оптимальной линейной

оценки математического ожидания $\mu^* = \sum_{i=1}^N a_i x(i)$ при различных объемах выборки N

Вид распределения		Нормальное	Симпсона	Двустороннее экспоненциальное (Лапласа)	Равномерное
$N = 2$	a_1	0,5	0,5	0,5	0,5
	a_2	0,5	0,5	0,5	0,5
	$\sigma(\mu^*)/\sigma_x$	0,705	0,705	0,705	0,705
$N = 3$	a_1	0,333	0,395	0,148	0,5
	a_2	0,333	0,21	0,704	0
	a_3	0,333	0,395	0,148	0,5
	$\sigma(\mu^*)/\sigma_x$	0,575	0,57	0,545	0,55
$N = 4$	a_1	0,25	0,338	0,047	0,5
	a_2	0,25	0,162	0,453	0
	a_3	0,25	0,162	0,453	0
	a_4	0,25	0,338	0,047	0,5
	$\sigma(\mu^*)/\sigma_x$	0,5	0,495	0,455	0,45
$N = 5$	a_1	0,2	0,306	0,017	0,5
	a_2	0,2	0,119	0,221	0
	a_3	0,2	0,15	0,524	0
	a_4	0,2	0,119	0,221	0
	a_5	0,2	0,306	0,017	0,5
	$\sigma(\mu^*)/\sigma_x$	0,45	0,44	0,396	0,38

Значение коэффициента $B(N, l)$ и сравнение точностей различных оценок определения СКО для различных объемов выборки N

Объем выборки N	Номера подгрупп, входящих в оценку	$\sigma^*_x = B(N, l) \sum_{i=1}^l \Delta_i$			
		$B(N, l)$	$\sigma[\sigma^*_x]$	$\sigma[\sigma^*_{\text{опт}}]$	$\sigma[\sigma^*_{\text{прям}}]$
6	1, 2	0,262	0,330	0,325	0,323
7	1, 2	0,237	0,299	0,296	0,294
8	1, 2	0,220	0,276	0,273	0,272
9	1, 2	0,207	0,258	0,255	0,254
10	1, 2	0,197	0,243	0,240	0,239
11	1, 2, 4	0,161	0,230	0,227	0,226
12	1, 2, 4	0,152	0,219	0,217	0,216
13	1, 2, 4	0,146	0,209	0,207	0,206
14	1, 2, 4	0,140	0,200	0,199	0,198
15	1, 2, 4	0,135	0,193	0,191	0,191
16	1, 2, 4	0,131	0,186	0,185	0,184
17	1, 2, 3, 5	0,105	0,180	0,179	0,178
18	1, 2, 3, 5	0,102	0,175	0,173	0,173
19	1, 2, 3, 5	0,099	0,170	0,168	0,168
20	1, 2, 4, 6	0,105	0,165	0,164	0,163

Таблица П6

Значение весовых коэффициентов оптимальной линейной оценки среднеквадратического отклонения $\sigma^*_x = \sum_{i=1}^N b_i x_{(i)}$ при различных объемах выборки N

Вид распределения		Нормальное	Симпсона	Двустороннее экспоненциальное (Лапласа)	Равномерное
$N = 2$	b_1	-0,886	-0,875	-0,943	-0,866
	b_2	0,886	0,875	0,943	0,866
	$\sigma[\sigma^*_x]/\sigma_x$	0,758			0,708
$N = 3$	b_1	-0,591	-0,583	-0,622	-0,577
	b_2	0	0	0	0
	b_3	0,591	0,583	0,622	0,577
	$\sigma[\sigma^*_x]/\sigma_x$	0,525	0,492	0,660	0,448

Продолжение табл. П6

Вид распределения		Нормальное	Симпсона	Двустороннее экспоненциальное (Лапласа)	Равномерное
$N = 4$	b_1	-0,454	-0,472	-0,431	-0,481
	b_2	-0,11	0,054	0,3	0
	b_3	0,11	0,054	0,3	0
	b_4	0,454	0,472	0,431	0,481
	$\sigma[\sigma^*_x]/\sigma_x$	0,423	0,390	0,541	0,335
$N = 5$	b_1	-0,372	-0,399	-0,326	-0,433
	b_2	-0,135	-0,064	-0,317	0
	b_3	0	0	0	0
	b_4	0,135	0,064	0,317	0
	b_5	0,372	0,399	0,326	0,433
	$\sigma[\sigma^*_x]/\sigma_x$	0,365	0,33	0,48	0,262

СПИСОК ЛИТЕРАТУРЫ

1. Бендат Дж., Пирсол А. Пер. с англ. Г. В. Матушевского и В. Е. Привального. Измерение и анализ случайных процессов. М.: Мир, 1974.
2. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М.: Наука, 1968.
3. Боярский Э. А. Порядковые статистики. М.: Статистика, 1972.
4. Ван дер Варден Б. Л. Пер. с нем. Л. Н. Большева/ Под ред. Н. В. Смирнова. Математическая статистика. М.: Иностранная литература, 1960.
5. Венецкий И. Г., Кильдишев Г. С. Теория вероятностей и математическая статистика. М.: Статистика, 1975.
6. Виленкин С. Я. Об оценке среднего в стационарных процессах. — Теория вероятностей и ее применения, 1959, вып. 4, с. 451—453.
7. Гаек Я., Шидак З. Пер. с англ. Д. М. Чибисова/ Под ред. Л. Н. Большева. Теория ранговых критериев. М.: Наука, 1971.
8. Гмурман В. Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 1972.
9. Голд В., Рейдер И. Пер. с англ./ Под ред. А. М. Трахтмана. Цифровая обработка сигналов. М.: Советское радио, 1973.
10. Горбенко В. С. Вычисление ожидаемой относительной ошибки при экспериментальном определении одномерных функций, распределения вероятностей. — Изв. вузов. Сер. Радиотехника, 1962, № 5, с. 234—242.
11. Горяинов В. Т., Журавлев А. Г., Тихонов В. И. Примеры и задачи по статистической радиотехнике. М.: Советское радио, 1970.
12. Денисенко А. Н. Об одном методе нахождения распределения интервалов между нулями случайного процесса. — Радиотехника, 1972, № 12, с. 88—91.
13. Дженкинс Г., Ваттс Д. Пер. с англ. В. Ф. Писаренко. Спектральный анализ и его приложения. М.: Мир, 1971, вып. 1.
14. Долуханов М. П. Флуктуационные процессы при распространении радиоволн. М.: Связь, 1971.
15. Жовинский А. Н. Экспериментальное определение точности оценки центральной частоты спектра узкополосного процесса методом счета нулей. — Труды НИИР, 1972, № 3, с. 66—73. В надзаг.: Государственный научно-исследовательский институт радио.
16. Жовинский А. Н. Точность оценки интервала корреляции. — В кн.: Методы представления и аппаратный анализ случайных процессов и полей. Ленинград — Ереван: 1973, с. 18—23. В надзаг.: VI Всесоюзный симпозиум.
17. Жовинский А. Н. Оценивание интервала корреляции с позиции теории проверки гипотез. — Радиотехника и электроника, 1975, № 9, с. 1962—1964.

18. **Жовинский В. Н.** Определение среднеквадратического отклонения по осциллограмме. — Вопросы радиоэлектроники. Сер. общетехническая, 1969, вып. 10, с. 97—101.
19. **Жовинский В. Н., Арховский В. Ф.** Корреляционные устройства. М.: Энергия, 1974.
20. **Жовинский В. Н., Венгеров А. А., Кулаковский А. И.** Определение ошибок оператора при одномерном слежении. — В кн.: Методы разработки радиоэлектронной аппаратуры. М.: 1970, с. 201—204. В надзаг.: Московский Дом научно-техн. пропаганды им. Ф. Э. Дзержинского.
21. **Жовинский В. Н., Жовинский А. Н.** О корреляционном анализе одного класса нестационарных процессов. — В кн.: Преобразование и обработка информации. Изд-во Московского университета, 1972, с. 131—138.
22. **Жовинский В. Н., Жовинский А. Н.** Метод экспресс-оценки спектральной плотности случайного процесса. — Материалы V научно-технической конференции. Радиоизмерения. Каунас—Вильнюс; 1973, т. 1, с. 110—113. В надзаг.: Литовский научно-исследовательский институт научно-технической информации и технико-экономического анализа.
23. **Инженерно-технический справочник по электросвязи.** Радиорелейные линии. М.: Связь, 1971.
24. **Кендал М. Дж., Стьюарт А.** Пер. с англ. Л. И. Гальчука и А. Т. Терехина/ Под ред. А. Н. Колмогорова. Статистические выводы и связи. М.: Наука, 1973.
25. **Козлов Б. А., Ушаков И. А.** Справочник по расчету надежности аппаратуры радиоэлектроники и автоматики. М.: Советское радио, 1975.
26. **Кринецкий Е. И.** Системы самонаведения. М.: Машиностроение, 1970.
27. **Кузьмина Л. М.** Сглаживание измерительной информации. — Труды I Всесоюзной конференции по вычислительным системам. Автоматизация физико-технических исследований. Новосибирск: Наука, 1968, вып. 6.
28. **Левин Б. Р.** Теоретические основы статистической радиотехники. М.: Советское радио, 1966, т. 1.
29. **Лившиц Н. А., Пугачев В. Н.** Вероятностный анализ систем автоматического управления. М.: Советское радио, 1963, т. 1.
30. **Лихарев В. А.** Цифровые методы и устройства в радиолокации. М.: Советское радио, 1973.
31. **Малыревский Н. М.** К вопросу измерения кривой распределения вероятностей случайного процесса. — Изв. вузов. Сер. Радиотехника, 1962, № 2, с. 189—199.
32. **Мирский Г. Я.** Аппаратурное определение характеристик случайных процессов. М.: Энергия, 1972.
33. **Мирский Г. Я.** Измерение нормированной функции корреляции методом условного среднего знаковой функции. — В кн.: Методы представления и аппаратный анализ случайных процессов и полей. Ленинград—Сухуми; 1978, с. 10—15. В надзаг.: X Всесоюзный симпозиум.
34. **Мэгун Г.** Бодрствующий мозг. М.: Мир, 1965.
35. **О вычислении функции распределения случайного процесса по экспериментальным данным.** / Бабурин В. М., Матвеев Л. С., Рождественский Ю. Б., Сорокин Ю. И. — Автоматика и телемеханика, 1962, № 5, с. 571—580.

36. Олейник С. Ф. Основы теории сердечных шумов. М.: Медгиз, 1961.

37. Оуэн Д. Б. Сборник статистических таблиц. Пер. с англ. Л. Н. Большева и В. Ф. Котельниковой. М.: Вычислительный центр АН СССР, 1973. В надзаг.: Академия наук СССР.

38. Павловский З. Пер. с польск. В. Д. Меникера/ Под ред. Ф. Д. Лифшица. Введение в математическую статистику. М.: Статистика, 1967.

39. Полляк Ю. Г. Вероятностное моделирование на электронных вычислительных машинах. М.: Советское радио, 1971.

40. Пугачев В. С. Теория случайных функций и ее применение к задачам автоматического управления. М.: Физматгиз, 1960.

41. Романенко А. Ф., Сергеев Г. А. Вопросы прикладного анализа случайных процессов. М.: Советское радио, 1968, 255 с. с ил.

42. Румшиский Л. З. Математическая обработка результатов измерений. Справочное пособие. М.: Наука, 1971.

43. Соболев И. М. Численные методы Монте-Карло. М.: Наука, 1973.

44. Солодовников В. В. Статистическая динамика линейных систем. М.: Физматгиз, 1960.

45. Справочник по вероятностным расчетам/ Абегауз Г. Г., Тронь А. П., Копенкин Ю. Н., Коровина И. А. — М.: Военное изд-во МО СССР, 1970.

46. Тихонов В. И. Выбросы случайных процессов. М.: Наука, 1970.

47. Феллер В. Пер. с англ. Р. Л. Добрушина, А. А. Юшкевича, С. А. Молчанова/ Под ред. Е. Б. Дынкина. Введение в теорию вероятностей и ее приложения. М.: Мир, 1967, т. 1.

48. Халиков Д. М. Обработка результатов статистического моделирования при исследовании сложных систем с помощью ЦВМ.— Вопросы кибернетики и вычислительной математики. Ташкент: Фан Узбекской ССР, вып. 24, 1969, с. 92—95. В надзаг.: Академия наук Узбекской ССР. Ин-т кибернетики с вычислительным центром.

49. Хан Г., Шапиро С. Пер. с англ. Е. Г. Коваленко/ Под ред. В. В. Налимова. Статистические модели в инженерных задачах. М.: Мир, 1969.

50. Химмельблау Д. Пер. с англ. В. Д. Скаржинского/ Под ред. В. Г. Горского. Анализ процессов статистическими методами. М.: Мир, 1973.

51. Шор Л. Б., Кузьмин Ф. И. Таблицы для анализа и контроля надежности. М.: Советское радио, 1968.