

Министерство образования и науки Российской Федерации

Федеральное агентство по образованию

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

А.В. Сикан

М Е Т О Д Ы СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

2-е издание

*Допущено Министерством образования и науки Российской Федерации
в качестве учебника для студентов высших учебных заведений, обучающихся по
специальности «Гидрология» направления подготовки «Гидрометеорология»*



Санкт-Петербург

2007

УДК 519.24: 556.5

Сикан А.В. Методы статистической обработки гидрометеорологической информации. Учебник. Специальность «Гидрология» направления подготовки «Гидрометеорология». – СПб.: изд. РГГМУ. 2007. – 279 с.

ISBN 5-86813-029-4

Рецензенты: д-р техн. наук, проф. Ю.Б. Виноградов (кафедра гидрологии суши СПбГУ), д-р физ.-мат. наук С.А. Кондратьев (Институт озераедения РАН), д-р техн. наук, проф. А.В. Рождественский (Государственный гидрологический институт).

Излагаются методы статистической обработки гидрометеорологической информации. Рассматриваются применяемые в гидрологии функции распределения, методы оценки однородности и случайности гидрологических рядов, техника расчета основных статистических характеристик, практические приемы использования аппарата парной и множественной линейной корреляции, излагаются методы анализа случайных процессов и способы их применения в практике гидрологических исследований. В книге содержится большое количество примеров, позволяющих закрепить излагаемый теоретический материал.

Предназначена для студентов гидрометеорологических специальностей и специалистов-гидрологов.

The book considers methods of statistical processing of hydrometeorological information. Distribution curves, methods of estimating homogeneity and random character of hydrological time series, calculation methods of basic statistical characteristics, practical methods of using pair and multiple linear correlations are discussed; methods of analysis of random processes and ways of their application in practical hydrological studies are presented. The book contains many examples, which helps to consolidate the theoretical material under study.

The book is designed for students of hydrometeorology and experts in hydrology.

ISBN 5-86813-029-4

© А.В. Сикан, 2007

© Российский государственный гидрометеорологический университет (РГГМУ), 2007

Российский государственный
гидрометеорологический университет

БИБЛИОТЕКА

195196, СПб, Малоохтинский пр., 98

1795

ПРЕДИСЛОВИЕ

В настоящее время статистические методы обработки гидрометеорологической информации являются важнейшей составной частью прикладной гидрологии. Методы математической статистики и теории вероятностей используются при проведении гидрологических, водохозяйственных и гидроэнергетических расчетов, при составлении гидрологических прогнозов, в процессе экологического мониторинга, при оценке экономической эффективности водохозяйственных проектов и т. д.

При подготовке учебника автор стремился в доступной и компактной форме изложить основные методы статистического анализа и обработки гидрометеорологических данных, используемые в гидрологической практике.

Данная книга является вторым изданием учебного пособия «Методы статистической обработки гидрометеорологической информации». Первая книга была написана в 2001 г. в соавторстве с безвременно ушедшим из жизни В.С. Дружининым.

По сравнению с первым изданием в книгу включены две дополнительные главы – 6 и 7, частично переработана глава 5, увеличен объем приложений.

Книга состоит из семи глав. Содержит одиннадцать приложений, список условных обозначений и сокращений, предметный указатель.

В первой главе приводятся краткие сведения из теории вероятностей, необходимые для понимания излагаемых в дальнейшем расчетных методов.

Вторая глава целиком посвящена описанию аналитических функций распределения, используемых в практике гидрологических расчетов; анализируются их достоинства и недостатки. Наряду с традиционными распределениями здесь рассматриваются четырехпараметрическое распределение Джонсона и H -распределение, предложенное в 1986 г. А. В. Ежовым (ГГИ); опубликованы ординаты этого распределения.

В третьей главе излагаются методы расчета оценок параметров распределения гидрологических характеристик и методы построения эмпирических кривых обеспеченностей. В книге впервые предложены практическая реализация графоаналитического метода

оценки параметров распределения для кривой Крицкого – Менкеля и новая методика оценки параметров распределения Джонсона.

Четвертая глава посвящена интервальной оценке параметров распределения и проверке статистических гипотез. Здесь излагаются методы проверки рядов на случайность и однородность. Рассматриваются наиболее часто используемые критерии согласия.

В пятой главе излагаются основы парной и множественной линейной корреляции. Рассматриваются приемы, позволяющие оценить надежность получаемых по эмпирическим данным уравнений регрессии.

В шестой главе рассматриваются методы теории случайных процессов, применяемые при анализе гидрологических рядов.

В седьмой главе в краткой форме излагаются основные требования, предъявляемые нормативными документами к статистической обработке гидрометеорологических данных.

В тексте содержится большое количество примеров, позволяющих практически закрепить теоретический материал.

Книга ориентирована на студентов-гидрологов, изучающих дисциплину "Методы статистической обработки гидрометеорологической информации". Вместе с тем она может быть полезной и инженерам-гидрологам, а также другим специалистам, работающим с гидрометеорологическими данными.

Автор выражает глубокую благодарность профессору А.В. Рождественскому, который внимательно прочитал рукопись, и сделал ряд ценных замечаний.

СПИСОК УСЛОВНЫХ ОБОЗНАЧЕНИЙ И СОКРАЩЕНИЙ

Условные обозначения

- $F(x)$ – интегральная функция распределения
 $f(x)$ – дифференциальная функция распределения
(функция плотности вероятности)
 p – значение вероятности
 $P(x)$ – функция обеспеченностей
 Mo – мода
 Me – медиана
 α_s – начальный момент s -го порядка
 μ_s – центральный момент s -го порядка
 m_x – математическое ожидание
 σ – среднеквадратическое отклонение (стандарт)
 σ^* и S – выборочное среднеквадратическое отклонение
 D и σ^2 – дисперсия
 D^* и S^2 – выборочная дисперсия
 C_v – коэффициент вариации
 C_s – коэффициент асимметрии
 E_x – эксцесс
 k – модульный коэффициент
 t – стандартная нормированная величина
 $\Gamma(\cdot)$ – гамма-функция
 ν – число степеней свободы
 r – коэффициент парной корреляции
 R – коэффициент множественной корреляции
 P – оператор вероятности
 M – оператор математического ожидания
 D – оператор дисперсии
 σ – оператор среднеквадратического отклонения

Сокращения

СВ – случайная величина

МО – математическое ожидание

СКО – среднеквадратическое отклонение

ОДЗ – область допустимых значений

МНК – метод наименьших квадратов

СП – случайный процесс

ПНСП – периодически нестационарный СП

ПКСП – периодически коррелированный СП

АР – модель авторегрессии

СС – модель скользящего среднего

АРСС – модель авторегрессии и скользящего среднего

АРПСС – модель авторегрессии и проинтегрированного скользящего среднего

ВВЕДЕНИЕ

Многие гидрологические характеристики (среднегодовой, максимальный и минимальный расходы воды; слой стока за половодье или паводок; продолжительность половодья или межени; дата начала и окончания ледостава на реке или озере и др.) определяются огромным числом факторов, степень влияния каждого из которых учесть практически невозможно. При этом конкретное значение характеристики есть результат случайного сочетания этих факторов. С учётом этого, сама исследуемая характеристика должна рассматриваться как случайная величина (или случайный процесс), и для ее определения могут быть использованы методы теории вероятностей и математической статистики. Кроме того, в практике гидрологических расчетов довольно часто встречаются задачи, когда необходимо определить значения гидрологических величин, которые будут встречаться в будущем, например в период эксплуатации того или иного гидротехнического сооружения. Решение такого рода задач также невозможно без привлечения вероятностных (стохастических) методов.

Широкое применение методов теории вероятностей в гидрологической практике началось относительно недавно. Основу для этого заложили работы А. Хазена и А. Фостера, опубликованные в первой четверти XX века. В России толчок для развития этого направления дала работа Д.Л. Соколовского (1930 г.) "Применение кривых распределения к установлению вероятных колебаний годового стока рек европейской части СССР". Большой вклад в развитие стохастической гидрологии внесли также работы Г.А. Алексеева, Е.Г. Блохинова, Г.П. Калинина, С.Н. Крицкого, М.Ф. Менкеля, А.В. Рождественского и др.

К настоящему времени накоплен достаточно большой опыт применения вероятностных методов в практике гидротехнического, дорожного и водохозяйственного проектирования.

В книге рассматриваются основные приемы обработки и анализа гидрологических данных, используемые в российской и мировой гидрологической практике. Излагаемый в книге материал соответствует программе дисциплины "Методы статистической обработки гидрометеорологической информации", которая читается на гидрологическом факультете РГГМУ.

1. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. Случайные величины и функции распределения

Величина, значение которой меняется от опыта к опыту случайным образом, называется *случайной величиной* (СВ).

В отличие от неслучайных (детерминированных) величин для СВ нельзя предсказать точно, какое она примет значение в определенных условиях, а можно только указать *закон распределения* СВ.

Закон распределения СВ считается заданным, если:

- указано множество возможных значений СВ;
- указан способ количественного определения вероятности попадания СВ в любую область из множества возможных значений.

Вероятность попадания СВ в заданную область, например в интервал $[a, b]$, можно определить следующим образом:

$$P_{(a,b)} = \lim_{N \rightarrow \infty} \left(\frac{m}{N} \right), \quad (1.1)$$

где m – число наблюдений случайной величины, оказавшихся в заданной области; N – общее число наблюдений.

Аналитическими выражениями законов распределения случайных величин являются *функции распределения* – интегральная и дифференциальная.

Интегральная функция распределения $F(x)$ случайной величины X показывает вероятность того¹, что случайная величина не превысит некоторого заданного числа x (рис. 1.1), т. е.

$$F(x) = P\{X \leq x\}.$$

Следовательно, вероятность того, что значение случайной величины X заключено между x_1 и x_2 , равна разности значений функций распределения, вычисленных в этих двух точках:

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1). \quad (1.2)$$

¹ Здесь и далее символ P будет обозначать оператор вероятности, а p – конкретное значение вероятности.

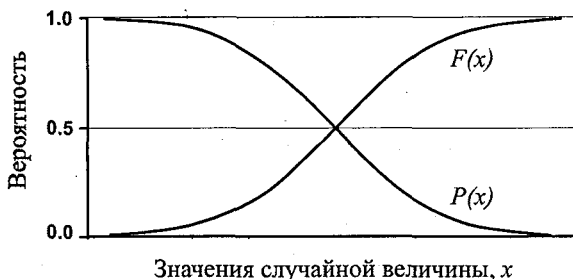


Рис. 1.1. Интегральная функция распределения $F(x)$ и функция обеспеченностей $P(x)$ случайной величины X .

Аналогично

$$P\{X > x\} = P\{+\infty > X > x\} = 1 - F(x). \quad (1.3)$$

В российской гидрологической практике вместо функции $F(x)$ чаще используется так называемая функция обеспеченностей $P(x)$, которая определяется выражением (1.3), но с включением в интервал значения x :

$$P(x) = 1 - F(x) = P\{X \geq x\}. \quad (1.3a)$$

Таким образом, функция обеспеченностей СВ X показывает вероятность превышения некоторого заданного числа x .

Интегральная функция распределения $F(x)$ и функция обеспеченностей $P(x)$ обладают следующими свойствами:

1. $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow -\infty} P(x) = 1$.
2. $\lim_{x \rightarrow +\infty} F(x) = 1$; $\lim_{x \rightarrow +\infty} P(x) = 0$.
3. $F(x) \geq 0$ для всех x ; $P(x) \geq 0$ для всех x .
4. $F(x_2) \geq F(x_1)$, если $x_2 > x_1$; $P(x_2) \leq P(x_1)$, если $x_2 > x_1$.

Если функция $F(x)$ дифференцируема для всех значений случайной величины X , то закон распределения вероятностей может быть выражен и в виде дифференциальной функции распределения вероятностей

$$f(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}, \quad (1.4)$$

где $\Delta x > 0$.

Таким образом, значение функции $f(x)$ приближенно равно отношению вероятности попадания случайной величины X в интервал $(x, x + \Delta x)$ к длине Δx этого интервала, когда Δx — бесконечно малая величина. Поэтому функцию $f(x)$ называют также функцией плотности распределения вероятностей (или короче — функцией *плотности вероятности*).

Основные свойства функции $f(x)$:

$$1) f(x) \geq 0; \quad 2) \lim_{|x| \rightarrow \infty} f(x) = 0; \quad 3) \int_{-\infty}^{+\infty} f(x) dx = 1; \quad 4) \int_{-\infty}^x f(z) dz = F(x).$$

С помощью дифференциальной функции распределения можно вычислить вероятность попадания случайной величины в любую заданную область из множества возможных значений, в частности (рис. 1.2):

$$a) P\{X \leq a_1\} = \int_{-\infty}^{a_1} f(x) dx;$$

$$б) P\{X > a_2\} = \int_{a_2}^{+\infty} f(x) dx;$$

$$в) P\{a_1 < X \leq a_2\} = \int_{a_1}^{a_2} f(x) dx.$$

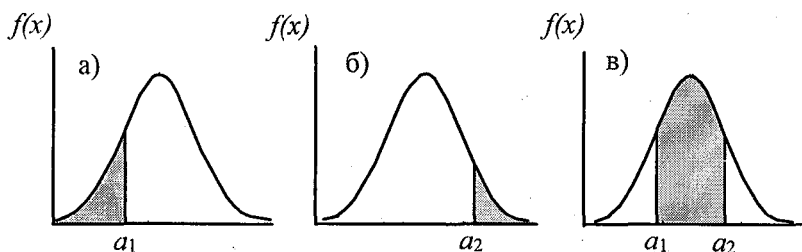


Рис. 1.2. Графики функции плотности вероятности; заштрихованная площадь соответствует вероятности попадания СВ X в заданный интервал.

Как интегральная, так и дифференциальная функции распределения являются исчерпывающими характеристиками случайной величины.

1.2. Дискретные и непрерывные случайные величины

Следует различать дискретные и непрерывные случайные величины. Дискретная СВ X может принимать только конечное или счетное множество значений: $x_1, x_2, x_3, \dots$ Например, бросая игральную кость, мы генерируем дискретную случайную величину, которая может принимать только шесть конкретных значений: 1, 2, 3, ..., 6. В другом случае можно рассматривать дискретную СВ, которая принимает любые неотрицательные целочисленные значения. Здесь мы уже имеем бесконечное множество возможных значений СВ X , но это множество является счетным.

Непрерывная СВ может принимать любые значения из некоторого замкнутого или открытого интервала, в том числе и бесконечного. Например, если область возможных значений непрерывной СВ ограничена интервалом $[0, +1]$, то СВ может принимать несчетное число значений. Перечислить их можно, только превратив непрерывную СВ в дискретную, вводя шаг дискретности. Если шаг дискретности 0,1, то СВ X округляется до одной десятой и будет принимать значения 0, 0,1; 0,2, ..., 0,9; 1,0. Если шаг дискретности 0,001, то СВ X будет принимать значения 0, 0,001; 0,002; ..., 0,999; 1,000.

Трактовка функции распределения, которая приведена в п. 1.1, в полной мере может быть отнесена только к непрерывным случайным величинам.

Интегральная функция распределения дискретной СВ X в практических ситуациях представляет собой ступенчатую функцию со скачками в точках $x_1, x_2, x_3, \dots$ (рис. 1.3).

Интегральная функция распределения $F(x)$ дискретной СВ не дифференцируема, и вместо функции плотности вероятности используется ее дискретный аналог, который называют рядом распределения.

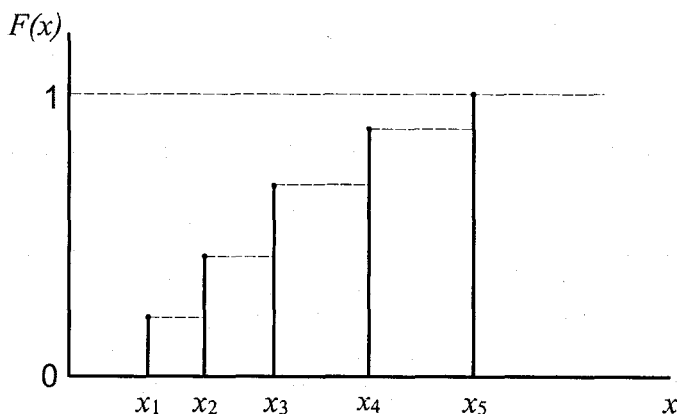


Рис. 1.3. Интегральная функция распределения дискретной случайной величины.

Ряд распределения дискретной случайной величины можно представить в виде таблицы, где для каждого значения СВ из области возможных значений указана его вероятность p_x (табл. 1.1).

Таблица 1.1

Ряд распределения случайной величины X

$X \dots$	x_1	x_2	x_3	$\dots$	x_n
$p_x \dots$	p_1	p_2	p_3	$\dots$	p_n

На основании такой таблицы можно построить гистограмму распределения вероятностей дискретной СВ (рис. 1.4).

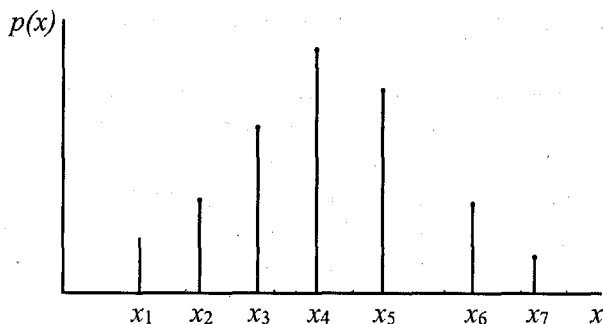


Рис. 1.4. Гистограмма распределения вероятностей дискретной случайной величины.

Для ряда распределения дискретной СВ должно выполняться равенство $\sum_{i=1}^n p(x_i) = 1$, что эквивалентно свойству 3 плотности вероятности.

В гидрометеорологической практике встречаются и дискретные, и непрерывные случайные величины. Например, число дней с отрицательной температурой в третьей декаде марта можно рассматривать как дискретную СВ с областью возможных значений: 0, 1, 2, 3, ..., 10. А максимальный в году расход воды можно рассматривать как непрерывную СВ, область возможных значений которой ограничена интервалом $[0, +\infty)$.

1.3. Случайная величина и время

В гидрологической практике исследователь имеет дело с рядом наблюдений за различными гидрометеорологическими характеристиками. Например, 32-летний ряд наблюдений за максимальными расходами дождевых паводков выглядит так, как показано в табл. 1.2, т.е. мы имеем отсчёты времени (с шагом в один год) и соответствующие паводочные расходы воды.

Таблица 1.2

Максимальные расходы воды дождевых паводков ($\text{м}^3/\text{с}$);
р. Луга – ст. Толмачево; $F = 6350 \text{ км}^2$, $n = 32$

Год	Расход	Год	Расход
1954	71,4	1970	84,5
1955	22,1	1971	37,5
1956	32,7	1972	24,0
1957	145	1973	22,1
1958	46,0	1974	130
1959	29,9	1975	21,3
1960	28,4	1976	58,2
1961	73,8	1977	58,7
1962	121	1978	130
1963	25,0	1979	22,0
1964	31,4	1980	54,0
1965	17,3	1981	100
1966	54,8	1982	78,1
1967	29,9	1983	39,2
1968	104	1984	66,0
1969	27,5	1985	84,4

Таким образом, максимальный паводочный расход можно представить, как функцию времени, и такой подход вполне допустим. Но если мы пытаемся описать вероятностную структуру гидрологического ряда в рамках модели случайной величины, то время рассматривается только в качестве формального счетчика опытов: 32 года – 32 опыта. При этом последовательность расположения расходов не имеет значения.

Безусловно, случайная величина является наиболее простой вероятностной моделью, но она позволяет решить очень многие практические задачи.

Более сложные вероятностные модели, где учитывается фактор времени, описываются в *теории случайных процессов*.

1.4. Числовые характеристики случайных величин

Как отмечалось выше, исчерпывающей характеристикой любой случайной величины является ее функция распределения (интегральная или дифференциальная). Но основные свойства СВ (хотя и не все) могут быть описаны более компактно с помощью нескольких числовых характеристик (параметров распределения).

1.4.1. Мода. Медиана. Математическое ожидание.

На практике одной из наиболее важных числовых характеристик является *центр распределения*. Однако определить его можно по-разному. Существует, по крайней мере, три параметра, характеризующих положение центра распределения (или короче – три характеристики положения): мода, медиана и математическое ожидание.

Модой M_o непрерывной СВ X называется такое ее значение, которому соответствует максимум плотности вероятности (рис. 1.5, а).

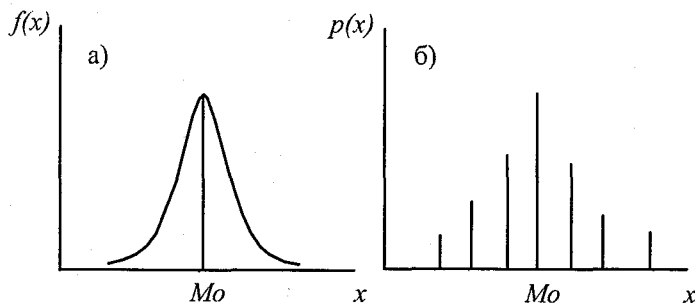


Рис. 1.5. Мода непрерывной (а) и дискретной (б) случайных величин.

В случае дискретного распределения модой называется наиболее вероятное значение случайной величины (рис. 1.5, б). Если распределение имеет более чем один локальный максимум, такое распределение называется многомодальным (полимодальным).

Медианой Me непрерывной СВ X называется такое ее значение, при котором

$$P\{X < Me\} = P\{X > Me\} = 0,5. \quad (1.5)$$

Можно также сказать, что Me — это такое значение СВ, при котором значение функции обеспеченностей равно значению интегральной функции распределения (рис. 1.6).

$$P(Me) = F(Me) = 0,5. \quad (1.6)$$

Для дискретных СВ медиана определяется неоднозначно и практически не употребляется.

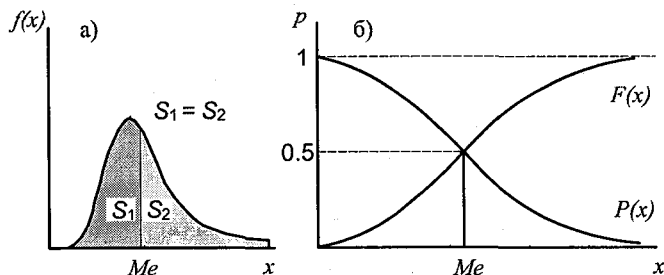


Рис. 1.6. Положение медианы на графиках дифференциальной (а) и интегральной (б) функций распределения.

Математическое ожидание (МО) случайной величины определяется следующими формулами:

для дискретной СВ
$$m_x = \sum_i x_i p_i; \quad (1.7)$$

для непрерывной СВ
$$m_x = \int_{-\infty}^{+\infty} x f(x) dx. \quad (1.8)$$

Как следует из (1.8), МО можно трактовать как центр тяжести плотности вероятности.

Математическое ожидание является наиболее употребляемой характеристикой положения, и в дальнейшем в качестве центра распределения мы будем подразумевать именно его. При этом следует отметить, что термин *математическое ожидание* применим не только к некоторой случайной величине X , но и к случайным величинам, которые являются функциями X , например, к X^2 или $\ln(X)$. В этом случае в качестве символа МО мы будем использовать обозначение $M[\cdot]$. Таким образом, для СВ X можно записать также $m_x \sim M[X]$.

Кроме термина *математическое ожидание* используется и другой – *генеральное среднее*, в этом случае для обозначения МО используется символ $\bar{X}_N$, где $N \rightarrow \infty$.

Если мода, медиана и математическое ожидание совпадают, то распределение является симметричным (см. рис. 1.7 а). Если МО расположено правее медианы, то распределение имеет положительную асимметрию (рис. 1.7 б), в противном случае – отрицательную.

1.4.2. Моменты случайной величины

Большинство числовых характеристик, используемых на практике, связаны с понятием момента СВ. Различают начальные и центральные моменты СВ.

Начальный момент S -го порядка СВ определяется формулой

$$\alpha_S = M[X^S], \quad (1.9)$$

или
$$\alpha_S = \int_{-\infty}^{+\infty} x^S f(x) dx; \quad (1.9a)$$

центральный момент S -го порядка СВ X – формулой

$$\mu_S = M[(X - m_x)^S], \quad (1.10)$$

или
$$\mu_S = \int_{-\infty}^{+\infty} (x - m_x)^S f(x) dx. \quad (1.10a)$$

Из формулы (1.9), в частности, следует, что МО есть первый начальный момент, т.е. $m_x = M[X^1] = \alpha_1$.

Начальные и центральные моменты связаны следующими соотношениями:

$$\left. \begin{aligned} \mu_1 &= 0; \\ \mu_2 &= \alpha_2 - \alpha_1^2; \\ \mu_3 &= \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3; \\ \mu_4 &= \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4. \end{aligned} \right\} \quad (1.11)$$

1.4.3. Дисперсия. Среднеквадратическое отклонение. Коэффициент вариации

Вторую группу наиболее часто используемых на практике параметров составляют параметры, характеризующие степень рассеяния СВ относительно центра распределения. К ним относятся дисперсия, среднеквадратическое отклонение и коэффициент вариации.

Дисперсия СВ X представляет собой второй центральный момент, т.е.

$$D_x = \mu_2 = M[(X - m_x)^2], \quad (1.12)$$

в частности, для непрерывной СВ X дисперсия определяется формулой

$$D_x = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x) dx. \quad (1.12a)$$

Среднеквадратическое отклонение (СКО) СВ X есть квадратный корень из дисперсии

$$\sigma_x = \sqrt{D_x} \text{ или } \sigma_x^2 = D_x. \quad (1.13)$$

Эту величину также называют *стандартным отклонением*, или *стандартом*. Для описания рассеяния положительных СВ можно использовать безразмерную характеристику – коэффициент вариации.

Коэффициент вариации C_v СВ X есть отношение СКО к МО:

$$C_v = \frac{\sigma_x}{m_x} = \frac{\sqrt{D_x}}{m_x}. \quad (1.14)$$

Как следует из формул (1.12) – (1.14), дисперсия, СКО и коэффициент вариации взаимосвязаны. На практике обычно использует-

ся только одна (любая) из перечисленных характеристик. Выбор того или иного параметра зависит от специфики задачи. Влияние коэффициента вариации (а следовательно, дисперсии и СКО) на функцию распределения иллюстрируется на рис. 1.8 а.

1.4.4. Асимметрия и эксцесс

Помимо уже перечисленных числовых характеристик рассмотрим еще две: коэффициент асимметрии и эксцесс.

Коэффициент асимметрии C является безразмерным параметром и характеризует степень симметричности рассеяния относительно математического ожидания (см. рис. 1.7). Коэффициент асимметрии определяется формулой

$$C_s = \frac{\mu_3}{\sigma_x^3} = \frac{M[(X - m_x)^3]}{\sigma_x^3}. \quad (1.15)$$

Для симметричных распределений коэффициент асимметрии равен нулю.

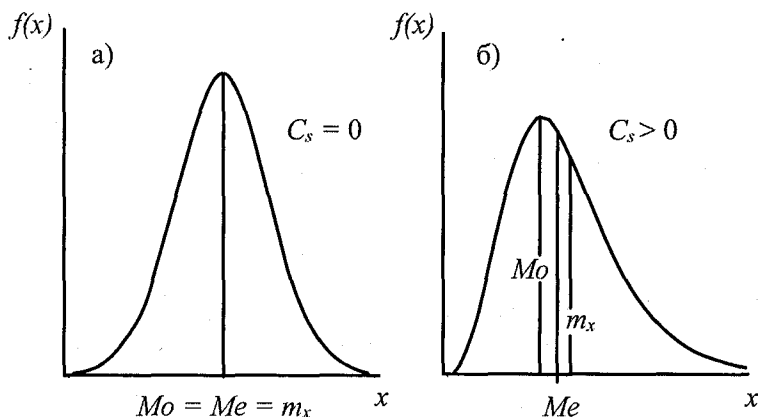


Рис. 1.7. Графики функций плотности вероятности для симметричного (а) и несимметричного (б) распределений.

Эксцесс E_x также является безразмерным параметром и определяется формулой

$$E_x = \frac{\mu_4}{\sigma_x^4} - 3 = \frac{M[(x - m_x)^4]}{\sigma_x^4} - 3. \quad (1.16)$$

Эксцесс позволяет оценить наличие островершинности, или наоборот туповершинности, функции плотности вероятности СВ X относительно нормального закона распределения (см. п. 2.1), для которого $E_x = 0$ (см. рис. 1.8, б).

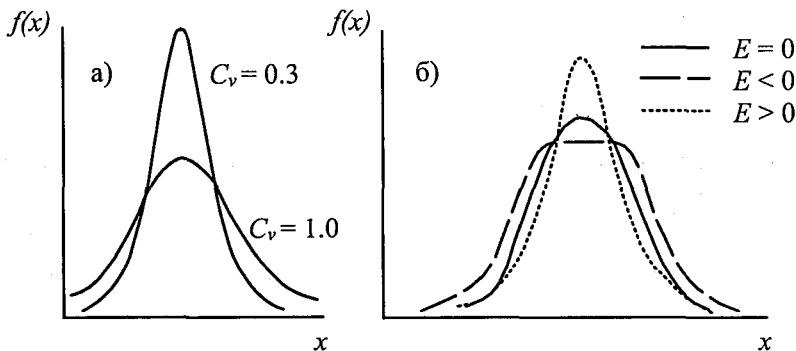


Рис. 1.8. Влияние коэффициента вариации (а) и эксцесса (б) на форму функции плотности вероятности.

Как видно из формул (1.15) – (1.16), коэффициент асимметрии зависит от третьего, а эксцесс от четвертого центральных моментов. Теоретически доказано, что погрешность оценки моментов по эмпирическим данным резко возрастает с увеличением порядка момента, поэтому в гидрологической практике эксцесс используется крайне редко.

1.4.5. Свойства числовых характеристик случайных величин

В данном параграфе приводятся без доказательства основные свойства МО и дисперсии. Доказательства приводимых соотношений можно найти в [20, 25].

Свойства математического ожидания.

1. Математическое ожидание постоянной величины равно самой этой величине:

$$M[c] = c, \quad (1.17)$$

где c – const.

2. Постоянный множитель можно выносить за знак МО:

$$M[cX] = cM[X]. \quad (1.18)$$

3. Математическое ожидание суммы независимых случайных величин равно сумме их МО

$$\mathbf{M}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbf{M}[X_i], \quad (1.19)$$

в частности, для двух случайных величин:

$$\mathbf{M}[X + Y] = \mathbf{M}[X] + \mathbf{M}[Y].$$

4. Математическое ожидание линейной функции от СВ выражается формулой

$$\mathbf{M}[aX + b] = a\mathbf{M}[X] + b, \quad (1.20)$$

где a и b — const. Данное свойство является следствием свойств 1–3.

5. Математическое ожидание произведения независимых СВ равно произведению их МО:

$$\mathbf{M}\left[\prod_{i=1}^N X_i\right] = \prod_{i=1}^N \mathbf{M}[X_i]. \quad (1.21)$$

Свойства дисперсии.

1. Дисперсия постоянной величины равна нулю¹

$$\mathbf{D}[c] = 0, \quad (1.22)$$

где $c = \text{const.}$

2. Постоянную величину можно вынести за знак дисперсии, возводя ее в квадрат

$$\mathbf{D}[cX] = c^2 \mathbf{D}[X]. \quad (1.23)$$

Из выражения (1.23) также следует:

$$\sigma[cX] = c \cdot \sigma[X]. \quad (1.24)$$

3. Дисперсия суммы случайных величин равна сумме их дисперсий:

$$\mathbf{D}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbf{D}[X_i]. \quad (1.25)$$

4. Дисперсия линейной функции СВ определяется выражением

$$\mathbf{D}[aX + b] = a^2 \mathbf{D}[X]. \quad (1.26)$$

Данное свойство является следствием свойств 1–3.

¹ Здесь и далее $\mathbf{D}$ и σ — соответственно операторы дисперсии и СКО.

1.5. Стандартные преобразования случайной величины

Во многих случаях оказывается удобным рассматривать вместо СВ X другую случайную величину, полученную из СВ X на основе элементарного преобразования. В гидрологической практике наиболее часто используется замена СВ X модульными коэффициентами и замена СВ стандартной нормированной случайной величиной.

5 Модульным коэффициентом называется отношение значения СВ к ее математическому ожиданию:

$$k_i = x_i / m_x. \quad (1.27)$$

6 Стандартная нормированная величина может быть получена из СВ по формуле

$$t_i = (x_i - m_x) / \sigma_x, \quad (1.28)$$

или с учетом формул (1.14), (1.27)

$$t_i = (k_i - 1) / C_v \quad (1.28a)$$

Из свойств МО и дисперсии, изложенных в п. 1.4.5, следует, что формула (1.27) преобразует СВ X с МО, равным m_x , в случайную величину с МО, равным единице, а преобразование (1.28) приводит к СВ с нулевым МО и $\sigma_t = D_t = 1$.

1.6. Квантили распределения

В п. 1.1 мы определили интегральную функцию распределения $F(x)$, по которой для любого x можно установить вероятность его непревышения. Во многих практических случаях необходимо решить обратную задачу: по заданной вероятности непревышения $F(x) = p'$ определить величину x'_p . Для обозначения x'_p в этом случае в математической статистике используется специальный термин — *квантиль*.

Таким образом, p -квантилем или $p\%$ -квантилем (если p в %) называется значение случайной величины x'_p , соответствующее заданному значению вероятности непревышения $F(x) = p'$.

По аналогии с квантилями в гидрологической практике используются p -ординаты кривой обеспеченностей¹.

Ординатой кривой обеспеченностей будем называть такое значение СВ X (обозначим его x_p), которое соответствует заданной вероятности превышения $P(x) = p$.

Как следует из (3.1a), $P(x) = 1 - F(x)$, и следовательно, p и p' связаны соотношением $p = 1 - p'$ или (если p в %) $p = 100 - p'$.

¹ Для того чтобы подчеркнуть отличие квантилей от ординат кривой обеспеченностей, будем отмечать квантили штрихом.

2. АНАЛИТИЧЕСКИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ, ИСПОЛЬЗУЕМЫЕ В ГИДРОЛОГИИ

Во многих случаях в качестве математической модели для описания гидрометеорологических явлений может использоваться случайная величина. В частности, в качестве СВ можно рассматривать среднегодовой расход воды, тогда среднегодовые расходы за отдельные годы, например за 30 лет, следует трактовать как последовательность значений СВ, полученную в результате 30 природных опытов.

Точно так в качестве СВ могут рассматриваться максимальный и минимальный расходы воды, годовая сумма осадков, снегозапасы на водосборе и т. д. Кроме того, в качестве случайных величин рассматриваются погрешности измерений гидрометеорологических характеристик (уровня воды, расхода, слоя осадков и т. д.), а также ошибки расчета различных числовых параметров, определяемых по эмпирическим данным.

Как уже отмечалось, каждая случайная величина исчерпывающим образом характеризуется своим законом распределения. Аналитическим выражением закона распределения является функция распределения (дифференциальная или интегральная), которую записывают в виде параметрического выражения, где в качестве параметров обычно используют числовые характеристики случайной величины (МО, СКО и т. д.). При такой записи для каждого значения СВ соответствующее значение функции распределения однозначно определяется параметрами аналитического выражения (параметрами распределения).

В гидрологической практике, как правило, рассматриваются случайные величины, функции распределения которых зависят от небольшого числа параметров – обычно двух или трех.

В одних ситуациях закон распределения можно вывести из теоретических соображений, в других он не известен, и аналитическое выражение функции распределения является лишь аппроксимацией истинного распределения. В последнем случае можно предложить несколько аналитических выражений. Исследователь вынужден производить выбор подходящего, опираясь на практический опыт и некоторые априорные соображения.

Ниже рассматриваются аналитические функции распределения, наиболее часто используемые в практике гидрологических расчетов.

2.1. Нормальное распределение

В природе и различных областях человеческой деятельности весьма распространены СВ, представляющие собой сумму большого числа независимых или слабо зависимых СВ, дисперсии которых малы по сравнению с дисперсией всей суммы. Как следует из центральной предельной теоремы [7, 14], распределение таких СВ при весьма общих дополнительных условиях хорошо аппроксимируется нормальным распределением. Этим объясняется весьма широкое распространение последнего. Нормальное распределение применяется и в тех случаях, когда истинный закон распределения известен, но вычисления по этому закону затруднены, а аппроксимация его нормальным законом не приводит к большим ошибкам.

Функция плотности вероятности нормального распределения определяется выражением

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[-\frac{(x - m_x)^2}{2\sigma_x^2} \right], \quad (2.1)$$

где π – число «пи». Таким образом, в общем случае нормальное распределение является двухпараметрическим, т.е. зависит от двух параметров: математического ожидания (m_x) и среднеквадратического отклонения (σ_x).

График плотности вероятности нормального распределения представлен на рис. 1.7 а. Как видно на рисунке, нормальное распределение является симметричным и, следовательно, для него коэффициент асимметрии равен нулю ($C_s = 0$), а мода, медиана и МО совпадают. Область возможных значений СВ, подчиняющейся нормальному распределению – интервал $(-\infty, +\infty)$.

Интегральная функция распределения имеет вид

$$F(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^x \exp \left[-\frac{(z - m_x)^2}{2\sigma_x^2} \right] dz, \quad (2.2)$$

где z – переменная интегрирования.

Интеграл, стоящий в правой части выражения (2.2), нельзя выразить через элементарные функции, поэтому ординаты нормального закона распределения обычно представляют в виде таблиц. Чтобы не публиковать множество таблиц, для различных сочетаний m_x и σ_x вместо СВ X рассматривают нормированную СВ t (табл. 2.1), значения которой связаны со значениями X формулой (1.28).

Так как у нормированной СВ t : $m_t = 0$ и $\sigma_t = 1$, функции распределения для нее с учетом формул (2.1) – (2.2), будут определяться выражениями

$$f(x) = (1/\sqrt{2\pi}) \exp(-t^2/2), \quad (2.3)$$

$$F(x) = (1/\sqrt{2\pi}) \int_{-\infty}^t \exp(-z^2/2) dz. \quad (2.4)$$

Таблица 2.1

Квантили нормированной нормально распределенной случайной величины t
(значения $F(t)$ даны в %)

$F(t)$	t	$F(t)$	t
0,01	-3,72	60	0,25
0,1	-3,09	70	0,52
0,5	-2,58	75	0,67
1	-2,33	80	0,84
2	-2,02	90	1,28
2,5	-1,96	95	1,64
3	-1,88	97	1,88
5	-1,64	97,5	1,96
10	-1,28	98	2,02
20	-0,84	99	2,33
25	-0,67	99,5	2,58
30	-0,52	99,9	3,09
40	-0,25	99,99	3,72
50	0,00		

Переход от нормированных значений случайной величины к исходным в соответствии с выражением (1.28) производится по формуле

$$x_p = t_p \sigma_p + m_x, \quad (2.5)$$

а для модульных коэффициентов – по формуле

$$k_p = t_p C_v + 1. \quad (2.5a)$$

Пример 2.1. Известно, что СВ X подчиняется нормальному распределению с параметрами $m_x = 10$, $\sigma_x = 5$. Требуется определить значение x'_p , соответствующее вероятности непревышения 95 % (т.е. при $F(x) = 95\%$).

Решение.

1. По табл. 2.1 определяем 95%-ный квантиль нормированной СВ t : $t'_{95} = 1,64$.

2. По формуле (2.5) определяем $x'_{95} = 1,64 \cdot 5 + 10 = 18,2$.

Пример 2.2. Для СВ X из предыдущего примера требуется определить значение x_p , соответствующее вероятности превышения 1 % (при $P(x) = 1\%$).

Решение.

1. Так как в табл. 2.1 приводятся ординаты интегральной функции распределения (а не функции обеспеченностей), сначала определяем соответствующую вероятность непревышения для нормированной СВ t :

$$F(t) = 100 - P(t) = 100 - 1 = 99\%.$$

2. По табл. 2.1 определяем 99%-ный квантиль нормированной СВ t : $t'_{99} = 2,33$.

3. По формуле (2.5) определяем x_1 :

$$x_1 = x'_{99} = 2,33 \cdot 5 + 10 = 21,65.$$

Несмотря на широкое распространение, нормальное распределение не универсально. В частности, если в качестве СВ рассматривать среднегодовые расходы воды, то налицо два несоответствия:

- расходы воды всегда положительны, а область возможных значений для нормальной СВ определяется интервалом $(-\infty, +\infty)$.
- для нормального закона распределения $C_s = 0$, а распределение среднегодовых расходов воды (и целого ряда других гидрологических характеристик), как правило, имеет умеренную положительную асимметрию.

2.2. Закон равномерной плотности

Пусть возможные значения СВ X лежат в интервале $[a, b]$ и нет оснований для того, чтобы отдать предпочтение какому-либо из этих значений. При этих условиях СВ X подчиняется закону равномерной плотности (рис. 2.1). Это распределение называют также равномерным или прямоугольным.

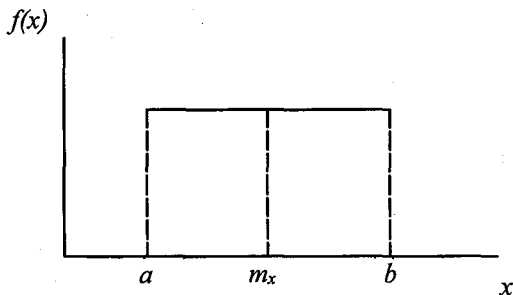


Рис. 2.1. Дифференциальная функция распределения закона равномерной плотности.

Закон равномерной плотности определяется двумя параметрами: началом a и концом b интервала изменения СВ X .

Плотность вероятности равна:

$$f(x) = \begin{cases} 0 & \text{при } x \leq a; \\ 1/(b-a) & \text{при } a < x \leq b; \\ 0 & \text{при } x > b. \end{cases} \quad (2.6)$$

Интегральная функция распределения определяется равенством:

$$F(x) = \begin{cases} 0 & \text{при } x \leq a; \\ (x-a)/(b-a) & \text{при } a < x \leq b; \\ 1 & \text{при } x > b. \end{cases} \quad (2.7)$$

МО, дисперсия и СКО соответственно равны:

$$m_x = \frac{a+b}{2}; \quad (2.8)$$

$$D_x = \frac{(a-b)^2}{12}; \quad (2.9)$$

$$\sigma_x = \frac{b-a}{2\sqrt{3}}. \quad (2.10)$$

Медиана совпадает с МО. Закон равномерной плоскости моды не имеет, так как все значения плотности вероятности равны между собой.

Закон равномерной плотности играет важную роль при моделировании искусственных гидрологических рядов, поскольку значения любой случайной величины можно получить преобразованием значений СВ, равномерно распределенной на отрезке $[0, 1]$.

2.3. Логарифмически нормальное распределение

Многие гидрологические переменные, которые можно рассматривать как случайные величины, обладают правой, т.е. положительной асимметрией. Это объясняется тем что гидрологические величины обычно превышают нуль, либо некоторый другой нижний предел, но теоретически не ограничены верхним пределом. В таких случаях распределение случайной величины не соответствует нормальному закону.

Однако часто асимметричное распределение можно привести к нормальному путем замены переменной на ее логарифм.

Неотрицательная СВ X называется распределенной логарифмически нормально, если ее логарифм $Z = \ln(X)$ распределен по нормальному закону (рис. 2.2).

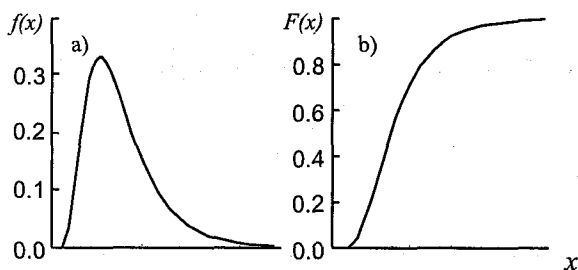


Рис. 2.2. Дифференциальная (а) и интегральная (б) функции распределения логнормального закона при $m_z = 1$ и $\sigma_z = 0,5$.

Интегральная функция распределения для СВ X в этом случае имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u \exp\left(-\frac{s^2}{2}\right) ds & \text{при } x > 0. \end{cases} \quad (2.11)$$

где $u = (z - m_z)/\sigma_z$; $z = \ln(x)$; s — переменная интегрирования.

Дифференцируя выражение (2.11), получим функцию плотности вероятности логнормального распределения (см. рис. 2.2):

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ \frac{1}{\sigma_z x} f(u) = \frac{1}{\sigma_z x \sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) & \text{при } x > 0. \end{cases} \quad (2.12)$$

Таким образом, данное распределение определяется двумя параметрами: m_z и σ_z . Величина m_z представляет собой МО случайной величины Z , а σ_z — ее среднеквадратическое отклонение.

Дисперсия СКО и МО случайных величин X и Z связаны соотношениями

$$D_z = \sigma_z^2 = \ln\left(\frac{\sigma_x^2}{m_x^2} + 1\right); \quad (2.13)$$

$$m_z = \ln(m_x) - \frac{\sigma_z^2}{2} = \ln(m_x) - \frac{\ln(\sigma_x^2/m_x^2 + 1)}{2}. \quad (2.14)$$

Если закон распределения СВ X не известен и логнормальное распределение используется для аппроксимации истинного распределения X , то значения m_z и σ_z для СВ Z могут отличаться от m_z и σ_z , полученных по формулам (2.13), (2.14) через m_x и σ_x . При этом расхождение будет тем больше, чем больше отклонение истинного распределения СВ X от логнормального.

Поскольку в гидрологической практике истинное распределение анализируемой СВ, как правило, не известно, то допускается два варианта расчета:

- 1) оценка m_z и σ_z производится по ряду значений СВ Z ;
- 2) по ряду значений СВ X производится оценка m_x и σ_x , а затем по формулам (2.13), (2.14) определяются m_z и σ_z .

Коэффициент асимметрии логарифмически нормального распределения определяется по формуле

$$C_s = 3C_v + C_v^3, \quad (2.15)$$

где $C_v = \sigma/m_x$. Так как для большинства гидрологических переменных $C_v < 1$, то наиболее хорошего соответствия эмпирических данных с логнормальным распределением можно ожидать при $C_s/C_v = 3 - 3,5$.

Мода и медиана для СВ X , имеющей логнормальное распределение, определяются равенствами:

$$Mo = \exp(m_z - \sigma_z^2); \quad (2.16)$$

$$Me = \exp(m_z). \quad (2.17)$$

Для определения ординат кривой обеспеченности логнормального закона используют таблицы стандартного нормального распределения, но в качестве случайной величины рассматривают $Z = \ln(X)$.

Пример 2.3. Определить максимальный паводочный расход однопроцентной обеспеченности (Q_1) на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя логнормальный закон распределения. Привести два варианта решения.

Решение 1.

1. Проводим преобразование исходного ряда по формуле $z_i = \ln Q_i$.

2. По ряду значений СВ Z приближенно определяем m_z и σ_z (методика приближенной оценки параметров распределения излагается в главе 3). Получаем: $m_z \approx 3,87$; $\sigma_z \approx 0,64$.

3. Так как таблица стандартного нормального распределения (см. табл. 2.1) приводится для интегральной функции распределения, заменяем обеспеченность на вероятность непревышения: $F(t) = 100 - P(t) = 100 - 1 = 99\%$.

4. По табл. 2.1 определяем 99 %-ный квантиль нормированной нормально распределенной СВ t : $t'_{99} = 2,33$.

5. По формуле (2.5) определяем z_1 : $z_1 = z'_{99} = \sigma_z t'_{99} + m_z = 2,33 \times 0,64 + 3,87 = 5,36$.

6. Так как $z_1 = \ln Q_1$, то $Q_1 = \exp(z_1) = \exp(5,36) = 213 \text{ м}^3/\text{с}$.

Решение 2.

1. Непосредственно по исходному ряду (см. табл. 1.2) приближенно определяем m_Q и σ_Q , получаем: $m_Q \approx 58,44$; $\sigma_Q \approx 37,19$.

2. По формулам (2.13), (2.14) вычисляем m_z и σ_z :

$$\sigma_z^2 = \ln[(37,19)^2 / (58,44)^2 + 1] = 0,34; \quad \sigma_z = 0,58;$$

$$m_z = \ln(58,44) - 0,34/2 = 3,90.$$

3. Дальнейший расчет ведется аналогично первому варианту:

$$t'_{99} = 2,33; \quad z_1 = z'_{99} = 2,33 \cdot 0,58 + 3,90 = 5,25;$$

$$Q_1 = \exp(5,25) = 191 \text{ м}^3/\text{с}.$$

Заметим, что полученные по формулам (2.13), (2.14) значения m_z и σ_z несколько отличаются от значений этих параметров, полученных по первому варианту, а расчетные расходы (213 и 191) отличаются на 12 %. Это косвенным образом подтверждает, что в данном случае истинное распределение может быть аппроксимировано логнормальным, однако такая аппроксимация не будет идеальной.

В практике гидрологических исследований логнормальное распределение наиболее часто используется при расчетах максимального паводочного стока. Такая практика имеет место в Болгарии, Великобритании, России, США, Японии и ряде других стран.

Наряду со стандартным логнормальным распределением используются различные его модификации. В частности, Г.А. Алексеев (1960 г.) предложил вместо преобразования $Z = \ln(X)$ использовать преобразование $Z = \ln(X - a)$, где a — дополнительный (третий) параметр.

Полученное на основе этого преобразования распределение иногда называют трехпараметрическим логнормальным распределением. В отличие от стандартного логнормального распределения, которое имеет нулевой нижний предел, это распределение имеет нижний предел, равный a .

Параметр a связан с коэффициентом асимметрии соотношением

$$C_{s,x} = \frac{3C_{v,x}}{1 - k_0} + \left(\frac{C_{v,x}}{1 - k_0} \right)^3, \quad (2.18)$$

где k_0 — минимальный модульный коэффициент, $k_0 = a/m_x$. При $k_0 = 0$ выражение (2.18) превращается в (2.15).

Распределение является более гибким по сравнению со стандартным, однако возможности его применения в гидрологии ограничены. Действительно, для того чтобы кривая распределения не уходила в область отрицательных значений, необходимо, чтобы значение k_0 удовлетворяло условию $k_0 \geq 0$. В то же время минимальный модульный коэффициент всегда меньше единицы. Окончательно получаем $0 \leq k_0 < 1$. Таким образом, как видно из (2.18), логнормальное распределение с положительным нижним пределом ($a > 0$) будет иметь более высокую асимметрию, чем обычное логнормальное, т.е.

$$C_s > 3C_v + C_v^3.$$

С учетом сказанного это распределение можно рекомендовать только для случайных величин с большой асимметрией $C_s > 3,5C_v$, но такие ситуации в гидрологической практике довольно редки.

Пример 2.4. Определить максимальный расход дождевого паводка однопроцентной обеспеченности Q_1 на р. Луге в створе ст. Толмачево (см. табл.1.2), используя трехпараметрическое логнормальное распределение. Расчет вести по второму варианту.

Решение.

1. Непосредственно по исходному ряду приближенно определяем МО, СКО и коэффициент вариации: $m_Q \approx 58,4$; $\sigma_Q \approx 37,2$; $C_v \approx 0,64$.

2. Поскольку расчёт коэффициента асимметрии по эмпирическим данным производится с большой ошибкой, принимаем соотношение C_s/C_v осредненным по группе рек исследуемого района. В данном случае $C_s/C_v = 2,5$. Следовательно, принимаем $C_{s,Q} \approx C_{v,Q} \cdot 2,5 \approx 0,64 \cdot 2,5 \approx 1,60$.

3. В зависимости от $C_{s,Q}$ по формуле (2.18) находим минимальный модульный коэффициент k_0 . Решаем задачу методом последовательных приближений (табл. 2.2), т.е., задаваясь различными значениями k_0 при $C_v = 0,64$, вычисляем теоретическое значение коэффициента асимметрии $C_{s,T}$. И в качестве расчётного принимаем такое значение k_0 , при котором $C_{s,T} = C_{s,Q}$. Как видно из табл. 2.2, $k_0 = -0,3$.

Таблица 2.2

Зависимость $k_0 = f(C_{s,T})$ при $C_{v,Q} = 0,64$

$k_0, \dots, \dots$	-0,5	-0,4	-0,3	-0,2	-0,1	0,0	0,1
$C_{s,T}, \dots, \dots$	1,36	1,47	1,60	1,75	1,94	2,18	2,49

4. Зная k_0 , находим параметр a : $a = k_0 \bar{Q} = 58,4(-0,3) = -17,5$.

Замечание. Знак минус указывает на то, что при больших обеспеченностях кривая будет уходить в область отрицательных значений. Этого и следовало ожидать, так как асимметрия ряда не велика и не выполняется условие $C_{s,Q} > 3C_{v,Q} + C_{v,Q}^3$.

5. Для СВ U , которая связана с Q соотношением $u_i = Q_i - a$, рассчитываем числовые характеристики:

$$\bar{U} = \bar{Q} - a = 58,4 - (-17,5) = 58,4 + 17,5 = 75,9; \quad \sigma_u = \sigma_x = 37,2.$$

6. По формулам (2.13) и (2.14) вычисляем m_z и σ_z , где $Z = -\ln(U)$: $\sigma_z^2 = \ln[(37,2)^2 / (75,9)^2 + 1] = 0,22$; $\sigma_z = 0,47$; $m_z = \ln(75,9) - (0,22/2) = 4,22$.

7. По табл. 2.1 определяем 99 %-ный квантиль нормированной нормально распределённой СВ t : $t'_{99} = 2,33$.

8. По формуле (2.5) определяем

$$z_1 = z'_{99} = \sigma_z t'_{99} + m_z = 2,33 \cdot 0,47 + 4,22 = 5,32.$$

9. Так как $z_1 = \ln(Q_1 + a)$, то $Q_1 = \exp(z_1) - a = \exp(5,32) - 17,5 = 187 \text{ м}^3/\text{с}$.

Ещё одна модификация логнормального распределения будет рассмотрена в п. 2.7.

2.4. Закон распределения крайних членов выборки (распределение Гумбеля)

Закон распределения Гумбеля¹ разрабатывался для случая, когда в качестве случайной величины рассматриваются экстремальные характеристики гидрометеорологического режима, например максимальный в году срочный расход, минимальный в году среднесуточный расход, максимальный в году суточный слой осадков и т.д. Функция обеспеченностей Гумбеля определяется выражением

¹ В литературе встречается также "распределение Гамбела".

$$P(x) = P\{X \geq x\} = 1 - \exp[-\exp(-y)], \quad (2.19)$$

где y – безразмерная величина, связанная с x выражением

$$y = \alpha(x - q), \quad (2.20)$$

где q – мода случайной величины X . Параметр q определяется в зависимости от среднего значения и СКО исходного ряда по формуле

$$q = \bar{x} - 0,45\sigma_x, \quad (2.21)$$

а α можно выразить как

$$\alpha = 1,28 / \sigma_x. \quad (2.22)$$

На практике вместо уравнения (2.20) используется уравнение, полученное при решении (2.20) относительно x :

$$x_p = q + (1/\alpha)y_p. \quad (2.23)$$

Значения y_p можно получить из выражения (2.19) после двукратного логарифмирования

$$y_p = -\ln \{-\ln [(100 - p)/100]\}, \quad (2.24)$$

где p – расчетная обеспеченность в процентах. В табл. 2.3 приводятся значения y_p для основных опорных обеспеченностей.

Таблица 2.3

Значения y_p для различных обеспеченностей

$p \%$	y_p	$p \%$	y_p	$p \%$	y_p
0,01	9,09	10	2,25	80	-0,48
0,1	6,89	20	1,50	90	-0,83
0,5	5,29	30	1,03	95	-1,10
1	4,60	50	0,37	99,0	-1,53
5	2,97	70	-0,19	99,9	-1,93

Уравнения (2.21) и (2.22) связывают параметры q и α с $\bar{X}$ и σ_x при $n \rightarrow \infty$. Для конечных выборок из n членов Гумбель предложил следующие формулы:

$$(1/\alpha) = \sigma_x / \sigma_y; \quad (2.25)$$

$$q = \bar{x} - \bar{y}(1/\alpha). \quad (2.26)$$

Параметры $\bar{y}$ и σ_y определяются в зависимости от длины анализируемого ряда (табл. 2.4).

С учетом формулы (2.20) выражение функции плотности вероятности распределения Гумбеля имеет вид

$$f(x) = \alpha \exp \{-\alpha(x - q) - \exp[-\alpha(x - q)]\}. \quad (2.27)$$

Как видно из (2.27), для распределения Гумбеля область возможных значений СВ X — интервал $(-\infty, +\infty)$. Распределение является двухпараметрическим, т.е. однозначно определяется параметрами $\bar{x}$ и σ_x .

Таблица 2.4

Средние значения параметров $\bar{y}$ и σ_y при различном числе членов ряда n (по Гумбелю)

n	$\bar{y}$	σ_y	n	$\bar{y}$	σ_y
20	0,524	1,063	50	0,548	1,161
22	0,527	1,076	52	0,549	1,164
24	0,530	1,086	54	0,550	1,167
26	0,532	1,096	56	0,551	1,170
28	0,534	1,105	58	0,552	1,172
30	0,536	1,112	60	0,552	1,175
32	0,538	1,119	65	0,554	1,180
34	0,540	1,126	70	0,555	1,185
36	0,541	1,131	75	0,556	1,190
38	0,542	1,136	80	0,557	1,194
40	0,544	1,141	85	0,558	1,197
42	0,545	1,146	90	0,559	1,201
44	0,546	1,150	95	0,559	1,204
46	0,547	1,154	100	0,560	1,206
48	0,548	1,157	∞	0,577	1,282

Теоретические исследования, проведенные Г.А. Алексеевым, показали, что при $n \rightarrow \infty$ третий центральный момент $\mu_3 \approx 2,404$. Следовательно,

$$C_{s,y} = C_{s,x} = \mu_3 / \sigma_y^3 = 2,404 / (1,282)^3 = 1,14. \quad (2.28)$$

Таким образом, коэффициент асимметрии для распределения Гумбеля является постоянным.

Это обстоятельство и то что график функции $P(x)$ при больших значениях обеспеченности уходит в область отрицательных значений, существенно ограничивает возможность применения распределения Гумбеля в практике гидрологических расчетов.

В настоящее время распределение Гумбеля используется главным образом при расчетах максимальных расходов дождевых паводков. Такая практика, в частности, существует в Нидерландах, США, Франции, Японии и ряде других стран.

Пример 2.5. Определить максимальный паводочный расход 1%-ный обеспеченности $Q_{1\%}$ на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя закон распределения Гумбеля.

Решение.

1. По ряду значений СВ Q приближенно определяем m_Q и σ_Q :

$$m_Q \approx \bar{x} = 58,4; \quad \sigma_Q \approx 37,2.$$

2. По формуле (2.24) или по табл. 2.3 определяем $y_{1\%} = 4,60$.

3. В зависимости от длины ряда n по табл. 2.4 определяем $\bar{y}$ и σ_y . Так как в данном случае $n = 32$, то $\bar{y} = 0,54$; $\sigma_y = 1,12$.

4. По формулам (2.25) и (2.26) вычисляем параметры $(1/\alpha)$ и q :

$$1/\alpha = \sigma_Q / \sigma_y = 37,2 / 1,12 = 33,2; \quad q = 58,4 - 0,54 \cdot 33,2 = 40,5.$$

5. По формуле (2.23) вычисляем расчетный расход воды:

$$Q_{1\%} = 40,5 + 33,2 \cdot y_{1\%} = 40,5 + 33,2 \cdot 4,60 = 193 \text{ м}^3/\text{с}.$$

2.5. Распределение Пирсона III типа (непрерывное биномиальное распределение)

Для аппроксимации закона распределения гидрометеорологических величин с одномодальной асимметричной функцией плотности вероятности А. Фостер предложил использовать общес дифференциальное уравнение Пирсона:

$$\frac{dy}{dx} = \frac{y(z+a)}{b_0 + b_1 z + b_2 z^2}, \quad (2.29)$$

где Z – случайная величина, связанная с исходной СВ X соотношением $z = x/m_x - 1 = k - 1$; k – модульный коэффициент; y – ордината функции плотности вероятности СВ Z ; a – расстояние от центра распределения (m_k) до моды (МО); b_0, b_1, b_2 – параметры, изменяя которые можно получить различные типы кривых распределения (рис. 2.3).

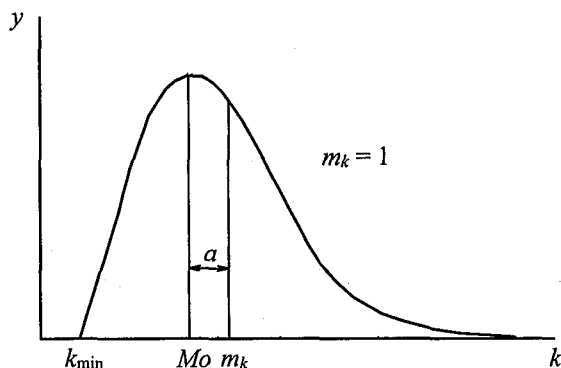


Рис. 2.3. Схема построения функции плотности вероятности Пирсона III типа для случая $C_S/C_v > 2$ ($k_{\min} > 0$).

В практике гидрологических расчетов наибольшее распространение получила кривая *Пирсона III типа*, для которой $b_2 = 0$, и уравнение (2.29) принимает вид

$$\frac{dy}{dz} = \frac{y(z+a)}{b_0 + b_1 z}. \quad (2.30)$$

Данному уравнению соответствует целое семейство функций. Для того чтобы $y(z)$ действительно являлась искомой функцией плотности вероятности, необходимо ввести ряд дополнительных условий:

1. $\int_{z_{\min}}^{\infty} y(z) dz = 1$;
2. $y(z) = 0$ при $z = z_{\min}$;
3. $y(z) = 0$ при $z \rightarrow \infty$.

Интегрируя (2.30) и переходя от СВ Z к модульным коэффициентам, после ряда преобразований можно получить выражение для функции плотности вероятности:

$$y(k) = \begin{cases} 0 & \text{при } k \leq k_{\min} \\ \frac{\beta^\alpha (k - k_{\min})^{\alpha-1}}{\Gamma(\alpha)} \exp[-\beta(k - k_{\min})] & \text{при } k > k_{\min} \end{cases}, \quad (2.31)$$

где $\Gamma(\cdot)$ – гамма-функция; α и β – параметры распределения, связанные с C_v и C_s случайной величины X соотношениями:

$$\alpha = (2/C_s)^2; \quad (3.32)$$

$$\beta = 2/(C_s \cdot C_v). \quad (2.33)$$

Заметим, что коэффициенты вариации и асимметрии не изменяются при замене значений СВ X на модульные коэффициенты.

Минимальное значение модульного коэффициента в выражении (2.31) определяется формулой

$$k_{\min} = 1 - 2C_v/C_s. \quad (2.34)$$

Из (2.34) следует, что

$$C_s = 2C_v \text{ при } k_{\min} = 0,$$

$$C_s > 2C_v \text{ при } k_{\min} > 0,$$

$$C_s < 2C_v \text{ при } k_{\min} < 0.$$

Таким образом, дифференциальная кривая распределения Пирсона III типа при $C_s = 2C_v$ начинается с нуля; при $C_s > 2C_v$ с некоторого положительного числа и, наконец, при $C_s < 2C_v$ уходит в область отрицательных чисел.

Мода функции (2.31) при $C_s < 2$ выражается формулой

$$Mo = 1 - (C_s C_v)/2. \quad (2.35)$$

При больших значениях C_s кривая Пирсона III типа имеет весьма специфический вид. Из (2.32) следует, что при $C_s = 2$ $\alpha = 1$ и функция (2.31) превращается в экспоненту, а при $C_s > 2$ $\alpha < 1$ и функция (2.31) превращается в гиперболу.

Зная C_v и C_s , можно получить численные значения параметров $k_{\min}$, α , β и записать выражение для вычисления обеспеченностей модульных коэффициентов

$$P(k) = \int_k^{\infty} y(s) ds = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_k^{\infty} (s - k_{\min})^{\alpha-1} \exp[-\beta(s - k_{\min})] ds, \quad (2.36)$$

где s — переменная интегрирования.

В случае $C_s = 2C_v$ имеем: $k_{\min} = 0$, $\alpha = 1/C_v^2$, $\beta = 1/C_v^2$ и уравнения существенно упрощаются:

$$y(k) = \frac{\alpha^{\alpha}}{\Gamma(\alpha)} k^{\alpha-1} \exp(-\alpha k); \quad (2.37)$$

$$P(k) = \frac{\alpha^{\alpha}}{\Gamma(\alpha)} \int_k^{\infty} s^{\alpha-1} \exp(-\alpha s) ds. \quad (2.38)$$

Выражение (2.37) называют двухпараметрическим гамма-распределением, или Γ -распределением.

Кривая обеспеченности Пирсона III типа (2.36) в общем случае является трехпараметрической и однозначно определяется параметрами C_v и C_s ; третий параметр m_x необходимо знать для перехода от модульных коэффициентов к значениям СВ X .

Кривая имеет нижний предел $k_{\min}$ и не ограничена верхним пределом. При $C_s \rightarrow 0$ кривая Пирсона III типа стремится к нормальному распределению.

Численное решение уравнений (2.36), (2.38) — довольно трудоемкая задача, поэтому ординаты кривой обеспеченности Пирсона III типа обычно представляют в виде таблиц (приложение 1).

В таблицах даны значения нормированной случайной величины в зависимости от обеспеченности и коэффициента асимметрии. Переход от нормированных ординат к значениям искомой величины производится по формулам (2.5), (2.5а).

Кривая распределения Пирсона III типа имеет особое значение для гидрологов как первая кривая, широко внедренная в практику гидрологических расчетов. Впервые таблицы распределения Пирсона III типа были опубликованы А. Фостером в 1924 г.

В России приоритет внедрения этой кривой принадлежит Д.Л. Соколовскому (1930 г.). Таблицы А. Фостера были уточнены и расширены С. И. Рыбкиным (1938 г.), а затем и другими авторами.

В настоящее время кривая Пирсона остается одной из наиболее популярных в мировой практике гидрологических расчетов.

Наряду с обычной кривой Пирсона III типа в США и ряде других стран используется логарифмическая кривая Пирсона III типа. В этом случае считается, что распределению Пирсона подчиняется не исходная СВ X , а ее логарифм $\ln(X)$. Такой подход позволяет вывести кривую из отрицательной области при соотношении $C_s/C_v < 2$.

Пример 2.6. Определить максимальный паводочный расход 1%-ной обеспеченности ($Q_{1\%}$) на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя кривую обеспеченностей Пирсона III типа.

Решение.

1. По ряду значений СВ Q приближенно определяем $m_x \approx \bar{Q} = 58,4$; $\sigma_Q \approx 37,2$; $C_v \approx 0,64$.

2. Поскольку по эмпирическим данным значение C_s определяется с большой погрешностью, примем соотношение C_s/C_v , осредненным для всего рассматриваемого района. Исследования, проведенные на кафедре гидрологии суши, показали, что для паводочного стока рек Лужского бассейна $C_s/C_v = 2,5$. Теперь, зная C_v и C_s/C_v , определяем коэффициент асимметрии: $C_s = 2,5 \cdot 0,64 = 1,60$.

3. По приложению 1 находим нормированную ординату кривой обеспеченности Пирсона III типа: $t_{1\%} = 3,39$.

4. По формуле (2.5) переходим от $t_{1\%}$ к $Q_{1\%}$:

$$Q_{1\%} = t_{1\%} \sigma_Q + \bar{Q} = 3,39 \cdot 37,2 + 58,4 = 185 \text{ м}^3/\text{с}.$$

Можно действовать и иначе. Сначала по формуле (2.5а) получаем модульный коэффициент $k_{1\%} = t_{1\%} C_v + 1 = 3,39 \cdot 0,64 + 1 = 3,17$. Затем, умножая его на $\bar{x}$, переходим к расходу:

$$Q_{1\%} = k_{1\%} \bar{x} = 3,17 \cdot 58,4 = 185 \text{ м}^3/\text{с}.$$

2.6. Распределение Крицкого – Менкеля (трехпараметрическое гамма-распределение)

Кривая распределения вероятностей Пирсона III типа, широко используемая в практике гидрологических расчетов, обладает, как было показано выше, одним существенным недостатком: при $C_s < 2C_v$, она уходит в область отрицательных значений.

В связи с этим неоднократно предпринимались попытки получения кривой для аппроксимации закона распределения гидрологических величин с диапазоном изменения $(0 \leq x < \infty)$, и пригодной для всех реально встречающихся на практике соотношений C_s и C_v .

Удачное решение рассматриваемой задачи в середине сороковых годов предложили С.Н. Крицкий и М.Ф. Менкель. В качестве исходной модели они приняли кривую Пирсона III типа при $C_s = 2C_v$ (двухпараметрическое Γ -распределение):

$$G(z) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \int_0^z s^{\alpha-1} e^{-\alpha s} ds, \quad (2.39)$$

где $G(z)$ – интегральная функция Γ -распределения; s – переменная интегрирования; $\bar{z} = 1$; $C_s = 2C_v$; $\alpha = 1/C_{v,z}$.

Авторы трансформировали аргумент z в новую переменную

$$k = az^b, \quad (2.40)$$

где a и b – параметры. При этом предполагалось, что МО новой переменной равно единице, т.е.

$$M[k] = M[az^b] = 1. \quad (2.41)$$

Подставляя в выражение (2.39) $z = (k/a)^{1/b}$ и имея в виду, что $f(k) = dG(k)/d(k)$, Крицкий и Менкель получили новое распределение с плотностью вероятности вида

$$f(k) = \frac{\alpha^\alpha}{a^{\alpha/b} \Gamma(\alpha) b} e^{-\alpha(k/a)^{1/b}} k^{(\alpha/b)-1}. \quad (2.42)$$

Начальный момент i -го порядка этого распределения связан с параметрами α , a и b соотношением

$$\mu_i = \frac{\Gamma(\alpha + ib)a^i}{\Gamma(\alpha)\alpha^{ib}}. \quad (2.43)$$

Из выражения (2.43), следует, что

$$M[k] = \mu_1 = \frac{\Gamma(\alpha + b)a}{\Gamma(\alpha)\alpha^b}. \quad (2.44)$$

А так как по условию $M[k] = 1$, то, приравняв (2.44) единице, получаем выражение для параметра a :

$$a = \frac{\Gamma(\alpha)\alpha^b}{\Gamma(\alpha + b)}. \quad (2.45)$$

Подставляя (2.45) в (2.42), можно получить выражение кривой плотности вероятности Крицкого – Менкеля, записанное через Γ -функцию. Это выражение определяется двумя параметрами α и b , которые с учетом (2.43) могут быть выражены через второй и третий начальные моменты. В свою очередь, μ_2 и μ_3 можно выразить через C_v и C_s (см. п. 1.4).

Таким образом, полученное распределение будет двухпараметрическим, однако, чтобы в дальнейшем перейти от модульных коэффициентов k к искомой СВ X , необходимо знать третий параметр – m_x . С учетом сказанного распределение Крицкого–Менкеля называют также трехпараметрическим Γ -распределением.

Поскольку функцию обеспеченностей Крицкого–Менкеля нельзя выразить через элементарные функции, ее ординаты представляют в виде таблиц (приложение 2). Таблицы составлены в модульных коэффициентах и позволяют определить значение $k_{p\%}$ в зависимости от (C_s/C_v) , C_v и расчетной обеспеченности p (%). Впервые такие таблицы были опубликованы Д.В. Коренистовым (1948 г.). Наиболее полные таблицы кривой обеспеченности Крицкого – Менкеля даны в [16]. Они охватывают диапазон изменения C_s/C_v от 0,5 до 6,0 и достаточно широкий диапазон C_v .

Действующие в настоящее время в России нормативные документы рекомендуют кривую Крицкого – Менкеля в качестве стандартной кривой при проведении гидрологических расчетов.

В заключение еще раз перечислим основные особенности данной кривой.

1. Кривая плотности вероятности является одномодальной с положительной асимметрией.

2. Нижним пределом кривой всегда является нуль.

3. Кривая Крицкого – Менкеля не ограничена верхним пределом.

4. При $C_s = 2C_v$, кривая превращается в двухпараметрическое Γ -распределение, т.е. совпадает с кривой Пирсона III типа.

Пример 2.7. Определить максимальный расход дождевого паводка 1%-ный обеспеченности $Q_{1\%}$ на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя кривую обеспеченностей Крицкого – Менкеля.

Решение.

1. По эмпирическим данным определяем:

$$m_x \approx \bar{Q} = 58,4; C_v \approx 0,64; C_s / C_v = 2,5 \text{ (см. пример 2.5).}$$

2. По таблицам (приложение 2) определяем модульный коэффициент 1%-ный обеспеченности. При этом, если таблицы составлены только для целых соотношений C_s/C_v , необходимо провести интерполяцию.

Для $C_s/C_v = 2$ при $C_v = 0,64$ и $p = 1\%$ имеем $k_{1\%} = 3,05$.

Для $C_s/C_v = 3$ при $C_v = 0,64$ и $p = 1\%$ имеем $k_{1\%} = 3,24$.

Следовательно, для $C_s/C_v = 2,5$ $k_{1\%} = 3,15$.

3. Умножая модульный коэффициент на $\bar{Q}$, получаем расчетный расход $Q_{1\%} = k_{1\%} \cdot \bar{Q} = 3,15 \cdot 58,4 = 184 \text{ м}^3/\text{с}$.

2.7. Распределение Джонсона

В начале 70-х годов XX в. грузинские гидрологи Г.Г. Сванидзе и Г.Л. Григолия предложили использовать для аппроксимации закона распределения гидрологических величин кривую распределения, ограниченную с двух сторон. В качестве такой кривой они рекомендовали S_b -распределение Джонсона [13, 21].

В рамках этого подхода предполагается, что если исходную СВ X преобразовать по формуле

$$z = \ln \left(\frac{x-a}{b-x} \right), \quad (2.46)$$

то новая СВ Z будет иметь нормальное распределение, где a и b соответственно нижний и верхний пределы СВ X .

Плотность распределения Джонсона имеет вид:

$$f(x) = \frac{b-a}{\sigma_z \sqrt{2\pi}(x-a)(b-x)} \exp \left\{ -\frac{1}{2\sigma_z^2} \left[\ln \left(\frac{x-a}{b-x} \right) - m_z \right]^2 \right\}. \quad (2.47)$$

Как видно из (2.47), это распределение является четырехпараметрическим и помимо параметров a и b содержит еще два параметра: m_z – математическое ожидание СВ X и σ_z – СКО СВ Z .

При расчете ординат кривой обеспеченности Джонсона используют таблицы нормированной нормально распределенной СВ t (см. табл. 2.1), по аналогии с тем, как это делалось в первом варианте для логарифмически нормального распределения (см. пример 2.3), с той лишь разницей, что для расчета параметров m_z и σ_z (при известных a и b) исходный ряд преобразуется по формуле (2.46). Если все четыре параметра известны, то по табл. 2.1 определяется нормированная ордината t_p , затем вычисляется z_p и с учетом (2.46) определяется значение x_p :

$$x_p = \frac{b \exp(z_p) + a}{\exp(z_p) + 1}. \quad (2.48)$$

В гидрологической практике ситуации, когда верхний и нижний пределы известны заранее, довольно редки. В качестве примера здесь можно привести кривую обеспеченности коэффициентов паводочного стока, где $a = 0$, $b = 1$. Однако гораздо чаще верхний и нижний пределы СВ не известны, поэтому параметры и приходится определять методом последовательных приближений на основе наилучшего соответствия эмпирической и аналитической кривых обеспеченностей.¹

¹ Об эмпирических кривых обеспеченности см. п. 3.1. Практические приемы определения параметров a и b описаны в п. 3.1.5.

Пример 2.8. Определить максимальный расход дождевого паводка 1%-ный обеспеченности $Q_{1\%}$ на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя кривую обеспеченности Джонсона, если параметры a и b известны (определены методом последовательных приближений: $a = 11$, $b = 490$).

Решение.

1. Производим преобразование исходного ряда по формуле (2.46): $z_i = \ln [(Q_i - 11)/(490 - Q_i)]$.

2. По ряду значений СВ Z приближенно определяем m_z и σ_z (методика оценки параметров распределения по эмпирическим данным излагается в главе 3): $m_z = -2,54$; $\sigma_z = 0,95$.

3. Так как таблица стандартного нормального распределения (см. табл. 2.1) приводится для интегральной функции распределения, заменяем обеспеченность на вероятность непревышения: $F(t) = 100 - P(t) = 100 - 1 = 99\%$.

4. По табл. 2.1 определяем 99 %-ный квантиль для СВ t : $t'_{99} = 2,33$.

5. По формуле (2.5) определяем

$$z_1 = z'_{99} = \sigma_z t'_{99} + m_z = 0,95 \cdot 2,33 - 2,54 = -0,33.$$

6. По формуле (2.47) определяем $Q_{1\%}$:

$$Q_1 = (490 e^{-0,33} + 11)/(e^{-0,33} + 1) = 363/1,72 = 211 \text{ м}^3/\text{с}.$$

В заключение отметим, что, несмотря на всю привлекательность данного распределения (интуитивно мы понимаем, что верхний и нижний пределы существуют), оно не нашло широкого распространения в гидрологической практике, так как очень редко удастся надежно оценить параметры a и b по эмпирическим данным.

В то же время использование кривых, не ограниченных верхним пределом, не противоречит физической сути гидрологических явлений, потому что недостоверно высокие значения СВ (например, расхода) при правильном выборе кривой обеспеченности будут соответствовать очень малой вероятности превышения (менее 0,01 %), т.е. будут практически невероятными.

По мнению авторов, использование кривой Джонсона может быть эффективным лишь в том случае, когда хотя бы один из параметров (a или b) известен априори или с достаточной точностью может быть назначен, исходя из физических соображений.

2.8. Графическое представление функций распределения на клетчатке вероятности

В тех случаях, когда закон распределения исследуемой гидрологической величины не известен, для его аппроксимации используется одно из перечисленных выше распределений.

Выбранный закон распределения задается функцией обеспеченности, которую в гидрологии принято называть аналитической кривой обеспеченности.

Аппроксимация считается достаточно надежной, если аналитическая кривая хорошо согласуется с эмпирическими данными.¹

Для удобства проведения такого анализа используются специальные клетчатки вероятности. На клетчатке вероятности по оси абсцисс откладываются значения обеспеченности (в %). По оси ординат могут откладываться либо значения исследуемой СВ, либо ее модульные коэффициенты (1.27), либо ее нормированные значения (1.28).

Масштабы по осям на клетчатке вероятности трансформируются таким образом, чтобы график теоретической кривой обеспеченности представлял собой прямую линию. Очевидно, что клетчатки вероятности могут быть построены для распределений, имеющих только два изменяемых параметра: обычно используются среднее значение и СКО (или коэффициент вариации). Любые дополнительные параметры, такие как коэффициент асимметрии, должны быть постоянными. Иными словами, для трехпараметрических распределений необходимо иметь спрямляющую клетчатку для каждого соотношения C_s/C_v .

Наиболее распространенной является клетчатка, спрямляющая нормальный закон распределения. Напомним, что нормальный закон является двухпараметрическим и для него $C_s = 0$.

Нормальная клетчатка может также использоваться для изображения кривых обеспеченности с умеренной асимметрией. В этом случае теоретическая кривая полностью не выпрямляется. Кривые с положительной асимметрией будут иметь вогнутую форму, а с отрицательной – выпуклую (рис. 2.4).

¹ О технике построения эмпирических кривых обеспеченностей см. п. 3.1.

Для построения клетчаток вероятности можно использовать аналитические и графические схемы. Более простым и наглядным является графический прием. Практические приемы построения спрямляющих клетчаток вероятности для различных функций распределения приводятся в работе [20].

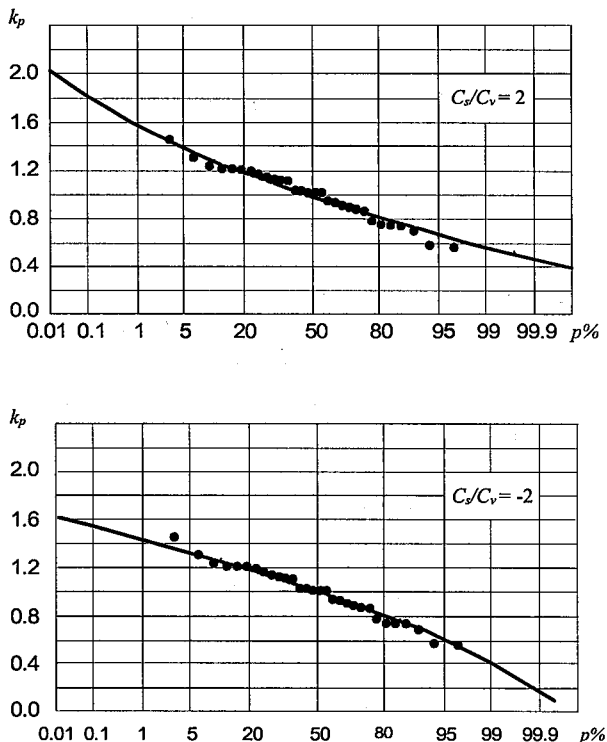


Рис. 2.4. Кривые обеспеченности Пирсона III типа на клетчатке с умеренной асимметричностью, построенные для одного и того же ряда ($n = 30$; $\bar{Q} = 98.5$; $C_v = 0.22$), при различных соотношениях C_s/C_v .

При построении клетчатки, спрямляющей нормальный закон распределения, в качестве исходной принимается кривая обеспеченности в модульных коэффициентах с параметрами: $\bar{k} = 1$, $C_v = 1$, $C_s = 0$. Данная клетчатка имеет равномерный масштаб по оси ординат и неравномерный масштаб по оси абсцисс. Если на такой клетчатке построить график нормальной функции обеспеченности

с другим значением коэффициента вариации, то этот график все равно будет иметь вид прямой линии, но угол наклона изменится. При $C_v < 1$ угол наклона уменьшится, при $C_v > 1$ – увеличится.

В российской гидрологической практике в основном используются нормальные клетчатки (их также называют клетчатками с умеренной асимметричностью) и клетчатки, спрямляющие кривую обеспеченности Крицкого – Менкеля. Образцы этих клетчаток для различных соотношений C_s и C_v даны в приложении к пособию [16].

2.9. Комментарии к разделу

В этом разделе были рассмотрены распределения, наиболее часто используемые в практике гидрологических расчетов. Основные характеристики этих функций приводятся в табл. 2.5.

Таблица 2.5

Основные характеристики функций распределения, используемых в гидрологических расчетах

Тип распределения	Число параметров	Область изменения аргумента	Примечание
Равномерное	2	$[a, b]$	Прямоугольное
Нормальное	2	$(-\infty, +\infty)$	Симметричное
Логнормальное...	2	$[0, +\infty)$	Асимметричное
Трехпараметрическое логнормальное	3	$[a, +\infty)$	Асимметричное, $a \geq 0$ при $C_s \geq 3C_v + C_v^3$
Гумбеля	2	$(-\infty, +\infty)$	Асимметричное
Пирсона III типа	3	$[a, +\infty)$	Асимметричное, $a \geq 0$ при $(C_s / C_v) \geq 2$
Крицкого–Менкеля	3	$[0, +\infty)$	Асимметричное
S_b -Джонсона	4	$[a, b]$	Асимметричное, $a \leq x_{\min}; b > x_{\max}$

Для аппроксимации законов распределения конкретных гидрологических величин выбор типа функции распределения следует производить с учетом области изменения ее аргумента. Так, для заведомо положительных величин (расход воды, слой осадков и т.д.) наиболее подходящими будут кривые логнормального распределения Крицкого–Менкеля и Пирсона III типа, имеющие нижний

предел, но не ограниченные сверху. В то же время для температуры воздуха или воды в большей степени подходят кривые с диапазоном изменения аргумента $(-\infty, +\infty)$ и т.д.

Помимо перечисленных распределений в российской гидрологической литературе [17, 20 и др.] есть ссылки на распределения Р.Д. Гудрича (1927 г.) и Г.Н. Бровковича (1941 г.), однако, в настоящее время эти распределения практически не используются.

Из современных разработок можно отметить H -распределение А.В. Ежова [18]. В основу вывода этого распределения были заложены следующие принципы:

- 1) максимально возможная гибкость функции распределения при соблюдении некоторых естественных ограничений.
- 2) максимальная согласованность состава ограничений с выбранной функцией распределения.

С позиций теории информации, сказанное можно сформулировать в виде критерия максимума условной энтропии искомого распределения при данных ограничениях.

Полученная А.В. Ежовым на основе этого подхода функция распределения является трехпараметрической с областью изменения аргумента $[0, +\infty)$. Распределение обладает рядом положительных свойств. В частности, для него совпадают оценки параметров для метода моментов и метода наибольшего правдоподобия.¹ Однако, как показали исследования, проведенные на кафедре гидрологии суши РГГМУ, H -распределение накладывает весьма жесткие ограничения на параметры распределения. Необходимо чтобы функция плотности вероятности этого распределения была одномодальной и при этом удовлетворяла условию $f(0) = 0$, требуется, чтобы C_s и C_s/C_v удовлетворяли системе ограничений

$$\begin{cases} 0 < C_v \leq 1,0; \\ 0 < (C_s/C_v) \leq 2; \\ (C_s/C_v) \geq (2,7C_v - 0,7). \end{cases}$$

Кроме того, его ординаты (приложение 3) в интервале реально используемых обеспеченностей весьма незначительно отличаются от ординат кривой Крицкого – Менкеля при тех же значениях C_s и C_v .

¹ Об оценках параметров распределения см. п. 3.2.

Все перечисленные выше распределения являются непрерывными. Дискретные распределения в гидрологии используются весьма редко. Здесь можно назвать распределение Пуассона и дискретное биномиальное распределение.

Следует отметить, что распределение Пирсона III типа можно получить из дискретного биномиального распределения на основе предельного перехода, устремив шаг дискретности к нулю и при $n \rightarrow \infty$ (n – число опытов). Поэтому распределение Пирсона III типа называют также непрерывным биномиальным распределением.

В заключение приведем сравнительный анализ рассмотренных в этом разделе функций распределения. В табл. 2.6 представлены максимальные паводочные расходы воды различной обеспеченности, полученные с использованием семи типов функций распределения для створа р. Луга – ст. Толмачево.

Как видно из табл. 2.6, выбор типа распределения оказывает ощутимое влияние на значение расчетного расхода. При одной и той же обеспеченности расходы могут существенно различаться. Причем надо иметь в виду, что наиболее заметным это различие становится при больших значениях коэффициента вариации ($C_v > 0,5$).

Таблица 2.6

Максимальные расходы дождевых паводков различной обеспеченности на р. Луге – ст. Толмачево

Распределение	Обеспеченность, p %								
	0,1	1	10	25	50	75	90	99	99,9
Нормальное...	172	145	106	83,3	58,4	33,5	10,8	-28,3	-55,1
Логнормальное 1-ный вариант..., 2-ный вариант...,	338 290	213 191	109 104	73,7 73,0	47,9 49,4	31,2 33,4	21,1 23,5	10,8 12,8	6,81 8,42
Трехпараметрическое логнормальное...	267	187	106	75,7	50,5	32,2	19,8	5,26	-1,28
Гумбеля...	269	193	115	82,5	52,3	29,4	12,9	-10,3	-23,6
Пирсона III типа ($C_v/C_v = 2,5$)...	258	185	108	75,5	49,1	31,2	21,6	13,8	12,3
Крицкого – Менкеля ($C_v/C_v = 2,5$)...	266	184	108	77,7	49,9	31,0	19,6	8,94	4,55
S_b -Джонсона ($a = 11, b = 490$)	293	211	112	73,1	46,0	30,2	21,9	15,1	13,1

Сопоставляя кривые, можно отметить, что в интервале обеспеченностей от 10 до 75 % все кривые дают весьма близкие результаты, а кривые Пирсона III типа и Крицкого – Менкеля практически совпадают.

В зоне малых обеспеченностей ($p < 1\%$) расхождение становится более ощутимым и может достигать нескольких десятков процентов; но наибольшее расхождение кривых наблюдается в зоне больших обеспеченностей. Так, кривая Пирсона III типа (при $C_s/C_v = 2.5$) имеет нижний предел, см. формулу (3.34):

$$Q_{\min} = \bar{Q} k_{\min} = \bar{Q}(1 - 2C_v/C_s) = 58,4 \cdot 0,2 = 11,7 \text{ м}^3/\text{с}$$

и при любых значениях обеспеченности не может дать расчетный расход ниже этого значения. Это можно сказать и о кривой Джонсона, для которой нижний предел $b = 11$. В то же время кривая Крицкого – Менкеля имеет нулевой нижний предел и дает расчетный расход значительно ниже, чем две предыдущие кривые.

Нормальная кривая и кривая Гумбеля не ограничены нижним пределом и при больших значениях обеспеченностей вообще уходят в область отрицательных значений. Чтобы избежать разночтений при аппроксимации функций распределения гидрологических величин, в России введен стандарт, закрепленный в нормативных документах. В качестве стандартной кривой обеспеченности рекомендуется использовать кривую Крицкого – Менкеля. Однако при достаточном обосновании допускается применять кривую распределения Пирсона III типа или другие распределения.

3. ПОСТРОЕНИЕ КРИВЫХ ОБЕСПЕЧЕННОСТЕЙ И ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ПО ЭМПИРИЧЕСКИМ ДАННЫМ

При решении многих практических задач функция распределения СВ не может быть определена теоретическим путем. В таких случаях используются результаты наблюдений за СВ, позволяющие (при достаточном их числе и надлежащей обработке) определить с известной степенью достоверности вид функции распределения и оценить ее числовые характеристики. Для решения такого рода задач используются методы математической статистики.

Известный в этой области математик А. Уайльд определяет математическую статистику как совокупность методов, которые дают нам возможность принимать оптимальные решения в условиях неопределенности.

В качестве более строгого определения можно привести следующее: математическая статистика – раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов.

Фундаментальными понятиями статистической теории являются понятия *генеральной совокупности* и *выборки*.

Генеральная совокупность обычно интерпретируется как совокупность всех возможных значений случайной величины, которые в принципе могут иметь место при данных условиях.

Смысл этого понятия состоит в том, что предполагается существование некоторых вполне определенных свойств, неслучайных закономерностей, присущих данной совокупности тех свойств, которые и должны быть выяснены исследователем. Фактически эти свойства являются объективным отображением вероятностных свойств изучаемого объекта, которые могут быть охарактеризованы с помощью соответствующих законов распределения вероятностей или связанных с ними числовых параметров.

Считается, что указанные свойства не изменяются во времени и присущие генеральной совокупности неслучайные закономерности сохраняют постоянным свой характер, т.е. являются устойчивыми.

Выборка – это конечный набор значений случайной величины, полученный в результате наблюдений.

В гидрологии в качестве выборок можно рассматривать ряды наблюдений за среднегодовыми расходами воды, максимальными и минимальными расходами и уровнями воды, среднегодовой, максимальной и минимальной температурой воздуха и т.д.

Число элементов выборки называется ее *объемом*.

Выборка называется *репрезентативной*, если она достаточно полно характеризует генеральную совокупность.

Смысл статистических методов заключается в том, чтобы по выборке ограниченного объема n , т.е. по некоторой части генеральной совокупности, высказать обоснованное суждение о ее свойствах в целом.

Такое суждение можно составить, произведя специальную обработку результатов наблюдений. В частности, можно построить эмпирическую кривую распределения (см. п. 3.1) и вычислить числовые характеристики эмпирического распределения (выборки).

В силу случайности выборки ее числовые характеристики являются случайными величинами, отличаясь этим от достоверных числовых характеристик теоретического распределения, которому подчиняется генеральная совокупность.

Полученные по эмпирическим данным числовые характеристики случайной величины X (такие как m_x^*, σ_x^*, D_x^* и т. д.) принято называть статистическими оценками числовых характеристик (или эмпирическими, или выборочными характеристиками)¹.

При оценивании одного и того же параметра в зависимости от применяемой методики можно получить различные оценки. Для того чтобы выбрать наилучшую из них, необходимо сформулировать некоторые требования к свойствам оценок. Наиболее важными из этих свойств являются состоятельность, несмещенность и эффективность.

Состоятельность. Оценка $G^* = f(x_1, x_2, x_3, \dots, x_n)$ неизвестного параметра G называется состоятельной, если по мере роста числа

¹ Для обозначения статистических оценок будет использоваться символ (*): σ_x^*, D_x^* и т.д.

наблюдений n она стремится по вероятности к оцениваемому значению G , т.е.

$$\lim_{n \rightarrow \infty} \mathbf{P} \{ |G - G^*| < \varepsilon \} = 1, \quad (3.1)$$

где ε — сколь угодно малое число.

Состоятельность оценки гарантирует исследователю увеличение точности оценивания с ростом n и то, что хотя бы в пределе (при $n \rightarrow \infty$) он может получить точное значение G .

Несмещенность. Оценка $G^* = f(x_1, x_2, x_3, \dots, x_n)$ неизвестного параметра G называется несмещенной, если при любом объеме выборки n результат ее осреднения по всем возможным выборкам данного объема приводит к точному (истинному) значению оцениваемого параметра, т.е. $M[G^*] = G$.

Говоря другими словами, несмещенность означает отсутствие систематической погрешности при оценивании параметра.

Если оценка является смещенной, но причина смещенности известна, то ее можно устранить, введя поправку на смещенность в формулу оценки. Полученная на основе такой корректировки оценка будет сходиться к истинному значению оцениваемого параметра только при $n \rightarrow \infty$, т.е. будет являться асимптотически несмещенной.

Следует отметить, что асимптотически несмещенные оценки не обязательно более точные, чем смещенные, однако они являются более удобными с математической точки зрения (состоятельность, инвариантность относительно линейных преобразований и т.д.).

Эффективность. Оценка $G^* = f(x_1, x_2, x_3, \dots, x_n)$ называется эффективной, если среди всех оценок параметра G она обладает наименьшей мерой случайного разброса относительно истинного значения оцениваемого параметра, т.е. $D[G^*] = D_{\min}$.

Эффективная оценка, следовательно, имеет минимальную случайную погрешность и в этом смысле является наиболее точной.

Конкретные методы расчета оценок числовых характеристик СВ рассматриваются в п. 3.2.

3.1. Эмпирические кривые обеспеченностей

В практике строительного, дорожного и водохозяйственного проектирования часто требуется определить расход воды (или иную

гидрологическую характеристику) заданной обеспеченности, или что то же самое, заданной вероятности ежегодного превышения.

Такая задача легко решается, если закон распределения случайной величины известен. Однако чаще в практике закон распределения исследуемой СВ не известен, и представление о нем необходимо составить по эмпирическим данным.

Функция обеспеченностей, построенная по эмпирическим данным, называется *эмпирической кривой обеспеченностей*.

Возможны два способа построения эмпирической кривой обеспеченностей. Первый из них можно рекомендовать в том случае, когда имеется достаточно продолжительный ряд значений СВ.

Рассмотрим первый способ на примере данных наблюдений за среднегодовыми расходами на р. Днепре у пгт Лоцманская Каменка с 1818 по 1962 г. (табл. 3.1).

Определяем амплитуду (размах R) колебаний среднегодовых расходов как разницу между наибольшим ($Q_{\max}$) и наименьшим ($Q_{\min}$) значениями расходов: $R = Q_{\max} - Q_{\min} = 3040 - 717 = 2323 \text{ м}^3/\text{с}$.

Разбиваем амплитуду на k равных интервалов. Величину k можно приблизительно оценить по формуле

$$k \approx 5 \lg(n). \quad (3.2)$$

В данном случае $n = 145$, и следовательно, $k = 5 \lg(145) = 11$. Таким образом, примем в качестве первого приближения $k = 11$.

Определяем длину расчетного интервала: $l = R/k = 2323/11 = 211$. Для удобства расчета величину l можно округлить (но так, чтобы значение l изменилось не более чем на 10 – 15 %). Окончательно принимаем $l = 200$.

В качестве левой границы первого интервала примем значение, большее или равное $Q_{\max}$. С учетом того, что в данном случае значение l округлено до сотен, в качестве такой границы принимаем 3100. Тогда значение, соответствующее правой границе первого интервала, будет равно $3100 - 200 = 2900$. Последующие интервалы будут иметь границы: 2900 – 2700; 2700 – 2500; 2500 – 2300 и т.д.

Для того чтобы границы интервалов не пересекались, проведем корректировку левых границ всех интервалов. В результате получим: 3099 – 2900; 2899 – 2700; 2699 – 2500 и т.д. Дальнейший расчет сведен в табл. 3.2.

Таблица 3.1

Среднегодовые расходы воды р. Днепра у пгт Лоцманская Каменка

Год	$Q \text{ м}^3/\text{с}$	Год	$Q \text{ м}^3/\text{с}$	Год	$Q \text{ м}^3/\text{с}$	Год	$Q \text{ м}^3/\text{с}$
1818	1480	1855	1900	1891	1470	1927	1730
1819	1600	1856	1500	1892	1120	1928	1840
1820	2400	1857	1300	1893	1820	1929	1910
1821	1740	1858	1400	1894	1460	1930	1180
1822	1370	1859	1000	1895	2420	1931	2480
1823	1200	1860	1800	1896	1990	1932	2620
1824	1150	1861	2400	1897	1590	1933	2630
1825	1800	1862	1400	1898	1190	1934	1860
1826	1200	1863	900	1899	1220	1935	1620
1827	1300	1864	1000	1900	1670	1936	1560
1828	2080	1865	1700	1901	1320	1937	1620
1829	2500	1866	1600	1902	1690	1938	1490
1830	2100	1867	2000	1903	1620	1939	1080
1831	2060	1868	2000	1904	1080	1940	1680
1832	1550	1869	1200	1905	1740	1941	2300
1833	1340	1870	1400	1906	2060	1942	2470
1834	1750	1871	2000	1907	2100	1943	1010
1835	1000	1872	1780	1908	2220	1944	1380
1836	1030	1873	1100	1909	1700	1945	1480
1837	1680	1874	1520	1910	1160	1946	1310
1838	1940	1875	900	1911	1310	1947	1530
1839	1540	1876	2100	1912	1890	1948	1660
1840	1930	1877	3040	1913	1800	1949	1290
1841	1930	1878	1960	1914	1610	1950	1060
1842	870	1879	2450	1915	1820	1951	1480
1843	1100	1880	1640	1916	2050	1952	1240
1844	2000	1881	1670	1917	2400	1953	1970
1845	3000	1882	1060	1918	1300	1954	874
1846	1000	1883	2070	1919	1910	1955	1520
1847	1400	1884	1340	1920	1480	1956	1740
1848	1200	1885	1160	1921	717	1957	1500
1849	2400	1886	1550	1922	1730	1958	2420
1850	1800	1887	1330	1923	1710	1959	1320
1851	1740	1888	2060	1924	1820	1960	1250
1852	1650	1889	1940	1925	874	1961	1330
1853	2100	1890	1250	1926	1890	1962	1510
1854	1800						

Из таблицы видно, что для покрытия амплитуды колебаний расходов воды в расчетном створе потребовалось (при $l = 200$) не 11, а 12 интервалов (что не имеет принципиального значения).

Во втором столбце табл. 3.2 приводим число расходов воды, попавших в каждую градацию.

Таблица 3.2

**Сгруппированные данные годового стока р. Днепра
у пгт Лоцманская Каменка**

Интервал расходов, Q $\text{м}^3/\text{с}$	Число случаев попадания в интервал (повторяемость, m_i)	Относитель- ная частота, $p^* \%$	Число превышений правой грани- цы интервала	Эмпириче- ская обеспечен- ность, $\%$
3099 – 2900	2	1,38	2	1,38
2899 – 2700	0	0	2	1,38
2699 – 2500	3	2,07	5	3,45
2499 – 2300	10	6,90	15	10,4
2299 – 2100	5	3,45	20	13,8
2099 – 1900	20	13,80	40	27,6
1899 – 1700	23	15,86	63	43,5
1699 – 1500	25	17,24	88	60,7
1499 – 1300	24	16,56	112	77,3
1299 – 1100	17	11,73	129	89,0
1099 – 900	12	8,28	141	97,3
899 – 700	4	2,76	145	100
Сумма	145	100		

В третьем столбце даем относительную частоту попадания расходов воды в каждый интервал: $p^* = (m_i/n)100\%$, где m_i – число значений расхода воды в i -том интервале (повторяемость).

Четвертый столбец рассчитываем, последовательно суммируя данные второго столбца. Первое из полученных чисел будет соответствовать числу случаев, когда расход воды равен или превышает 2900 $\text{м}^3/\text{с}$, второе – когда расход равен или превышает 2700 $\text{м}^3/\text{с}$, третье – когда расход равен или превышает 2500 $\text{м}^3/\text{с}$ и т.д.

В пятом столбце число соответствующих превышений выразим в процентах от общего числа наблюдений.

По данным третьего столбца можно построить график эмпирических частот, по данным пятого столбца – график эмпирической кривой обеспеченностей (рис. 3.1).

В гидрологии значения СВ принято откладывать по вертикальной оси, а значения вероятности p – по горизонтальной. Для гистограммы эмпирических частот p следует трактовать как вероятность попадания в заданный интервал, а для кривой обеспеченностей – как вероятность превышения соответствующего значения СВ.

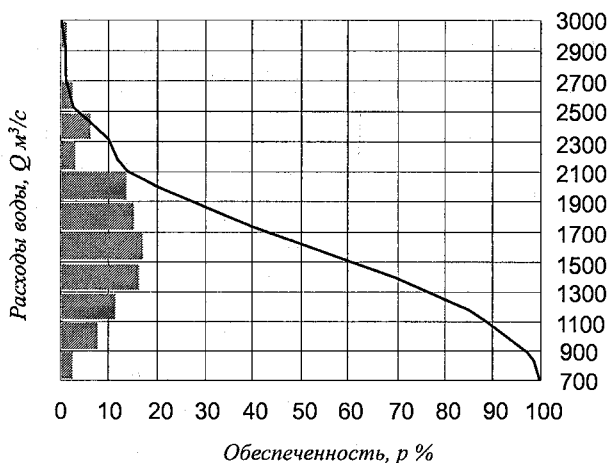


Рис. 3.1. Гистограмма эмпирических частот и эмпирическая кривая обеспеченностей годового стока р. Днепра у пгт Лоцманская Каменка.

В дополнение к сказанному можно сделать три замечания.

1. Если ступенчатый график эмпирических частот сгладить плавной линией, то мы получим эмпирическую функцию плотности вероятности.

2. Если суммирование эмпирических частот по интервалам вести не от больших значений к меньшим, а наоборот, то вместо кривой обеспеченностей мы получим интегральную кривую распределения. В зарубежной литературе эту кривую называют кривой накопленных частот.

3. В ряде случаев при построении кривых обеспеченностей по оси абсцисс удобнее откладывать не значения СВ, а ее нормированные значения или модульные коэффициенты.

Хотя изложенный способ построения эмпирической кривой обеспеченностей является довольно простым и наглядным, в практике гидрологических расчетов он применяется крайне редко, так как реальные ряды наблюдений за гидрологическими характеристиками, как правило, не превышают нескольких десятков лет.

При анализе рядов небольшого объема используется иной метод построения эмпирической кривой обеспеченности. При использовании этого метода члены эмпирического ряда ранжируются, т.е. располагаются в убывающем или возрастающем порядке. В гидрологии принято располагать члены ряда в убывающем порядке.

Допустим, имеется ряд величин какой-либо характеристики гидрологического режима, расположенных в убывающем порядке: $x_1 > x_2 > x_3 > \dots > x_m > \dots > x_N$. Тогда теоретическая вероятность превышения для m -ного члена ряда может быть выражена формулой

$$P\{X \geq x_m\} = \lim_{N \rightarrow \infty} (m/N). \quad (3.3)$$

В то же время длина реальной выборки всегда конечна, $n < \infty$. Заменяя N на n и используя формулу (3.3), можно попытаться приблизительно оценить вероятность превышения для каждого члена имеющейся выборки

$$p_m = P\{X \geq x_m\} \approx (m/n)100 \%, \quad (3.4)$$

где m — порядковый номер x_m в ранжированном ряду; p_m — обеспеченность (в %) m -ного члена ранжированного ряда.

В соответствии с формулой (3.4) обеспеченность первого (самого большого) члена ранжированного ряда будет равна $p_1 = (1/n)100 \%$, второго — $p_2 = (2/n)100 \%$ и т. д. Обеспеченность последнего члена ряда будет равна $p_n = (n/n)100 = 100 \%$. Таким образом, если принять данную формулу, получится, что последний член ранжированного ряда представляет собой абсолютный минимум и СВ X никогда не примет значение меньше чем x_n .

Данный парадокс возникает в связи с тем, что мы заменили N на n . В действительности можно получить бесконечное множество выборок из генеральной совокупности длиной n , каждая из которых будет иметь свой максимум и свой минимум. В этом смысле эмпирическая обеспеченность m -го члена ранжированного ряда сама будет являться случайной величиной, и в качестве расчетного значения разумно принять ее математическое ожидание, моду или иную устойчивую характеристику. С учетом сказанного в настоящее время разработано около десятка формул для расчета эмпирической обеспеченности.

Формула А. Хазена:

$$p_m = \frac{m-0,5}{n} 100 \% . \quad (3.5)$$

Формула С.Н. Крицкого – М.Ф. Менкеля (Вейбула):

$$p_m = \frac{m}{n+1} 100 \% . \quad (3.6)$$

Формула Н.Н. Чегодаева:

$$p_m = \frac{m-0,3}{n+0,4} 100 \% . \quad (3.7)$$

Формула (3.5) предполагает замену ступенчатого графика эмпирических частот сглаженной кривой, проходящей через середины ступенек графика. Обеспеченность первого члена ряда по зависимости (3.5) составит $(1/2n) 100 \%$. Такая оценка недостаточно обоснована, поскольку в этом случае повторяемость наблюдаемого максимума относится к периоду, вдвое превышающему период наблюдений.

Формула (3.6) соответствует математическому ожиданию эмпирической обеспеченности, а формула (3.7) – медианному значению эмпирической обеспеченности.

Среди перечисленных формул в определенном смысле наилучшей является формула (3.6), так как получаемая по ней оценка эмпирической обеспеченности является состоятельной, несмещенной и эффективной.

Однако при выводе формулы эмпирической обеспеченности допустимо рассуждать и по-иному. Значение СВ, расположенное в ранжированном ряду (длиной n) на m -ом месте, тоже является случайной величиной, и можно рассчитать обеспеченность, соответствующую МО или моде m -ного члена. Но в такой постановке задача разрешима лишь в том случае, когда закон распределения исходной СВ X известен. Приведем примеры таких формул.

Для распределения Гумбеля, а также экспоненциального распределения:

$$p_m = \frac{m-0,44}{n+0,12} 100 \% ; \quad (3.8)$$

для нормального распределения:

$$p_m = \frac{m - 3/8}{n + 1/4} 100 \% . \quad (3.9)$$

Таким образом, различие этих двух подходов состоит в том, что в первом случае в качестве СВ мы рассматриваем обеспеченность m -ного члена ранжированного ряда, а во втором в качестве СВ рассматривается значение m -ного члена и затем вычисляется обеспеченность его МО или моды. Следует отметить, что оба подхода при выводе формул эмпирической обеспеченности дают одинаковые результаты лишь в том случае, если теоретическая кривая обеспеченности представляет собой линейную зависимость СВ X от обеспеченности, как это имеет место для закона распределения равномерной плотности.

Учитывая сказанное, можно констатировать, что выбор формулы для расчета эмпирической обеспеченности содержит элемент субъективизма и зависит от традиций и специфики решаемых задач. Так, например, в гидротехнической практике за критерий надежности работы сооружений обычно принимается ежегодная вероятность нарушения нормальной работы установки. В этих условиях наиболее приемлемой при расчетах паводочного стока является формула (3.6), поскольку она дает большую надежность, особенно для малых значений обеспеченности. В США эта формула рекомендуется в качестве основной расчетной формулы.

В то же время, например, во Франции используется формула (3.5), а в Англии – формула (3.8).

Грингортен (1963 г.) предложил универсальную формулу для расчета эмпирической обеспеченности:

$$p_m = \frac{m - a}{n + 1 - 2a} 100 \% . \quad (3.10)$$

При $a = 0.5$ мы получаем формулу (3.5), при $a = 0,3$ – формулу (3.7), при $a = 0$ – формулу (3.6), при $a = 0,44$ – формулу (3.8) и при $a = 3/8$ – формулу (3.9). Сам же Грингортен рекомендовал устанавливать значение параметра a в зависимости от длины выборки по табл. 3.3.

Таблица 3.3

Зависимость параметра a от длины выборки n по Грингортену

$n \dots$	10	20	30	40	50	80	90	100
$a \dots$	0,448	0,443	0,442	0,441	0,440	0,440	0,439	0,439

В России действующие в настоящее время нормативные документы рекомендуют в качестве основной расчетной формулы выражение (3.6).

Пример 3.1. Построить эмпирическую кривую обеспеченностей максимальных паводочных расходов для р. Луга – ст. Толмачево (см. табл. 1.2).

Решение.

1. Ранжируем исходный ряд, т.е. располагаем расходы в убывающем порядке (табл. 3.4)

2. Для каждого расхода воды в ранжированном ряду рассчитываем эмпирическую обеспеченность по формуле (3.6). Так как длина ряда $n = 32$, обеспеченность первого члена ранжированного ряда равна $p_1 = (1/33)100\% = 3,03\%$; второго – $p_2 = (2/33)100\% = 6,06\%$ и т.д.

3. По данным граф 5 и 6 на клетчатке вероятности строим эмпирическую кривую обеспеченностей (рис. 3.2).

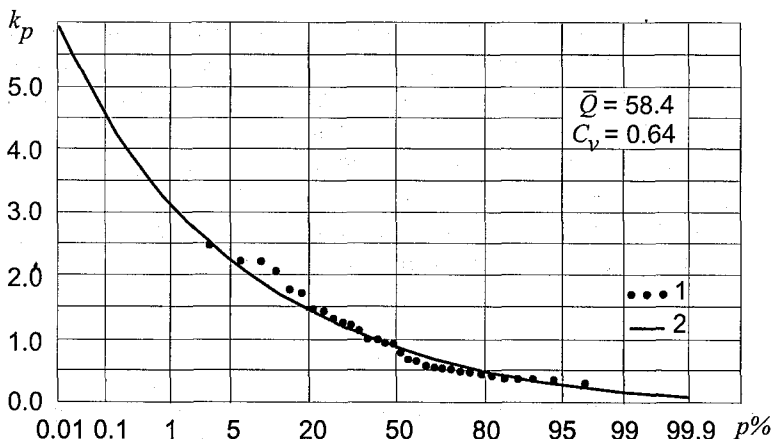


Рис. 3.2. Эмпирическая кривая обеспеченностей (1) и аналитическая кривая обеспеченностей Крицкого – Менкеля при $C_s/C_v = 2,5$ (2) максимальных расходов дождевых паводков; р. Луга – ст. Толмачево.

Таблица 3.4

Расчет ординат эмпирической кривой обеспеченностей максимальных расходов дождевых паводков р. Луги – ст. Толмачево ($\bar{Q} = 58,4 \text{ м}^3/\text{с}$)

m_i	Годы	Расходы, $\text{м}^3/\text{с}$	Ранжированные		Обеспеченность, %
			расходы, $\text{м}^3/\text{с}$	модульные коэффициенты	
1	1954	71,4	145	2,48	3,03
2	1955	22,1	130	2,22	6,06
3	1956	32,7	130	2,22	9,09
4	1957	145	121	2,07	12,1
5	1958	46,0	104	1,78	15,2
6	1959	29,9	100	1,71	18,2
7	1960	28,4	84,5	1,45	21,2
8	1961	73,8	84,4	1,44	24,2
9	1962	121	78,1	1,34	27,3
10	1963	25,0	73,8	1,26	30,3
11	1964	31,4	71,4	1,22	33,3
12	1965	17,3	66,0	1,13	36,4
13	1966	54,8	58,7	1,00	39,4
14	1967	29,9	58,2	1,00	42,4
15	1968	104	54,8	0,94	45,5
16	1969	27,5	54,0	0,92	48,5
17	1970	84,5	46,0	0,79	51,5
18	1971	37,5	39,2	0,67	54,6
19	1972	24,0	37,5	0,64	57,6
20	1973	22,1	32,7	0,56	60,6
21	1974	130	31,4	0,54	63,6
22	1975	21,3	29,9	0,51	66,7
23	1976	58,2	29,9	0,51	69,7
24	1977	58,7	28,4	0,49	72,7
25	1978	130	27,5	0,47	75,8
26	1979	22,0	25,0	0,43	78,8
27	1980	54,0	24,0	0,41	81,8
28	1981	100	22,1	0,38	84,9
29	1982	78,1	22,1	0,38	87,9
30	1983	39,2	22,0	0,38	90,9
31	1984	66,0	21,3	0,36	93,9
32	1985	84,4	17,3	0,30	97,0

Масштаб по оси ординат при построении эмпирической кривой обеспеченностей рекомендуется выбирать так, чтобы точка, соответствующая максимальному значению расхода, располагалась не слишком близко к верхнему краю клетчатки.

Как видно из приведенного примера, эмпирическая кривая обеспеченностей позволяет составить представление о распределении исследуемой СВ в диапазоне обеспеченностей от $100/(n+1)$ до $(100n)/(n+1)$ %, в данном случае от 3,03 % до 97,0 %. В то же время довольно часто требуется рассчитать значения расходов для обеспеченностей, которые гораздо меньше или гораздо больше указанных.

Например, при проектировании гидротехнических сооружений первого класса требуется иметь расход обеспеченностью 0,01 %. Для решения подобных задач необходимо уметь экстраполировать эмпирическую кривую в область больших и малых значений. Очевидно, что экстраполяция на глаз здесь не возможна, поскольку результаты такой экстраполяции будут носить слишком субъективный характер.

В практике гидротехнических расчетов для сглаживания и экстраполяции эмпирических кривых обеспеченностей используются аналитические кривые из числа перечисленных в главе 2. Выбранная аналитическая кривая по отношению к эмпирической кривой исполняет роль своего рода гидрологического лекала.

В такой постановке задача гидролога сводится к решению двух основных задач:

- 1) произвести выбор аналитической кривой;
- 2) оценить по эмпирическим данным параметры распределения (см. п. 3.2). Обычно требуется оценить МО, СКО (или C_v), C_s (или C_s/C_v).

Если эта задача решена, то на клетчатку вероятности помимо эмпирической кривой обеспеченностей наносится и аналитическая кривая (см. рис. 3.2), что позволяет проконтролировать степень согласованности аналитической кривой и эмпирических точек.

При построении графика аналитической кривой на клетчатку вероятности наносят 20 – 25 точек (координаты точек определяют по соответствующим таблицам) и затем соединяют точки плавной линией.

Если аналитическая кривая хорошо согласуется с эмпирической кривой обеспеченностей, то в дальнейших расчетах используются ординаты аналитической кривой.

При массовых расчетах нет необходимости строить графики эмпирической и аналитической кривых обеспеченностей для каждого створа. Можно провести такой анализ только для нескольких опорных пунктов. И если вопрос о выборе типа аналитической кривой обеспеченностей решен, то для остальных створов достаточно по эмпирическим данным оценить параметры распределения (см. п. 3.2.) и выписать из таблиц ординаты аналитической кривой для тех обеспеченностей, которые диктуются условиями конкретной задачи.

3.2. Методы расчета оценок параметров распределения

При статистической обработке гидрологических данных используются несколько методов расчета оценок параметров распределения. Эти методы можно условно разделить на три группы: аналитические, графоаналитические и графические.

Графические и графоаналитические методы, как видно из названия, сочетают в себе элементы аналитического расчета и графического построения.

В аналитических методах оценки параметров распределения представляют собой числа, полученные путем подстановки выборочных значений $x_1, x_2, x_3, \dots, x_n$ СВ X в теоретическую формулу оцениваемого параметра. При этом желательно, чтобы оценка удовлетворяла требованиям состоятельности, несмещенности и эффективности.

К числу аналитических методов относятся метод моментов и метод наибольшего правдоподобия.

3.2.1. Метод моментов

Метод основан на использовании моментов эмпирического распределения, которые являются состоятельными оценками соответствующих теоретических моментов (см. п. 1.4).

При замене теоретических моментов эмпирическими вместо $N \rightarrow \infty$ используется конечное число значений случайной величины n , а вероятность (p_i) того, что СВ X примет значение x_i , заменяется частотой $(p_i = 1/n)$.

Эмпирические начальный (α_s^*) и центральный (μ_s^*) моменты S -ного порядка определяются формулами:

$$\alpha_s^* = \frac{1}{n} \sum_{i=1}^n x_i^s, \quad (3.11)$$

$$\mu_s^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^s. \quad (3.12)$$

Как следует из (3.11), выборочную оценку математического ожидания (выборочное среднее $\bar{x}$) можно вычислить по формуле

$$m_x^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.13)$$

Эта оценка является состоятельной и несмещенной.

Для обозначения выборочной дисперсии наряду с D^* используется символ S^2 . Выборочная дисперсия представляет собой второй центральный момент и как следует из (3.12) определяется выражением

$$D^* = S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (3.14)$$

Так как математическое ожидание D^* равно не D , а величине $[(n-1)/n]D$, оценка (3.14) является состоятельной, но смещенной оценкой теоретической дисперсии. В связи с этим (особенно при $n < 40$) рекомендуется применять несмещенную состоятельную оценку дисперсии

$$S_H^2 = \frac{n}{n-1} S_C^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (3.15)$$

где S_H^2 и S_C^2 — соответственно несмещенная и смещенная оценки дисперсии; $n/(n-1)$ — поправка на смещенность. Отсюда несмещенная оценка СКО и коэффициента вариации выражается формулами:

$$\sigma^* = S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3.16)$$

$$C_v^* = \frac{S}{\bar{x}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (k_i - 1)^2}, \quad (3.17)$$

где $k_i = x_i / \bar{x}$ – модульный коэффициент.

Оценка коэффициента асимметрии определяется выражением

$$C_s^* = \frac{\mu_3^*}{S^3} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{nS^3} = \frac{\sum_{i=1}^n (k_i - 1)^3}{n(C_v^*)^3}. \quad (3.18)$$

Данная оценка также является состоятельной, но смещенной. Это смещение можно устранить, вводя поправочный коэффициент $n^2/[(n-1)(n-2)]$:

$$C_{s,H}^* = \frac{n^2 C_{s,C}^*}{(n-1)(n-2)} = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)S^3} = \frac{n \sum_{i=1}^n (k_i - 1)^3}{(n-1)(n-2)(C_v^*)^3}. \quad (3.19)$$

Оценки числовых характеристик, связанных с моментами более высоких порядков, в практике гидрологических расчетов обычно не используются, так как при существующей в настоящее время длине гидрологических рядов эти оценки не удовлетворяют необходимой точности.

Строго говоря, для достоверной оценки уже третьего центрального момента (а следовательно, и C_s) необходимы выборки в несколько сотен членов. Поэтому на практике вместо выборочного значения коэффициента асимметрии рекомендуется использовать районное соотношение C_s/C_v . Имея районное соотношение C_s/C_v и рассчитав выборочную оценку коэффициента вариации несложно получить и оценку C_s .

Для получения районного соотношения C_s/C_v в пределах гидрологически однородного района выделяется несколько десятков опорных створов с наиболее продолжительными рядами наблюдений. Для этих створов по формулам (3.17), (3.19) производится расчет C_v^* и C_s^* и вычисляется значение C_s^*/C_v^* . В качестве районного соотношения C_s/C_v принимается среднее значение по всем опорным створам.

При ограниченности исходных данных, когда невозможно получить районное соотношение C_s/C_v , подбор значения C_s/C_v для каждой выборки производят индивидуально, ориентируясь на наилуч-

шее соответствие аналитической и эмпирической кривых обеспеченностей. Эту процедуру удобнее производить с использованием набора спрямляющих клетчаток вероятностей для различных сочетаний C_s/C_v (см. п. 3.2.4).

К достоинствам метода моментов можно отнести то, что оценки параметров не зависят от закона распределения исследуемой случайной величины; а также то, что расчетные формулы достаточно просты и позволяют получить искомые параметры в явном виде. Поэтому метод моментов получил наибольшее распространение в практике гидрологических расчетов.

В то же время нужно иметь в виду, что оценки дисперсии, коэффициента вариации и коэффициента асимметрии имеют отрицательную смещенность. Это приводит к тому, что при больших значениях коэффициента вариации ($C_v > 0,5$) достоверность моментных оценок ощутимо снижается, а введение поправочных коэффициентов становится неэффективным. Поэтому при $C_v > 0,5$ для расчета оценок параметров распределения рекомендуется использовать метод наибольшего правдоподобия.

Пример 3.2. По ряду наблюдений за максимальными расходами дождевых паводков за период 1954 – 1985 гг. на р. Луге – ст. Толмачево (см. табл. 3.4) оценить параметры распределения методом моментов.

Решение.

При оценке параметров этим методом используются формулы (3.13), (3.17), (3.19). В табл. 3.5 приведены необходимые для этого промежуточные расчеты.

Таблица 3.5

Таблица для расчета выборочных параметров распределения максимальных паводочных расходов, р. Луга – ст. Толмачево

№ п/п	$Q_i, \text{м}^3/\text{с}$	$k_i = Q_i / \bar{Q}$	$k_i - 1$	$(k_i - 1)^2$	$(k_i - 1)^3$
1	71,4	1,22	0,22	0,0484	0,01065
2	22,1	0,38	-0,62	0,3844	-0,23833
3	32,7	0,56	-0,44	0,1936	-0,08518
...	...	...	...	...	...
30	39,2	0,67	-0,33	0,1089	-0,03594
31	66,0	1,13	0,13	0,0169	0,00219
32	84,4	1,45	0,45	0,2025	0,09112
Сумма	1868	32	0,00	12,698	6,857
Среднее	58,4	1	0		

По данным табл. 3.5 и вышеназванным формулам рассчитываем оценки параметров распределения:

$$\bar{Q} = \frac{\sum_{i=1}^n Q_i}{n} = \frac{1868}{32} = 58,4 \text{ м}^3/\text{с};$$

$$C_v^* = \sqrt{\frac{\sum_{i=1}^n (k_i - 1)^2}{n-1}} = \sqrt{\frac{12.698}{31}} = 0,64;$$

$$C_s^* = \frac{n \sum_{i=1}^n (k_i - 1)^3}{(n-1)(n-2)C_v^3} = \frac{32 \cdot 6,857}{31 \cdot 30 \cdot 0,64^3} = 0,90; \quad C_s/C_v = 1,41.$$

3.2.2. Метод наибольшего правдоподобия

Для нахождения оценки методом наибольшего правдоподобия необходимо прежде всего построить *функцию правдоподобия*. Чтобы понять, как это делается, рассмотрим выборку значений случайной величины X объемом в n членов. Предположим, что функция плотности вероятности СВ X имеет достаточно простой вид: $f(x, G)$, т.е. зависит от одного параметра G . Суть метода состоит в том, чтобы найти такое значение параметра G , при котором вероятность получить в результате n опытов именно данную выборку $(x_1, x_2, x_3, \dots, x_n)$ являлась бы максимальной. С математической точки зрения эта задача сводится к нахождению максимума некоторой функции $L(x_i, G)$, которая и называется функцией правдоподобия. Функция $L(x_i, G)$ представляет собой совместную плотность вероятности вектора $X = (x_1, x_2, x_3, \dots, x_n)$ при данном G :

$$L(x_i, G) = f(x_1, G) f(x_2, G) f(x_3, G) \dots f(x_n, G) = \prod_{i=1}^n f(x_i, G). \quad (3.20)$$

Согласно правилам нахождения экстремумов, для определения максимума функции $L(x_i, G)$ нужно решить уравнение

$$\frac{\partial L(x_i, G)}{\partial G} = 0. \quad (3.21)$$

Для упрощения решения обычно используют логарифмическую функцию правдоподобия $\ln[L(x_i, G)]$. Учитывая что эта функция имеет максимум при том же значении G , получаем

$$\frac{1}{L} \frac{\partial L}{\partial G} = \frac{\partial \ln L}{\partial G} = \sum_{i=1}^n \frac{\partial [\ln f(x_i, G)]}{\partial G} = 0. \quad (3.22)$$

Если функция распределения зависит от нескольких параметров, следует взять частные производные по каждому из них:

$$\frac{\partial \ln L}{\partial G_j} = 0 \quad (j = 1, 2, 3, \dots, m).$$

В качестве примера рассмотрим оценку методом наибольшего правдоподобия параметров нормального распределения (см. п. 2.1), для которого функция плотности вероятности имеет вид:

$$f(x) = [1/(\sigma_x \sqrt{2\pi})] \exp[-(x - m_x)^2 / (2\sigma_x^2)],$$

где m_x и σ_x – искомые параметры распределения.

Образует функцию правдоподобия (3.20), предварительно найдем логарифм функции $f(x)$:

$$\ln f(x) = -\ln \sigma - 0,5 \ln(2\pi) - (1/2\sigma^2)(x_i - m_x)^2,$$

$$L = \sum_{i=1}^n \ln f(x) = -\sum_{i=1}^n \ln \sigma - 0,5 \sum_{i=1}^n \ln(2\pi) - (1/2\sigma^2) \sum_{i=1}^n (x_i - m_x)^2.$$

Далее находим уравнение правдоподобия применительно к оценке параметра m_x :

$$\frac{\partial L}{\partial m_x} = -\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m_x) = 0,$$

откуда, учитывая что $\sum_{i=1}^n m_x = nm_x$, получаем $m_x^* = \frac{1}{n} \sum_{i=1}^n x_i$.

Таким образом, статистической оценкой параметра m_x (математического ожидания) в данном случае является среднее арифметическое значение ряда $(x_1, x_2, x_3, \dots, x_n)$.

Для получения оценки параметра σ_x образуем новое уравнение правдоподобия

$$\frac{\partial L}{\partial \sigma} = \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - m_x)^2 - \frac{n}{\sigma} = 0.$$

Откуда получим

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^n (x_i - m_x)^2}{n}},$$

т.е. выражение среднеквадратического отклонения.

Из приведенного анализа следует, что применительно к нормальному закону распределения оценки параметров, полученные методом наибольшего правдоподобия, совпадают с моментными оценками. Для других распределений такое совпадение не является обязательным.

Более сложная система уравнений правдоподобия получается для трехпараметрического гамма-распределения. На практике параметры этого распределения C_v и C_s определяются на основе приближенного метода наибольшего правдоподобия с помощью специально разработанных номограмм [16] как функция вспомогательных статистик λ_2 и λ_3 :

$$\lambda_2 = \frac{\sum_{i=1}^n \lg k_i}{n-1}, \quad \lambda_3 = \frac{\sum_{i=1}^n k_i \lg k_i}{n-1}, \quad (3.23)$$

где $k_i = x_i / \bar{x}$ – модульные коэффициенты; $\bar{x}$ – среднее арифметическое ряда. Пример такой номограммы для диапазона коэффициентов вариации от 0,4 до 0,7 приводится на рис. 3.3.

Считается, что полученные таким образом оценки параметров трехпараметрического гамма-распределения являются состоятельными, эффективными и несмещенными.

В качестве особенностей метода наибольшего правдоподобия (МНП) можно отметить то, что в нем наибольший вес придается

средним членам выборки, имеющим наибольшую вероятность, в отличие от метода моментов, где самый большой вклад вносят крайние члены. МНП дает состоятельные и наиболее эффективные оценки. Полученные по этому методу оценки могут быть незначительно смещены, но это смещение легко устраняется путем введения соответствующих поправок. Однако перечисленные свойства проявляются только в случае достаточно больших выборок, при малых n с оценкам МНП могут конкурировать другие оценки.

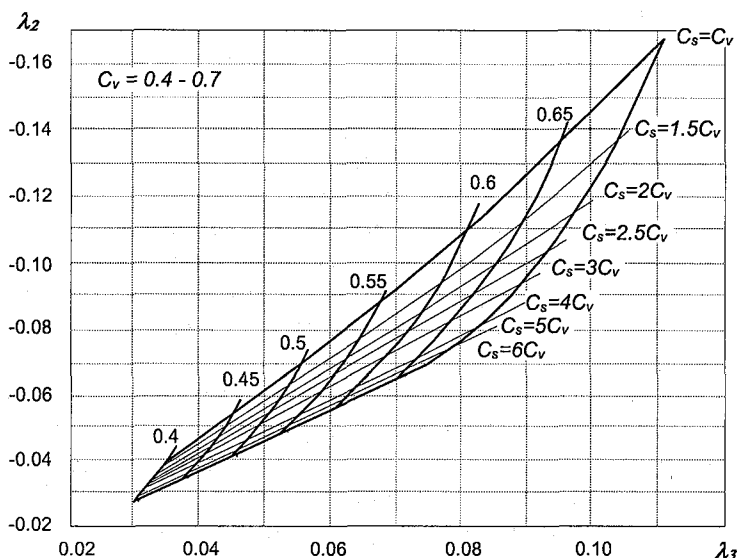


Рис. 3.3. Номограмма для определения параметров трехпараметрического гамма-распределения C_s и C_v методом приближенного наибольшего правдоподобия.

К недостаткам метода можно отнести то, что для его применения необходимо точно знать аналитическое выражение закона распределения. В гидрологической практике это не всегда возможно. К тому же не всегда удастся получить решение в аналитическом виде, и определение максимума функции правдоподобия придется производить численными методами или строить вспомогательные номограммы.

В России номограммы для метода наибольшего правдоподобия разработаны применительно к распределению Крицкого – Менкеля. Однако на практике не редки ситуации, когда выборочные λ_2 и λ_3

приводят к выходу за пределы номограмм, т.е. решение отсутствует. В этом случае можно использовать сокращенный метод наибольшего правдоподобия.

При реализации сокращенного метода статистика λ_3 не рассчитывается, а вместо нее используется районное соотношение C_s/C_v . Зная λ_2 и C_s/C_v , по номограмме можно определить коэффициент вариации.

Сокращенный метод наибольшего правдоподобия используется и применительно к кривой Пирсона III типа. Для этого распределения зависимость коэффициента вариации от параметров λ_2 и C_s/C_v представлена в виде таблицы (приложение 9).

Пример 3.3. По ряду наблюдений за максимальными расходами дождевых паводков за период 1954 – 1985 гг. на р. Луге – ст. Толмачево (см. табл. 3.4) оценить параметры распределения методом наибольшего правдоподобия.

Решение.

1. Для расчета среднего значения ряда и вспомогательных статистик λ_2 , λ_3 заполняем табл. 3.6.

Таблица 3.6

Вспомогательная таблица для расчета оценок параметров распределения методом наибольшего правдоподобия

№ п/п	$Q, \text{м}^3/\text{с}$	k	$\lg(k)$	$k \lg(k)$
1	71,4	1,22	0,086	0,1054
2	22,1	0,38	-0,4202	-0,1435
3	32,7	0,56	-0,2518	-0,1410
...	...	...	...	...
30	39,2	0,67	-0,1739	-0,1165
31	66,0	1,13	0,0531	0,0600
32	84,4	1,45	0,1614	0,2340
Сумма	1868		-2,697	2,573
Среднее	58,4			

2. По формулам (3.23) рассчитываем статистики λ_2 и λ_3 :

$$\lambda_2 = \frac{-2,697}{31} = -0,087, \quad \lambda_3 = \frac{2,573}{31} = 0,083.$$

3. Используя номограмму [16] (см. рис. 3.3) в зависимости от λ_2 и λ_3 определяем параметры распределения:

$$C_v = 0,67, \quad C_s/C_v = 2,8.$$

3.2.3. Графоаналитический метод (метод квантилей)

Графоаналитический метод расчета оценок параметров распределения включает в себя элементы графического и аналитического анализа.

При использовании этого метода необходимо заранее знать, какая из аналитических кривых будет применяться для аппроксимации распределения исследуемой СВ.

3.2.3.1. Графоаналитический метод на основе кривой Пирсона III типа

Методика, применяемая в настоящее время в России, была разработана Г.А. Алексеевым (1960 г.) и предполагает использование кривой Пирсона III типа. Рассмотрим основные этапы расчета.

Сначала на клетчатку вероятности наносятся точки эмпирической кривой обеспеченностей (рис. 3.4). При этом по оси ординат следует откладывать значения исследуемой СВ, а не модульные коэффициенты. Затем в поле точек проводится сглаженная кривая.

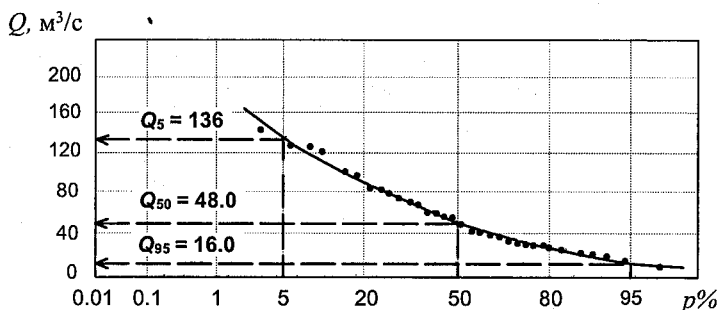


Рис. 3.4. Схема определения опорных ординат по сглаженной эмпирической кривой обеспеченностей ряда максимальных паводочных расходов на р. Луге – ст. Толмачево.

По сглаженной кривой определяются три опорные ординаты (для обеспеченностей – 5, 50 и 95 %): x_5 , x_{50} , x_{95} .

Дальнейший расчет сводится к тому, чтобы найти такие значения параметров распределения ($\bar{x}$, C_v^* , C_s^*), при которых аналитическая кривая будет проходить через точки x_5 , x_{50} , x_{95} . Для решения этой задачи трижды воспользуемся формулой (2.5):

$$x_5 = \sigma_x^* t_5 + \bar{x}; \quad (3.24)$$

$$x_{50} = \sigma_x^* t_{50} + \bar{x}; \quad (3.25)$$

$$x_{95} = \sigma_x^* t_{95} + \bar{x}, \quad (3.26)$$

где t_5, t_{50}, t_{95} – нормированные ординаты кривой обеспеченностей Пирсона III типа (см. приложение 1). Таким образом, мы имеем три неизвестных параметра и три уравнения.

Для расчета оценки коэффициента асимметрии сначала рассчитаем так называемый коэффициент скошенности S :

$$S = (x_p + x_{100-p} - 2x_{50}) / (x_5 - x_{95}). \quad (3.27)$$

В нашем случае обеспеченность первой ординаты $p = 5\%$, следовательно:

$$S^* = (x_5 + x_{95} - 2x_{50}) / (x_5 - x_{95}). \quad (3.27a)$$

Подставляя выражения (3.24) – (3.26) в формулу (3.27 а), получим

$$S^* = (t_5 + t_{95} - 2t_{50}) / (t_5 - t_{95}), \quad (3.27б)$$

т.е. коэффициент скошенности не изменится при замене x_5, x_{50}, x_{95} на нормированные ординаты. Но при фиксированном значении p для нормированного распределения Пирсона III типа имеет место однозначная зависимость коэффициента скошенности от коэффициента асимметрии (см. приложение 1).

Следовательно, зная коэффициент скошенности (3.27а), по приложению 1 можно определить C_s^* .

Вычитая из уравнения (3.24) соответственно левую и правую части уравнения (3.26), получим выражение для оценки СКО:

$$\sigma_x^* = (x_5 - x_{95}) / (t_5 - t_{95}). \quad (3.28)$$

Оценку МО находим из уравнения (3.25)

$$\bar{x} = x_{50} - \sigma_x^* t_{50}. \quad (3.29)$$

С учетом формулы (1.14) находим коэффициент вариации

$$C_v^* = \sigma_x^* / \bar{x}. \quad (3.30)$$

Таким образом, аналитическая часть расчета сводится к последовательному использованию формул (3.27а), (3.28) – (3.30).

Пример 3.4. По ряду наблюдений за максимальными расходами дождевых паводков на р. Луге – ст. Толмачево (см. табл. 1.2) оценить параметры распределения графоаналитическим методом. В качестве аналитической кривой обеспеченностей использовать кривую Пирсона III типа.

Решение.

1. Ранжируем ряд и по формуле (3.6) рассчитываем ординаты эмпирической кривой обеспеченностей (см. табл. 3.4).

2. На клетчатке с умеренной асимметричностью строим эмпирическую кривую обеспеченностей (см. рис. 3.4).

3. По сглаженной эмпирической кривой обеспеченностей определяем опорные ординаты $Q_5^* = 136$, $Q_{50}^* = 48$, $Q_{95}^* = 16$.

4. По формуле (3.27а) рассчитываем коэффициент скошенности $S^* = (136 + 16 - 2 \cdot 48) / (136 - 16) = 0,47$.

5. В зависимости от коэффициента скошенности по приложению 1 определяем три нормированные ординаты кривой Пирсона III типа и коэффициент асимметрии:

$$t_5 = 1,98; t_{50} = -0,26; t_{95} = -1,07; C_s^* = 1,67.$$

6. По формуле (3.28) вычисляем оценку СКО:

$$\sigma_q^* = (136 - 16) / (1,98 + 1,07) = 39,3.$$

7. По формуле (3.29) вычисляем оценку МО:

$$\bar{Q} = 48 - 39,3(-0,26) = 58,2.$$

8. По формуле (3.30) вычисляем оценку C_v :

$$C_v^* = 39,3 / 58,2 = 0,68.$$

9. В результате получаем:

$$\bar{Q} = 58,2; C_v^* = 0,68; C_s^* = 1,67; C_s^* / C_v^* = 2,46.$$

Если на основе этих параметров требуется рассчитать ординаты аналитической кривой обеспеченностей, то расчет ведется по схеме, изложенной в п. 2.5 (см. пример 2.6).

3.2.3.2. Графоаналитический метод на основе трехпараметрической логнормальной кривой

У трехпараметрического логнормального распределения третий параметр (a) не влияет на форму кривой обеспеченностей, а лишь приводит к сдвигу по оси X . Поэтому для данного распределения имеет место однозначная зависимость коэффициента скошенности S от коэффициента асимметрии (табл. 3.7). И, следовательно, можно применить подход, аналогичный тому, который применялся для кривой Пирсона III типа.

Таблица 3.7

Значения коэффициента скошенности S , вычисленного на основе опорных ординат t_5, t_{50}, t_{95} логнормального распределения в зависимости от коэффициента асимметрии C_s

C_s	S	Нормированные ординаты t_p			C_s	S	Нормированные ординаты t_p		
		5 %	50 %	95 %			5 %	50 %	95 %
0	0	1,64	0	-1,64	1,9	0,41	1,89	-0,23	-1,12
0,1	0,03	1,67	-0,02	-1,62	2,0	0,42	1,89	-0,24	-1,10
0,2	0,06	1,70	-0,04	-1,59	2,1	0,44	1,89	-0,24	-1,08
0,3	0,09	1,72	-0,06	-1,56	2,2	0,45	1,89	-0,25	-1,06
0,4	0,11	1,75	-0,07	-1,53	2,3	0,46	1,88	-0,25	-1,04
0,5	0,14	1,77	-0,09	-1,49	2,4	0,48	1,88	-0,26	-1,02
0,6	0,16	1,79	-0,10	-1,46	2,5	0,49	1,88	-0,26	-1,00
0,7	0,19	1,81	-0,11	-1,43	2,6	0,50	1,87	-0,26	-0,99
0,8	0,21	1,82	-0,13	-1,40	2,7	0,51	1,87	-0,27	-0,97
0,9	0,23	1,84	-0,14	-1,37	2,8	0,51	1,86	-0,27	-0,96
1,0	0,25	1,85	-0,15	-1,34	2,9	0,52	1,86	-0,27	-0,95
1,1	0,27	1,86	-0,16	-1,31	3,0	0,53	1,85	-0,28	-0,93
1,2	0,29	1,87	-0,17	-1,29	3,2	0,55	1,84	-0,28	-0,90
1,3	0,31	1,88	-0,18	-1,26	3,4	0,56	1,83	-0,29	-0,88
1,4	0,33	1,88	-0,19	-1,23	3,6	0,57	1,81	-0,29	-0,86
1,5	0,35	1,89	-0,20	-1,21	3,8	0,58	1,80	-0,29	-0,84
1,6	0,37	1,89	-0,21	-1,18	4,0	0,59	1,78	-0,29	-0,82
1,7	0,38	1,89	-0,22	-1,16	4,5	0,62	1,75	-0,30	-0,78
1,8	0,39	1,89	-0,22	-1,14	5,0	0,64	1,71	-0,30	-0,74

Методики отличаются только тем, что здесь требуется дополнительно рассчитать параметр сдвига a .

Как показал Г.А. Алексеев, параметр a связан с опорными ординатами формулой

$$a = \frac{x_5 x_{95} - x_{50}^2}{x_5 + x_{95} - 2x_{50}}. \quad (3.31)$$

Дальнейший расчет ведется по стандартной схеме.

- По формуле (3.27а) рассчитывается коэффициент скошенности S^* .

- В зависимости от S^* по таблице логнормального распределения (табл. 3.7) определяются три нормированные ординаты и коэффициент асимметрии.

- По формулам (3.28) – (3.30) рассчитываются σ_x^* , $\bar{x}$, $C_{v,x}^*$.

Пример 3.5. По ряду расходов из предыдущего примера оценить параметры распределения графоаналитическим методом. В качестве аналитической кривой обеспеченностей использовать трехпараметрическую логнормальную кривую.

Решение.

Пункты 1 – 4 – аналогичны пунктам примера 3.4.

5. В зависимости от S по табл. 3.7 определяем три нормированные ординаты логнормального распределения и коэффициент асимметрии:

$$t_5 = 1,88; t_{50} = -0,26; t_{95} = -1,03; C_s^* = 2,35.$$

6. По формуле (3.28) вычисляем оценку СКО:

$$\sigma_Q^* = (136 - 16) / (1,88 + 1,03) = 41,2.$$

7. По формуле (3.29) вычисляем оценку МО:

$$\bar{Q} = 48 - 41,2 \cdot (-0,26) = 58,7.$$

8. По формуле (3.30) вычисляем оценку C_v :

$$C_v^* = 41,2 / 58,7 = 0,70.$$

В результате получаем:

$$\bar{Q} = 58,7; C_v^* = 0,70; C_s^* = 2,35; C_s^* / C_v^* = 3,36.$$

Если на основе полученных параметров требуется рассчитать ординаты аналитической кривой обеспеченностей, то дополнительно по формуле (3.31) определяется параметр a :

$$a = (136 \cdot 16 - 48^2) / (136 + 16 - 2 \cdot 48) = -2,29.$$

Дальнейший расчет ведется по схеме, изложенной в п. 2.3. (см. пример 2.4).

3.2.3.3. Графоаналитический метод на основе кривой Крицкого – Менкеля

Хотя кривая Крицкого – Менкеля рекомендуется нормативными документами как основная расчетная кривая, до последнего времени она не использовалась для оценки параметров распределения методом квантилей. Причиной этого является отсутствие для распределения Крицкого – Менкеля однозначной связи между коэффициентом асимметрии C_s и коэффициентом скошенности S .

В 1996 г. на кафедре гидрологии суши РГГМУ была предложена методика, учитывающая специфику данной кривой.

В отличие от стандартной схемы, наряду с обычным коэффициентом скошенности S , рассчитывается коэффициент S_2

$$S_2 = 2x_{50} / (x_5 - x_{95}) = 2k_{50} / (k_5 - k_{95}). \quad (3.32)$$

Как показали исследования, при фиксированном значении C_s/C_v коэффициент S_2 однозначно зависит от S . Данная зависимость представлена в виде номограммы на рис. 3.5.

Таким образом, рассчитав S и S_2 , можно по номограмме оценить соотношение C_s/C_v . Значение C_s/C_v определенное по номограмме, следует округлить до 0,5.

Большое значение коэффициента скошенности ($S > 0,6$) указывает на то, что $C_v > 1$. В этой ситуации выборочная оценка соотношения C_s/C_v является крайне ненадежной, и следует отказаться от ее расчета графоаналитическим методом (можно принять районное значение C_s/C_v).

При фиксированном C_s/C_v имеет место однозначная зависимость между коэффициентом вариации C_v и коэффициентом скошенности S . Эта зависимость представлена в табл. 3.8. При известных значениях C_s/C_v и S по этой таблице можно определить коэффициент вариации.

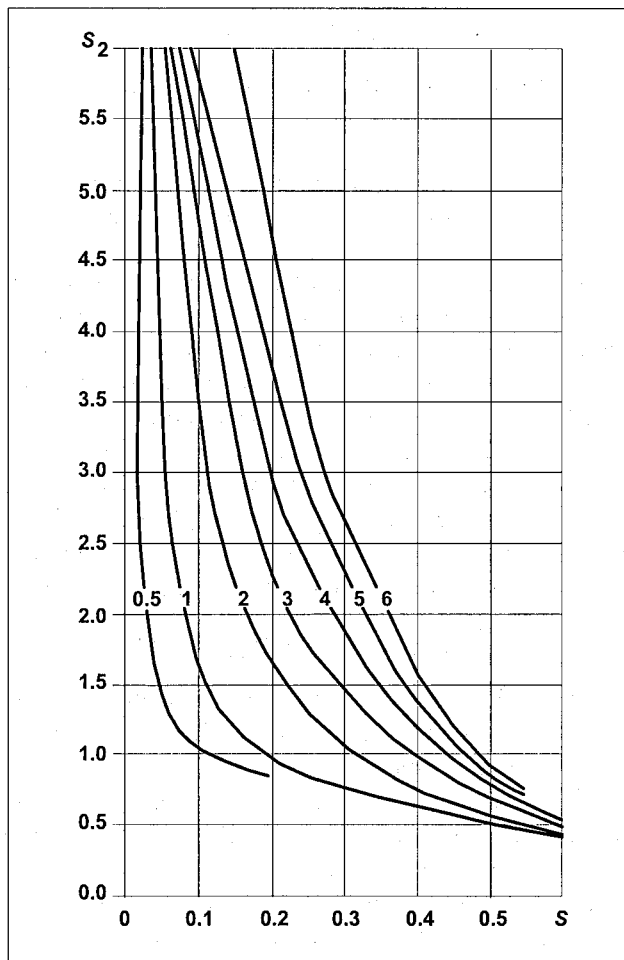


Рис. 3.5. Номограмма для определения отношения C_s/C_v в зависимости от коэффициента скошенности S и параметра S_2 для кривой обеспеченностей Крицкого–Менкеля.

Для расчета среднего значения необходимо по таблице ординат кривой обеспеченностей Крицкого–Менкеля (см. приложение 2) определить модульный коэффициент 50 %-ной обеспеченности (k_{50}). Тогда

$$\bar{x} = x_{50}/k_{50}. \quad (3.33)$$

Таблица 3.8

Значения коэффициента скошенности S в зависимости от C_v
и соотношения C_s/C_v для кривой обеспеченностей Крицкого – Менкеля

C_v	Соотношение C_s/C_v						
	0,5	1	2	3	4	5	6
0,1	0,027	0,036	0,055	0,061	0,068	0,091	0,152
0,2	0,030	0,056	0,113	0,163	0,198	0,238	0,270
0,3	0,041	0,084	0,169	0,229	0,286	0,326	0,356
0,4	0,050	0,113	0,226	0,298	0,348	0,378	0,402
0,5	0,074	0,156	0,279	0,357	0,403	0,427	0,443
0,6	0,114	0,206	0,335	0,409	0,450	0,464	0,481
0,7	0,197	0,270	0,390	0,461	0,488	0,489	0,508
0,8		0,358	0,445	0,504	0,527	0,538	0,539
0,9		0,456	0,506	0,546	0,561	0,563	0,571
1,0		0,562	0,565	0,586	0,595		
1,1		0,663	0,618	0,623	0,625		
1,2		0,754	0,667	0,658	0,652		
1,3		0,823	0,715	0,689	0,679		
1,4		0,871	0,760	0,719	0,704		
1,5		0,904	0,796	0,746	0,726		
1,6		0,930	0,823	0,772	0,748		
1,7		0,952	0,856	0,795	0,768		
1,8		0,969	0,882	0,817	0,786		
1,9		0,983	0,902	0,837	0,803		
2,0		0,996	0,920	0,855	0,819		

Пример 3.6. Оценить параметры распределения графоаналитическим методом, используя данные примера 3.4. В качестве аппроксимирующей кривой использовать кривую обеспеченностей Крицкого – Менкеля.

Решение.

Пункты 1 – 4 аналогичны пунктам примера 3.4.

5. По формуле (3.32) определяем коэффициент S_2 :

$$S_2 = (2 \cdot 48) / (136 - 16) = 0,80.$$

6. Для значений $S = 0,47$ и $S_2 = 0,80$ по номограмме (см. рис. 3.5) определяем соотношение (C_s/C_v) . В данном случае (с округлением до целых) $C_s/C_v = 3$.

7. По табл. 3.8 для $S = 0,47$ и $C_s/C_v = 3$ определяем коэффициент вариации. В данном случае $C_v = 0,72$.

8. По таблицам кривой обеспеченностей Крицкого – Менкеля (см. приложение 2) при $C_s/C_v = 3$ и $C_v = 0,72$ определяем k_{50} : $k_{50} = 0,815$.

9. По формуле (3.33) определяем среднее значение $\bar{x} = 48,0 / 0,815 = 58,9 \text{ м}^3/\text{с}$.

Расчет ординат аналитической кривой обеспеченностей производится по схеме, изложенной в п. 2.6 (см. пример 2.7).

3.4.3.4. Комментарии

В табл. 3.9 даны значения параметров распределения, рассчитанные графоаналитическим методом с использованием трех типов кривых распределения.

Таблица 3.9

Параметры распределения максимальных расходов дождевых паводков
р. Луги – ст. Толмачево, рассчитанные графоаналитическим методом
с использованием разных аналитических кривых

Кривая обеспеченностей	$\bar{x}$	C_v^*	C_s^*	C_s^*/C_v^*	Доп. параметры
Пирсона III типа	58,2	0,68	1,67	2,46	$S = 0,47$
Трехпараметрическая логнормальная	58,7	0,70	2,35	3,36	$S = 0,47, a = -2,29$
Крицкого – Менкеля	58,9	0,72	2,16	3,0	$S = 0,47, S_2 = 0,80$

Как видно из таблицы, все три варианта дают близкие значения *среднего и коэффициента вариации*. В то же время значения *коэффициента асимметрии* различаются весьма значительно, так как именно коэффициент асимметрии (который связан с коэффициентом скошенности S) обеспечивает главным образом нужное соотношение опорных ординат, но для различных типов кривых это соотношение будет достигаться при различных C_s .

На аналитических кривых обеспеченностей различие в асимметрии будет проявляться прежде всего в области больших и малых значений, т.е. на концах кривых.

Хотя графоаналитический метод является достаточно простым и наглядным, его следует применять с осторожностью, так как проведение сглаженной эмпирической кривой всегда носит субъективный характер и требует определенных практических навыков.

3.2.4. Графический метод

При использовании графического метода для расчета оценок параметров распределения необходимо заранее знать, какая из аналитических кривых будет применяться для аппроксимации закона распределения исследуемой СВ.

Этот метод может применяться как для двухпараметрических, так и для трехпараметрических кривых распределения.

При использовании трехпараметрической кривой на первом этапе применяется набор спрямляющих клетчаток выбранного теоретического распределения для различных сочетаний C_s и C_v .

На каждой клетчатке строится эмпирическая кривая обеспеченностей в модульных коэффициентах. В качестве расчетного значения принимается такое соотношение C_s/C_v , при котором на соответствующей клетчатке эмпирическая кривая обеспеченностей превращается в прямую линию. Как отмечалось в п. 2.8, наклон прямой линии, в которую превращается кривая обеспеченностей на спрямляющей клетчатке, зависит от коэффициента вариации. С учетом этого на клетчатке вероятности можно построить шкалу C_v . Обычно для контроля такая шкала наносится в левом верхнем и правом нижнем углах клетчатки (рис. 3.6).

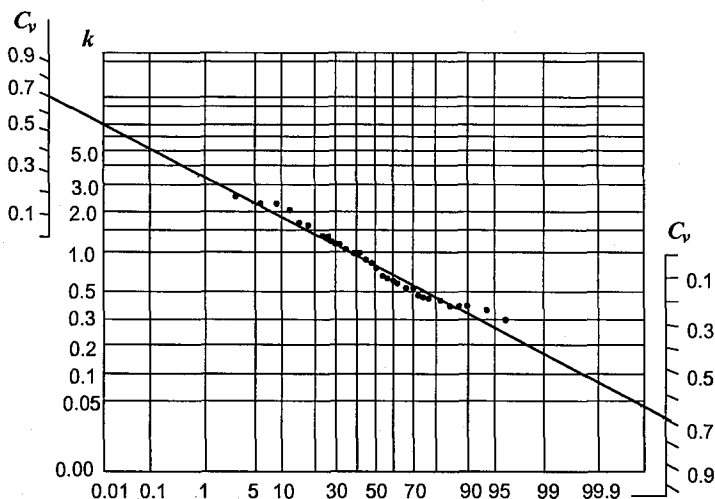


Рис. 3.6. Кривая обеспеченностей максимальных паводочных расходов на р. Луга – ст. Толмачево, построенная на клетчатке, спрямляющей кривую Крицкого – Менкеля при соотношении $C_s/C_v = 3$.

В качестве расчетного C_v^* принимается такое значение, при котором отсчеты по верхней и нижней шкалам совпадают. При этом аппроксимирующая прямая достаточно хорошо соответствует эмпирическим точкам.

Таким образом графический метод позволяет найти оценки коэффициента вариации и соотношения C_s/C_v . Оценка МО рассчитывается аналитически по формуле (3.13).

Пример 3.7. По ряду наблюдений за максимальными годовыми расходами воды на реке Луге – ст. Толмачево (см. табл. 1.2) оценить параметры распределения графическим методом. В качестве аналитической кривой обеспеченностей использовать кривую Крицкого – Менкеля.

Решение.

1. По формуле (3.13) рассчитываем оценку МО: $\bar{Q} = 58,4$.
2. По формуле $k_i = Q_i / \bar{Q}$ рассчитываем модульные коэффициенты.
3. Ранжируем ряд модульных коэффициентов и по формуле (3.6) рассчитываем ординаты эмпирической кривой обеспеченностей (см. табл. 3.4).
4. На клетчатках, спрямляющих кривую Крицкого – Менкеля, пять раз строим эмпирическую кривую обеспеченностей (для $C_s/C_v = 1; 1,5; 2,0; 3,0; 4,0$).
5. В качестве расчетного принимаем значение $C_s/C_v = 3$, так как на этой клетчатке точки эмпирической кривой группируются вдоль прямой линии (см. рис. 3.6).
6. В поле точек проводим прямую линию и по угловой шкале определяем коэффициент вариации. В данном случае $C_v^* = 0,70$.

Таким образом, окончательно получаем:

$$\bar{Q} = 58,4; C_v^* = 0,70; (C_s/C_v)^* = 3$$

и, следовательно, $C_s^* = 2,1$.

Метод позволяет оценить C_v с меньшей точностью, чем аналитические методы. В то же время оценка C_s/C_v может оказаться более надежной, поскольку при использовании аналитических методов в случае короткой выборки несколько «тяжелых» точек могут существенно исказить оценку отношения C_s/C_v .

При использовании двухпараметрических распределений C_z однозначно определяется коэффициентом вариации и требуется только одна клетчатка. Если эмпирические точки на такой клетчатке лягут в виде прямой линии, расчет C_z производится по изложенной выше схеме. Если точки будут образовывать вогнутую или выпуклую линии, то следует сделать вывод о том, что данная двухпараметрическая кривая не подходит для аппроксимации закона распределения исследуемой СВ.

3.2.5. Оценка параметров распределения Джонсона

Распределение Джонсона (см. п. 2.7) является одной из модификаций логнормального распределения. Вывод этого распределения базируется на том, что СВ Z имеет нормальное распределение, где $Z = \ln[(X - a)/(b - X)]$; a и b — соответственно нижний и верхний пределы СВ X . Данное распределение является четырехпараметрическим и требует в общем случае оценки четырех параметров: m_z , σ_z , a и b .

Если параметры a и b известны, то, преобразовав члены выборки по формуле $z_i = \ln[(x_i - a)/(b - x_i)]$, можно легко рассчитать оценки m_z и σ_z методом моментов по формулам (3.13) и (3.16).

Таким образом, основная проблема состоит в оценке параметров a и b . Оценка этих параметров на основе аналитического решения в зависимости от третьего и четвертого моментов не может дать надежного результата, так как эти числовые характеристики при небольшой длине выборок крайне неустойчивы. На практике параметры a и b находят методом последовательных приближений¹.

На первом этапе устанавливаются диапазоны возможных значений параметров. Так, для положительных величин

$$0 \leq a < x_{\min}, \quad (3.34)$$

$$x_{\max} < b < M, \quad (3.35)$$

где $x_{\min}$ и $x_{\max}$ — соответственно минимальное и максимальное наблюдаемые значения. В качестве M принимается такое значение, которое заведомо является невероятным, т.е. никогда не будет превышено. Например, можно принять $M = 5x_{\max}$. Эта процедура необходима, чтобы заранее отсеять физически не обоснованные решения.

¹ Учитывая большой объем вычислений метод может быть реализован только на ЭВМ.

Затем для нескольких вариантов параметра a , например, $a_0 = 0$, $a_1 = 0,1x_{\min}$, $a_2 = 0,2x_{\min}$, $a_3 = 0,3x_{\min}$ и т.д., производится оптимизация параметра b , т.е. для каждого a_j находят такое значение параметра $b = b_j$, при котором имеет место наибольшее соответствие эмпирической и аналитической кривых обеспеченностей. После чего выбирается наилучшая пара значений a и b (для которой степень соответствия эмпирической и аналитической кривых максимальна).

Степень соответствия эмпирической и аналитической кривых можно оценивать по-разному. Так, Г.Г. Сванидзе [21] рекомендует использовать для этой цели критерий согласия¹ *Хи-квадрат*, а В.А. Шелутко [25] – критерий согласия $n\omega^2$.

Мы предлагаем другой вариант. В качестве оптимального будем принимать такое значение параметра b , при котором коэффициент асимметрии СВ Z равен или близок нулю. Такой подход, на наш взгляд, является наиболее естественным, так как по условию СВ Z должна быть нормально распределенной, т.е. иметь нулевую асимметрию.

Однако как в первом случае (при использовании критериев согласия), так и во втором основная проблема оптимизации состоит в том, что можно получить несколько решений (несколько пар a и b), при которых степень согласия эмпирической и аналитической кривых обеспеченностей будет достаточно высокой, и выбор окончательного решения носит субъективный характер. В то же время формальный подход на основе абсолютного минимума критерия качества² может вообще привести к абсурдным результатам из-за выборочного характера исходных данных. В силу указанных причин физический смысл параметров a и b , как абсолютного минимума и абсолютного максимума, размывается, и они превращаются в формальные параметры распределения. Именно поэтому четырехпараметрическое распределение Джонсона не нашло широкого распространения в практике гидрологических расчетов.

Если хотя бы один из параметров (a или b) известен, проблем с оптимизацией не возникает.

¹ О критериях согласия см. п. 4.7.

² Критерий качества – любой числовой показатель, характеризующий степень отклонения эмпирических данных от оптимизируемой аналитической кривой.

Пример 3.8. Для ряда максимальных расходов дождевых паводков р. Луги у ст. Толмачево (см. табл. 1.2) оценить параметры распределения. В качестве аппроксимирующей кривой использовать кривую обеспеченностей Джонсона.

Решение.

1. Ориентируясь на систему ограничений (3.34) – (3.35), приближенно определяем область возможных значений параметров a и b : $0 \leq a < 17,3$; $145 < b < (145 \cdot 5 = 725)$.

2. Для различных значений параметра a ($a_0 = 0, a_1 = 1, a_2 = 2, \dots, a_{17} = 17$) находим такие значения параметра b ($b_0, b_1, b_2, \dots, b_{17}$), при которых для преобразованного ряда значений $z_i = \ln [(x_i - a) / (b - x_i)]$ выборочный коэффициент асимметрии близок к нулю. Результаты такой оптимизации даны в табл. 3.10.

Таблица 3.10

Результаты оптимизации параметров a и b распределения Джонсона для максимальных паводочных расходов на р. Луге у ст. Толмачево

Оптимизируемый параметр		$C_{s,z}^*$	Оптимизируемый параметр		$C_{s,z}^*$
a	b		a	b	
0	Нет решения	–	11	490	0,00
5	Нет решения	–	12	365	0,00
10	770	0,00	15	202	0,00

3. В качестве расчетных параметров принимаем $a = 11$; $b = 490$, так как это ближайшее к нулю значение a , при котором $b < M$.

Замечание. Поскольку значение M назначено достаточно произвольно, не исключено, что решением будет пара $a = 10$; $b = 770$ или пара $a = 12$; $b = 365$. Таким образом, в ситуации, когда дополнительная информация о параметрах a и b отсутствует, выбор окончательного решения всегда будет достаточно субъективным.

4. Преобразуем исходный ряд по формуле

$$z_i = \ln [(x_i - 11) / (490 - x_i)].$$

5. Используя формулы (3.13) и (3.16), по ряду значений z_i вычисляем $\bar{z}$ и σ_z . В данном случае $\bar{z} = -2,54$; $\sigma_z = 0,95$. Дальнейший расчет ведется по схеме, изложенной в п. 2.7 (см. пример 2.8).

3.3. Оценка погрешностей выборочных параметров распределения

Оценка параметра распределения некоторой СВ X , полученная по выборке тем или иным методом, сама представляет собой случайную величину, обладающую определенным разбросом. Имеется в виду, что можно получить сколько угодно выборок объемом n из одной и той же генеральной совокупности, и все они будут давать различные значения оцениваемого параметра. Поэтому точечная оценка параметра G в виде конкретного значения G^* не дает полного представления об искомом параметре без оценки погрешности ее вычисления. Мерой случайной погрешности для выборочного параметра G^* может служить среднеквадратическое отклонение σ_G (абсолютная погрешность) или относительное среднеквадратическое отклонение $\varepsilon_G = (\sigma_G/G)100\%$ (относительная погрешность)¹. Формулы для расчета погрешностей зависят от того, каким методом производилась оценка параметра.

Метод моментов. Абсолютная погрешность выборочного среднего определяется по формуле

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}, \quad (3.36)$$

где σ – СКО выборки; n – объем выборки (длина ряда).

Относительная погрешность выборочного среднего:

$$\varepsilon_{\bar{x}} = \frac{\sigma_{\bar{x}}}{\bar{x}} 100 = \frac{\sigma}{x\sqrt{n}} 100 = \frac{C_v}{\sqrt{n}} 100\%, \quad (3.37)$$

где C_v – выборочный коэффициент вариации.

Для расчета абсолютной и относительной погрешностей коэффициента вариации используются формулы

$$\sigma_{c_v} = \frac{C_v \sqrt{1 + aC_v^2}}{\sqrt{2n}}, \quad (3.38)$$

$$\varepsilon_{c_v} = \frac{\sqrt{1 + aC_v^2}}{\sqrt{2n}} 100\%, \quad (3.39)$$

¹ Иногда вместо термина “погрешность” используется термин “ошибка”, а также термины “стандартная погрешность” и “стандартная ошибка”.

где $a = 2$ для нормального распределения и $a = 1$ для двухпараметрического гамма-распределения.

Напомним, что у двухпараметрического гамма-распределения соотношение $C_s/C_v = 2$, а у нормального распределения $C_s = 0$. Так как гидрологические ряды имеют, как правило, умеренную положительную асимметрию, эти формулы рекомендуется использовать при $a = 1$ [19].

Помимо формулы (3.38) в российской гидрологической практике достаточно широкое распространение получила формула Е.Г. Блохинова:

$$\sigma_{c_v} = \frac{C_v}{n + 4C_v^2} \sqrt{\frac{n(1 + C_v^2)}{2}}. \quad (3.40)$$

Данная формула также получена для двухпараметрического гамма-распределения путем введения эмпирической поправки в формулу (3.38).

Для оценки погрешности коэффициента асимметрии предложены следующие формулы:

формула С.Н. Крицкого и М.Ф. Менкеля

$$\sigma_{c_s} = \sqrt{(6/n)(1 + 6C_v^2 + 5C_v^4)}, \quad (3.41)$$

$$\varepsilon_{c_s} = \frac{1}{C_s} \sqrt{(6/n)(1 + 6C_v^2 + 5C_v^4)} 100 \% ; \quad (3.42)$$

формула А.Ш. Резниковского

$$\sigma_{c_s} = \sqrt{(6/n)(1 + C_v^2)}, \quad (3.43)$$

$$\varepsilon_{c_s} = \frac{1}{C_s} \sqrt{(6/n)(1 + C_v^2)} 100 \% . \quad (3.44)$$

Формулы (3.41) и (3.43) разработаны для случая $C_s = 2C_v$, причем формула (3.41) получена теоретическим путем, а формула (3.43) на основании материалов статистического моделирования.

Для распределений, у которых коэффициент асимметрии близок к нулю, формулы (3.41) – (3.44) не применимы, так как множитель $1/C_s$ стремится к бесконечности. В этом случае относительная погрешность C_s вообще не вычисляется.

Метод наибольшего правдоподобия.

Абсолютная и относительная погрешность среднего значения и коэффициента асимметрии определяется так же, как и для метода моментов: по формулам (3.36) – (3.37); (3.41) – (3.42).

Для приближенной оценки погрешности коэффициента вариации в настоящее время рекомендуется применять формулы

$$\sigma_{C_v} = \frac{C_v}{\sqrt{2n}} \sqrt{\frac{3}{3 + C_v^2}}, \quad (3.45)$$

$$\varepsilon_{C_v} = \sqrt{\frac{3}{2n(3 + C_v^2)}} 100 \%. \quad (3.46)$$

В гидрологической практике принято считать, что расчет параметров распределения выполнен надежно, если погрешность среднего значения не превышает 10%, а коэффициента вариации 15%.

Ошибка коэффициента асимметрии при имеющейся в настоящее время длине рядов довольно высока и может достигать нескольких десятков процентов, поэтому величина C_s назначается в соответствии с рекомендациями, изложенными в п. 3.2.1.

Метод квантилей.

При использовании этого метода абсолютные и относительные погрешности параметров распределения не рассчитываются. Критерием правильности подбора аналитической кривой в этом случае служит условие выполнения неравенства

$$|\bar{x} - \bar{x}_r| < 0,02\bar{x}, \quad (3.47)$$

где $\bar{x}$ – среднее значение ряда, полученное по выборке; $\bar{x}_r$ – среднее значение, полученное путем расчета графоаналитическим методом.

При выполнении неравенства (3.47) предполагается, что оценка параметров аналитической кривой обеспеченностей произведена с достаточной точностью.

3.4. Экономические аспекты определения расчетных вероятностей

Как уже отмечалось в п. 3.1 одной из основных задач гидрологических расчетов при строительстве различного рода сооружений, связанных с использованием водных ресурсов, является определение той или иной гидрологической характеристики расчетной вероятности превышения (обеспеченности). В качестве примера рассмотрим в общих чертах схему определения расчетных обеспеченностей максимальных расходов паводков.

3.4.1. Цели определения расчетных вероятностей

При проектировании любого гидротехнического сооружения необходимо учитывать возможность риска, связанного с прохождением паводков, так как всегда существует вероятность образования такого расхода воды, который превысит проектные возможности сооружения и причинит ущерб как самому сооружению, так и окружающей местности. Поэтому первым шагом при проектировании является изучение повторяемости паводков и получения кривой обеспеченностей максимальных расходов воды. Следующим шагом является определение оптимальной расчетной обеспеченности максимального расхода воды, гарантирующего нормальную работу сооружения в период его строительства и эксплуатации.

При этом расчетный паводок, закладываемый в проект, должен удовлетворять следующим условиям.

1. Сооружение должно быть осуществимо с экономической точки зрения. Это означает, что затраты на строительство и эксплуатацию его не должны превышать прямые и косвенные затраты в случае возможного ущерба и каких-либо неблагоприятных воздействий для жизни людей и самого сооружения.

2. Средние годовые затраты на сооружение должны быть сведены к минимуму. К средним годовым затратам относятся: а) годовые расходы на строительство сооружения пропорционально продолжительности его эксплуатации; б) годовые расходы на эксплуатацию и ремонт сооружения; в) средний годовой ущерб, причиняемый паводками в проектных условиях (при строительстве сооружения или при эксплуатации), включая расходы на его восстановление в случае повреждения во время прохождения паводка.

3.4.2. Определение оптимальных расчетных вероятностей

В практике России средние годовые затраты U на строительство гидротехнических сооружений включают ряд отдельных элементов [17].

К первому элементу U_1 относится годовая стоимость строительства новых сооружений, рассчитанных на срок службы T (лет):

$$U_1 = \frac{K_0 + \alpha Q_p}{T}, \quad (3.48)$$

где K_0 – доля расходов на строительство сооружения независимо от расчетного максимального расхода воды; α – коэффициент, представляющий собой увеличение расходов на строительство при каждом приращении расхода воды; Q_p – максимальный расход воды, соответствующей ежегодной вероятности превышения p .

Второй элемент U_2 представляет собой годовые затраты на эксплуатацию и текущий ремонт сооружения (U_2 принимается пропорционально общим затратам на строительство, $K_0 + \alpha Q_p$)

$$U_2 = r(K_0 + \alpha Q_p), \quad (3.49)$$

где r – коэффициент пропорциональности между годовыми затратами на эксплуатацию сооружения и общей стоимостью строительства.

Третий элемент U_3 выражает среднюю годовую стоимость ущерба в проектных условиях, включая повреждение сооружения:

$$U_3 = p[Y_0 + \eta(K_0 + \alpha Q_p)], \quad (3.50)$$

где p – вероятность превышения максимального расхода Q_p ; Y_0 – годовая стоимость ущерба, не считая ущерба, наносимого сооружению при прохождении максимального расхода расчетной вероятности превышения; η – коэффициент пропорциональности между затратами на восстановление проектного сооружения (в случае его повреждения) и общими затратами на строительство.

В формулах (3.48) – (3.50) переменная Q_p обозначает максимальный расход воды, соответствующей вероятности превышения p , определяемый по кривой обеспеченностей, построенной для места проектирования сооружения. Аналитически Q_p можно выразить в виде уравнения

$$Q_p = A\bar{Q}p^{-n}, \quad (3.51)$$

где $\bar{Q}$ – средний многолетний максимальный расход воды; A и n – константы, значения которых зависят от характеристик распределения максимальных паводочных расходов.

Подставляя $(A\bar{Q}p^{-n})$ вместо Q_p в уравнения (3.48) – (3.50) и суммируя U_1 , U_2 и U_3 , чтобы получить U , находим общую годовую сумму затрат, отнесенную к одному сооружению:

$$U = K_0\left(\frac{1}{T} + r\right) + p(\eta K_0 + Y_0) + p^{-n}A\bar{Q}\alpha\left(\frac{1}{T} + r\right) + p^{1-n}A\bar{Q}\eta\alpha. \quad (3.52)$$

Минимальная общая годовая стоимость соответствует обеспеченности p , при которой производная dU/dp становится равной нулю:

$$\frac{dU}{dp} = \eta K_0 + Y_0 - np^{-(n+1)}A\bar{Q}\alpha\left(\frac{1}{T} + r\right) + (1-n)p^{-n}A\bar{Q}\eta\alpha = 0. \quad (3.53)$$

Уравнение (3.53) решается относительно p ; получаемое значение представляет собой оптимальную вероятность паводка, которую следует использовать при расчете.

Практика гидротехнического проектирования США в основном та же; главное отличие заключается в том, что величины U_1 , U_2 и U_3 не определяются по генерализованным связям, а рассчитываются на основе индивидуального изучения каждого отдельного проекта гидротехнического сооружения.

Не вдаваясь в детали определения элементов общих годовых затрат U , отметим, что решение задачи сводится к определению минимума функции, связывающей годовые затраты с вероятностью превышения максимального расхода паводка.

Расчет производится по следующей схеме.

1. Строится график, связывающий общие годовые затраты $(U_1 + U_2 + U_3)$ с заданной вероятностью ежегодного превышения (рис. 3.7). При этом вероятность превышения откладывается на оси абсцисс по схеме, принятой в США, т.е. с уменьшением слева направо. Крайняя левая точка соответствует вероятности превышения наименьшего расхода, который вызовет разрушения гидротехнического сооружения в предпроектных условиях. Численное значение этой точки на оси ординат соответствует $U_{3(0)}$, или общему годовому ущербу в предпроектных условиях (когда сооружение еще не построено).

2. Нанесенные на график точки соединены плавной кривой. Оптимальная расчетная вероятность соответствует значению, при котором общие годовые затраты составляют минимум, т.е. точке на графике, в которой уклон кривой равен нулю. По форме кривой видно, что на довольно большом участке вероятностей разница в годовых затратах незначительная.

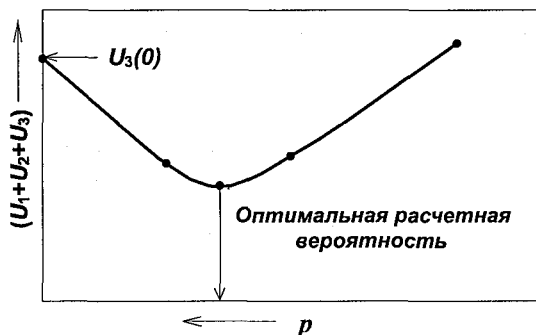


Рис. 3.7. График для оценки оптимальной расчетной вероятности превышения максимальных расходов воды.

Описанный анализ производится для каждого значительного сооружения.

Расчетные уравнения, о которых речь шла выше, включают экономические параметры, определение которых часто затруднительно, а порой вообще невозможно. В силу этого, как правило, технико-экономическое обоснование расчетных вероятностей превышения паводочного стока сводится к общим методическим построениям. В ряде стран вошло в практику определение расчетных вероятностей превышения на основе: а) типа и значимости гидротехнических сооружений и б) степени экономического развития окружающей территории, которая бы потерпела потенциальный ущерб от паводка в случае недостаточной разработки проекта гидротехнического сооружения.

В табл. 3.11 приводится расчетная вероятность превышения, используемая для каждого класса капитальности гидротехнических сооружений.

В табл. 3.12 приводятся используемые в Российской Федерации классы капитальности гидроэнергетических, навигационных и других гидротехнических сооружений.

Таблица 3.11

Связь расчетной вероятности превышения максимальных паводков с классом капитальности гидротехнического сооружения (практика России)

Класс капитальности гидротехнического сооружения	Расчетная вероятность, %	Период повторяемости *, число лет
I	0,01	10 000
II	0,1	1000
III	0,5	200
IV	1,0	100

Примечание. * Период повторяемости в годах является эквивалентом вероятности превышения.

Таблица 3.12

Классы капитальности гидроэнергетических, навигационных и других гидротехнических сооружений

Гидротехническое сооружение	Номер класса	
	Основные сооружения	Второстепенные сооружения
Первая категория		
Гидроэлектростанции мощностью более 1 млн. кВт	I	III
Гидротехнические сооружения на основных судоходных внутренних водных путях	II	III
Речные порты с годовым грузооборотом более 3 млн. т	II	III
Вторая категория		
Гидроэлектростанции мощностью от 301 тыс. до 1 млн. кВт	II	III
Гидротехнические сооружения на судоходных внутренних водных путях	III	IV
Речные порты с годовым грузооборотом от 701 тыс. до 3 млн т	III	IV
Третья категория		
Гидроэлектростанции мощностью от 50 до 300 тыс. кВт	III	IV
Гидротехнические сооружения на внутренних водных путях местного значения	III	IV
Речные порты с годовым грузооборотом от 150 до 700 тыс. т	III	IV
Четвертая категория		
Гидроэлектростанции мощностью менее 50 тыс. кВт	IV	IV
Гидротехнические сооружения на малых внутренних водотоках местного значения	IV	IV
Речные порты и пристани с годовым грузооборотом менее 150 тыс. т	IV	IV

4. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Для понимания материала, который излагается в этой главе, рассмотрим три теоремы математической статистики. Суть этих теорем состоит в определении закона распределения для случайной величины, которая является функцией других случайных величин.

4.1. Распределение χ^2 (хи-квадрат)

Теорема 4.1. Если X_i – независимые случайные величины, подчиняющиеся нормальному распределению, у которых m_x равно нулю, а σ_x равно единице, то случайная величина $\chi^2 = \sum_{i=1}^{\nu} X_i^2$ подчиняется распределению χ^2 (хи-квадрат) с ν степенями свободы.

Распределение χ^2 определяется одним параметром ν , который называется числом степеней свободы (значение ν равно числу независимых случайных величин под знаком суммы).

Плотность вероятности равна

$$f(\chi^2) = \begin{cases} \left[2^{\frac{\nu}{2}} \Gamma(\nu/2) \right]^{-1} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0, \end{cases} \quad (4.1)$$

где $\Gamma(\cdot)$ – гамма-функция; x – значение случайной величины χ^2 .

График плотности вероятности распределения хи-квадрат приведен на рис. 4.1.

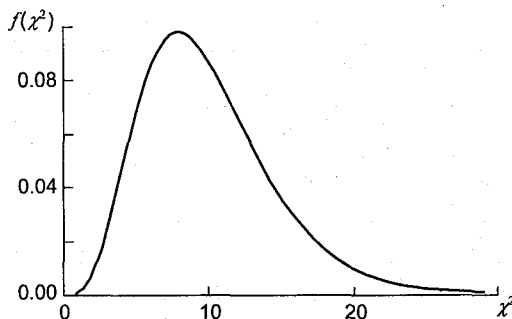


Рис. 4.1. Функция плотности вероятности распределения хи-квадрат при числе степеней свободы $\nu = 10$.

Математическое ожидание и дисперсия распределения χ^2 равны:

$$m_x = \nu; \quad (4.2)$$

$$D_x = 2\nu. \quad (4.3)$$

Медиана может быть определена приближенным равенством

$$Me \approx \nu - 0,66. \quad (4.4)$$

Мода при $\nu \geq 2$ равна

$$Mo = \nu - 2. \quad (4.5)$$

При $\nu = 1$ мода отсутствует, так как $f_\nu = \infty$ при $x = 0$.

При увеличении значения ν распределение χ^2 медленно приближается к нормальному.

В приложении 4 приводятся квантили распределения χ^2 для ν от 1 до 30. При $\nu > 30$ можно использовать аппроксимацию

$$\chi_p^2(\nu) \approx \frac{1}{2} \left(\sqrt{2\nu - 1} + t'_p \right)^2, \quad (4.6)$$

где t'_p — квантиль нормального распределения с $m_t = 0$ и $\sigma_t = 1$ (см. табл. 2.1); p — вероятность непревышения.

Эта аппроксимация не годится при p , близких к 0 или 100 %. Лучшая аппроксимация в этом случае дается формулой

$$\chi_p^2(\nu) \approx \nu \left[1 - 2/(9\nu) + t'_p \sqrt{2/(9\nu)} \right]^3. \quad (4.7)$$

Из теоремы 4.1, в частности, следует, что величина $(n-1)[S_x^2/\sigma_x^2]$ имеет распределение χ^2 с $(n-1)$ степенями свободы (S_x^2 и σ_x^2 — соответственно выборочная и теоретическая дисперсии).

4.2. *t*-распределение (Стьюдента)

Теорема 4.2. Если Z — нормированная нормально распределенная СВ, а U — независимая от Z СВ, подчиненная распределению χ^2 с ν степенями свободы, тогда СВ $t = Z\sqrt{\nu/U}$ подчиняется распределению Стьюдента (У. Госсета) с ν степенями свободы.

Распределение Стьюдента называют также *t*-распределением (приложение 5). Плотность вероятности этого распределения определяется равенством

$$f_v(t) = c(v) [1 + t^2 / v]^{-(v+1)/2}, \quad (4.8)$$

где $c(v)$ – параметр, зависящий от числа степеней свободы:

$$c(v) = \Gamma[(v+1)/2] / [\sqrt{v\pi} \Gamma(v/2)]; \quad (4.9)$$

$\Gamma(\cdot)$ – гамма-функция; π – число «пи».

Распределение Стьюдента симметрично. График функции плотности вероятности приведен на рис. 4.2.

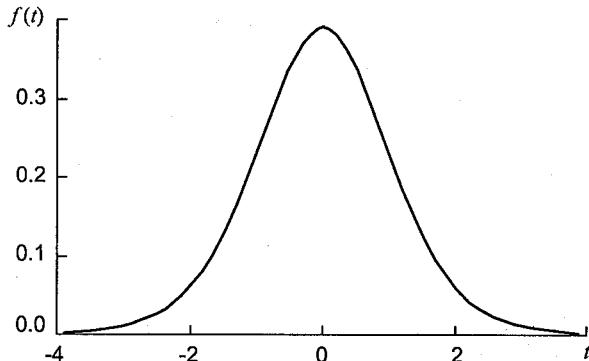


Рис. 4.2. Функция плотности вероятности распределения Стьюдента при числе степеней свободы $v = 10$.

Математическое ожидание m_t , дисперсия D_t и среднее квадратическое отклонение σ_t равны:

$$m_t = 0, \quad v > 1; \quad (4.10)$$

$$D_t = \sigma_t^2 = v/(v-2), \quad v > 2. \quad (4.11)$$

При $v = 1$ распределение Стьюдента приводит к распределению Коши, у которого не существует математического ожидания и дисперсии. С увеличением v распределение Стьюдента асимптотически приближается к нормальному распределению с параметрами $m_t = 0$ и $\sigma_t = 1$.

Из теоремы 4.2, в частности, следует, что величина $(\bar{x} - m_x)/(S/\sqrt{n})$ имеет распределение Стьюдента, где $\bar{x}$ и S – выборочные среднее и СКО, n – длина выборки.

4.3. F-распределение (Фишера)

Теорема 4.3. Если Z и U независимые СВ, обладающие χ^2 -распределением с ν_1 и ν_2 степенями свободы, то случайная величина $F = (Z/\nu_1) / (U/\nu_2)$ имеет распределение Фишера с ν_1 и ν_2 степенями свободы. Это распределение также называется F -распределением (приложение 6).

Плотность вероятности F -распределения имеет вид:

$$f(F) = \begin{cases} c_1(\nu_1, \nu_2) F^{\frac{\nu_1-2}{2}} / \left(1 + \frac{\nu_1}{\nu_2} F\right)^{\frac{\nu_1+\nu_2}{2}} & \text{при } x > 0; \\ 0 & \text{при } x \leq 0, \end{cases} \quad (4.12)$$

где $c_1(\nu_1, \nu_2)$ – параметр, зависящий от ν_1 и ν_2 :

$$c_1(\nu_1, \nu_2) = \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \Gamma\left[\frac{\nu_1 + \nu_2}{2}\right] / \left[\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right)\right]. \quad (4.13)$$

График плотности вероятности $f(F)$ приведен на рис. 4.3.

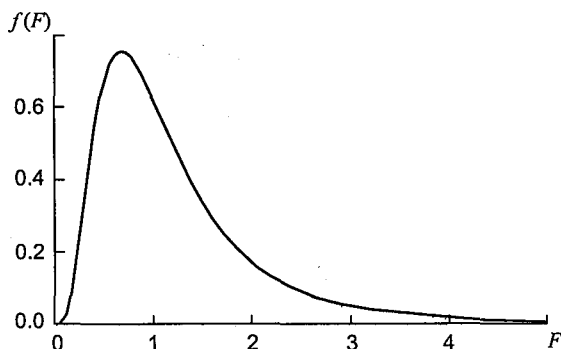


Рис. 4.3. График плотности вероятности F -распределения при числе степеней свободы $\nu_1 = 10$ и $\nu_2 = 10$.

Математическое ожидание, дисперсия и мода соответственно равны:

$$m_F = \nu_2 / (\nu_2 - 2), \quad \nu_2 > 2; \quad (4.14)$$

$$D_F = 2\nu_2^2(\nu_1 + \nu_2 - 2)/[\nu_1(\nu_2 - 2)^2(\nu_2 - 4)], \quad \nu_2 > 4; \quad (4.15)$$

$$Mo = \nu_2(\nu_1 - 2)/[\nu_1(\nu_2 + 2)], \quad \nu > 1. \quad (4.16)$$

Из теоремы следует, что отношение выборочных дисперсий S_1^2 / S_2^2 двух выборок длиной m и n будет иметь F -распределение с числом степеней свободы соответственно $\nu_1 = (m - 1)$ и $\nu_2 = (n - 1)$.

4.4. Интервальные оценки параметров распределения

Точечные оценки представляют собой числа, полученные путем подстановки выборочных значений СВ $x_1, x_2, \dots, x_n$ в формулу для оцениваемого параметра. Например, точечной оценкой математического ожидания m_x является выборочное среднее $\bar{x}$, а точечной оценкой дисперсии D является выборочная дисперсия D^* . Однако точечные оценки параметров не дают информации о степени близости оценки G^* к оцениваемому параметру G . Подчеркнем, что несмещенная оценка G^* является случайной величиной, которая имеет математическое ожидание, равное G , но по отдельным выборкам из генеральной совокупности величина G^* может существенно отличаться от G как в большую, так и в меньшую сторону.

Более информативный способ оценивания неизвестных параметров состоит не в определении единичного точечного значения, а в построении интервала, в котором с заданной вероятностью окажется оцениваемый параметр, т.е. в построении так называемой интервальной оценки параметра G .

Интервальной оценкой параметра G называется интервал, границы которого l_1^* и l_2^* являются функциями выборочных значений $x_1, x_2, \dots, x_n$, и который с заданной вероятностью p накрывает оцениваемый параметр G :

$$P\{l_1^* < G \leq l_2^*\} = p. \quad (4.17)$$

Интервал (l_1^*, l_2^*) называется доверительным интервалом, а величина p — доверительной вероятностью. В качестве p наиболее часто используются значения: 0,9; 0,95 и 0,99.

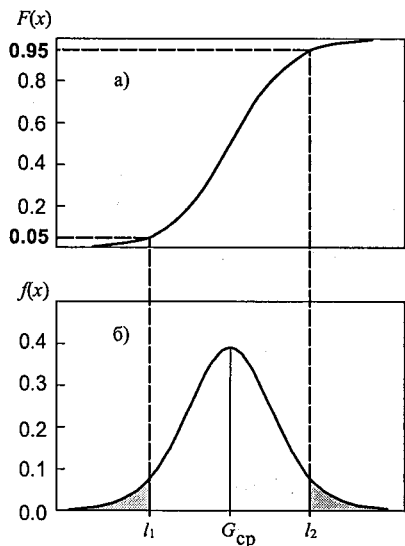


Рис. 4.4. 90 %-ный доверительный интервал (l_1, l_2) для параметра G относительно интегральной функции распределения (а) и функции плотности вероятности (б).

Используя функцию распределения выборочных значений параметра G , можно записать вероятности непревышения для l_1 и l_2 :

$$P\{G^* \leq l_1\} = F(l_1) = (1 - p)/2; \quad (4.18)$$

$$P\{G^* \leq l_2\} = F(l_2) = p + (1 - p)/2 = (1 + p)/2. \quad (4.19)$$

Например, если рассматривается 90 %-ный доверительный интервал ($p = 0,9$), то $F(l_1) = 0,05$, $F(l_2) = 0,95$ или соответственно 5 и 95 % (рис. 4.4 а). На рис. 4.4 б показан тот же доверительный интервал на графике функции плотности вероятности. Не заштрихованная площадь на этом рисунке составляет 90 % от общей площади подграфика.

4.4.1. Интервальная оценка математического ожидания

Для построения интервальной оценки математического ожидания воспользуемся следствием из теоремы 4.2. В соответствии с этой теоремой величина $(\bar{x} - m_x)/(S/\sqrt{n})$ имеет распределение Стьюдента с числом степеней свободы $\nu = n - 1$. Таким образом, можно записать, что с вероятностью p она попадает в интервал

$$t'_{(1-p)/2} \leq [(\bar{x} - m_x)/(S_x/\sqrt{n})] < t'_{(1+p)/2},$$

где $t'_{(1-p)/2}$ и $t'_{(1+p)/2}$ – квантили распределения Стьюдента, соответствующие вероятностям непревышения $(1 - p)/2$ и $(1 + p)/2$. Однако учитывая, что распределение Стьюдента симметрично относительно нуля $t_{(1-p)/2} = -t_{(1+p)/2}$ и, следовательно,

$$-t_{(1+p)/2} \leq [(\bar{x} - m_x)/(S_x/\sqrt{n})] < t_{(1+p)/2}.$$

После преобразования получаем:

$$\bar{x} - \frac{t_{(1+p)/2} S_x}{\sqrt{n}} \leq m_x < \bar{x} + \frac{t_{(1+p)/2} S_x}{\sqrt{n}}. \quad (4.20)$$

Пример 4.1. По выборке длиной $n = 37$ лет вычислены среднее значение и среднеквадратическое отклонение $\bar{x} = 12,0$; $S_x = 6,00$. Найти интервальную оценку математического ожидания (m_x) в виде 90%-ного доверительного интервала.

Решение.

1. Для 90 %-ного доверительного интервала $p = 0,9$, следовательно, $t'_{(1+p)/2} = t'_{95}$.

По таблице t -распределения Стьюдента (см. приложение 5) при $\nu = 36$ получаем $t'_{95} = 1,69$.

2. Подставляем это значение в формулу (4.20):

$$12 - (1,69 \cdot 6,00) / \sqrt{36} \leq m_x < 12 + (1,69 \cdot 6,00) / \sqrt{36}.$$

3. В окончательном виде 90 %-ный доверительный интервал для m_x будет иметь вид $10,31 \leq m_x < 13,69$.

4.4.2. Интервальная оценка дисперсии

Для построения интервальной оценки дисперсии воспользуемся теоремой 4.1. Из этой теоремы следует, что величина $[(n-1)S_x^2] / \sigma_x^2$ имеет распределение χ^2 с $(n-1)$ степенями свободы, и, следовательно, можно записать:

$$\chi_{(1-p)/2}^2 \leq \frac{(n-1)S_x^2}{\sigma_x^2} < \chi_{(1+p)/2}^2, \quad (4.21)$$

где $S_x^2 = D^*$ – выборочная дисперсия; $\sigma_x^2 = D$ – фактическая дисперсия; n – длина выборки.

После преобразований из (4.21) получаем:

$$\frac{(n-1)S_x^2}{\chi_{(1+p)/2}^2} < D \leq \frac{(n-1)S_x^2}{\chi_{(1-p)/2}^2}. \quad (4.22)$$

Из выражения (4.22) можно получить также интервальную оценку среднеквадратического отклонения:

$$S_x \sqrt{(n-1)/\chi_{(1+p)/2}^2} < \sigma \leq S_x \sqrt{(n-1)/\chi_{(1-p)/2}^2}. \quad (4.22a)$$

Пример 4.2. По выборке длиной $n = 26$ лет установлено, что $D^* = 147$. Найти интервальную оценку дисперсии в виде 95 %-ного доверительного интервала.

Решение.

1. Для 95 %-ного доверительного интервала $p = 0,95$, следовательно, $\chi_{(1+p)/2}^2 = \chi_{97,5}^2$; $\chi_{(1-p)/2}^2 = \chi_{2,5}^2$. По таблице распределения χ^2 (см. приложение 4) при $\nu = n - 1 = 25$ получаем $\chi_{97,5}^2 = 40,65$, $\chi_{2,5}^2 = 13,12$.

2. Подставляем эти значения в формулу (4.22): $(25 \cdot 147) / 40,65 < D \leq (25 \cdot 147) / 13,2$, в окончательном виде: $90,4 < D \leq 280$.

Пример 4.3. Дано: $n = 54$; $D^* = 312$. Требуется найти интервальную оценку дисперсии в виде 90 %-ного доверительного интервала.

Решение.

1. При $p = 0,9$ $\chi_{(1+p)/2}^2 = \chi_{95}^2$, $\chi_{(1-p)/2}^2 = \chi_5^2$. Так как в данном случае число степеней свободы $\nu = n - 1 = 53$, т.е. больше 30, вместо таблицы, приведенной в приложении 4, используем формулу (4.7): $\chi_p^2 \approx 53 \left[1 - 2 / (9 \cdot 53) + t_p \sqrt{2 / (9 \cdot 53)} \right]^3$.

2. Значения квантилей нормального распределения определяем по табл. 2.1: $t'_{95} = 1,64$; $t'_5 = -1,64$; подставляя их в формулу, получаем: $\chi_{95}^2 = 70,9$; $\chi_5^2 = 37,3$.

3. Подставляем эти значения в формулу (4.22): $(53 \cdot 312) / 70,9 < D \leq (53 \cdot 312) / 37,3$; $233 < D \leq 443$.

4.5. Статистическая гипотеза.

Критерий статистической гипотезы

Статистическая гипотеза есть некоторое предположение относительно свойств генеральной совокупности, из которой извлечена выборка. Например, мы можем предположить, что математическое ожидание m_x генеральной совокупности равно 12. Тогда среднее значение выборки $x_1, x_2, x_3, \dots, x_n$ (если n достаточно велико) должно мало отличаться от 12. В этом случае говорят: «Мы приняли нулевую гипотезу $H_0: m_x = 12$ ». В качестве альтернативных гипотез

тез (H_1) здесь можно использовать следующие: $H_1: m_x \neq 12$ или $H_1: m_x > 12$ или $H_1 < 12$.

Для проверки гипотез используются специальные тесты, которые называются *критериями*.

Критерий (тест) статистической гипотезы — это правило, позволяющее принять или отвергнуть данную гипотезу.

Для реализации теста используются определенные функции результатов $g(x_1, x_2, x_3, \dots, x_n)$, называемые *статистиками*.

Критерий и статистика часто имеют одинаковые названия, например критерий Стьюдента (t -критерий) базируется на использовании статистики Стьюдента, а критерий Фишера (F -критерий) — на статистике Фишера.

Продолжим рассмотрение примера. Итак, мы предполагаем, что $m_x = 12$. В то же время в нашем распоряжении имеется конкретная выборка с параметрами $\bar{x}$, S_x . Если наше предположение верно, то разница $(\bar{x} - m_x)$ должна быть достаточно малой. Эту разницу можно рассматривать в качестве анализируемой статистики. Однако на практике используют другую статистику: $t = (\bar{x} - m_x) / (S_x / \sqrt{n})$. Статистика t является более удобной, так как заранее известно (теорема 4.2), что она подчинена распределению Стьюдента. На использовании этой статистики базируется критерий Стьюдента, который можно применить для проверки нашей нулевой гипотезы $H_0: m_x = 12$. С этой целью для конкретной реализации рассчитывают эмпирическое значение статистики Стьюдента: t^* . Если, например, $n = 36$, $\bar{x} = 11,0$, $S = 5,00$, то $t^* = (\bar{x} - m_x) / (S_x / \sqrt{n}) = (11 - 12) / (5 / \sqrt{36}) = 1,2$.

Величина t является случайной величиной, и для различных выборок данной длины значение t^* будет различным. Область возможных значений (ОВЗ) этой статистики — вся числовая ось. ОВЗ делится на две части: область принятия гипотезы и критическую область (рис. 4.5). Если t^* попадает в область принятия гипотезы, H_0 не опровергается, если в критическую область, H_0 опровергается.

Область принятия гипотезы называется *доверительной областью* (*доверительным интервалом*).

Вероятность по произвольной выборке получить t^* , которая попадает в доверительную область, геометрически равна незаштрихованной площади на рис. 4.5. Эта вероятность называется *доверительной вероятностью* (p_d).

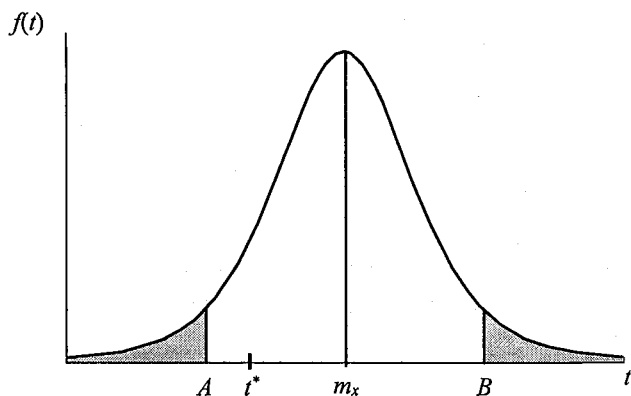


Рис. 4.5. Область принятия гипотезы: $[A, B]$ и критическая область: $(-\infty, A) + (B, +\infty)$ для статистики Стьюдента.

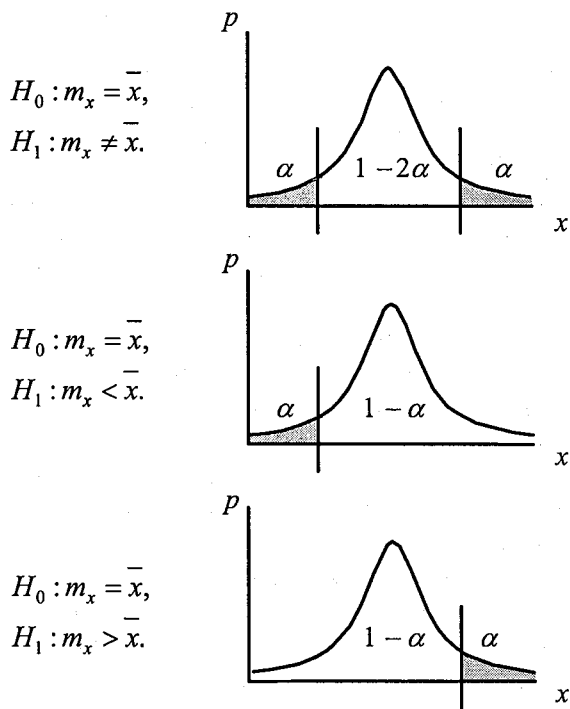


Рис. 4.6. Уровень значимости и доверительная вероятность для различных альтернативных гипотез.

Вероятность попадания t^* в критическую область выражается равенством $\alpha = 1 - p_d$. Вероятность α называют *уровнем значимости*.

Если критическая область состоит из двух частей, то вместо α пишут $2\alpha = 1 - p_d$, где 2α указывает на то, что уровень значимости двусторонний. Геометрически уровень значимости равен заштрихованной площади (см. рис. 4.5). Различие между односторонним и двусторонним уровнями значимости показано на рис. 4.6.

При использовании статистических критериев всегда есть возможность совершить ошибку. Типы возможных ошибок представлены в табл. 4.1.

Таблица 4.1

Типы возможных ошибок при проверке статистических гипотез

Гипотеза	Объективно верна	Объективно не верна
Принимается	Правильное решение	Ошибка II рода
Отвергается	Ошибка I рода	Правильное решение

С этой точки зрения *уровень значимости можно трактовать как вероятность совершить ошибку I рода*. Поэтому на практике в качестве уровня значимости принимается такое достаточно малое значение вероятности, которое в данном случае характеризует практически невероятное (маловероятное) событие. Вообще говоря, назначение уровня значимости не является математической задачей и устанавливается обычно, исходя из тех последствий, которые возможны вследствие совершения ошибки при принятии или отклонении гипотезы. В гидрологической практике наиболее часто используются уровни значимости 5 и 10 %.

Если для нашего примера принять уровень значимости $2\alpha = 10\%$, то, используя таблицу распределения Стьюдента (см. приложение 5) при $\nu = n - 1 = 35$, получаем теоретическое значение t -статистики: $t = 1,69$. Это означает, что доверительная область представлена отрезком $[-1,69, +1,69]$.

Так как $t^* = -1,20$ и, значит, попадает в доверительную область, наша нулевая гипотеза не опровергается, т.е. с высокой долей вероятности можно считать, что m_x действительно равно 12, а разница между m_x и $\bar{x}$ является статистически не значимой.

Следует особо подчеркнуть, что при проверке статистических гипотез нужно избегать категорических формулировок типа: "гипо-

теза верна” или “гипотеза неверна”. Если значение анализируемой статистики не попадает в критическую область, говорят, что нулевая гипотеза H_0 не опровергается при принятом уровне значимости α ; если попадает, говорят, что H_0 опровергается при уровне значимости α , т.е. расхождение эмпирических данных с нулевой гипотезой статистически значимо.

Хотя в данном случае речь шла о конкретном критерии (Стьюдента), основные принципы проверки нулевой гипотезы сохраняются и при использовании других критериев. Различие состоит лишь в том, как строится анализируемая статистика и какому распределению она подчиняется.

Методика проверки статистических гипотез, изложенная в данном разделе, широко применяется в практике гидрологических расчетов, однако это не единственный подход к решению задач такого типа.

Таблица 4.2

Традиционная интерпретация уровней значимости

α	Значимость несоответствия эмпирических данных и нулевой гипотезы H_0
$> 0,10$	Данные хорошо согласуются с H_0
0,05	Возможна значимость. Есть некоторые сомнения в истинности H_0
0,02	Значимость. Довольно сильный довод против H_0
0,01	Высокая значимость. Гипотеза H_0 почти наверняка опровергается

Второй подход заключается в том, что при анализе той или иной статистики границы доверительной области не фиксируются путем назначения уровня значимости. Вместо этого определяется вероятность, соответствующая данному значению выборочной статистики. Таким образом, решается обратная задача: каким должен быть предельный уровень значимости, чтобы нулевая гипотеза не опровергалась. В этом случае говорят, что гипотеза не опровергается на таком-то уровне значимости, при этом результат интерпретируется в соответствии с табл. 4.2 [22].

4.6. Критерии, используемые для проверки однородности гидрологических рядов

Если в качестве математической модели для описания статистической структуры гидрологических рядов рассматривать случайную величину, то статистические характеристики гидрологического ряда не должны изменяться, т.е. все элементы данной выбор-

ки должны относиться к одной генеральной совокупности. По сути это эквивалентно предположению о неизменности условий формирования стока. Если в результате природных катаклизмов или антропогенной деятельности условия формирования стока изменятся, то это неминуемо приведет и к изменению статистических характеристик ряда, например, среднего или дисперсии.

Если статистические характеристики отдельных частей ряда существенно отличаются, то такой ряд является *неоднородным* (нестационарным). При этом надо иметь в виду, что изменение условий формирования стока может сказаться на одних характеристиках стока и почти не отразиться на других. Например, строительство водохранилища многолетнего регулирования наверняка окажет влияние на характеристики максимального стока, но может практически не оказать влияния на среднегодовые расходы.

Статистические методы анализа гидрологических данных, которые были изложены в предыдущих главах, применимы только к однородным рядам, поэтому перед проведением любых статистических расчетов необходимо осуществить проверку однородности исходных гидрологических рядов.

Для проверки однородности гидрологических рядов используются критерии двух типов: *параметрические* и *непараметрические*. В параметрических критериях при построении анализируемой статистики используют выборочные оценки параметров распределения. При этом считается, что исходная выборка относится к генеральной совокупности с известным типом распределения (как правило, принимается нормальный закон распределения).

Непараметрические критерии базируются на использовании непараметрических статистик.

Статистика $g(x_1, x_2, x_3, \dots, x_n)$ является непараметрической, если ее распределение не зависит от распределения X . Поэтому наряду с термином "непараметрический" используется более точный: "свободный от распределения". Хотя непараметрические критерии часто менее эффективны, чем стандартные (параметрические), потеря эффективности в этом случае компенсируется более широкими возможностями их применения.

Из рассматриваемых в этом разделе критериев к числу параметрических относятся приведенные ниже критерии Стьюдента и Фишера.

4.6.1. Критерий Стьюдента для проверки значимости различия средних значений двух выборок

Пусть $(x_1, x_2, x_3, \dots, x_n)$ и $(y_1, y_2, y_3, \dots, y_n)$ – выборки длиной m и n из нормальных распределений с неизвестными параметрами m_x, σ_x и m_y, σ_y , но при этом известно, что $\sigma_x = \sigma_y$, т.е. они имеют одинаковое, хотя и неизвестное СКО, обозначим его буквой σ (без индекса).

Если предположить, что эти выборки относятся к одной генеральной совокупности, то разность должна быть близка к нулю, или говоря языком статистики, различие средних должно быть статистически незначимым. На основе этой разности построим статистику.

$$t = \frac{\bar{x} - \bar{y}}{\sigma_{\bar{x} - \bar{y}}}, \quad (4.23)$$

где $\sigma_{\bar{x} - \bar{y}}$ – СКО разности $(\bar{x} - \bar{y})$. Как следует из теоремы 4.2, эта статистика подчиняется распределению Стьюдента при $\nu = (m + n - 2)$. В математической статистике доказано, что

$$\sigma_{\bar{x} - \bar{y}} = S \sqrt{(m + n) / mn}, \quad (4.24)$$

где S – эмпирическая оценка $\sigma_{\bar{x} - \bar{y}}$. Значение S определяется в зависимости от выборочных значений S_x и S_y .

$$S = \sqrt{\frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2}}. \quad (4.25)$$

В окончательном виде выражение для статистики t имеет вид:

$$t = \frac{\bar{x} - \bar{y}}{S} \sqrt{\frac{mn}{m+n}}. \quad (4.26)$$

В практике гидрологических расчетов эта статистика используется для проверки однородности гидрологических рядов по среднему значению. Исходный ряд делится на две части (две выборки). Если дата возможного нарушения стока не известна, ряд делится пополам. При этом предполагается, что если условия формирования стока не изменились, то разница двух выборочных средних должна

быть незначимой. Уровень значимости обычно принимается $2\alpha = 5\%$ или $2\alpha = 10\%$.

Пример 4.4. Дана выборка длиной 43 года. Требуется провести проверку ряда на однородность по среднему значению при уровне значимости $2\alpha = 10\%$.

Решение.

1. Так как дата возможного нарушения стока неизвестна, делим ряд примерно пополам: $m = 22$, $n = 21$.

2. Определяем выборочное среднее и СКО для первой половины ряда: $\bar{x} = 20,4$; $S_x = 9,50$.

3. Определяем выборочное среднее и СКО для второй половины ряда: $\bar{y} = 32,7$; $S_y = 11,4$.

4. Вычисляем общую дисперсию по формуле (4.25) и СКО: $S^2 = [(22-1)(9,50)^2 + (21-1)(11,4)^2] / (22+21-2) = 109,6$; $S = \sqrt{109,6} = 10,5$.

5. Тогда выборочное значение статистики t с учетом (4.26) равно: $t^* = [(20,4 - 32,7) / 10,5] \sqrt{(22 \cdot 21) / (22 + 21)} = -3,84$.

6. По приложению 5 определяем теоретическое значение t при уровне значимости $2\alpha = 10\%$ и числе степеней свободы $\nu = 43 - 2 = 41$. Это значение равно $t_{1-\alpha} = 1,68$. Таким образом, доверительная область для t^* определена отрезком $[-1,68, +1,68]$. Так как в данном случае t^* попадает в критическую область, т.е. выходит за пределы доверительной области, гипотеза об однородности ряда по среднему значению опровергается при уровне значимости $2\alpha = 10\%$.

Критерий Стьюдента является так называемым стандартным критерием и рекомендуется в большинстве нормативных документов в качестве одного из официальных тестов на однородность. Однако при его использовании следует учитывать три момента.

1. При построении критерия предполагалось, что анализируемые выборки относятся к нормальным совокупностям, а большинство гидрологических рядов, как правило, имеют небольшую положительную асимметрию, что повышает риск совершить ошибку (это замечание относится и к большинству других параметрических критериев).

2. При построении критерия предполагалось, что анализируемые выборки имеют одинаковую (хотя и неизвестную) дисперсию, поэтому перед использованием критерия Стьюдента следует проверить ряд на однородность по дисперсии.

3. В классической статистике длина выборок предполагается значительно большей, чем та, которую мы имеем на практике. Поэтому нельзя исключить ситуацию, когда гипотеза об однородности ряда опровергается из-за недостаточной длины этого ряда. Например, если ряд включает серию маловодных и серию многоводных лет, то при разрезке пополам в одну выборку могут попасть расходу маловодной фазы, а в другую – многоводной.

4.6.2. Критерий равенства двух дисперсий (Критерий Фишера)

Если $x_1, x_2, x_3, \dots, x_m$ и $y_1, y_2, y_3, \dots, y_n$ – выборки из нормальных совокупностей с параметрами m_x, σ_x и m_y, σ_y и если $\sigma_x = \sigma_y = \sigma$, то, как следует из теорем 4.1 и 4.3, отношение их выборочных дисперсий S_x^2 / S_y^2 подчиняется распределению Фишера с числом степеней свободы $\nu_1 = m - 1$ и $\nu_2 = n - 1$. Следовательно, при нулевой гипотезе $H_0: S_x^2 = S_y^2$ и уровне значимости 2α доверительная область для отношения S_x^2 / S_y^2 определяется выражением

$$F_\alpha(\nu_1, \nu_2) \leq (S_x^2 / S_y^2) < F_{1-\alpha}(\nu_1, \nu_2) \quad (4.27)$$

или
$$1 / F_{1-\alpha} \leq (S_x^2 / S_y^2) < F_{1-\alpha}. \quad (4.28)$$

Распределение Фишера, как известно, несимметрично, и для того чтобы сократить объем таблиц, их составляют только для значений $F > 1$, а при сравнении S_x^2 и S_y^2 в числитель всегда подставляют большую дисперсию. В этом случае доверительная область при уровне значимости 2α определяется выражением

$$1 \leq (S_x^2 / S_y^2) < F_{1-\alpha}. \quad (4.29)$$

Этот критерий используется для проверки однородности гидрологических рядов по дисперсии. Исходный ряд делится на две части, затем оцениваются дисперсии для каждой из частей ряда и

вычисляется эмпирическое значение статистики Фишера $F^* = S_1^2 / S_2^2$, где $S_1^2 > S_2^2$. Полученное значение F^* сравнивается с табличным значением $F_{1-\alpha}$. Если при принятом уровне значимости оказывается, что $F^* < F_{1-\alpha}$, то расхождение дисперсий считается незначимым и гипотеза об однородности ряда по дисперсии не опровергается.

Пример 4.5. Пусть дана выборка длиной 43 года. Проверить ряд на однородность при уровне значимости $2\alpha = 10\%$.

Решение.

1. Разбиваем ряд на две примерно равные части: $m = 22, n = 21$.

2. Находим выборочную дисперсию для первой части ряда: $S_m^2 = 370$.

3. Находим выборочную дисперсию для второй части ряда: $S_n^2 = 422$.

4. Вычисляем эмпирическое значение статистики F по формуле $F^* = S_n^2 / S_m^2$; так как $S_n > S_m$ получаем $F^* = 442/370 = 1,19$.

5. По таблице F -распределения (см. приложение 6) для $2\alpha = 10\%$ при $\nu_1 = n - 1 = 20$ и $\nu_2 = m - 1 = 21$ определяем теоретическое значение статистики Фишера $F_{1-\alpha} = F_{95} = 2,09$. Так как $F^* < F_{1-\alpha}$, гипотеза об однородности ряда по дисперсии не опровергается.

Критерий Фишера (так же как и критерий Стьюдента) относится к категории стандартных критериев и рекомендуется в большинстве нормативных документов в качестве официального теста на однородность.

4.6.3. Рангово-суммарный критерий Уилкоксона – Манна – Уитни

Данный критерий был предложен Уилкоксоном и в дальнейшем усовершенствован Манном и Уитни.

Критерий используется для проверки нулевой гипотезы о том, что две независимые выборки принадлежат к совокупностям, которые имеют идентичные функции распределения.

Критерий Уилкоксона – Манна – Уитни относится к категории непараметрических критериев и не подразумевает непосредственного расчета выборочных параметров функции распределения.

Достоинством критерия является то, что он не требует обязательной принадлежности выборок к нормальным совокупностям.

Рассмотрим схему построения статистики Уилкоксона. Пусть даны две выборки из совокупностей X и Y длиной m и n ($m \leq n$). Объединим эти выборки в один ряд и расположим все значения в возрастающем порядке так, что $y_1 < y_2, y_2 < x_1, x_1 < x_2$ и т.д. (табл. 4.3)

Таблица 4.3

Определение рангов (по Уилкоксону)

Ранжированный объединенный ряд	y_1	y_2	x_1	x_2	x_3	y_3	x_4	y_4	x_5	y_5	y_6
Ранг	1	2	3	4	5	6	7	8	9	10	11

Каждому значению нового ряда присвоим порядковый номер (ранг). Теперь выпишем значения полученных рангов отдельно для каждой из выборок (табл. 4.4) и вычислим ранговые суммы ω_1 и ω_2 .

Таблица 4.4

Вычисление ранговых сумм (по Уилкоксону)

Выборка 1 ...	x_1	x_2	x_3	x_4	x_5		Сумма рангов ω_1
Ранг	3	4	5	7	9		28
Выборка 2	y_1	y_2	y_3	y_4	y_5	y_6	Сумма рангов ω_2
Ранг	1	2	6	8	10	11	38

Если расчет выполнен правильно, то должно выполняться равенство

$$\omega_1 + \omega_2 = N(N+1)/2, \quad (4.30)$$

где $N = m + n$.

В данном случае $\omega_1 + \omega_2 = 28 + 38 = 66$; $N(N+1) = 66$.

В качестве анализируемой статистики ω^* Уилкоксон рассматривал ранговую сумму для меньшей выборки, в данном случае $\omega^* = \omega_1$. Символически формулу для статистики ω можно записать в виде

$$\omega^* = r_1 + r_2 + r_3 + \dots + r_m, \quad (4.31)$$

где r_i – ранг x_i .

Для статистики Уилкоксона разработаны таблицы, позволяющие определить доверительный интервал для ω (в зависимости от m, n и уровня значимости 2α). Однако в настоящее время более широкое распространение получила статистика, предложенная Манном и Уитни, так называемая U -статистика.

Для определения U вычислим:

$$U_1 = mn + m(m+1)/2 - \omega_1; \quad (4.32)$$

$$U_2 = mn + m(m+1)/2 - \omega_2. \quad (4.33)$$

В качестве анализируемой статистики U^* можно использовать любое из полученных значений (U_1 или U_2). Обычно в качестве U^* принимается меньшее значение. При правильном расчете должно выполняться равенство

$$U_1 + U_2 = mn. \quad (4.34)$$

Отметим, что U_1 представляет собой так называемое *число инверсий* для первой выборки; а U_2 – *число инверсий* для второй выборки. Вообще говоря, их можно получить, не вычисляя ранговые суммы ω_1 и ω_2 .

Рассмотрим в качестве объединенного ряда ряд из табл. 4.3. Если значение x_i предшествует значению y_j , то говорят, что x_i образует инверсию с y_j . В данном случае x_1 образует инверсию с y_3, y_4, y_5 и y_6 – всего четыре инверсии. Еще по четыре инверсии образуют x_2 и x_3 ; x_4 образует три инверсии: с y_4, y_5 и y_6 – и т. д. Если обозначить через U_1 общее число инверсий, то для выборки из X получим:

$$U_1 = 4 + 4 + 4 + 3 + 2 = 17.$$

Аналогичный расчет для выборки из Y дает

$$U_2 = 5 + 5 + 2 + 1 = 13.$$

Если расчет провести по формулам (4.32) и (4.33), получим те же результаты:

$$U_1 = (5 \cdot 6) + (5 \cdot 6)/2 - 28 = 17,$$

$$U_2 = (5 \cdot 6) + (6 \cdot 7)/2 - 38 = 13.$$

При этом принимают $U^* = U_2 = 13$, так как $U_2 < U_1$.

Какому из методов расчета U -статистики отдать предпочтение – вопрос традиций и личного вкуса.

Распределение U -статистики Манна – Уитни является симметричным с МО и дисперсией:

$$m_U = (mn)/2, \quad (4.35)$$

$$D_U = \sigma_U^2 = mn(m+n+1)/12. \quad (4.36)$$

При $m \geq 8$ и $n \geq 8$ функция распределения нормированной величины статистики U может быть с достаточной точностью аппроксимирована стандартным нормальным распределением. Следовательно, доверительный интервал для статистики U при уровне значимости 2α имеет вид:

$$m_U - t_{1-\alpha}\sigma_U \leq U < m_U + t_{1-\alpha}\sigma_U, \quad (4.37)$$

где $t_{1-\alpha}$ – квантиль стандартного нормального распределения (см. табл. 2.1); m_U и σ_U – параметры, определяемые по формулам (4.35) и (4.36).

U -критерий – один из наиболее строгих непараметрических критериев, так как статистика U – очень сложная функция от среднего значения, эксцесса и асимметрии. В то же время надо учитывать, что U -критерий наиболее чувствителен к различию выборок по характеристикам положения (например, моды или среднего), но весьма слабо реагирует на разницу дисперсий.

Критерий Уилкоксона – Манна – Уитни можно использовать для проверки однородности гидрологических рядов. В этом случае исходный ряд разбивается на две выборки, длиной m и n .

Пример 4.6. Дан ряд максимальных паводочных расходов на р. Плюсса у дер. Брод за 1936 – 1955 гг., $N = 17$ (табл. 4.5, графа 3). Проверить его однородность с использованием критерия Уилкоксона – Манна – Уитни при уровне значимости $2\alpha = 10\%$.

Таблица 4.5

Расчет ранговых сумм (по Уилкоксону)
для ряда максимальных паводочных расходов р. Плюсы – дер. Брод

Ранжированный ряд	Ранг	Исходный ряд	Ранг	Ранжированный ряд	Ранг	Исходный ряд	Ранг
1	2	3	4	1	2	3	4
11,5	1	53,4	6	67,4	9	38,4	4
27,0	2	99,2	14	69,0	10	67,4	9
30,5	3	27,0	2	78,0	11	162	17
38,4	4	11,5	1	91,6	12	64,0	8
49,6	5	95,6	13	95,6	13	30,5	3
53,4	6	49,6	5	99,2	14	122	15
60,0	7	69,0	10	122	15	137	16
64,0	8	78,0	11	137	16	91,6	12
				162	17	60,0	7
Сумма	62	Сумма	91				

Решение.

1. Располагаем ряд в возрастающем порядке и каждому значению ранжированного ряда ставим в соответствии порядковый номер – ранг (см. табл. 4.5, графы 1, 2).

2. Для каждого члена исходного (неранжированного) ряда выписываем его ранг (см. табл. 4.5, графа 4).

3. Делим исходный ряд на две выборки $m = 8$ и $n = 9$ и для каждой выборки определяем ранговую сумму ω_1 и ω_2 : $\omega_1 = 62$; $\omega_2 = 91$.

4. По формуле (4.30) проверяем правильность расчета.

$$\omega_1 + \omega_2 = 62 + 91 = 153; \quad N(N+1)/2 = 17(17+1)/2 = 153.$$

5. По формулам (4.32) и (4.33) вычисляем U_m и U_n : $U_m = 8 \cdot 9 + (8+1)/2 - 62 = 46$; $U_n = 8 \cdot 9 + 9 \cdot (9+1)/2 - 91 = 26$.

6. Проверяем правильность расчета по формуле (4.34): $U_m + U_n = 46 + 26 = 72$; $mn = 72$.

7. В качестве U^* принимаем $U^* = U_n = 26$, поскольку $U_n < U_m$.

8. По формулам (4.35) и (4.36) вычисляем МО и СКО: $m_U = mn/2 = 36$; $\sigma_U = \sqrt{8 \cdot 9 \cdot (8+9+1)/12} = 10,4$.

9. Так как условие $m \geq 8$ и $n \geq 8$ выполнено, по табл. 2.1 определяем ординату стандартного нормального распределения при уровне значимости $2\alpha = 10\%$: $t_{1-\alpha} = t_{0,95} = 1,64$.

10. С учетом формулы (4.37) определяем доверительный интервал для статистики U : $(36 - 1,64 \cdot 10,4) \leq U < (36 + 1,64 \cdot 10,4)$; округляя до целых, получаем $19 \leq U < 53$.

11. Так как $U^* = 26$, т.е. эмпирическое значение статистики U попадает в доверительный интервал, гипотеза об однородности ряда при уровне значимости $2\alpha = 10\%$ не опровергается.

4.6.4. Ранговый критерий рассеяния Зигеля – Тьюки

Поскольку критерий Уилкоксона со статистикой Манна–Уитни является слабо чувствительным к различию выборок по характеристикам рассеяния Зигель и Тьюки предложили другой ранговый критерий. Формально критерий Зигеля – Тьюки позволяет проверить ту же нулевую гипотезу: две независимые выборки относятся к одной и той же генеральной совокупности. Однако этот критерий является более чувствительным именно к различию дисперсий выборок.

Критерий Зигеля – Тьюки относится к категории непараметрических критериев, так как распределение его тестовой статистики не зависит от типа распределения исходных выборок.

Для построения этой статистики анализируемые выборки объединяются в один ряд длиной $N = m + n$ и ранжируются в порядке возрастания (табл. 4.6), так что $y_1 < y_2 < x_1 < x_2 < x_3 < y_3$ и т.д.

Таблица 4.6

Определение рангов (по Зигелю – Тьюки)

Ранжированный объединенный ряд	y_1	y_2	x_1	x_2	x_3	y_3	x_4	y_4	x_5	y_5	y_6
Ранг	1	4	5	8	9	-	10	7	6	3	2

Таким образом, на первом этапе обработка данных аналогична той, что использовалась для критерия Уилкоксона – Манна – Уитни. Но присваивание рангов происходит иначе. Первый ранг получает наименьшее значение (в данном случае – y_1), второй и третий – два самых больших значения (y_6 и y_5), ранги 4 и 5 получают следующие наименьшие значения, 6 и 7 – следующие наибольшие и т.д.

Если общее число наблюдений четное, то среднее значение получает наивысший ранг, если нечетное – оно не получает никакого ранга, а длину соответствующей выборки следует сократить на единицу. В данном случае y_3 не получает ранга, а длина выборки Y будет $n = 6 - 1 = 5$.

При правильном расчете должно выполняться равенство

$$R_1 + R_2 = (m + n)(m + n + 1)/2, \quad (4.38)$$

где R_1 и R_2 – ранговые суммы для выборок из X и Y .

В данном случае: $R_1 + R_2 = 38 + 17 = 55$; $(m + n)(m + n + 1)/2 = (5 + 5)(5 + 5 + 1)/2 = 55$.

Если анализируемые выборки удовлетворяют условию: m и $n > 9$ или $m > 2$, $n > 20$, то в качестве тестовой статистики можно использовать величину

$$Z = \frac{2R_1 - m(m + n + 1) + 1}{\sqrt{m(m + n + 1)(n/3)}}, \quad (4.39)$$

где R_1 – сумма рангов меньшей выборки; m и n – соответственно длины меньшей и большей выборок. При этом если $2R_1 > m(m + n + 1)$, то в числителе выражения (4.39) $+1$ заменяют на -1 .

При соблюдении названных условий распределение статистики Z с достаточной точностью аппроксимируется стандартным нормальным распределением.

Критерий Зигеля – Тьюки можно использовать для проверки однородности гидрологических рядов в отношении характеристик рассеяния. В этом случае исходный гидрологический ряд разбивается на две выборки, длиной m и n .

Данный критерий следует применять вместо критерия Фишера, если распределение исходного ряда существенно отличается от нормального.

Пример 4.7. Дан ряд расходов воды длиной $N = 17$ лет (табл. 4.7, графа 3). Проверить однородность ряда с использованием критерия Зигеля – Тьюки при уровне значимости $2\alpha = 10\%$.

Решение.

1. Располагаем ряд в возрастающем порядке и каждому значению ранжированного ряда присваиваем ранг по Зигелю – Тьюки (см. табл. 4.7, графы 1, 2).

2. Для каждого члена исходного (неранжированного) ряда выписываем его ранг (см. табл. 4.7, графа 4).

3. Делим исходный ряд на две выборки. Длина первой выборки $m = 8$. Длина второй выборки $n = 9 - 1 = 8$, поскольку общее число членов объединенного ряда нечетное и значение 67,4 исключается (не получает ранга).

Замечание: здесь мы немного отклонились от требования m и $n > 9$.

Таблица 4.7

Расчет ранговых сумм (по Зигелю – Тьюки) для ряда максимальных паводочных расходов р. Плюссы – дер. Брод (1936 – 1955 гг.)

Ранжированный ряд	Ранг	Исходный ряд	Ранг	Ранжированный ряд	Ранг	Исходный ряд	Ранг
11,5	1	53,4	12	67,4	–	38,4	8
27,0	4	99,2	7	69,0	15	67,4	–
30,5	5	27,0	4	78,0	14	162	2
38,4	8	11,5	1	91,6	11	64,0	16
49,6	9	95,6	10	95,6	10	30,5	5
53,4	12	49,6	9	99,2	7	122	6
60,0	13	69,0	15	122	6	137	3
64,0	16	78,0	14	137	3	91,6	11
				162	2	60,0	13
Сумма			72	Сумма			64

4. Для каждой выборки определяем ранговую сумму $R_1 = 72$, $R_2 = 64$.

5. По формуле (4.38) проверяем правильность расчета рангов: $R_1 + R_2 = 72 + 64 = 136$; $(m + n)(m + n + 1)/2 = (8 + 8)(8 + 8 + 1)/2 = 136$.

6. По формуле (4.39) вычисляем статистику Z^* :

$$Z^* = \frac{2 \cdot 72 - 8(8 + 8 + 1) + 1}{\sqrt{8(8 + 8 + 1)(8/3)}} = \frac{9,0}{19,0} = 0,47.$$

7. По табл. 2.1 определяем ординату стандартного нормально-го распределения при уровне значимости $2\alpha = 10\%$: $t_{1-\alpha} = t_{95} = 1,64$, и строим доверительный интервал для Z : $-1,64 \leq Z < +1,64$.

8. Так как $Z^* = 0,47$, т.е. эмпирическое значение статистики Z попадает в доверительный интервал, гипотеза об однородности ряда при уровне значимости $2\alpha = 10\%$ не опровергается.

4.7. Критерии согласия

Одной из основных задач при определении расчетных гидрологических характеристик является выбор аналитического выражения для аппроксимации закона распределения исследуемой гидрологической величины.

Выбор считается обоснованным, если имеет место хорошее соответствие эмпирической и аналитической кривых обеспеченностей. На практике такой анализ нередко производится визуально, что, вносит элемент субъективизма. Более объективно решить поставленную задачу можно, используя критерии согласия. При этом реализуется схема, аналогичная той, которая применялась при проверке гипотез об однородности гидрологических рядов: выдвигаются нулевая и альтернативная гипотезы; назначается уровень значимости; вычисляется эмпирическое значение тестовой статистики; по результатам расчетов принимается решение.

В данном случае в качестве нулевой принимается гипотеза о соответствии (согласии) аналитической и эмпирической функций распределения, а в качестве альтернативной — их несоответствие.

Степень согласия оценивается с помощью специальных статистик. Различные критерии (тесты) используют для оценки степени согласия разные статистики.

В гидрологической практике наиболее часто применяются критерий χ^2 (Пирсона), критерий Колмогорова и критерий $n\omega^2$ (Крамера – Мизеса – Смирнова).

4.7.1. Критерий χ^2 (Пирсона)

Критерий χ^2 был предложен в начале XX в. Карлом Пирсоном и в настоящее время является наиболее распространенным критерием согласия. Для его применения область допустимых значений (ОДЗ) исследуемой СВ X разбивается на k интервалов. Число интервалов можно приблизительно определить по формуле (3.2) в зависимости от длины имеющейся выборки.

При назначении границ интервалов наиболее удобно использовать следующую схему.

1. Выбрать аналитическую кривую для аппроксимации закона распределения исследуемой случайной величины.
2. Оценить по имеющейся выборке параметры распределения.
3. Построить на клетчатке вероятности аналитическую кривую обеспеченностей.

4. Разбить ось обеспеченностей на k равных интервалов.

5. Используя аналитическую кривую, определить границы интервалов для СВ X в зависимости от границ интервалов на оси обеспеченностей. Например, если $k = 5$, то границы интервалов на оси обеспеченностей будут соответствовать значениям 0, 20, 40, 60, 80, 100 %; а ось X разобьется на интервалы: $x_{20} \leq x$; $x_{40} \leq x < x_{20}$; $x_{60} \leq x < x_{40}$; $x_{80} \leq x < x_{60}$; $x < x_{80}$.

При реализации данной схемы интервалы по оси X не будут равновеликими, но зато вероятность попадания значения СВ X в любой интервал будет одинаковой $p = 1/k$. Одинаковым будет и теоретическое число случаев попадания значения СВ X в каждый интервал $m = n/k$ (n – длина выборки). Таким образом, принятая схема разбивки ОДЗ СВ X на интервалы обеспечивает следующие теоретические соотношения:

$$\begin{cases} m_1 = m_2 = m_3 \dots = m_k = m = n/k, \\ p_1 = p_2 = p_3 \dots = p_k = p = m/n. \end{cases} \quad (4.40)$$

В качестве меры расхождения между эмпирическими данными и аналитической функцией распределения используется тестовая статистика

$$\chi^2 = n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i}, \quad (4.41)$$

где p_i^* и p_i — соответственно эмпирическая и теоретическая вероятности попадания значения СВ в i -й интервал; n — длина выборки; k — число интервалов.

К. Пирсон показал, что закон распределения данной статистики не зависит от вида исходного распределения и при достаточно большом n хорошо аппроксимируется распределением хи-квадрат (см. приложение 4) с числом степеней свободы ($\nu = k - r - 1$); где r — число параметров исходного распределения, определяемых по эмпирическим данным.

Замечание. Следует отметить, что здесь мы сталкиваемся с некоторой терминологической путаницей. Хотя статистика Пирсона называется χ^2 (хи-квадрат), ее выборочное распределение лишь приближенно совпадает с распределением χ^2 .

Учитывая (4.40) и то, что $p_i^* = m^*/n$, выражение (4.41) можно привести к более удобному для расчетов виду:

$$\begin{aligned} \chi^2 &= n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i} = n \sum_{i=1}^k \left(\frac{m_i^*}{n} - \frac{m}{n} \right)^2 \frac{n}{m} = \sum_{i=1}^k \frac{(m_i^* - m)^2}{m} = \\ &= \frac{1}{m} \left[\sum_{i=1}^k (m_i^*)^2 + \sum_{i=1}^k \left(\frac{n}{k} \right)^2 - 2 \sum_{i=1}^k m_i^* \frac{n}{k} \right] = \frac{1}{m} \left[\sum_{i=1}^k (m_i^*)^2 + \frac{n^2}{k} - 2 \frac{n^2}{k} \right]. \end{aligned}$$

В итоге получаем:

$$\chi^2 = \frac{1}{m} \sum_{i=1}^k (m_i^*)^2 - n, \quad (4.42)$$

где m_i^* и m — соответственно эмпирическое и теоретическое число случаев попадания значения СВ X в i -тый интервал.

Из выражения (4.41) видно, что чем больше значение статистики χ^2 , тем больше расхождение между эмпирической и аналитической кривыми. Поэтому при использовании критерия χ^2 (Пирсона) назначают односторонний уровень значимости (обычно $\alpha = 5\%$ или $\alpha = 10\%$).

Гипотеза о соответствии (согласии) эмпирической и аналитической кривых обеспеченностей не опровергается, если эмпирическое значение статистики $(\chi^2)^*$, определенное по формуле (4.41) или

(4.42), не превышает теоретическое значение χ^2 , соответствующее принятому уровню значимости (α), т.е. если $(\chi^2)^* \leq \chi^2_{1-\alpha}$.

Критерий χ^2 может быть применен при выяснении вопроса о лучшем соответствии одной из нескольких аналитических кривых распределения одному и тому же эмпирическому ряду. При этом меньшее значение χ^2 будет свидетельствовать о лучшем соответствии данной функции распределения эмпирическим данным.

Следует подчеркнуть, что использование критерия Пирсона для выборок небольшого объема может приводить к неправильным выводам, а ряды наблюдений за гидрологическими характеристиками обычно не превышают нескольких десятков лет. Это накладывает определенные ограничения на использование данного критерия. Принято считать, что критерий согласия χ^2 допустимо применять при $n > 50$. При этом желательно, чтобы число разрядов (интервалов) было равно 8 – 12 и в каждом разряде было не менее 5 элементов.

Пример 4.8. На рис. 4.7 приведены эмпирическая кривая обеспеченностей средних годовых расходов р. Тихвинки – дер. Горелуха ($n = 98$) и аналитическая кривая трехпараметрического гамма-распределения (Крицкого – Менкеля) с параметрами $\bar{Q} = 19,5 \text{ м}^3/\text{с}$, $C_v = 0,25$, $C_s/C_v = 2$. Оценить степень соответствия эмпирической и аналитической кривой с помощью критерия согласия χ^2 (Пирсона) при уровне значимости $\alpha = 5 \%$.

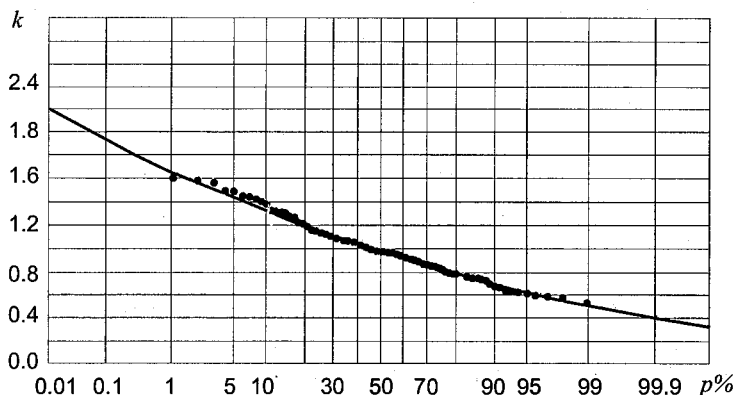


Рис. 4.7. Эмпирическая и аналитическая (Крицкого – Менкеля) кривые обеспеченностей среднегодовых расходов воды для створа р. Тихвинка – дер. Горелуха в модульных коэффициентах: $n = 98$; $\bar{Q} = 19,5$; $C_v = 0,25$; $C_s = 2 C_v$.

Решение.

1. Для решения этой задачи разбиваем всю ось обеспеченностей на 10 равных интервалов $k \approx 5 \lg(n) \approx 10$.

2. Находим значения ординат аналитической кривой обеспеченностей на границах интервалов. Для первого интервала (0 – 10 %) верхняя граница будет равна бесконечности, поскольку функция распределения не ограничена сверху. Для второго интервала (10–20%) верхняя граница равна нижней первого интервала и т.д.; для последнего интервала нижняя граница равна нулю, так как функция распределения в качестве нижнего предела имеет нуль (табл. 4.8).

Таблица 4.8

Исходные данные для расчета критерия согласия χ^2

Интервал k	Обеспеченность, %	Границы интервалов	Число случаев попадания расхода воды в интервал m_i^*	$(m_i^*)^2$
1	0 – 10	$\infty - 1,33$	11	121
2	10 – 20	1,33 – 1,20	8	64
3	20 – 30	1,20 – 1,11	10	100
4	30 – 40	1,11 – 1,05	8	64
5	40 – 50	1,05 – 0,98	13	169
6	50 – 60	0,98 – 0,92	11	121
7	60 – 70	0,92 – 0,85	10	100
8	70 – 80	0,85 – 0,79	6	36
9	80 – 90	0,79 – 0,70	11	121
10	90 – 100	0,70 – 0,00	10	100
Сумма				996

3. Для каждого интервала подсчитываем число случаев попадания в него расхода воды (m_i^*), табл. 4.9. Если значение расхода приходится на границу интервала, ее относят к верхнему интервалу. Значение m_i^* заносим в графу 4 таблицы 4.8.

4. Теоретическое число случаев попадания расхода воды в каждый интервал определяется по выражению $m = n / k$ и в данном случае $m = 98/10 = 9,8$. Эмпирические данные, приведенные в графе 4 табл. 4.8, отличаются от теоретических. Расхождения эти обусловлены либо случайными колебаниями выборочных частот, либо

несоответствием эмпирической и аналитической кривых обеспеченностей (что и предстоит выяснить).

5. Используя формулу (4.42), рассчитаем эмпирическое значение χ^2 : $(\chi^2)^* = \frac{1}{m} \sum_{i=1}^{10} (m_i^*)^2 - n = \frac{1}{9.8} 996 - 98 = 3,63$.

6. По приложению 4 находим теоретическое значение χ^2 при числе степеней свободы $\nu = k - r - 1 = 10 - 3 - 1 = 6$ и уровне значимости $\alpha = 5\%$. В данном случае $\chi_{95}^2 = 12,6$.

7. Поскольку $(\chi^2)^* < \chi_{95}^2$ можно сделать вывод о том, что гипотеза о соответствии эмпирической и аналитической кривых обеспеченностей при уровне значимости $\alpha = 5\%$ не опровергается.

Таблица 4.9

Ранжированный ряд модульных коэффициентов годового стока для створа р. Тихвинка – д. Горелуха

№ п/п	№ в группе	k	№ п/п	№ в группе	k
1	1	1,61	78	1	0,78
2	2	1,58	79	2	0,78
3	3	1,57	80	3	0,78
4	4	1,51	81	4	0,78
5	5	1,48	82	5	0,76
6	6	1,45	83	6	0,75
7	7	1,44	84	7	0,75
8	8	1,42	85	8	0,75
9	9	1,40	86	9	0,75
10	10	1,38	87	10	0,73
11	11	1,34	88	11	0,70
12	1	1,32	89	1	0,68
13	2	1,32	90	2	0,67
14	3	1,31	91	3	0,64
15	4	1,29	92	4	0,64
16	5	1,29	93	5	0,62
17	6	1,27	94	6	0,62
18	7	1,22	95	7	0,59
19	8	1,22	96	8	0,59
20	1	1,18	98	10	0,53
...	...	...			

4.7.2. Критерий Колмогорова

В качестве меры отличия эмпирических данных от аналитической функции распределения А.М. Колмогоров принимает наибольшее по абсолютной величине расхождение между эмпирической $P^*(x)$ и аналитической $P(x)$ функциями обеспеченностей:

$$\Delta = \max |P^*(x) - P(x)|. \quad (4.43)$$

Вычисления по критерию Колмогорова производятся следующим образом. Для каждого значения СВ X , полученного по выборке, вычисляются $P^*(x)$, $P(x)$ и их разности. Затем выбирается наибольшая по модулю разность Δ и рассчитывается статистика:

$$\lambda^* = \Delta \sqrt{n}, \quad (4.44)$$

где n — объем выборки. Колмогоров доказал, что функция обеспеченностей статистики λ при достаточно больших n ($n > 40$) может быть аппроксимирована выражением

$$P\{\lambda^* > \lambda\} \approx P(\lambda) = 2 \sum_{k=1}^n (-1)^{k-1} \exp(-2k^2\lambda^2). \quad (4.45)$$

Координаты функции (4.45) представлены в приложении 7.

Если значение $P(\lambda)$ больше принятого уровня значимости, то гипотеза о соответствии эмпирической и аналитической кривых обеспеченностей не опровергается.

Расчетная схема для критерия согласия Колмогорова достаточно проста. Вместе с тем этот критерий имеет определенные недостатки, в частности, при его применении учитывается лишь наибольшее расхождение между эмпирической и аналитической функциями распределения, т.е. используется далеко не вся информация, заключенная в рядах наблюдений. Оценка согласия по одной точке (особенно при небольшой длине рядов) может в отдельных случаях плохо отражать соответствие эмпирических данных теоретическому закону распределения. К тому же наибольшая разность Δ обычно отмечается в средней части кривой распределения, в то время как в гидрологической практике нас чаще интересуют ее левая и правая части. Кроме того, данный критерий не учитывает числа параметров, входящих в теоретическую функцию распределения, хотя на

практике оценку каждого из них приходится производить по эмпирическим данным.

Отмеченные недостатки не исключают возможности применения критерия Колмогорова в гидрологических расчетах. Статистика λ может быть применена и для сравнения степени соответствия нескольких аналитических функций распределения применительно к одному эмпирическому материалу аналогично тому, как это было указано при использовании критерия χ^2 .

Пример 4.9. В качестве исходных будем использовать данные из примера 4.8. Оценить с помощью критерия Колмогорова степень согласия эмпирических данных с аналитической кривой Крицкого – Менкеля (рис. 4.7) при уровне значимости $\alpha = 5\%$.

Решение.

1. Для каждого из 98 значений исходного ряда рассчитываем абсолютное значение разницы между эмпирической и теоретической обеспеченностью. Фрагмент такого расчета представлен в табл. 4.10.

Таблица 4.10

Расчет разностей между эмпирической и теоретической обеспеченностями

№ п/п	k_i	$P^* \%$	$P \%$	$ P^* - P $
...	...	...	...	...
40	1,03	40,40	42,14	1,74
41	1,02	41,41	43,72	2,31
42	1,01	42,42	45,22	2,80
43	1,00	43,43	46,70	3,27
44	0,99	44,44	48,19	3,75
45	0,99	45,45	48,19	2,74
46	0,98	46,46	49,68	3,22
47	0,98	47,47	49,68	2,21
48	0,98	48,48	49,68	1,20
49	0,98	49,49	49,68	0,19
...	...	...	...	...

2. Как видно из таблицы, наибольшая разница между P^* и P составила 3,75 %, следовательно $\Delta = 0,0375$.

3. Рассчитываем эмпирическое значение статистики Колмогорова: $\lambda^* = \Delta\sqrt{n} = 0,0375\sqrt{98} = 0,367$.

4. По приложению 7 определяем ординату функции обеспеченностей Колмогорова: $P(\lambda^*) = 0,999$.

5. Так как $0,999 \gg 0,05$, то есть $P(\lambda^*)$ намного больше принятого уровня значимости, то гипотеза о соответствии эмпирических точек кривой Крицкого – Менкеля не опровергается.

4.7.3. Критерий $n\omega^2$ (Крамера – Мизеса – Смирнова)

Тестовой статистикой данного критерия является средний квадрат отклонений между аналитической $P(x)$ и эмпирической $P^*(x)$ функциями обеспеченностей по всем значениям случайной величины X :

$$n\omega^2 = \sum_{i=1}^n \left[P(x) - P^*(x) \right]^2 + \frac{1}{12n}, \quad (4.46)$$

где $P(x)$ рассчитывается по формуле (3.5).

В работе [13] приводится модификация этого критерия, учитывающая особенности построения эмпирических функций обеспеченностей гидрологических характеристик. Суть этого предложения состоит в том, что для расчета эмпирической обеспеченности вместо формулы (3.5) используется формула (3.6). В этом случае выражение для тестовой статистики имеет вид¹:

$$(n+1)\omega^2 = \sum_{i=1}^n \left[P(x) - \frac{m}{n+1} \right]^2 + \frac{n}{12(n+1)^2}. \quad (4.47)$$

При $n > 40$ распределение статистики $n\omega^2$ не зависит от вида исходного теоретического распределения и близко к некоторому предельному распределению (табл. 4.11).

Таблица 4.11

Значения верхних доверительных границ статистики $n\omega^2$
в зависимости от уровня значимости α %

α , %	20	10	5	3	2	1	0.1
$n\omega^2$	0,241	0,347	0,461	0,549	0,620	0,744	1,168

Если эмпирическое значение тестовой статистики, вычисленное с использованием выражения (4.47), оказывается больше теоретического значения $n\omega^2$ при уровне значимости α %, то гипотеза

¹ Хотя в левой части формулы (4.47) стоит выражение $(n+1)\omega^2$, мы будем по-прежнему в качестве названия критерия использовать термин $n\omega^2$.

о соответствии эмпирической и аналитической функций обеспеченностей опровергается.

Следует отметить, что, по сравнению с предыдущими, использование данного критерия более трудоемко, но это компенсируется наиболее полным использованием информации, заключенной в фактических рядах наблюдений.

Применение этого критерия предполагает знание не только типа распределения, но и параметров, обычно оцениваемых по выборке, что может приводить к признанию нуль-гипотезы даже в случае недостаточного согласования аналитических и эмпирических функций обеспеченностей.

Пример 4.10. Имеется ряд наблюдений за средними годовыми расходами на р. Саба – дер. Райково с 1945 по 1988 г. ($n = 44$). Расчет оценок параметров распределения выполнен методом моментов: $\bar{Q} = 9,09$, $C_v^* = 0,29$, $C_s/C_v = 2$. Оценить степень соответствия эмпирических данных с аналитической кривой трехпараметрического гамма-распределения Крицкого – Менкеля (рис. 4.8) с помощью критерия $\rho\omega^2$ при $\alpha = 5\%$.

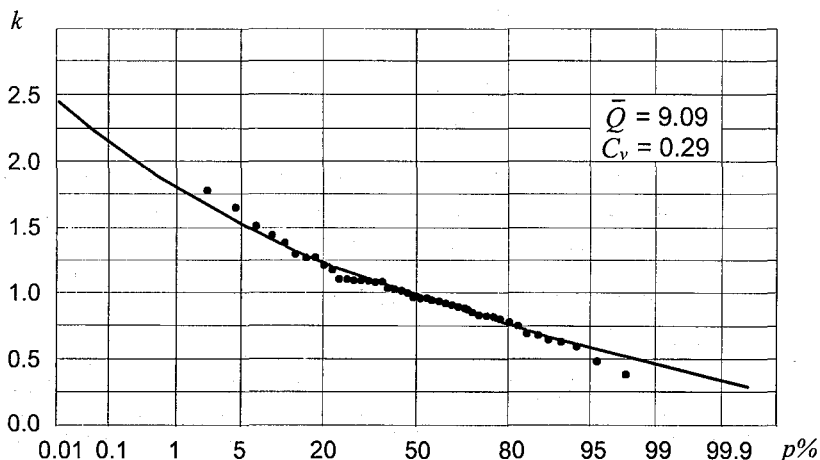


Рис. 4.8. Эмпирическая и аналитическая (Крицкого – Менкеля) кривые обеспеченностей среднегодовых расходов; р. Саба – д. Райково.

Решение.

1. В табл. 4.12 приведены результаты расчета квадратов разностей Δp между эмпирической p^* и теоретической p обеспеченности для всех членов выборки.

2. Эмпирическое значение тестовой статистики $[(n + 1)\omega^2]^*$, рассчитанное по формуле (4.47) и равное 0.04468, сравнивается с табличным (табл. 4.11) значением $n\omega^2$ при $\alpha = 5\%$: $(n\omega^2)_{5\%} = 0,461$.

3. Так как $[(n + 1)\omega^2]^* \ll (n\omega^2)_{5\%}$, гипотеза о соответствии эмпирических данных аналитической кривой обеспеченностей Крицкого – Менкеля не опровергается.

Таблица 4.12

Расчет квадратов разностей Δp между эмпирическими p^* и теоретическими p обеспеченностями; значения $(\Delta p)^2$ даны в долях от единицы, т.е. $(\Delta p)^2 = [(p^* - p) / 100]^2$

№ п/п	k_i	$p^* \%$	$p \%$	$(\Delta p)^2$
1	1,78	2,22	1,73	0,00002
2	1,65	4,44	2,91	0,00023
3	1,51	6,67	6,07	0,00004
...	...	...	...	...
20	1,02	44,44	43,75	0,00005
21	1,00	46,67	46,25	0,00002
22	0,97	48,89	50,00	0,00012
...	...	...	...	...
42	0,60	93,33	92,48	0,00007
43	0,49	95,56	97,75	0,00048
44	0,38	97,78	99,60	0,00033

Сумма | 0,04287

4.8. Критерии случайности

Изложенные ранее приемы обработки эмпирических данных, строго говоря, пригодны лишь для статистических совокупностей, вероятностная структура которых может быть описана в рамках математической модели *случайной величины*. Для проверки, соответствует ли этому требованию конкретная выборка, служат критерии случайности. Проверка гидрологических рядов на случайность проводится в рамках общей схемы статистической проверки гипотез.

тез. В качестве нулевой принимается гипотеза о том, что имеющаяся выборка представляет собой последовательность независимых значений некоторой случайной величины.

Применение критериев случайности в общем случае основано на сопоставлении конкретных статистик эмпирического ряда с соответствующими теоретическими статистиками случайных совокупностей.

4.8.1. Критерии серий

Под *серией* в данном случае будем понимать всякий участок последовательности, состоящий из элементов одного и того же рода.

К серии из элементов *a* будем относить члены последовательности, значения которых превышают выборочное среднее (или медианное) значение, а к серии из элементов *b* — значения которых меньше среднего. Длиной серии назовем число элементов, входящих в серию. Обработанная таким образом выборка будет представлять собой последовательность серий различной продолжительности.

4.8.1.1. Критерий общего числа серий

Для проверки гипотезы о том, что данная совокупность сформирована из независимых значений случайной величины, используется статистика *R*, представляющая собой сумму серий из элементов *a* *r_a* и элементов *b* *r_b* (длина серий *i* при этом значения не имеет).

Для случайных совокупностей статистика $R = r_a + r_b$ имеет нормальное распределение с параметрами

$$m_R = \frac{n+1}{2}; \quad \sigma_R = \frac{\sqrt{n-1}}{2}. \quad (4.48)$$

Исходя из этого, задавшись уровнем значимости α , можно построить для *R* доверительный интервал (см. п. 4.4):

$$m_R - t_{1-\alpha} \sigma_R < R \leq m_R + t_{1-\alpha} \sigma_R, \quad (4.49)$$

где $t_{1-\alpha}$ — квантиль стандартного нормального распределения.

Если значение R^* , определенное по выборке, попадает в этот интервал, то гипотеза о случайности данной совокупности не опровергается.

Пример 4.11. Имеется ряд максимальных в году расходов на р. Систе – дер. Среднее Райково ($F = 573 \text{ км}^2$) за период 1946 – 1984 гг. ($n = 39$). Проверить гипотезу о том, что вероятностная структура гидрологического ряда соответствует модели *случайной величины* с помощью критерия общего числа серий при уровне значимости $2\alpha = 5 \%$.

Решение.

1. Рассчитываем средний расход: $\bar{Q}_{\text{max}} = 57,3 \text{ м}^3/\text{с}$. Присвоим значению расхода, имеющему значение, большее $57,3 \text{ м}^3/\text{с}$, индекс a , а меньшему – индекс b (табл. 4.13).

Таблица 4.13

Расчет общего числа серий R для ряда максимальных расходов воды на р. Систе – д. Среднее Райково за период 1946 – 1984 гг.

Год	$Q, \text{ м}^3/\text{с}$	Индекс	Число лет, i	Год	$Q, \text{ м}^3/\text{с}$	Индекс	Число лет, i
1946	66,0	a	1	1966	79,0	a	1
1947	46,1	b	1	1967	36,4	b	1
1948	61,1	a	1	1968	64,0	a	2
1949	40,1	b	2	1969	54,1	a	
1950	26,5	b		1970	33,0	b	1
1951	72,7	a	1	1971	50,4	a	1
1952	35,5	b	3	1972	29,4	b	4
1953	34,1	b		1973	19,4	b	
1954	18,0	b		1974	26,6	b	
1955	79,6	a	3	1975	23,6	b	
1956	68,3	a		1976	72,7	a	1
1957	68,7	a		1977	39,9	b	5
1958	45,1	b	1	1978	32,5	b	
1959	77,4	a	2	1979	29,4	b	
1960	81,5	a		1980	32,5	b	
1961	29,6	b	1	1981	35,7	b	3
1962	70,2	a	1	1982	58,2	a	
1963	40,8	b	1	1983	67,5	a	
1964	84,6	a	1	1984	67,4	a	
1965	49,4	b	1				

2. По данным табл. 4.13 подсчитаем число серий (независимо от их длины i), состоящих из элементов a и b . В данном случае $r_a = 12$, $r_b = 11$. Таким образом, общее число серий в данной выборке равно 23 ($R = 23$).

3. По формулам (4.48) получаем: $m_R = 20$, $\sigma_R = 3,08$. По табл. 2.1 определяем квантиль стандартного нормального распределения, соответствующий уровню значимости $2\alpha = 5\%$: $t'_{1-\alpha} = t'_{97,5} = 1,96$.

4. Подставив полученные значения в (4.49), рассчитаем доверительный интервал для R : $14 \leq R < 26$.

5. Так как общее число серий, полученное по выборке ($R^* = 23$), попадает в доверительный интервал, можно сделать заключение, что гипотеза о случайности ряда не опровергается.

4.8.1.2. Критерий наибольшей длины серий

Критерий *наибольшей длины серий* использует в качестве тестовой статистики наибольшую длину серии из элементов a или b : $K = i_{\max}$. Теоретически доказано, что для случайных независимых совокупностей значение K выражается формулой

$$K_\alpha = \frac{\lg \left[-\frac{n}{\ln(1-\alpha)} \right]}{\lg 2} - 1, \quad (4.50)$$

где α представляет собой вероятность (в долях единицы), с которой в выборке объемом n можно встретить хотя бы одну серию из элементов a или b длиной K и более.

При проверке нулевой гипотезы о случайности выборки эмпирическое значение статистики K^* сравнивается с теоретическим, рассчитанным по формуле (4.50) при уровне значимости α . Гипотеза не опровергается, если $K^* < K_\alpha$.

Пример 4.12. По данным примера 4.11 проверить гипотезу о том, что вероятностная структура гидрологического ряда соответствует модели случайной величины с помощью критерия наибольшей длины серий K при уровне значимости $2\alpha = 5\%$.

Решение.

1. Как видно из табл. 4.13, наибольшую длину имеет серия, состоящая из элементов b и приходящаяся на период 1977 – 1981 гг., т.е. $K = 5$.

2. По формуле (4.50) рассчитаем значение $K_\alpha = 8,6$.

3. Так как $K^* < K_\alpha$, гипотеза о случайности ряда при данном уровне значимости не опровергается.

4.8.2. Критерий числа повышений и понижений

Пусть имеется выборка СВ X : $x_1, x_2, x_3, \dots, x_n$. Если $x_{i-1} < x_i$ будем называть переход от x_{i-1} к x_i повышением и обозначать его знаком "+", если $x_{i-1} > x_i$ будем называть переход от x_{i-1} к x_i понижением и обозначать знаком "-".

Для случайных последовательностей число повышений (или понижений) распределяется асимптотически нормально с параметрами:

$$m_+ = m_- = n/2, \quad D_+ = D_- = (n+1)/12. \quad (9.51)$$

Отсюда, задавшись уровнем значимости α и учитывая, что математические ожидания и дисперсии числа повышений и понижений равны, можно построить доверительный интервал

$$m - t'_{1-\alpha} \sqrt{D} < k + t'_{1-\alpha} \sqrt{D}, \quad (9.52)$$

где k – число повышений (k_+) или понижений (k_-) в исследуемой выборке; $t'_{1-\alpha}$ – квантиль стандартного нормального распределения.

Если эмпирические значения k_+^* и k_-^* попадают в доверительный интервал, гипотеза о случайности выборки не опровергается.

Пример 4.13. Имеется ряд максимальных в году расходов воды на р. Систе – дер. Среднее Райково за период 1946 – 1984 гг. ($n = 39$). Проверить гипотезу о случайности ряда с помощью критерия числа повышений и понижений при уровне значимости $2\alpha = 5\%$.

Решение.

1. В табл. 4.14 приведен исследуемый ряд и в графе 3 индексами "+" и "-" отмечены повышения и понижения.

2. Из табл. 4.14 видно, что число повышений в данной выборке $k_+^* = 17$, число понижений $k_-^* = 21$.

3. По формулам (4.51) рассчитываем математическое ожидание и дисперсию для числа повышений и понижений. В данном случае $m_+ = m_- = 19,5$; $D_+ = D_- = 3,33$, $\sigma_+ = \sigma_- = 1,82$.

4. По табл. 2.1 определяем квантиль нормированного нормального распределения, соответствующий уровню значимости $2\alpha = 5\%$: $t_{1-\alpha} = t_{97,5} = 1,96$.

5. Подставив полученные значения в (4.52), рассчитаем доверительный интервал для k : $16 \leq k < 23$.

Таблица 4.14

Расчет числа повышений "+", понижений "-" и экстремумов "э"
для ряда максимальных в году расходов воды
в створе р. Сиса – д. Среднее Райково за период 1946 – 1984 гг.

Год	$Q, \text{ м}^3/\text{с}$	+/-	э	Год	$Q, \text{ м}^3/\text{с}$	+/-	э
1946	66,0			1966	79,0	+	э
1947	46,1	-	э	1967	36,4	-	э
1948	61,1	+	э	1968	64,0	+	э
1949	40,1	-		1969	54,1	-	
1950	26,5	-	э	1970	33,0	-	э
1951	72,7	+	э	1971	50,4	+	э
1952	35,5	-		1972	29,4	-	
1953	34,1	-		1973	19,4	-	э
1954	18,0	-	э	1974	26,6	+	э
1955	79,6	+	э	1975	23,6	-	э
1956	68,3	-	э	1976	72,7	+	э
1957	68,7	+	э	1977	39,9	-	
1958	45,1	-	э	1978	32,5	-	
1959	77,4	+		1979	29,4	-	э
1960	85,1	+	э	1980	32,5	+	
1961	29,6	-	э	1981	35,7	+	
1962	70,2	+	э	1982	58,2	+	
1963	40,8	-	э	1983	67,5	+	э
1964	84,6	+	э	1984	67,4	-	
1965	49,4	-	э				

6. Так как определенные по выборке k_+^* и k_-^* попадают в этот интервал, проверяемая гипотеза не опровергается.

4.8.3. Критерий числа экстремумов

Экстремум – элемент последовательности $x_1, x_2, x_3, \dots, x_n$, для которого выполняется одно из неравенств:

$$x_{i-1} < x_i > x_{i+1} \text{ — максимум;}$$

$$x_{i-1} > x_i < x_{i+1} \text{ — минимум.}$$

Для выборок, представляющих собой последовательности независимых значений случайной величины, число экстремумов распределено асимптотически нормально с параметрами:

$$m_3 = \frac{2n}{3}; \quad D_3 = \frac{16n-29}{90}. \quad (4.53)$$

Проверка гипотезы о случайности ряда производится так же, как в предыдущем случае (см. п. 4.8.2).

Пример 4.14. По данным примера 4.13 проверить гипотезу о случайности ряда с помощью критерия числа экстремумов при уровне значимости $2\alpha = 5\%$.

Решение.

1. Как видно из табл. 4.14, число экстремумов для данной выборки равно 26 ($k_3^* = 26$).

2. По формулам (4.53) рассчитываем значения m_3 и D_3 . В данном случае $m_3 = 26$; $D_3 = 6,6$; $\sigma_3 = 2,57$.

3. По табл. 2.1 определяем квантиль нормированного нормального распределения, соответствующий уровню значимости $2\alpha = 5\%$: $t_{1-\alpha} = t_{97,5} = 1,96$.

4. Рассчитываем доверительный интервал для числа экстремумов: $m_3 - \sigma_3 t_{1-\alpha} < k_3 < m_3 + \sigma_3 t_{1-\alpha} \Rightarrow 21 \leq k_3 < 31$.

5. Так как $k_3^* = 26$ и попадает в доверительный интервал, нулевая гипотеза не опровергается.

5. СТАТИСТИЧЕСКИЙ АНАЛИЗ ЗАВИСИМОСТЕЙ МЕЖДУ ГИДРОЛОГИЧЕСКИМИ ПЕРЕМЕННЫМИ

При изучении природных явлений часто встречаются ситуации, когда СВ Y зависит от других величин $(X_1, X_2, X_3, \dots, X_k)$, и по эмпирическим данным требуется найти аналитическое выражение зависимости $y = f(x_1, x_2, x_3, \dots, x_k)$.

Величину Y называют зависимой переменной, или предиктантом, а $(X_1, X_2, X_3, \dots, X_k)$ – независимыми переменными, или предикторами. В специальной литературе зависимую переменную иногда называют откликом, а независимые переменные – факторами.

Сначала рассмотрим ситуацию, когда Y зависит только от одного предиктора, т.е. $y = f(x)$. В этом случае данные наблюдений можно представить в виде простой таблицы (табл. 5.1) или графика (рис. 5.1).

Таблица 5.1

Минимальные 30-суточные (y) и минимальные суточные (x) расходы воды за период открытого русла ($\text{м}^3/\text{с}$) на р. Вороне у г. Воронежа ($n = 13$)

x	8,53	10,5	17,0	16,1	26,0	13,4	15,3	13,0	14,6	12,9	11,4	29,7	19,0
y	9,53	11,0	17,6	17,9	27,9	14,4	16,0	14,0	15,4	13,3	12,3	30,3	19,9

Обычно точки на графике не лежат на одной линии, даже если X и Y связаны функциональной зависимостью, а образуют так называемое поле рассеяния (или корреляционное поле). Причиной рассеяния при наличии функциональной зависимости является неточность измерений. Однако для гидрологической практики более характерна другая ситуация. Гидрологические явления, как правило, обусловлены весьма большим числом факторов, полный учет которых практически не возможен, а в большинстве случаев и не целесообразен, так как только некоторые из них (например, один или два) оказывают решающее влияние на зависимую переменную. Они-то и рассматриваются в качестве предикторов. В такой ситуации при любой точности измерений точки на графике будут образовывать поле рассеяния.

В обоих рассмотренных случаях мы имеем дело с так называемыми стохастическими (вероятностными) зависимостями, когда одному и тому же значению X соответствует множество значений Y . Но в первом случае исследователь должен по эмпирическим дан-

ным выявить функциональную зависимость $y = f(x)$, а во втором случае требуется найти некоторую осредненную кривую $\bar{y}(x)$, относительно которой будет наблюдаться минимальный разброс точек.

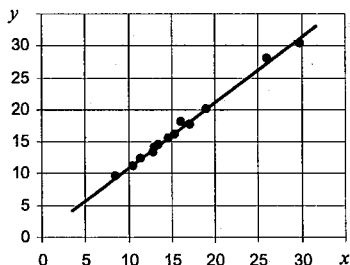


Рис. 5.1. График зависимости минимальных 30-суточных расходов (y) от минимальных суточных расходов (x) за период открытого русла на р. Вороне у г. Воронежа.

Линию, относительно которой наблюдается минимальный разброс точек, называют линией регрессии, а соответствующее ей аналитическое выражение – уравнением регрессии¹.

Вид зависимости $\bar{y}(x)$ может быть известен заранее – из физических или иных соображений. Тогда задача исследователя сводится к оценке параметров аналитического выражения по эмпирическим данным.

Например, если зависимость является линейной, т.е. $\bar{y}(x) = ax + b$, то достаточно найти такие a и b , при которых разброс точек относительно этой прямой будет наименьшим.

Если же вид $\bar{y}(x)$ не известен, то до определения параметров необходимо выдвинуть гипотезу о виде этой зависимости.

Параметры известной или предполагаемой зависимости оцениваются с помощью излагаемого ниже метода наименьших квадратов.

5.1. Метод наименьших квадратов.

Уравнение линейной регрессии для двух переменных

Рассмотрим ситуацию, когда зависимость $\bar{y}(x)$ является линейной. Линейные зависимости довольно часто используются в гидрологической практике, так как аппарат для них хорошо разработан. В пользу линейных моделей говорят и следующие факты.

1. Если вид зависимости $\bar{y}(x)$ не известен, то в качестве первой модели логично испытать линейную как наиболее простую.

¹ Строго говоря, регрессией Y на X называют условное МО СВ Y при фиксированном значении СВ $X = x$.

2. Если зависимость является слабо нелинейной, то область допустимых значений (ОДЗ) переменной X можно разбить на несколько интервалов, в пределах которых зависимость можно считать линейной.

3. Почти всегда переменные Y и X можно преобразовать так, что зависимость между новыми переменными станет линейной.

Искомым уравнением регрессии в случае линейной зависимости является следующее:

$$\tilde{y}_i = ax_i + b. \quad (5.1)$$

Суть метода наименьших квадратов (МНК) состоит в том, чтобы определить такие расчетные параметры a и b , при которых сумма квадратов отклонений наблюдаемых значений y_i от рассчитанных по формуле (5.1) будет иметь минимальное значение.

Сумма квадратов отклонений равна:

$$\sum_{i=1}^n (y_i - \tilde{y}_i)^2 = \sum_{i=1}^n [y_i - (ax_i + b)]^2. \quad (5.2)$$

Для того чтобы сумма была минимальной, необходимо, чтобы частные производные этого выражения по параметрам a и b равнялись нулю:

$$\frac{\partial}{\partial a} \left[\sum_{i=1}^n (y_i - ax_i - b)^2 \right] = 0; \quad (5.3)$$

$$\frac{\partial}{\partial b} \left[\sum_{i=1}^n (y_i - ax_i - b)^2 \right] = 0. \quad (5.4)$$

Продифференцировав выражения (5.3) и (5.4), приходим к двум уравнениям, содержащим параметры a и b (к так называемым нормальным уравнениям):

$$b \sum_{i=1}^n x_i + a \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i; \quad (5.5)$$

$$bn + a \sum_{i=1}^n x_i = \sum_{i=1}^n y_i. \quad (5.6)$$

Решая уравнения (5.5), (5.6) относительно параметров a и b , получаем:

$$a = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad (5.7)$$

$$b = \bar{y} - a\bar{x}, \quad (5.8)$$

где $\bar{x}$ и $\bar{y}$ — средние значения соответственно для выборок из X и Y .

Коэффициент a , который характеризует тангенс угла наклона искомой прямой, принято называть коэффициентом регрессии зависимости $\hat{y}(x)$. Из выражения (5.7) следует, что коэффициент регрессии можно представить в виде

$$a_{y/x} = r \frac{\sigma_y}{\sigma_x}, \quad (5.9)$$

где σ_y и σ_x — среднеквадратические отклонения выборок из Y и X ; r — выборочный коэффициент парной корреляции, определяемый по формуле

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (5.10)$$

Коэффициент r представляет собой эмпирическую меру линейной зависимости между Y и X . Он может принимать значения от -1 до $+1$. Если коэффициент корреляции по модулю равен единице, зависимость является функциональной (точки лежат на прямой без разброса). Знак «плюс» указывает на то, что зависимость прямая, знак «минус» — зависимость обратная. Чем ближе абсолютное значение r к нулю, тем слабее линейная зависимость между Y и X (рис. 5.2).

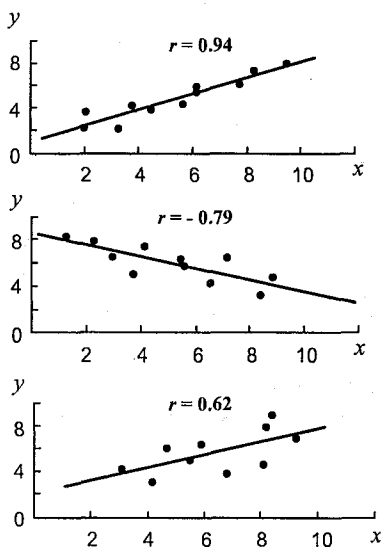


Рис. 5.2. Разброс точек на графиках зависимостей $\tilde{y}_i = ax_i + b$ при различных значениях коэффициента корреляции.

Приведем еще одну формулу для коэффициента парной корреляции

$$r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}. \quad (5.11)$$

В этой формуле $\text{cov}(x, y)$ – ковариация (второй смешанный момент),

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (5.12)$$

Следует отметить, что если при практических расчетах вместо σ_x и σ_y используются их несмещенные эмпирические оценки (см. п. 3.2), то при расчете r величину $1/n$ в формуле (5.12) необходимо заменить на $1/(n-1)$.

С учетом равенств (5.8) и (5.9) уравнение линейной регрессии можно представить в виде:

$$(y_i - \bar{y}) = r \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}). \quad (5.13)$$

Полученное равенство есть уравнение линейной регрессии y по x . Уравнение линейной регрессии x по y получается аналогично и имеет вид:

$$(x_i - \bar{x}) = r \frac{\sigma_x}{\sigma_y} (y_i - \bar{y}). \quad (5.14)$$

Уравнения (5.13) и (5.14) не эквивалентны и не выводятся одно из другого. Проиллюстрируем этот факт с помощью рисунка. Если по оси абсцисс откладывать значения X , а по оси ординат – значения Y , то уравнению (5.13) будет соответствовать минимальное отклонение точек от линии регрессии по вертикали (рис. 5.3), а уравнению (5.14) – по горизонтали.

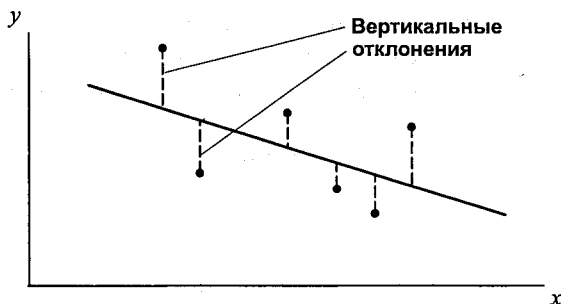


Рис. 5.3. Линия регрессии y по x и вертикальные отклонения, минимизирующие сумму квадратов в МНК.

Нетрудно показать, что прямые, соответствующие уравнениям (5.13) и (5.14), будут пересекаться (рис. 5.4) в точке с координатами $(\bar{x}, \bar{y})$. Угол между этими прямыми φ зависит от коэффициента корреляции. При слабой зависимости φ будет большим, при сильной – малым.

Угол φ будет максимальным ($\varphi = 90^\circ$), когда $r = 0$. Если $|r| = 1$ (функциональная зависимость), то угол φ равен нулю, т.е. линии сольются в одну.

Помимо уравнения регрессии в гидрологической практике для аппроксимации линейной зависимости между X и Y используется уравнение средней линии, которое соответствует биссектрисе угла φ (рис. 5.4). Уравнение средней линии можно получить как из (5.13), так и из (5.14), приравняв r единице.

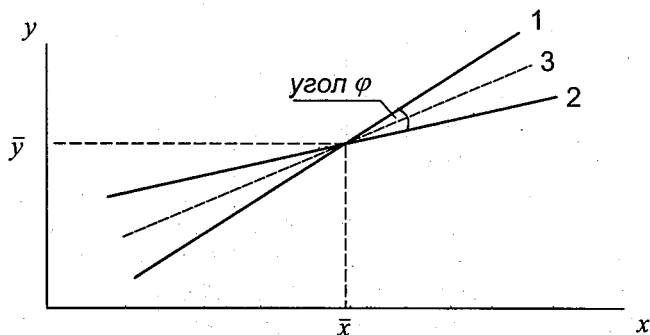


Рис. 5.4. Линии линейной регрессии: 1 – $y = f(x)$; 2 – $x = f(y)$; 3 – средняя линия.

Следует отметить, что аппроксимация по методу наименьших квадратов может сопровождаться дополнительными условиями. Например, можно потребовать, чтобы линия линейной регрессии проходила через начало координат. В этом случае свободный член b равен нулю, а коэффициент регрессии a в соответствии с МНК будет определяться формулой

$$a_{y/x} = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i}{x_i} \right). \quad (5.15)$$

При этом в связи с потерей одного параметра сумма квадратов отклонений для уравнения $y(x) = ax$ будет больше, чем для уравнения $y(x) = ax + b$.

Пример 5.1. По данным о минимальных 30-суточных (x) и минимальных суточных (y) расходах воды на р. Вороне у г. Воронежа за 13 лет (см. табл. 5.1) рассчитать параметры уравнения линейной регрессии.

Решение.

1. Производим расчет, используя вспомогательную табл. 5.2.

Таблица 5.2

Вспомогательная таблица для расчета параметров уравнения линейной регрессии для двух переменных

№ п/п	x_i	y_i	$\Delta x =$ $=(x_i - \bar{x})$	$\Delta y =$ $=(y_i - \bar{y})$	$(\Delta x)^2$	$(\Delta y)^2$	$\Delta x \Delta y$	$\Delta x + \Delta y$	$(\Delta x + \Delta y)^2$
1	8,53	9,53	-7,43	-7,36	55,20	54,17	54,68	-14,79	218,74
2	10,5	11,0	-5,46	-5,85	29,81	34,22	31,94	-11,31	127,92
3	17,0	17,6	1,04	0,71	1,08	0,50	0,74	1,75	3,06
4	16,1	17,9	0,14	1,01	0,02	1,02	0,14	1,15	1,32
5	26,0	27,9	10,04	11,01	100,8	121,2	110,5	21,05	443,10
6	13,4	14,4	-2,56	-2,45	6,55	6,00	6,27	-5,01	25,10
7	15,3	16,0	-0,66	-0,89	0,44	0,79	0,59	-1,55	2,44
8	13,0	14,0	-2,96	-2,89	8,76	8,35	8,55	-5,85	34,22
9	14,6	15,4	-1,36	-1,49	1,85	2,22	2,03	-2,85	8,12
10	12,9	13,3	-3,06	-3,59	9,36	12,89	10,99	-6,65	44,22
11	11,4	12,3	-4,56	-4,59	20,79	21,07	20,93	-9,15	83,72
12	29,7	30,3	13,74	13,41	188,8	179,8	184,2	27,15	737,12
13	19,0	19,0	3,04	3,01	9,24	9,06	9,15	6,05	36,60
Сумма	207,5	219,5	0,0	0,0	423,7	451,3	440,8	0,0	1765,6
Среднее	15,96	16,89							

$$\text{Контроль: } \sum (\Delta x)^2 + \sum (\Delta y)^2 + 2 \sum \Delta x \Delta y = \sum (\Delta x + \Delta y)^2 \rightarrow \\ 432,7 + 451,3 + 2 \cdot 440,8 = 1765,6 \rightarrow 1765,6 = 1765,6.$$

2. На основании таблицы (см. табл. 5.2) рассчитываем параметры уравнения линейной регрессии:

$$\bar{x} = \sum_{i=1}^n x_i / n = 15,96, \quad \bar{y} = \sum_{i=1}^n y_i / n = 16,89;$$

$$\sigma_x^* = \sqrt{\frac{\sum (\Delta x)^2}{n-1}} = \sqrt{\frac{427,7}{12}} = 6,00;$$

$$\sigma_y^* = \sqrt{\frac{\sum (\Delta y)^2}{n-1}} = \sqrt{\frac{451,3}{12}} = 6,13;$$

$$r = \frac{\sum (\Delta x \Delta y)}{\sqrt{\sum (\Delta x)^2 \sum (\Delta y)^2}} = \frac{440,8}{\sqrt{432,7 \cdot 451,3}} = 0,997;$$

$$a_{y/x} = r \frac{\sigma_y^*}{\sigma_x^*} = 0,997 \frac{6,13}{6,00} = 1,02;$$

$$b_{y/x} = \bar{y} - a\bar{x} = 16,89 - 1,02 \cdot 15,96 = 0,61;$$

$$y(x) = 1,02x + 0,61.$$

5.2. Линеаризация нелинейных зависимостей

Если в процессе анализа выясняется, что коэффициент корреляции между переменными Y и X близок к нулю, то это указывает на отсутствие между ними линейной зависимости. Однако не исключено, что зависимость между Y и X все-таки существует, но носит нелинейный характер. Чтобы это проверить, по эмпирическим данным строится график $y = f(x)$. И если точки на графике лежат с небольшим разбросом относительно некоторой кривой, то принимается нелинейная модель. Например, если точки на графике лежат так, как это показано на рис. 5.5 а, можно попытаться использовать для аппроксимации зависимости $y = f(x)$ уравнение экспоненты

$$y = c e^{ax}, \quad (5.16)$$

где a и c — параметры, которые должны определяться по эмпирическим данным.

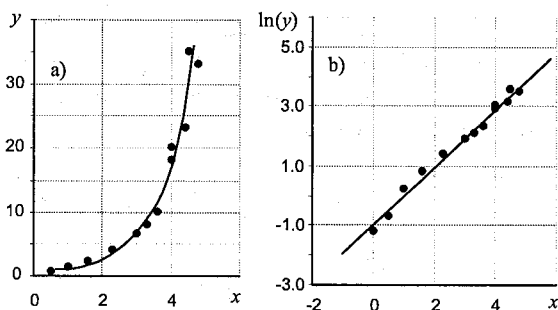


Рис. 5.5. График зависимости $\tilde{y}_i = c \exp(ax_i)$ без линейризации (а) и линейризованный на основе логарифмического преобразования (б).

Метод наименьших квадратов позволяет найти параметры и в случае нелинейной модели, однако можно существенно упростить расчеты, проведя линейризацию исходного выражения.

Прологарифмировав обе части выражения (5.16), получим

$$\ln(y) = \ln(c) + ax. \quad (5.17)$$

Вместо переменных y и x введем новые: $y^1 = \ln(y)$; $x^1 = x$. Выражение (5.17) перепишем в виде

$$y^1 = ax^1 + b, \quad (5.18)$$

где $b = \ln(c)$.

Как видно из (5.18), новые переменные y^1 и x^1 связаны линейной зависимостью (рис. 5.5 б), и для оценки параметров a и b можно применить методику, изложенную ранее (см. п. 5.1).

После того как параметры будут найдены, проводят обратное преобразование. В данном случае параметры выражения (5.16) следующие: $c = e^b$, а показатель степени a равен коэффициенту регрессии в выражении (5.18).

В табл. 5.3 приводятся примеры преобразований, применяемых при линейризации нелинейных зависимостей.

Таблица 5.3

Преобразования, применяемые при линейризации нелинейных зависимостей

Исходная зависимость	Преобразование		Уравнение в линейной форме
	Абсцисса	Ордината	
$y = c e^{ax}$	x	$\ln(y)$	$\ln(y) = ax + \ln(c)$
$y = c x^a$	$\ln(x)$	$\ln(y)$	$\ln(y) = a \ln(x) + \ln(c)$
$y = (a/x) + b$	$1/x$	y	$y = a(1/x) + b$
$y = x/(ax + b)$	x	x/y	$(x/y) = ax + b$
$y = c/(kx + m)$	x	$1/y$	$(1/y) = (k/c)x + (m/c)$

При использовании перечисленных преобразований следует иметь в виду, что параметры исходного выражения, полученные методом наименьших квадратов, будут смещенными, так как в данном случае минимизируется сумма квадратов отклонений преобразованных величин. Однако в большинстве практических ситуаций такой подход оказывается вполне приемлемым, так как ошибка аппроксимации, обусловленная смещенностью параметров, находится в пределах точности исходных данных.

5.3. Оценка точности уравнения линейной регрессии для двух переменных

Полученное на основе эмпирических данных уравнение линейной регрессии может использоваться в качестве прогностической зависимости или расчетной формулы, однако это является корректным только в случае, если данное уравнение обеспечивает необходимую точность расчета, или, проще говоря, является надежным.

Ранее отмечалось, что в качестве меры линейной зависимости между двумя переменными используется коэффициент парной корреляции. В соответствии с действующими в гидрологии нормативными документами зависимость может использоваться для практических расчетов, если $|r| \geq 0,7$.

Смысл этого неравенства будет более понятен, если рассмотреть величину r^2 , которую называют *коэффициентом детерминации*. Коэффициент детерминации характеризует долю общего разброса относительно среднего значения $\bar{y}$, объясняемую регрессией. Если зависимость функциональная и точки относительно линии регрессии лежат без разброса, то весь разброс относительно $\bar{y}$ объясняется регрессией, т.е. $r^2 = 1$. При $r = 0,7$ коэффициент детерминации $r^2 = 0,49 \approx 0,5$. Таким образом, в гидрологии зависимость может использоваться для практических расчетов, если регрессия объясняет 50 % или более разброса относительно $\bar{y}$. Коэффициент корреляции – важный, но не единственный показатель надежности уравнения регрессии. Рассмотрим еще ряд статистических характеристик, позволяющих судить о точности полученного уравнения.

– $\sigma_{y(x)}$ – *стандартная ошибка уравнения линейной регрессии*. По сути эта величина характеризует среднеквадратическое отклонение точек от принятой линии регрессии.

$$\sigma_{y(x)} = \sqrt{\sum_{i=1}^n (y_i - \tilde{y}_i)^2 / (n-2)}, \quad (5.19)$$

где y_i — наблюдаемая величина; $\tilde{y}_i$ — величина, рассчитанная по уравнению регрессии (в зависимости от x_i). Параметр $(n-2)$ называют числом степеней свободы.

Число степеней свободы равно числу наблюдений минус число параметров, определяемых по эмпирическим данным. В данном случае таких параметров два: коэффициент регрессии a и свободный член b .

Используя коэффициент корреляции, выражение (5.19) можно представить в виде

$$\sigma_{y(x)} = \sigma_y^* \sqrt{\frac{(1-r^2)(n-1)}{(n-2)}}, \quad (5.20)$$

где σ_y^* — несмещенная оценка СКО для ряда Y .

Иногда при практических расчетах пренебрегают величиной $\sqrt{(n-1)/(n-2)}$ и используют более простую формулу

$$\sigma_{y(x)} = \sigma_y^* \sqrt{1-r^2}. \quad (5.21)$$

— σ_r — стандартная ошибка коэффициента парной корреляции:

$$\sigma_r = (1-r^2) / \sqrt{n-1}. \quad (5.22)$$

При малой длине выборок ($n < 25$) В.Н. Романовский рекомендует вводить в формулу (5.22) поправочный коэффициент:

$$\sigma_r = \frac{1-r^2}{\sqrt{n-1}} \sqrt{1 + \frac{11r^2}{2n} + \frac{75r^2-13}{2n^2}}. \quad (5.23)$$

— σ_a — стандартная ошибка коэффициента регрессии:

$$\sigma_a = \sigma_{y(x)} / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (5.24)$$

С учетом выражения (5.20) формулу (5.24) можно представить в виде

$$\sigma_a = \frac{\sigma_y^*}{\sigma_x^*} \sqrt{\frac{1-r^2}{n-2}}, \quad (5.25)$$

где σ_x^* и σ_y^* — несмещенные оценки СКО соответственно для рядов X и Y .

— σ_b — стандартная ошибка свободного члена:

$$\sigma_b = \sigma_{y(x)} \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (5.26)$$

или с учетом формулы (5.20)

$$\sigma_b = \sigma_y^* \sqrt{\frac{1-r^2}{n-2}} \sqrt{1 + \left(\frac{\bar{x}}{\sigma_x^*} \right)^2}. \quad (5.27)$$

Сформулируем теперь полный набор требований, предъявляемых в гидрологии к уравнению линейной регрессии.

Уравнение можно рекомендовать для практических расчетов, если выполнены следующие условия:

$$n \geq 10; \quad |r| \geq 0,7; \quad |r| / \sigma_r \geq 2; \quad |a| / \sigma_a \geq 2. \quad (5.28)$$

Желательным (но не обязательным) является условие:

$$\frac{|b|}{\sigma_b} \geq 2.$$

При малых значениях параметра b это требование почти никогда не выполняется.

Помимо комплекса условий (5.28) при анализе уравнения линейной регрессии следует учитывать, что истинное значение y_i , соответствующее аргументу x_i , будет отличаться от $\hat{y}_i$, полученного по уравнению линейной регрессии (5.1), на некоторую величину ϵ_i . При этом предполагается, что ϵ_i является нормально распределенной случайной величиной с нулевым средним значением и среднеквадратическим отклонением $\sigma_{y(x)}$.

Рассмотренная система оценок позволяет, в большинстве случаев, составить правильное представление о надежности полученного уравнения регрессии и принять решение о возможности его применения для практических расчетов. Однако иногда такого анализа оказывается недостаточно. Проиллюстрируем это с помощью рис. 5.6.

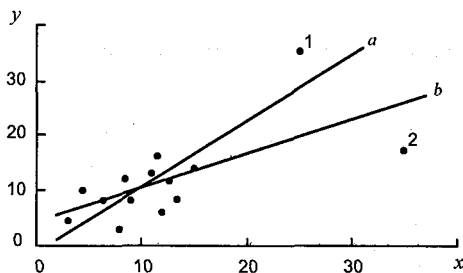


Рис. 5.6. Линии линейной регрессии $\hat{y}(x)$:

a – без учета точки 2 ($r = 0,83$); b – с учетом всех точек ($r = 0,66$)

Как видно на рисунке, в данном случае эмпирические точки расположены неравномерно. Точки основной группы образуют поле рассеяния в диапазоне значений X от 3 до 15, а две точки расположены в области больших значений X . Первая имеет координаты $x = 25$, $y = 35$, вторая $x = 35$, $y = 17$.

Значение коэффициента корреляции, полученное для точек основной группы значительно меньше 0,7 ($r = 0,45$). Но если к основной группе точек добавить точку № 1, коэффициент корреляции резко возрастет ($r = 0,82$). Использование всех данных, включая и точку № 2, вновь приводит к значительному изменению коэффициента корреляции ($r = 0,66$). Заметно изменится и положение линии регрессии.

Таким образом, здесь мы сталкиваемся с ситуацией, когда добавление одной или двух точек может существенно повлиять на параметры уравнения регрессии. Если в данном случае ориентироваться только на систему ограничений (5.28), можно принять ошибочное решение.

Действительно, как видно из табл. 5.4, во втором варианте, когда из совокупности данных исключена точка № 2, условие (5.28) выполняется.

Таблица 5.4

**Параметры уравнения линейной регрессии для зависимости,
представленной на рисунке 5.6**

Параметр	Вариант № 1 (основная группа без точек № 1 и 2)	Вариант № 2 (основная группа плюс точка № 1)	Вариант № 3 (все точки)
n	12	13	14
r	0,45	0,82	0,66
a	0,48	1,18	0,61
b	4,86	-1,34	4,10
$\sigma_{y(x)}$	3,63	4,82	6,13
σ_r	0,24	0,096	0,17
σ_a	0,30	0,25	0,20
σ_b	3,06	3,19	3,05
r/σ_r	1,88	8,54	3,88
a/σ_a	1,60	4,72	3,05
b/σ_b	1,59	0,42	1,34
Выполнение условий (5.28)	Не выполняются	Выполняются	Не выполняются

Основная причина возникших здесь трудностей состоит в том, что выборочные точки группируются крайне неравномерно. В подобных ситуациях следует провести расчет параметров уравнения регрессии с учетом и без учета «тяжелой» точки, и если они существенно различаются, уравнение не следует использовать в качестве расчетного, по крайней мере до тех пор, пока не появятся дополнительные данные. При этом надо понимать, что дополнительные данные могут как подтвердить, так и опровергнуть гипотезу о наличии высокой корреляции между X и Y , но в любом случае решение будет более обоснованным.

Ситуации, подобные той, которую мы рассмотрели, не так уж редки в гидрологической практике. Можно назвать, по крайней мере, две причины, по которым эмпирические точки могут группироваться очень неравномерно.

1. Метод наименьших квадратов подразумевает, что ряды X и Y являются выборками из нормальных совокупностей, что обеспечивает примерно равное количество больших и малых значений, а выборочная асимметрия гидрологических величин (особенно при небольших n) может быть весьма значительной.

2. Гидрологи, как правило, работают в условиях так называемого пассивного эксперимента и не могут по своему усмотрению изменять независимую переменную. В результате всегда есть риск получить нерепрезентативные выборки, особенно если длина выборки n недостаточно велика.

5.4. Интервальная оценка и оценка значимости параметров линейной регрессии для двух переменных

5.4.1. Интервальная оценка коэффициента парной корреляции. Z-преобразование Фишера

Коэффициент парной корреляции, определяемый по выборкам конечной длины, является случайной величиной. Если значение r не очень велико и длина выборок не превышает 40 лет, то распределение выборочных коэффициентов корреляции хорошо аппроксимируется нормальным законом со среднеквадратическим отклонением σ_r , определяемым по формуле (5.22). В этом случае доверительный интервал для истинного коэффициента корреляции можно представить в виде

$$r^* - t'_{1-\alpha} \sigma_r^* \leq r < r^* + t'_{1-\alpha} \sigma_r^*, \quad (5.29)$$

где r^* – выборочный коэффициент парной корреляции; $t'_{1-\alpha}$ – квантиль стандартного нормального распределения (см. табл. 2.1), соответствующий двустороннему уровню значимости 2α .

Пример 5.2. При $r^* = 0,82$; $\sigma_r^* = 0,096$; $n = 13$ (см. табл. 5.4, вариант № 2) найти интервальную оценку коэффициента корреляции в виде 90%-ного доверительного интервала.

Решение.

1. Определяем, какой двусторонний уровень значимости соответствует принятому доверительному интервалу: $2\alpha = 1 - 0,90 = 0,10$ (или 10 %).

2. Определяем α : $\alpha = 0,10/2 = 0,05$ (или 5 %).

3. По табл. 2.1 находим квантиль $t'_{1-\alpha}$: $t'_{1-\alpha} = t'_{0,95} = 1,64$.

4. Строим доверительный интервал для r : $0,82 - 1,64 \cdot 0,096 \leq r < 0,82 + 1,64 \cdot 0,096 \rightarrow 0,66 \leq r < 0,98$.

При больших значениях r и малой длине выборок ($r > 0,4$; $n < 40$), распределение выборочных значений коэффициента корреляции при

уменьшении n начинает все больше отклоняться от нормального, приобретая отрицательную асимметрию.

В этом случае при построении доверительного интервала рекомендуется применять Z -преобразование Фишера, основанное на использовании специальной статистики Z , функционально связанной с r выражением

$$Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right). \quad (5.30)$$

В отличие от r статистика Z имеет нормальное распределение даже при малых n . СКО Z определяется формулой

$$\sigma_z = 1/\sqrt{n-3}. \quad (5.31)$$

Схема построения интервальной оценки r при использовании преобразования Фишера следующая.

1. В зависимости от r по формуле (5.30) рассчитывается эмпирическое значение Z^* .

2. По формуле (5.31) вычисляется σ_z^* .

3. Строится доверительный интервал для Z :

$$Z^* - t_{1-\alpha} \sigma_z^* \leq Z < Z^* + t_{1-\alpha} \sigma_z^*. \quad (5.32)$$

4. Строится доверительный интервал для коэффициента корреляции путем обратного перехода от Z к r . Формулу обратного перехода от Z к r можно получить из (5.30):

$$r = (e^{2Z} - 1) / (e^{2Z} + 1). \quad (5.33)$$

Пример 5.3. По данным примера 5.2 найти интервальную оценку коэффициента корреляции в виде 90%-ного доверительного интервала (применить Z -преобразование Фишера).

Решение.

Последовательно реализуем изложенную выше схему.

$$1. Z^* = \frac{1}{2} \ln \left(\frac{1+0,82}{1-0,82} \right) = 1,157.$$

$$2. \sigma_z^* = 1/\sqrt{13-3} = 0,316.$$

$$3. (1,57 - 1,64 \cdot 0,316) \leq Z < (1,57 + 1,64 \cdot 0,316) \rightarrow 0,639 \leq Z < 1,675.$$

$$4. \frac{\exp(2 \cdot 0,639) - 1}{\exp(2 \cdot 0,639) + 1} \leq r < \frac{\exp(2 \cdot 1,675) - 1}{\exp(2 \cdot 1,675) + 1} \rightarrow 0,56 \leq r < 0,93.$$

Так как в данном примере величина r^* существенно больше 0,4, а $n < 40$, интервальные оценки r , полученные с учетом и без учета Z -преобразования Фишера, заметно различаются. При этом последняя оценка является более корректной.

Следует также отметить, что при $r < 0,4$ и $n > 40$ интервальная оценка r , полученная с учетом Z -преобразований Фишера, будет совпадать с оценкой, полученной без него. Поэтому Z -преобразование Фишера можно рекомендовать для любых сочетаний r и n .

Выборочный коэффициент корреляции можно использовать для проверки значимости линейной зависимости между X и Y . В этом случае выдвигается нулевая гипотеза $H_0: r = 0$. Гипотеза опровергается, если

$$\frac{|r^*|}{\sigma_r^*} > t'_{1-\alpha}, \quad (5.34)$$

и связь считается статистически значимой. Если условие (5.34) не выполняется, связь считается статистически незначимой, т.е. линейной зависимости между X и Y отсутствует.

Заметим, что при уровне значимости $2\alpha = 5\%$ получаем $t'_{1-\alpha} = t'_{97,5} = 1,96 \approx 2$, и выражение (5.34) в этом случае идентично третьему выражению системы ограничений (5.28).

5.4.2. Коэффициент ранговой корреляции Спирмэна

Если нужно оценить степень взаимосвязи между рядами $y_1, y_2, \dots, y_n$ и $x_1, x_2, \dots, x_n$, имеющими распределение существенно отличное от нормального, то можно использовать коэффициент ранговой корреляции Спирмэна r_s :

$$r_s = 1 - \frac{6 \sum_{i=1}^{i=n} \Delta_i^2}{n(n^2 - 1)}, \quad (5.35)$$

где n — длина выборок; Δ_i — разность рангов для пары значений y_i и x_i (ранжирование производится в возрастающем порядке).

Так же как для коэффициента парной корреляции, для коэффициента ранговой корреляции выполняется условие:

$$-1 \leq r_s \leq +1. \quad (5.36)$$

Выборочный коэффициент ранговой корреляции r_s^* обычно используется для проверки значимости линейной зависимости между Y и X . В этом случае выдвигается нулевая гипотеза $H_0: r_s = 0$, т.е. предполагается, что связь между Y и X отсутствует. Гипотеза опровергается, если

$$|r_s^*| > (r_s)_\alpha, \quad (5.37)$$

где $(r_s)_\alpha$ – критическое значение коэффициента ранговой корреляции при одностороннем уровне значимости α .

Для $n \leq 30$ критические значения $(r_s)_\alpha$ представлены в табл. 5.5.

Таблица 5.5

Критические значения коэффициента ранговой корреляции Спирмена, r_s

n	Уровень значимости, α %				
	0,5	1	2,5	5	10
5		0,9000	0,9000	0,8000	0,7000
6	0,9429	0,8857	0,8286	0,7714	0,6000
7	0,8929	0,8571	0,7450	0,6786	0,5357
8	0,8571	0,8095	0,6905	0,5952	0,4762
9	0,8167	0,7667	0,6833	0,5833	0,4667
10	0,7818	0,7333	0,6364	0,5515	0,4424
11	0,7545	0,7000	0,6091	0,5273	0,4182
12	0,7273	0,6713	0,5804	0,4965	0,3986
13	0,6978	0,6429	0,5549	0,4780	0,3791
14	0,6747	0,6220	0,5341	0,4593	0,3626
15	0,6536	0,6000	0,5179	0,4429	0,3500
16	0,6324	0,5824	0,5000	0,4265	0,3382
17	0,6152	0,5637	0,4853	0,4118	0,3260
18	0,5975	0,5480	0,4716	0,3994	0,3148
19	0,5825	0,5333	0,4579	0,3895	0,3070
20	0,5684	0,5203	0,4451	0,3789	0,2977
21	0,5545	0,5078	0,4351	0,3688	0,2909
22	0,5426	0,4963	0,4241	0,3597	0,2829
23	0,5306	0,4852	0,4150	0,3518	0,2767
24	0,5200	0,4748	0,4061	0,3435	0,2704
25	0,5100	0,4654	0,3977	0,3362	0,2646
26	0,5002	0,4564	0,3894	0,3299	0,2588
27	0,4915	0,4481	0,3822	0,3236	0,2540
28	0,4828	0,4401	0,3749	0,3175	0,2490
29	0,4744	0,4320	0,3685	0,3113	0,2443
30	0,4665	0,4251	0,3620	0,3059	0,2400

При $n \geq 30$ величина $(r_s \sqrt{n-1})$ с достаточной точностью описывается стандартным нормальным распределением. В этом случае нулевая гипотеза отвергается, если выполняется неравенство

$$|r_s^*| > \frac{t'_{1-\alpha}}{\sqrt{n-1}}, \quad (5.38)$$

где $t'_{1-\alpha}$ – квантиль стандартного нормального распределения при одностороннем уровне значимости α .

Пример 5.6.

В таблице 5.6 представлены среднегодовые расходы рек Уверь – д. Меглецы (y_i) и Мста – пос. Потерпелицы (x_i) за 30 лет (с 1959 по 1988 г.).

Требуется рассчитать коэффициент ранговой корреляции и оценить его значимость при $\alpha = 5\%$.

Решение.

1. Ряды y_i и x_i ранжируются в возрастающем порядке. Результат записывается в столбцы 4 и 5 табл. 5.6.

2. Каждому значению y_i и x_i в ранжированном ряду присваивается порядковый номер (ранг). Самый маленький расход получает первый ранг, следующий по величине расход получает второй ранг и т.д.

3. Каждому расходу исходного ряда ставится в соответствие его ранг. Эти ранги записываются в столбцы 6 и 7.

4. Рассчитывается разность рангов y_i и x_i (столбец 8).

5. Рассчитывается квадрат разности рангов Δ^2 (столбец 9).

6. Определяется $\sum_{i=1}^{i=n} \Delta_i^2$.

7. По формуле (5.35) рассчитывается коэффициент ранговой

$$\text{корреляции: } r_s = 1 - \frac{6 \sum_{i=1}^{i=n} \Delta_i^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 470}{30(30^2 - 1)} = 0,895.$$

8. По табл. 5.5 определяется критическое значение коэффициента ранговой корреляции. При $n = 30$ и $\alpha = 5\%$ получаем: $(r_s)_{5\%} = 0,306$.

Таблица 5.6

Расчет коэффициента ранговой корреляции

№ п/п	Исходные ряды		Ранжированные ряды		Ранг		Δ	Δ^2
	y_i	x_i	y_i	x_i	y_i	x_i		
1	15,2	79,0	8,32	40,4	16	18	-2	4
2	11,6	59,5	9,29	46,6	9	8	1	1
3	17,3	84,3	10,4	47,6	22	22	0	0
4	20,7	95,5	10,5	48,2	29	26	3	9
5	10,4	46,6	10,9	53,2	3	2	1	1
6	11,3	48,2	11,3	54,0	6	4	2	4
7	16,3	69,0	11,4	54,6	19	11	8	64
8	22,3	105	11,5	59,5	30	30	0	0
9	16,1	75,0	11,6	61,6	18	15	3	9
10	14,9	78,8	12,8	66,1	15	17	-2	4
11	15,6	72,3	13,0	69,0	17	13	4	16
12	10,5	53,2	13,8	69,9	4	5	-1	1
13	13,0	54,0	14,3	72,3	11	6	5	25
14	8,32	40,4	14,3	72,5	1	1	0	0
15	9,29	47,6	14,9	75,0	2	3	-1	1
16	12,8	66,1	15,2	77,6	10	10	0	0
17	11,4	54,6	15,6	78,8	7	7	0	0
18	14,3	69,9	16,1	79,0	13	12	1	1
19	18,7	99,8	16,3	79,1	24	28	-4	16
20	19,8	94,9	16,6	79,2	27	25	2	4
21	10,9	61,6	17,0	80,1	5	9	-4	16
22	11,5	72,5	17,3	84,3	8	14	-6	36
23	18,7	87,6	17,9	87,6	25	23	2	4
24	16,6	77,6	18,7	93,7	20	16	4	16
25	18,7	79,2	18,7	94,9	26	20	6	36
26	20,2	93,7	18,7	95,5	28	24	4	16
27	14,3	79,1	19,8	98,4	14	19	-5	25
28	17,9	98,4	20,2	99,8	23	27	-4	16
29	17,0	101	20,7	101	21	29	-8	64
30	13,8	80,1	22,3	105	12	21	-9	81
Сумма							0	470

9. В данном случае имеет место неравенство:

$$(|r_S^*| = 0,895) > ((r_S)_{5\%} = 0,306),$$

и, следовательно, зависимость между y_i и x_i является статистически значимой. Более того, связь является достаточно тесной, так как $r_S^* > 0,7$.

Поскольку в рассмотренном примере $n = 30$, оценку значимости можно выполнить иначе – с использованием неравенства (5.38). Для этого по табл. 2.1 определяем квантиль стандартного нормального распределения при $\alpha = 5\%$; получаем $t'_{1-\alpha} = t'_{95\%} = 1,64$. Следовательно $\frac{t'_{95\%}}{\sqrt{n-1}} = \frac{1,64}{\sqrt{30-1}} = 0,305$.

В итоге получаем практически тот же результат:

$$(|r_S^*| = 0,895) > ((r_S)_{5\%} = 0,305).$$

5.4.3. Интервальная оценка коэффициента регрессии

Если предположить, что разброс наблюдений относительно линии регрессии нормален, то доверительный интервал для коэффициента регрессии имеет вид:

$$a^* - t_{1-\alpha} \sigma_a < a \leq a^* + t_{1-\alpha} \sigma_a, \quad (5.39)$$

где a^* – эмпирическое значение коэффициента регрессии, определенное по формуле (5.7) или (5.9); σ_a – стандартная ошибка коэффициента регрессии, определенная по формуле (5.24) или (5.25); $t'_{1-\alpha}$ – квантиль распределения Стьюдента (см. приложение 5), соответствующий двустороннему уровню значимости 2α при числе степеней свободы $\nu = n - 2$.

Пример 5.4. При $a = 1,18$; $\sigma_a = 0,25$; $n = 13$ (см. табл. 5.4, вариант № 2) найти интервальную оценку коэффициента регрессии в виде 90%-ного доверительного интервала.

Решение.

1. Определяем, какой двусторонний уровень значимости соответствует принятому доверительному интервалу: $2\alpha = 1 - 0,90 = 0,10$ (или 10 %).

2. Определяем α : $\alpha = 0,10/2 = 0,05$ (или 5 %).

3. По таблице распределения Стьюдента (см. приложение 5) при $\nu = 13 - 2 = 11$ находим квантиль $t'_{1-\alpha}$: $t'_{1-\alpha} = t'_{95} = 1,80$.

4. Строим доверительный интервал для a :

$$1,18 - 1,80 \cdot 0,25 \leq a < 1,18 + 1,80 \cdot 0,25 \rightarrow 0,75 \leq a < 1,63.$$

При проверке значимости коэффициента регрессии выдвигается нулевая гипотеза $H_0: a = 0$. Гипотеза опровергается, если

$$|t_a^*| > t_{1-\alpha}, \quad (5.40)$$

где t_a^* — эмпирическое значение статистики Стьюдента, определяемое по формуле

$$t_a^* = a^* / \sigma_a. \quad (5.41)$$

Если равенство (5.40) выполняется, коэффициент регрессии считается статистически значимым, в противном случае коэффициент a^* является статистически незначимым и линейная связь между X и Y отсутствует.

Анализируя выражение (5.41), можно заключить, что при имеющейся длине рядов (20 — 60 членов) четвертое выражение в системе ограничений (5.28) приблизительно соответствует уровню значимости $2\alpha = 5\%$.

5.4.4. Интервальная оценка свободного члена

Если предположить, что разброс наблюдений относительно линии регрессии нормален, то доверительный интервал для свободного члена имеет вид:

$$b^* - t_{1-\alpha} \sigma_b < b \leq b^* + t_{1-\alpha} \sigma_b, \quad (5.42)$$

где b^* — эмпирическое значение свободного члена, определенное по формуле (5.8); σ_b — стандартная ошибка свободного члена, определенная по формуле (5.26) или (5.27); $t_{1-\alpha}$ — квантиль распределения Стьюдента (см. приложение 5), соответствующий двустороннему уровню значимости 2α при числе степеней свободы $\nu = n - 2$.

При проверке значимости свободного члена выдвигается нулевая гипотеза $H_0: b = 0$. Гипотеза опровергается, если

$$|t_b^*| > t'_{1-\alpha}, \quad (5.43)$$

где t_b^* – эмпирическое значение статистики Стьюдента, определяемое по формуле

$$t_b^* = b^* / \sigma_b. \quad (5.44)$$

Если равенство (5.43) выполняется, свободный член считается статистически значимым, в противном случае коэффициент b^* является статистически незначимым и для аппроксимации зависимости между X и Y вместо выражения $\tilde{y}_i = ax_i + b$ следует использовать выражение $\tilde{y}_i = ax_i$.

Пример 5.5. При $b = -1,34$; $\sigma_b = 3,19$; $n = 13$ (см. табл. 5.4, вариант № 2) оценить значимость свободного члена; уровень значимости принять $2\alpha = 10\%$.

Решение.

1. По формуле (5.44) рассчитываем эмпирическое значение статистики Стьюдента $t_b^* = |-1,34/3,19| = 0,42$.

2. По таблице распределения Стьюдента (см. приложение 5) при $\nu = 13 - 2 = 11$ находим квантиль $t_{1-\alpha}$: $t_{1-\alpha} = t_{0,95} = 1,80$.

3. Так как $0,42 < 1,80$, т.е. $t_b^* < t_{1-\alpha}$ нулевая гипотеза H_0 : $b = 0$ не опровергается – коэффициент b статистически незначим и для аппроксимации зависимости между X и Y можно рекомендовать уравнение $\tilde{y}_i = ax_i$.

5.4.5. F-критерий значимости регрессии

Довольно часто для проверки значимости линейной регрессии используется критерий, представляющий собой отношение среднего квадрата, обусловленного регрессией (с числом степеней свободы $\nu = 1$), к среднему квадрату относительно регрессии (с числом степеней свободы $\nu = n - 2$).

$$F_{y(x)}^* = \frac{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \tilde{y}_i)^2 / (n-2)} = \frac{(n-1)S_y^2 r^2}{\sigma_{y(x)}^2}, \quad (5.45)$$

где $\sigma_{y(x)}$ – стандартная ошибка уравнения линейной регрессии, определяемая по формуле (5.19) или (5.20); S_y – несмещенная оценка СКО для ряда Y ; r – коэффициент корреляции; n – длина рядов.

Доказано, что это отношение имеет распределение Фишера с числом степеней свободы $\nu_1 = 1$ и $\nu_2 = n - 2$. Связь считается значимой, если

$$F_{y(x)}^* > F_{1-\alpha}, \quad (5.46)$$

где $F_{1-\alpha}$ – теоретическое значение статистики Фишера (см. приложение 6) при уровне значимости α .

Принято считать, что связь можно использовать для практических расчетов, если значение $F_{y(x)}^*$ превосходит $F_{1-\alpha}$ по крайней мере в 4 – 5 раз (при $\alpha = 5 - 10 \%$).

5.4.6. Построение доверительного интервала для уравнения линейной регрессии

При использовании уравнения линейной регрессии произвольному взятому значению x_k соответствует значение $\hat{y}_k = ax_k + b$. При этом в силу выборочности исходных данных положение линии регрессии определяется с некоторой ошибкой и истинное значение $\hat{y}_{k,0}$ представляет собой случайную величину с МО, равным $\hat{y}_k$ и стандартным отклонением:

$$\sigma_{\text{рег}} = \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{S_x^2(n-1)}}, \quad (5.47)$$

где $\sigma_{y(x)}$ – стандартная ошибка уравнения линейной регрессии, определяемая по формуле (5.19) или (5.20); S_x – несмещенная оценка СКО для ряда X ; $\bar{x}$ – среднее значение для ряда X ; n – длина рядов.

Как видно из формулы (5.47), величина $\sigma_{\text{рег}}$ достигает минимума, когда $x_k = \bar{x}$, и возрастает по мере того, как мы "удаляем" x_k от $\bar{x}$ в любом направлении. Другими словами, чем больше разность между x_k и средним значением, тем больше ошибка, с которой мы будем предсказывать $\hat{y}_{k,0}$ по данному x_k .

С помощью формулы (5.47) можно построить доверительные пределы для уравнения регрессии (рис. 5.7):

$$\tilde{y}_{k,0} = \tilde{y}_k \pm t'_{1-\alpha} \sigma_{\text{рег}}, \quad (5.48)$$

где $\tilde{y}_k = ax_k + b$; $t'_{1-\alpha}$ – квантиль распределения Стьюдента, соответствующий уровню значимости 2α при числе степеней свободы $\nu = n - 2$ (в данном случае ν – это число степеней свободы, на котором основана оценка $\sigma_{y(x)}$).

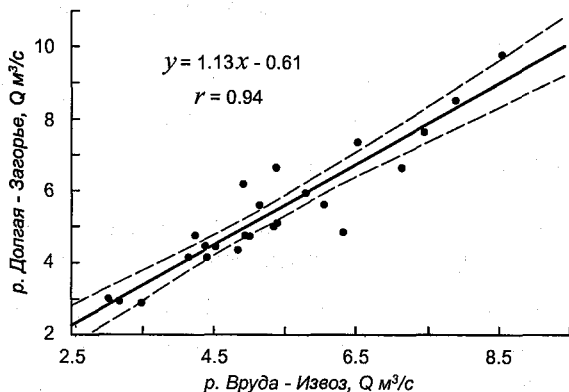


Рис. 5.7. График зависимости среднегодовых расходов воды на р. Долгой – дер. Загорье от среднегодовых расходов на р. Вруде – дер. Извоз. Штриховыми линиями обозначены 95%-ные доверительные пределы для уравнения регрессии.

5.5. Множественная линейная корреляция

При изучении многих гидрологических процессов требуется установить вид линейной зависимости между несколькими переменными. Для решения этой задачи привлекается аппарат множественной линейной корреляции. Сущность этого подхода состоит в распространении основных положений метода линейной корреляции двух переменных на случай зависимости интересующей нас переменной Y от произвольного числа аргументов. Основой для поиска такой зависимости служат материалы наблюдений за величиной Y и определяющими ее величинами $X_1, X_2, \dots, X_m$. Опираясь на данные указанных наблюдений, требуется найти параметры уравнения множественной линейной регрессии, которое согласно

принципу наименьших квадратов будет наилучшим образом описывать связь между зависимой переменной и предикторами. Уравнение линейной регрессии в этом случае будет иметь вид:

$$(y - \bar{y}) = a_1(x_1 - \bar{x}_1) + a_2(x_2 - \bar{x}_2) + \dots + a_m(x_m - \bar{x}_m), \quad (5.49)$$

где $a_1, a_2, \dots, a_m$ — коэффициенты регрессии; $\bar{y}, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ — средние значения соответственно независимой переменной и предикторов; m — число предикторов.

Для решения поставленной задачи нужно решить систему нормальных уравнений относительно коэффициентов $a_1, a_2, \dots, a_m$. Обычно эта задача решается на компьютере с использованием численных методов линейной алгебры. При небольшом числе предикторов можно получить аналитическое решение задачи методом Крамера. В этом случае для определения коэффициентов a_j необходимо рассчитать главный определитель системы D :

$$D = \begin{vmatrix} 1 & r_{0,1} & r_{0,2} & r_{0,3} & \dots & r_{0,j} & \dots & r_{0,m} \\ r_{1,0} & 1 & r_{1,2} & r_{1,3} & \dots & r_{1,j} & \dots & r_{1,m} \\ r_{2,0} & r_{2,1} & 1 & r_{2,3} & \dots & r_{2,j} & \dots & r_{2,m} \\ r_{3,0} & r_{3,1} & r_{3,2} & 1 & \dots & r_{3,j} & \dots & r_{3,m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ r_{i,0} & r_{i,1} & r_{i,2} & r_{i,3} & \dots & r_{i,j} & \dots & r_{i,m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ r_{m,0} & r_{m,1} & r_{m,2} & r_{m,3} & \dots & r_{m,j} & \dots & 1 \end{vmatrix}, \quad (5.50)$$

где r_{ij} — парные коэффициенты корреляции между переменными. Например, $r_{0,3}$ — это коэффициент корреляции между зависимой переменной и третьим предиктором, а $r_{2,1}$ — коэффициент корреляции между вторым и первым предикторами. При этом $r_{0,3} = r_{3,0}$, а $r_{2,1} = r_{1,2}$. На главной диагонали этого определителя стоят единицы, так как $r_{0,0} = r_{1,1} = r_{2,2} = \dots = r_{m,m} = 1$.

Коэффициенты регрессии будут определяться выражением

$$a_j = -(-1)^j \frac{\sigma_y}{\sigma_{x,j}} \frac{D_{0,j}}{D_{0,0}}, \quad (5.51)$$

где σ_y – среднее квадратическое отклонение независимой переменной; $\sigma_{x,j}$ – среднее квадратическое отклонение j -того предиктора; $D_{0,0}$ – определитель, получаемый из главного определителя системы (5.50) путем вычеркивания из него нулевой строки и нулевого столбца; $D_{0,j}$ – определитель, получаемый из главного определителя путем вычеркивания из него нулевой строки и j -го столбца¹.

Рассмотрим частный случай, когда число переменных равно трем (зависимая переменная и два предиктора). В этом случае:

$$D = \begin{vmatrix} 1 & r_{0,1} & r_{0,2} \\ r_{1,0} & 1 & r_{1,2} \\ r_{2,0} & r_{2,1} & 1 \end{vmatrix}; \quad (5.52)$$

$$D_{0,0} = \begin{vmatrix} 1 & r_{1,2} \\ r_{2,1} & 1 \end{vmatrix} = 1 - r_{2,1}r_{1,2} = 1 - r_{1,2}^2; \quad (5.53)$$

$$D_{0,1} = \begin{vmatrix} r_{1,0} & r_{1,2} \\ r_{2,0} & 1 \end{vmatrix} = r_{1,0} - r_{2,0}r_{1,2}; \quad (5.54)$$

$$D_{0,2} = \begin{vmatrix} r_{1,0} & 1 \\ r_{2,0} & r_{2,1} \end{vmatrix} = r_{1,0}r_{2,1} - r_{2,0}. \quad (5.55)$$

Подставляя выражения (5.53) – (5.55) в формулу (5.51), получаем:

$$a_1 = + \frac{\sigma_y}{\sigma_{x,1}} \frac{(r_{1,0} - r_{2,0}r_{1,2})}{1 - r_{1,2}^2}; \quad a_2 = - \frac{\sigma_y}{\sigma_{x,2}} \frac{(r_{1,0}r_{2,1} - r_{2,0})}{1 - r_{1,2}^2}.$$

При увеличении числа предикторов выражения для коэффициентов регрессии становятся все более громоздкими, поэтому метод Крамера обычно используют при числе предикторов $m \leq 3$.

В настоящее время разработано большое количество компьютерных программ, обеспечивающих быстрый и качественный расчет параметров множественной регрессии. Такая программа, в частности, входит в "пакет анализа" электронных таблиц Microsoft Excel.

¹ В линейной алгебре определитель $D_{i,j}$ называется минором определителя D .

5.5.1. Оценка точности уравнения множественной линейной регрессии

Так же как и в случае парной линейной корреляции, теснота связи при множественной линейной корреляции оценивается коэффициентом корреляции, который называют *общим*, или *сводным*, коэффициентом корреляции:

$$R = \sqrt{1 - \frac{D}{D_{0,0}}} . \quad (5.56)$$

Сводный коэффициент корреляции можно также рассчитать как парный коэффициент корреляции между фактическими значениями независимой переменной и значениями независимой переменной, полученными по уравнению множественной линейной регрессии при тех же значениях аргументов.

В отличие от парного коэффициента корреляции, который может изменяться от -1 до $+1$, коэффициент множественной корреляции всегда положителен и имеет пределы изменения от 0 до $+1$.

Стандартная ошибка коэффициента множественной корреляции определяется по формуле

$$\sigma_R = \frac{1 - R^2}{\sqrt{n - m - 1}} , \quad (5.57)$$

где n — длина анализируемых рядов; m — число предикторов.

Стандартную ошибку уравнения множественной линейной регрессии можно оценить по формуле

$$\sigma_{Y(X)} \approx \sigma_y^* \sqrt{1 - R^2} . \quad (5.58)$$

Стандартная ошибка j -того коэффициента регрессии определяется выражением

$$\sigma_{a_j} = \frac{\sigma_y^*}{\sigma_{x_j}^*} \sqrt{\frac{(1 - R^2) \Delta_{j,j}}{(n - m - 1) D_{0,0}}} , \quad (5.59)$$

где $\Delta_{j,j}$ — минор определителя $D_{0,0}$.

В соответствии с действующими нормативными документами уравнение множественной линейной регрессии должно удовлетворять требованиям, аналогичным (5.28), т.е.

$$R \geq 0,7; \quad (R/\sigma_R) \geq 2; \quad (a_j/\sigma_{a,j}) \geq 2; \quad n \geq 10. \quad (5.60)$$

При этом следует обратить внимание, что третье условие считается выполненным, если оно выполняется для каждого коэффициента регрессии в отдельности. Что касается четвертого условия, то оно в данном случае является слишком слабым. Практика показывает, что при использовании одного предиктора длина рядов n должна быть не менее 10, при двух предикторах минимальная длина рядов должна составлять 25 – 30, при четырех 50 – 60, при пяти 100 – 120 и т.д. Только в этом случае можно получить более или менее надежные оценки параметров уравнения регрессии. Поскольку продолжительность гидрологических рядов относительно невелика в практике гидрологических расчетов редко используются регрессионные модели, содержащие более четырех предикторов.

5.5.2. Выбор "наилучшего" уравнения регрессии

В расчетах по уравнению множественной регрессии необходимо использовать только те предикторы, которые дают объем информации, превышающий ошибку расчета. Неэффективные предикторы исключаются. В противном случае привлечение дополнительных данных может привести не к уменьшению ошибки, а к ее увеличению. Назовем несколько причин такой неэффективности.

- Связь между зависимой переменной и предиктором отсутствует или выражена очень слабо.

- Связь между зависимой переменной и предиктором существует, но в силу небольшой продолжительности рядов привлечение дополнительного предиктора приводит к тому, что коэффициенты регрессии становятся незначимыми. В частности, перестает выполняться требование $(a_j/\sigma_{a,j}) \geq 2$ для одного или нескольких коэффициентов регрессии.

- Наличие дублирующих предикторов. Эта ситуация возникает, когда несколько предикторов имеют высокий коэффициент взаимной корреляции. Определитель $D_{0,0}$ близок к нулю и, как следует

из формулы (5.51), система нормальных уравнений будет неопределенной, либо будет иметь неустойчивое решение.

Однако даже если неэффективные предикторы отсеяны, может оказаться, что из имеющегося набора предикторов можно получить несколько уравнений, удовлетворяющих набору требований (5.60). В этом случае необходимо выбрать "наилучшее" уравнение регрессии. Для решения этой задачи разработано несколько методов.

Метод всех возможных регрессий. При использовании этого метода рассчитываются все возможные уравнения регрессии, после чего исключаются те, которые не удовлетворяют системе ограничений (5.56). Из оставшихся уравнений наилучшим будет то, которое имеет самый высокий сводный коэффициент корреляции.

Метод исключения. Метод исключения более экономичен, чем метод всех регрессий, поскольку в нем делается попытка использовать только наилучшие регрессионные уравнения. Основные шаги этого метода сводятся к следующему:

1) рассчитывается регрессионное уравнение, включающее все переменные.

2) вычисляется значение частного F -критерия¹ для каждого предиктора в предположении, как будто это последняя переменная, введенная в регрессионное уравнение.

3) наименьшее значение частного F -критерия, обозначаемое, скажем, как F_L , сравнивается с заранее выбранным критическим значением, например F_0 , далее возможны два варианта:

– если $F_L < F_0$, то предиктор X_L исключается и производится пересчет уравнения регрессии с учетом оставшихся переменных; затем переходят к следующему шагу;

– если $F_L > F_0$, то уравнение регрессии оставляют без изменений и процедура поиска "наилучшего" уравнения завершается.

Шаговый регрессионный метод (метод включения). На первом этапе выбирается предиктор X , наиболее сильно скоррелированный с Y (предположим, что это X_1), и находят линейное регрессионное уравнение первого порядка $Y = f(X_1)$. Затем проверяется, значима ли эта переменная. Если это не так, то мы должны согласиться с выводом, что наилучшая модель выражается уравнением

¹ Частный F -критерий связан с проверкой гипотезы $H_0: a_j = 0$ против альтернативной $H_1: a_j \neq 0$ для любого отдельного коэффициента регрессии.

$Y = \bar{Y}$. В противном случае ищем второй предиктор для включения в модель. Обычно в модель включают тот предиктор, который имеет самый высокий коэффициент частной корреляции с зависимой переменной. Частный коэффициент корреляции характеризует индивидуальный вклад предиктора в общую регрессию в дополнение к тем вкладам, которые дают другие, включенные в уравнение предикторы. Так, если в уравнение множественной регрессии включается дублирующий предиктор, то его частный коэффициент корреляции с зависимой переменной будет низким. Частный коэффициент корреляции между зависимой переменной и j -м предиктором определяется по формуле

$$\eta_{0,j} = -\frac{A_{0,j}}{\sqrt{A_{0,0}A_{j,j}}}, \quad (5.61)$$

где A_{ij} — алгебраическое дополнение главного определителя (5.50) системы

$$A_{i,j} = (-1)^{i+j} D_{i,j}, \quad (5.62)$$

здесь D_{ij} — минор главного определителя системы, т.е. определитель получаемый из определителя D путем вычеркивания из него i -той строки и j -того столбца.

Заметим, что при наличии только двух предикторов формулы для частных коэффициентов корреляции имеют достаточно простой вид:

$$r_{01.2} = \frac{r_{01} - r_{02} \cdot r_{12}}{\sqrt{(1 - r_{02}^2)(1 - r_{12}^2)}}; \quad (5.63)$$

$$r_{02.1} = \frac{r_{02} - r_{01} \cdot r_{12}}{\sqrt{(1 - r_{01}^2)(1 - r_{12}^2)}}, \quad (5.64)$$

где $r_{01.2}$ и $r_{02.1}$ — частные коэффициенты корреляции зависимой переменной соответственно с первым и вторым предиктором; r_{01} , r_{02} , r_{12} — парные коэффициенты корреляции.

После включения в модель второго предиктора находится второе регрессионное уравнение $Y = f(X_1, X_2)$.

Полное уравнение проверяется на значимость. Отмечается улучшение сводного коэффициента корреляции и исследуются ча-

стные F -критерии для обеих переменных. Наименьшее из этих частных F -критериев (F_L) сравнивается с табличным значением F -критерия (F_0) при заданном уровне значимости. Если $F_L < F_0$, то соответствующий предиктор исключается из уравнения. Такая проверка "наименее полезного предиктора" проводится на каждом шаге. Может оказаться, что предиктор, который на предыдущем шаге был наилучшим кандидатом на включение в уравнение, на более позднем шаге оказывается ненужным. Это может быть вызвано теми связями, которые существуют между этой и другими переменными, содержащимися теперь в уравнении.

Процесс прекращается, если никакую из переменных, содержащихся в текущем уравнении, не удастся исключить из него, а ближайший наилучший предиктор-претендент не в состоянии занять место в уравнении.

5.5.3. Ложная корреляция

На практике при построении регрессионных моделей в число предикторов иногда включают не только переменные $Y, X_1, X_2, \dots, X_m$, но и их комбинации, например $(X_1 X_2); (X_1/X_2); (1/X_1)$ и т.д. Во многих случаях такой подход оправдан и позволяет повысить сводный коэффициент корреляции. Однако если в число предикторов включить комбинацию зависимой (Y) переменной с одной из независимых переменных (X_j) может возникнуть эффект *ложной корреляции*. Говоря другими словами, мы можем получить высокий коэффициент корреляции, хотя на самом деле никакой связи между Y и переменной X_j не существует.

Действительно, если дисперсия Y в несколько раз превышает дисперсию X , то при построении зависимостей типа $Y = f(YX)$, $Y = f(Y/X)$, $Y = f(Y + X)$ мы, по сути, коррелируем зависимую переменную с самой собой, а X выполняет роль слабого шума.

В частности, если коэффициент вариации Y превышает коэффициент вариации X , то даже при полном отсутствии связи между Y и X , коэффициент корреляции для зависимости $Y = f(YX)$ будет не менее 0,7 (см. приложение 8).

Неучет эффектов ложной корреляции при построении прогностических зависимостей может привести к серьезным ошибкам.

6. СЛУЧАЙНЫЕ ПРОЦЕССЫ

6.1. Основные понятия

Занимаясь гидрологическими расчетами или прогнозами, необходимо четко представлять, какая математическая модель используется для описания вероятностной структуры гидрологического ряда.

До сих пор мы почти всегда подразумевали, что в качестве такой модели используется модель случайной величины. Так, сорокалетний ряд среднегодовых расходов воды можно рассматривать как сорок значений случайной величины, полученных в результате сорока природных опытов. Таким образом можно считать, что этот ряд представляет собой сорокалетнюю выборку из генеральной совокупности.

Если предположить, что эта выборка является репрезентативной, то по ней (на основе методов математической статистики) можно составить удовлетворительное представление обо всей генеральной совокупности, в частности оценить параметры функции распределения. Однако, определить вид функции распределения по такой короткой выборке весьма сложно. Поэтому на практике либо аналитическое выражение функции распределения задается априори, либо функция распределения выбирается из нескольких вариантов по наилучшему соответствию эмпирическим данным с использованием критериев согласия.

Опыт показывает, что большинство гидрологических величин имеет асимметричное распределение, которое хорошо аппроксимируется трехпараметрическими кривыми Крицкого – Менкеля и Пирсона III типа.

В то же время многие теоретические методы, используемые в практике гидрологических расчетов, предполагают, что случайная величина имеет нормальное распределение. Если распределение случайной величины существенно отличается от нормального, то эти методы могут приводить к значительным ошибкам.

Таким образом, недостаточно констатировать, что для описания статистической структуры ряда выбрана модель случайной величины, необходимо еще указать тип функции распределения. При этом формальный подход к выбору функции распределения также может привести к ошибкам.

Например, расходы воды — это всегда положительные величины, абсолютный максимум для которых обычно не известен; коэффициент стока изменяется от нуля до единицы; уровни воды (в зависимости от нуля графика) могут быть как положительными, так и отрицательными. Если во всех этих случаях использовать в качестве функции распределения кривую Крицкого–Менкеля, то наилучшего соответствия эмпирическим данным следует ожидать для ряда расходов воды, так как кривая Крицкого–Менкеля разработана для случайных величин с положительной асимметрией и диапазоном значений от 0 до $+\infty$. В двух других случаях ошибки при аппроксимации закона распределения могут быть весьма значительными.

Однако еще больше шансов совершить ошибку, если модель случайной величины в принципе не подходит для описания вероятностной структуры гидрологического ряда. Такая ситуация возникает, когда имеет значение фактор времени, а значит и последовательность значений в ряде наблюдений.

Модель случайной величины этого не учитывает. В рамках модели случайной величины время выступает в качестве формального счетчика опытов.

Для описания временных последовательностей используется аппарат теории случайных процессов.

Теорией случайных процессов называется математическая наука, изучающая закономерности случайных явлений в динамике их развития.

Поясним это на следующем примере. Рассмотрим гидрографы для конкретного створа некоторой условной реки за 1971–1975 гг.

Каждый из пяти гидрографов представляет собой функцию времени на интервале один год, которую можно рассматривать как реализацию случайного процесса (рис. 6.1). В данном случае мы имеем 5 гидрографов, но очевидно, что их может быть сколько угодно.

Совокупность всех реализаций дает полное представление о случайном процессе.

Если рассматривать произвольный момент времени $t = t_i$, то значения расходов в момент t_i по всем реализациям образуют *сечение случайного процесса*. При этом для любого сечения расход — есть случайная величина.

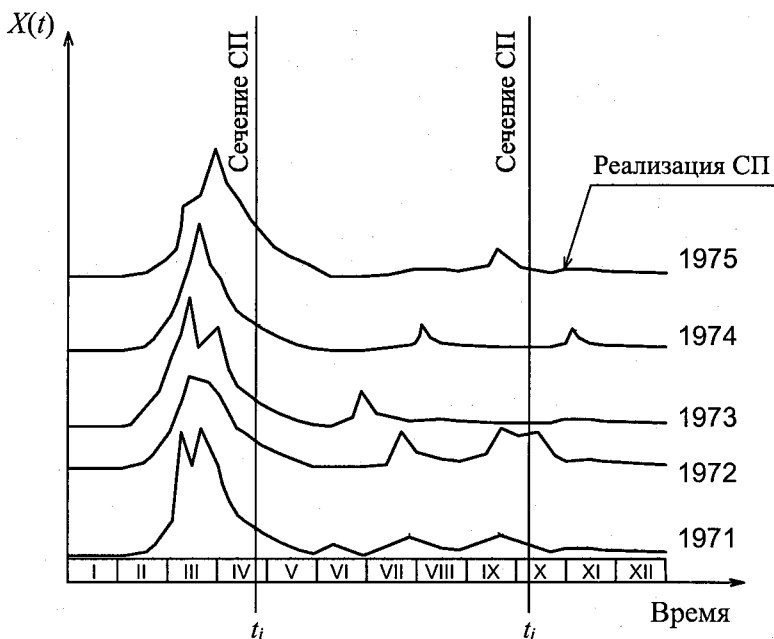


Рис. 6.1. Гидрографы за 1971 – 1975 гг.
как набор реализаций случайного процесса.

В каждом сечении случайная величина может иметь свою функцию распределения. В частности математическое ожидание и дисперсия могут быть различными в разных сечениях. В нашем примере как раз наблюдается такая ситуация, Средний многолетний расход за 15 марта будет, например, существенно выше среднего многолетнего расхода за 15 июля.

В рассмотренном примере мы имели дело с реализациями конечной длины. В общем случае длина реализаций может быть и бесконечной. Если вернуться к примеру со среднегодовыми расходами воды, то сорокалетний ряд можно рассматривать как часть одной бесконечной реализации. Эта реализация к тому же будет единственной.

Сформулируем теперь ряд определений.

Случайным процессом $X(t)$ называется процесс, значение которого при любом фиксированном $t = t_i$ является случайной величиной $X(t_i)$.

Реализацией случайного процесса $X(t)$ называется неслучайная функция $x(t)$, в которую превращается случайный процесс $X(t)$ в результате опыта.

Случайная величина $X(t_i)$, в которую обращается случайный процесс при $t = t_i$, называется *сечением случайного процесса*, соответствующим данному значению аргумента.

В настоящее время наряду с термином *случайный процесс* используется термин *случайная функция*. Хотя в ряде работ эти термины используются как синонимы, автору настоящей работы ближе точка зрения, в соответствии с которой, термин *случайная функция* трактуется как более общий. В качестве аргумента случайной функции может фигурировать любая переменная, а под случайным процессом будем подразумевать только случайную функцию времени.

6.2. Классификация случайных процессов "по времени" и "по состояниям"

Случайный процесс $X(t)$ называется процессом с дискретным временем, если система, в которой он протекает, может менять свои состояния только в моменты $t_1, t_2, \dots, t_j, \dots$, число которых конечно или счетно.

Случайный процесс $X(t)$ называется процессом с непрерывным временем, если переходы системы из состояния в состояние могут происходить в любой момент t наблюдаемого периода.

Случайный процесс $X(t)$ называется процессом с непрерывными состояниями, если его сечение в любой момент t представляет собой не дискретную, а непрерывную (или смешанную) величину.

Случайный процесс $X(t)$ называется процессом с дискретными состояниями, если в любой момент времени t множество его состояний конечно или счетно; другими словами, если его сечение в любой момент t характеризуется дискретной случайной величиной.

Таким образом, все случайные процессы можно разделить на четыре класса.

1. Процессы с дискретными состояниями и дискретным временем.
2. Процессы с дискретными состояниями и непрерывным временем.
3. Процессы с непрерывными состояниями и дискретным временем.

4. Процессы с непрерывными состояниями и непрерывным временем.

Большинство гидрологических процессов являются процессами с непрерывными состояниями и непрерывным временем. Например, расход воды может изменяться в любой момент времени и принимать любые значения из некоторого интервала, границы которого зависят от размера реки и климатических особенностей региона.

При этом надо учитывать, что на практике расходы воды осредняют за некоторый интервал времени (год, месяц, сутки и т.д.). Вводя шаг дискретности по времени, мы заменяем процесс с непрерывным временем на процесс с дискретным временем. При этом процесс остается непрерывным по состояниям.

Таким образом, производя измерения гидрологических процессов, мы чаще всего используем модель случайного процесса с дискретным временем и непрерывными состояниями.

Однако в гидрометеорологии используются и другие модели. Процесс количественного изменения облачности обычно представляют в виде процесса с дискретным временем и дискретными состояниями, так как наблюдения за облаками производятся в фиксированные сроки, а их количество округляется до целых баллов (по десятибалльной шкале).

6.3. Закон распределения и основные характеристики случайных процессов

Напомним, что исчерпывающей характеристикой любой случайной величины X — дискретной, непрерывной или смешанной является ее функция распределения $F = P\{X < x\}$, т.е. вероятность того, что случайная величина X примет значение, меньшее заданного x .

Рассмотрим теперь случайный процесс $X(t)$. Сечение случайного процесса $x(t)$ при любом фиксированном значении аргумента t представляет собой случайную величину, которая имеет закон распределения

$$F(t, x) = P\{X(t) < x\}. \quad (6.1)$$

Эта функция зависит от двух аргументов: во-первых, от значения t , для которого берется сечение; во-вторых, от значения x , меньше которого должна быть случайная величина $X(t)$. Функция

(6.1) называется одномерным законом распределения случайного процесса $X(t)$.

Для случайных процессов с непрерывными состояниями, у которых каждое сечение представляет собой непрерывную случайную величину, можно пользоваться дифференциальным законом распределения.

Если $F(t, x)$ имеет частную производную по x :

$$\frac{\partial F(t, x)}{\partial x} = f(t, x), \quad (6.2)$$

то она называется одномерной плотностью распределения или одномерным дифференциальным законом распределения случайного процесса $X(t)$.

Очевидно, что одномерный закон распределения не является исчерпывающей характеристикой случайного процесса. Функция (6.1) характеризует свойства любого, но отдельно взятого сечения. Она не дает представления о совместном распределении двух или более сечений.

Для иллюстрации этого факта рассмотрим два случайных процесса с примерно одинаковыми распределениями в каждом сечении (рис. 6.2). Как видно на рисунке, эти процессы имеют совершенно разную вероятностную структуру. Первый случайный процесс (а) имеет более плавный характер, для него характерна достаточно тесная зависимость между сечениями. Для второго (б) эта зависимость затухает довольно быстро с увеличением расстояния между сечениями.

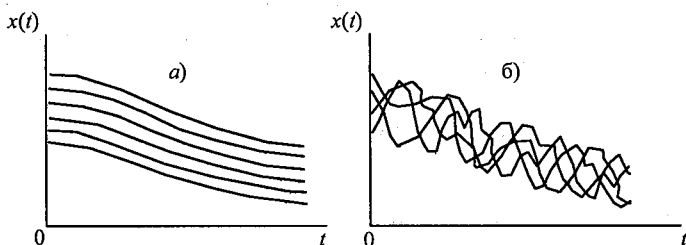


Рис. 6.2. Набор реализаций двух случайных процессов с одинаковыми одномерными законами распределения, но имеющие различную вероятностную структуру.

Более полной характеристикой случайного процесса будет двумерный закон распределения:

$$F(t_1, t_2, x_1, x_2) = P\{X(t_1) < x_1, X(t_2) < x_2\}. \quad (6.3)$$

Функция (6.3) дает представление о совместном распределении двух произвольно взятых сечений — для моментов времени t_1 и t_2 . Но это функция уже не двух, а четырех аргументов и ее использование связано с определенными трудностями как вследствие сложности экспериментального определения двумерных законов распределения, так и вследствие их громоздкости при решении прикладных задач.

В общем случае исчерпывающей характеристикой случайного процесса является n -мерный закон распределения. Однако существует большой класс случайных процессов (марковские процессы), для которых исчерпывающей характеристикой будет двумерный закон.

На практике, как правило, вместо многомерных законов распределения используют основные характеристики случайных процессов, которые описывают случайный процесс не полностью, а частично.

Так же как для случайной величины мы определяли математическое ожидание, дисперсию, начальные и центральные моменты, так и для случайного процесса мы будем определять основные характеристики, только для случайного процесса эти характеристики будут не числами, а функциями.

Математическим ожиданием случайного процесса $X(t)$ называется неслучайная функция $m_x(t)$, которая при любом значении аргумента t равна математическому ожиданию соответствующего сечения случайного процесса:

$$m_x(t) = M[X(t)] = \int_{-\infty}^{+\infty} x f_1(x, t) dx, \quad (6.4)$$

где $f_1(x, t)$ — одномерная плотность распределения случайного процесса $X(t)$.

Таким образом, математическое ожидание случайного процесса представляет собой некоторую "среднюю" функцию, вокруг которой происходит разброс случайного процесса (рис. 6.3).

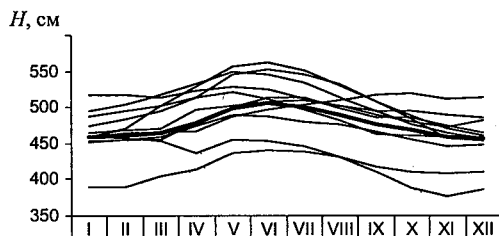


Рис. 6.3. Среднемесячные уровни Ладожского озера за 1987 – 1988 гг. (11 реализаций) и их математическое ожидание (жирная линия).

Если из случайного процесса $X(t)$ вычесть его математическое ожидание, то мы получим центрированный случайный процесс:

$$\overset{\circ}{X}(t) = X(t) - m_x(t). \quad (6.5)$$

Реализации центрированного случайного процесса $\overset{\circ}{X}(t)$ представляют собой отклонения случайного процесса $X(t)$ от его математического ожидания. Эти отклонения имеют как положительные, так и отрицательные значения, а в среднем равны нулю:

$$\mathbf{M} \left[\overset{\circ}{X}(t) \right] = \mathbf{M} [X(t)] - m_x(t) \equiv 0. \quad (6.6)$$

Дисперсией случайного процесса $X(t)$ называется неслучайная функция $D_x(t)$, которая при любом значении аргумента t равна дисперсии соответствующего сечения случайного процесса $X(t)$.

$$D_x(t) = \mathbf{D}[X(t)] = \mathbf{M} \{ [X(t) - m_x(t)]^2 \}. \quad (6.7)$$

Среднеквадратическим отклонением случайного процесса $X(t)$ называется неслучайная функция $\sigma_x(t)$, равная корню квадратному из дисперсии случайного процесса:

$$\sigma_x(t) = \sigma[X(t)] = \sqrt{D_x(t)}, \quad (6.8)$$

размерность функции $\sigma_x(t)$ равна размерности случайного процесса $X(t)$.

Математическое ожидание и дисперсия являются важными, но не исчерпывающими характеристиками случайного процесса, так как определяются только одномерным законом распределения и не позволяют учесть характер взаимосвязи между отдельными сечениями. Для учета этой взаимосвязи используется корреляционная функция случайного процесса.

Корреляционной (или ковариационной) функцией случайного процесса $X(t)$ называется неслучайная функция $K_x(t, t')$, которая при каждой паре значений аргументов t и t' равна ковариации соответствующих сечений $X(t)$ и $X(t')$:

$$K_x(t, t') = \mathbf{M}\{[X(t) - m_x(t)] \times [X(t') - m_x(t')]\}, \quad (6.9)$$

или

$$K_x(t, t') = \mathbf{M}\left[\dot{X}(t) \dot{X}(t')\right] = \mathbf{M}[X(t)X(t')] - m_x(t)m_x(t'). \quad (6.9a)$$

Отметим два важных свойства корреляционной функции.

1. При равенстве аргументов ($t = t'$) корреляционная функция равна дисперсии случайного процесса. Действительно,

$$K_x(t, t) = \mathbf{M}\left[\dot{X}(t) \dot{X}(t)\right] = \mathbf{M}\left\{[\dot{X}(t)]^2\right\} = D_x(t). \quad (6.10)$$

2. Корреляционная функция $K_x(t, t')$ симметрична относительно своих аргументов:

$$K_x(t, t') = K_x(t', t). \quad (6.11)$$

Наряду с корреляционной функцией используется нормированная корреляционная функция.

Нормированной корреляционной функцией $r_x(t, t')$ случайного процесса $X(t)$ называется функция, полученная делением корреляционной функции на произведение среднеквадратических отклонений $\sigma_x(t)$, $\sigma_x(t')$:

$$r_x(t, t') = \frac{K_x(t, t')}{\sigma_x(t)\sigma_x(t')} = \frac{K_x(t, t')}{\sqrt{D_x(t)D_x(t')}}. \quad (6.12)$$

Свойства нормированной корреляционной функции.

1. При равенстве аргументов t и t' нормированная корреляционная функция равна единице:

$$r_x(t, t) = 1. \quad (6.13)$$

2. Нормированная корреляционная функция $r_x(t, t')$ симметрична относительно своих аргументов:

$$r_x(t, t') = r_x(t', t). \quad (6.14)$$

3. Нормированная корреляционная функция по модулю не превышает единицу:

$$|r_x(t, t')| \leq 1. \quad (6.15)$$

До сих пор мы говорили о скалярных случайных процессах, но существуют и векторные случайные процессы, в которые в качестве составляющих могут входить два и более случайных процесса.

Так, например, как векторный случайный процесс можно рассматривать изменение расходов во времени одновременно в нескольких створах.

Для описания векторных случайных процессов необходимо знать математическое ожидание и корреляционную функцию для каждого скалярного процесса, входящего в векторный случайный процесс в качестве составляющей. Кроме того, необходимо иметь характеристику, отражающую взаимосвязь между отдельными составляющими векторного случайного процесса. В качестве такой характеристики используется взаимная корреляционная функция.

Взаимной корреляционной функцией $R_{ij}(t, t')$ двух случайных процессов $X_i(t)$ и $X_j(t')$ называется неслучайная функция двух аргументов t и t' , которая при каждой паре значений t и t' равна ковариации двух сечений случайных процессов $X_i(t)$ и $X_j(t')$:

$$R_{i,j}(t, t') = M \left[\dot{X}_i(t) \dot{X}_j(t') \right]. \quad (6.16)$$

Перечислим основные свойства взаимной корреляционной функции.

1. Взаимная корреляционная функция $R_{ij}(t, t')$ в общем случае не является симметричной относительно своих аргументов:

$$R_{i,j}(t, t') \neq R_{i,j}(t', t). \quad (6.17)$$

2. При одновременной перемене мест индексов и аргументов взаимная корреляционная функция не меняется:

$$R_{i,j}(t, t') = R_{j,i}(t, t'). \quad (6.18)$$

3. При равенстве индексов $i = j$ взаимная корреляционная функция равна корреляционной функции случайного процесса $X_i(t)$:

$$R_{i,i}(t,t') = M \left[\dot{X}_i(t) \dot{X}_i(t') \right] = K_i(t,t'). \quad (6.19)$$

Случайные процессы $X_i(t)$ и $X_j(t)$ называются некоррелированными, если их взаимная корреляционная функция $R_{ij}(t, t')$ равна нулю при любых значениях t и t' , если $i \neq j$.

6.4. Стационарные случайные процессы

К числу наиболее простых для изучения случайных процессов относятся стационарные случайные процессы, у которых все вероятностные характеристики не зависят от времени. В частности, у стационарного случайного процесса математическое ожидание и дисперсия из функций аргумента t превращаются в константы:

$$M[X(t)] = \int_{-\infty}^{+\infty} xf(t,x)dx = \int_{-\infty}^{+\infty} xf(x)dx = m_x = \text{const}, \quad (6.20)$$

$$D[X(t)] = \int_{-\infty}^{+\infty} (x - m_x)^2 f(t,x)dx = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)dx = D_x = \text{const}. \quad (6.21)$$

Это означает, что все сечения стационарного случайного процесса имеют одинаковое математическое ожидание и дисперсию (рис. 6.4).

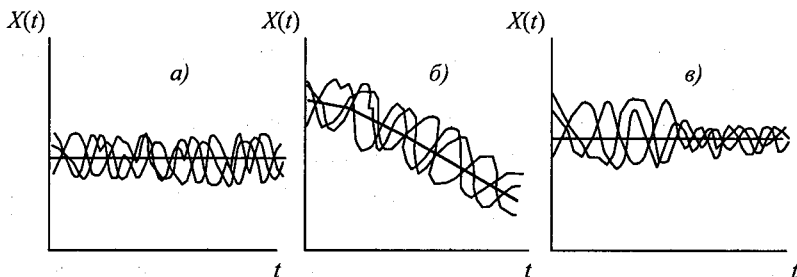


Рис. 6.4. Реализации стационарного (а) и нестационарных (б, в) случайных процессов: б – нестационарность по математическому ожиданию; в – нестационарность по дисперсии.

Различают стационарные случайные процессы в узком смысле и в широком смысле.

Случайный процесс называется *стационарным в узком смысле*, если его n -мерная плотность распределения не изменится при сдвиге всех его временных аргументов на одинаковую произвольную величину θ .

Поясним это на примере двумерной плотности распределения. Если стационарный случайный процесс $X(t)$ имеет двумерную плотность $f(t_1, t_2, x_1, x_2)$, то она не изменится, если заменить t_1 на $t_1 + \theta$, а t_2 на $t_2 + \theta$.

Это означает, что плотность распределения стационарного случайного процесса не зависит от того, в какие моменты t_1 и t_2 рассматриваются сечения, а зависит от расстояния между этими сечениями $t_1 - t_2 = \tau$.

Случайный процесс называется *стационарным в широком смысле*, если его математическое ожидание постоянно ($m_x = \text{const}$), а корреляционная функция – есть функция сдвига между аргументами: $K_x(t_1, t_2) = k_x(\tau)$.

Если стационарный случайный процесс является стационарным в узком смысле, то он является стационарным и в широком смысле (обратное утверждение не всегда верно).

Корреляционная функция стационарного случайного процесса обладает следующими свойствами.

1. Она является четной функцией своего аргумента:

$$k_x(\tau) = k_x(-\tau). \quad (6.22)$$

2. Значение корреляционной функции стационарного случайного процесса при нулевом сдвиге τ равно дисперсии случайного процесса:

$$D_x = K_x(t, t) = k_x(t - t) = k_x(0). \quad (6.23)$$

$$3. \quad |k_x(\tau)| \leq k_x(0). \quad (6.24)$$

Помимо корреляционной функции используется нормированная корреляционная функция стационарного случайного процесса, которую также называю *автокорреляционной функцией*:

$$r_x(\tau) = \frac{k_x(\tau)}{D_x} = \frac{k_x(\tau)}{k_x(0)}. \quad (6.25)$$

Она обладает практически теми же свойствами, что и корреляционная функция, у которой изменен масштаб по оси ординат.

$$1. \quad r_x(\tau) = r_x(-\tau). \quad (6.26)$$

$$2. \quad r_x(0) = 1. \quad (6.27)$$

$$3. \quad |r_x(\tau)| \leq 1. \quad (6.28)$$

6.5. Эргодические случайные процессы

Стационарный случайный процесс может обладать или не обладать эргодическим свойством. Эргодическое свойство состоит в том, что любая достаточно продолжительная реализация стационарного эргодического процесса является как бы "полномочным представителем" всей совокупности реализаций случайного процесса, и по ней можно составить представление о случайном процессе в целом.

Поясним это на примере конкретной числовой характеристики – математического ожидания. Для стационарного процесса математическое ожидание – это постоянная величина, и, следовательно, его можно оценить по любому сечению, имея достаточно много реализаций. Для стационарного эргодического процесса то же самое можно сделать по одной реализации, если она имеет достаточно большую продолжительность. Например, для стационарного эргодического процесса среднее значение по любому сечению можно заменить на среднее значение по одной достаточно продолжительной реализации.

Достаточным условием эргодичности стационарного случайного процесса является следующее:

$$\lim_{\tau \rightarrow \infty} k_x(\tau) = 0, \quad (6.29)$$

т.е. требуется, чтобы при увеличении сдвига между сечениями корреляционная функция затухала и в пределе при $\tau = \infty$ равнялась нулю.

Однако стационарный случайный процесс может быть и неэргодическим. Примеры эргодического и неэргодического случайных процессов представлены на рис. 6.5.

Неэргодичность случайного процесса может быть вызвана, например тем, что в качестве слагаемого случайного процесса фигурирует случайная величина.

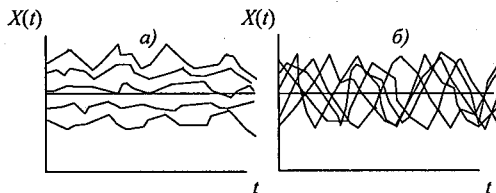


Рис. 6.5. Реализации неэргодического (а) и эргодического (б) стационарных случайных процессов.

Рассмотрим случайный процесс $U(t)$:

$$U(t) = X(t) + v, \quad (6.30)$$

где $X(t)$ – эргодический случайный процесс, а v – случайная величина с математическим ожиданием и дисперсией m_v и D_v .

Из теории известно, что в этом случае

$$m_U(t) = m_x + m_v, \quad (6.31)$$

$$k_U(\tau) = k_x(\tau) + D_x. \quad (6.32)$$

Отсюда получаем:

$$\lim_{\tau \rightarrow \infty} k_U(\tau) = \lim_{\tau \rightarrow \infty} [k_x(\tau) + D_v] = \lim_{\tau \rightarrow \infty} k_x(\tau) + \lim_{\tau \rightarrow \infty} D_v = D_v, \quad (6.33)$$

следовательно, не выполняется условие (6.29) и случайный процесс $U(t)$ является неэргодическим.

На практике, имея всего одну реализацию случайного процесса (например, сорокалетний ряд среднегодовых расходов воды), мы вынуждены принимать гипотезу о стационарности и эргодичности. Так как только в этом случае оценки статистических характеристик, полученные по данной реализации, мы можем считать характеристиками всего случайного процесса.

При этом следует иметь в виду, что выбор вероятностной модели не должен быть формальным. Необходимо руководствоваться как общими соображениями, так и эмпирическими данными, для того чтобы модель была непротиворечивой и адекватно описывала исследуемый процесс.

6.6. Элементарные случайные процессы

Как уже отмечалось, случайный процесс в результате опыта превращается на некотором интервале от $t = t_0$ до $t = t_0 + \theta$ в неслучайную функцию времени, которая называется реализацией случайного процесса на этом интервале. В общем случае (даже если характеристики случайного процесса известны) мы не можем точно предсказать, как будет развиваться случайный процесс в будущем при $t > (t_0 + \theta)$, если опыт еще не завершен. Однако существует класс простейших случайных процессов, для которых можно точно сказать, как будет выглядеть данная реализация при дальнейшем продолжении опыта. Такие случайные процессы называются элементарными случайными процессами.

Элементарным случайным процессом (э. с. п.) будем называть такую функцию аргумента t , для которой зависимость от t представлена обычной неслучайной функцией, в которую в качестве аргумента входит одна или несколько обычных случайных величин.

Таким образом в каждой реализации элементарного случайного процесса случайная величина является константой, или другими словами – каждая случайная константа порождает свою реализацию.

В качестве примера рассмотрим э. с. п. $Y(t)$:

$$Y(t) = Xe^{-t}, \quad (t > 0), \quad (6.34)$$

где X – положительная непрерывная случайная величина. Семейство реализаций э. с. п. $Y(t)$ показано на рис. 6.6.

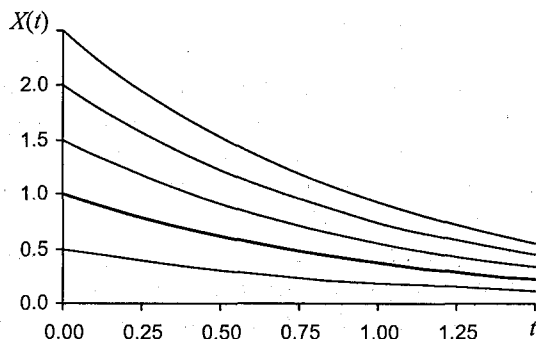


Рис. 6.6. Семейство реализаций элементарного случайного процесса $Y(t) = Xe^{-t}$.

Каждая их реализаций представляет собой кривую с ординатами, пропорциональными кривой e^{-t} (жирная линия); отдельные реализации (тонкие линии) различаются между собой масштабом по оси ординат.

Заметим, что хотя данный процесс довольно прост по структуре, он не является стационарным.

С некоторой натяжкой модель данного э. с. п. можно использовать для описания ветви спада половодья. Если предположить, что в створе наблюдений ветвь спада половодья является устойчивой и существенно не искажается осадками в течение, допустим, двух месяцев после прохождения максимума, то ее можно аппроксимировать выражением

$$Q(t) = Q_n e^{-at}, \quad (6.35)$$

где a – районный параметр ($a > 0$); Q_n – расход воды в начальный момент времени $t = t_0$.

В качестве t_0 следует принять постоянную дату, более позднюю, чем дата самого позднего из всех наблюдаемых максимумов половодья. Тогда выражение (6.35) можно трактовать как элементарный случайный процесс, где a – неслучайная величина, Q_n – случайная величина из интервала $(0, \infty)$, представляющая собой расход воды в i -том году на дату t_0 .

В качестве функции распределения Q_n можно использовать, например, кривую Крицкого – Менкеля, а ее параметры оценить по эмпирическим данным.

6.7. Каноническое разложение случайных процессов

Любой случайный процесс $X(t)$ может быть представлен в виде его разложения, т.е. в виде суммы элементарных процессов:

$$X(t) = \varphi_0(t) + \sum_{k=0}^{\infty} V_k \varphi_k(t), \quad (6.36)$$

где V_k – случайные величины; $\varphi_k(t)$ – неслучайные функции.

Такое представление широко используется на практике в силу того, что оно дает возможность проводить довольно просто различные преобразования случайных процессов.

Из теории случайных процессов известно, что разложение (6.36) можно построить множеством способов. В частности, *кано-*

ническим разложением случайного процесса $X(t)$ называется выражение вида:

$$X(t) = m_x(t) + \sum_{k=1}^{\infty} V_k \varphi_k(t). \quad (6.37)$$

В этом выражении $m_x(t) = \mathbf{M}[X(t)]$ представляет собой математическое ожидание случайного процесса $X(t)$; $V_1, \dots, V_k$ – некоррелированные и центрированные случайные величины с дисперсиями $D_1, \dots, D_k$; $\varphi_1(t), \dots, \varphi_k(t)$ – неслучайные функции аргумента t .

Случайные величины $V_1, \dots, V_k$ называются коэффициентами канонического разложения, а неслучайные функции $\varphi_1(t), \dots, \varphi_k(t)$ – координатными функциями канонического разложения.

Выпишем основные характеристики случайного процесса, заданного своим каноническим разложением:

$$\mathbf{M}[X(t)] = m_x(t); \quad (6.38)$$

$$K_x(t, t') = \sum_{k=1}^{\infty} \varphi_k(t) \varphi_k(t') D_k. \quad (6.39)$$

Выражение (6.39) называется каноническим разложением корреляционной функции случайного процесса $X(t)$. Из выражения (6.39) получаем каноническое разложение дисперсии случайного процесса $X(t)$:

$$D_k(t) = K_x(t, t) = \sum_{k=1}^{\infty} \varphi_k^2(t) D_k. \quad (6.40)$$

6.8. Спектральное разложение стационарного случайного процесса

Стационарный случайный процесс может быть представлен каноническим разложением вида:

$$X(t) = m_x(t) + \sum_{k=0}^{\infty} (V_k \cos \omega_k t + U_k \sin \omega_k t), \quad (6.41)$$

где V_k и U_k – некоррелированные и центрированные случайные величины с дисперсиями $\mathbf{D}[V_k] = \mathbf{D}[U_k] = D_k$; ω – неслучайная величина (частота).

В этом случае каноническое разложение корреляционной функции определяется выражением:

$$K_x(t, t') = \sum_{k=0}^{\infty} D_k \cos \omega_k(t - t') = \sum_{k=0}^{\infty} D_k \cos \omega_k(\tau) = k_x(\tau), \quad (6.42)$$

Каноническое разложение (6.41) называется *спектральным разложением* стационарного случайного процесса. Спектральное разложение (6.41) может быть также представлено в виде:

$$X(t) = m_x + \sum_{k=0}^{\infty} Z_k \cos(\omega_k t - \theta_k), \quad (6.43)$$

где θ_k — фаза гармонического колебания элементарного стационарного случайного процесса, являющаяся случайной величиной равномерно распределенной в интервале $(\theta, 2\pi)$; Z_k — амплитуда гармонического колебания элементарного стационарного случайного процесса — тоже случайная величина.

Случайные величины θ_k и Z_k зависимы. Для них имеют место соотношения:

$$Z_k \cos \theta_k = V_k; \quad Z_k \sin \theta_k = U_k. \quad (6.44)$$

Таким образом, стационарный случайный процесс может быть представлен в виде суммы гармонических колебаний со случайными амплитудами Z_k и случайными фазами θ_k на различных неслучайных частотах ω_k .

Так как корреляционная функция стационарного случайного процесса $X(t)$ является четной функцией своего аргумента, т.е. $k_x(\tau) = k_x(-\tau)$, то ее на интервале $(-T, T)$ можно разложить в ряд Фурье по четным (косинусам) гармоникам:

$$k_x(\tau) = \sum_{k=0}^{\infty} D_k \cos \omega_k \tau, \quad (6.45)$$

где ω_k и D_k определяются выражениями:

$$\omega_k = k\omega_1, \quad \omega_1 = 2\pi/(2T) = \pi/T; \quad (6.46)$$

$$D_0 = \frac{1}{2T} \int_{-T}^{+T} k_x(\tau) d\tau, \quad D_k = \frac{1}{T} \int_{-T}^{+T} k_x(\tau) \cos \omega_k \tau. \quad (6.47)$$

Подставляя в выражение (6.45) $\tau = 0$, получим каноническое разложение для дисперсии стационарного случайного процесса:

$$D_x = k(0) = \sum_{k=0}^{\infty} D_x \cos(\omega_k \cdot 0) = \sum_{k=0}^{\infty} D_k. \quad (6.48)$$

Таким образом, дисперсия стационарного случайного процесса, представленная своим каноническим разложением (6.41), равна сумме дисперсий всех гармоник его спектрального разложения.

Используя формулы (6.45) – (6.47) при $k = 1, 2, 3, \dots$, можно построить график изменения дисперсий D_k по частотам ω_k (рис. 6.7). Зависимость $D_k = f(\omega_k)$ в этом случае называется дискретным спектром дисперсий или просто – дискретным спектром тационарного случайного процесса.

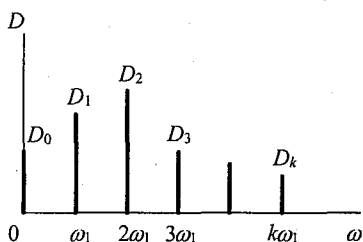


Рис. 6.7. Дискретный спектр стационарного случайного процесса.

Если увеличивать интервал разложения $(-T, T)$, то интервал между частотами будет сокращаться, так как

$$\Delta\omega = \omega_1 = \pi/T. \quad (6.49)$$

В пределе при $T \rightarrow \infty$ (т.е. при $\Delta\omega \rightarrow 0$) мы перейдем к непрерывному спектру.

В качестве характеристики распределения дисперсии по частотам непрерывного спектра используется спектральная плотность — $S_x(\omega)$, которую можно рассматривать, как предел отношения дисперсии, приходящейся на каждую частоту к $\Delta\omega$:

$$S(\omega) = \lim_{\Delta\omega \rightarrow 0} \frac{D_k}{\Delta\omega}. \quad (6.50)$$

В этом случае выражение (6.45) примет вид:

$$k_x(\tau) = \int_0^{\infty} S_x(\tau) \cos \omega \tau d\tau. \quad (6.51)$$

Таким образом, корреляционная функция и спектральная плотность стационарного случайного процесса связаны косинус-пре-

образованием Фурье. Следовательно, спектральная плотность стационарного случайного процесса может быть выражена через корреляционную функцию по формуле

$$S_x(\omega) = \frac{2}{\pi} \int_0^{\infty} k_x(\tau) \cos \omega \tau d\tau. \quad (6.52)$$

Спектральная плотность обладает следующими свойствами.

1. Она является неотрицательной функцией частоты:

$$S(\omega) \geq 0. \quad (6.53)$$

2. Интеграл от спектральной плотности в пределах от 0 до ∞ равен дисперсии стационарного случайного процесса:

$$D_x = \int_0^{\infty} S_x(\omega) d\omega. \quad (6.54)$$

По аналогии с нормированной корреляционной функцией вводится нормированная спектральная плотность:

$$s_x(\omega) = S_x(\omega)/D_x. \quad (6.55)$$

Нормированная корреляционная функция и нормированная спектральная плотность также связаны между собой преобразованием Фурье:

$$r_x(\tau) = \int_0^{\infty} s_x(\omega) \cos \omega \tau d\omega; \quad (6.56)$$

$$s_x(\omega) = \frac{2}{\pi} \int_0^{\infty} r_x(\tau) \cos \omega \tau d\tau. \quad (6.57)$$

Интеграл от нормированной спектральной плотности в пределах от 0 до ∞ равен единице:

$$\int_0^{\infty} s_x(\omega) d\omega = 1. \quad (6.58)$$

Таким образом, если на графике нормированной спектральной плотности (рис. 6.8) рассматривать полосу частот от ω_1 до ω_2 , то площадь подграфика, заключенная в интервале от ω_1 до ω_2 , будет

равна доле дисперсии случайного процесса, приходящейся на эти частоты. При этом общая площадь подграфика будет равна единице.

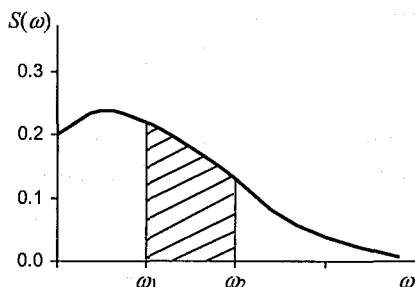


Рис. 6.8. Нормированная спектральная плотность стационарного случайного процесса; заштрихованная площадь численно равна доле дисперсии, приходящейся на полосу частот от ω_1 до ω_2 .

Следует отметить, что в специальной литературе встречаются различные определения спектральной плотности, отличающиеся постоянным множителем (табл. 6.1). Достоинством первого определения является то, что для действительных случайных процессов при его использовании выполняются условия (6.54) и (6.58).

Таблица 6.1

Различные определения спектральной плотности

№	Спектральная плотность	Корреляционная функция
1	$S_x(\omega) = \frac{2}{\pi} \int_0^{\infty} k_x(\tau) \cos \omega \tau \, d\tau$	$k_x(\tau) = \int_0^{\infty} S_x(\omega) \cos \omega \tau \, d\omega$
2	$S_x(\omega) = \frac{1}{\pi} \int_0^{\infty} k_x(\tau) \cos \omega \tau \, d\tau$	$k_x(\tau) = 2 \int_0^{\infty} S_x(\omega) \cos \omega \tau \, d\omega$
3	$S_x(\omega) = 4 \int_0^{\infty} k_x(\tau) \cos \omega \tau \, d\tau$	$k_x(\tau) = \frac{1}{2\pi} \int_0^{\infty} S_x(\omega) \cos \omega \tau \, d\omega$
4	$S_x(\omega) = 2 \int_0^{\infty} k_x(\tau) \cos \omega \tau \, d\tau$	$k_x(\tau) = \frac{1}{\pi} \int_0^{\infty} S_x(\omega) \cos \omega \tau \, d\omega$

При аналитических расчетах нередко используется спектральная плотность в комплексной форме $\tilde{S}(\omega)$. В этом случае рассматриваются как положительные, так и отрицательные частоты, а $\tilde{S}(\omega)$ и $S(\omega)$ связаны соотношением:

$$\begin{cases} \tilde{S}(\omega) = S_x(\omega)/2 & \text{при } \omega \geq 0, \\ \tilde{S}(\omega) = S_x(-\omega)/2 & \text{при } \omega < 0. \end{cases} \quad (6.59)$$

Спектральная плотность в комплексной форме обладает следующими свойствами.

1. $\tilde{S}(\omega)$ является неотрицательной функцией частоты, т.е.

$$\tilde{S}_x(\omega) \geq 0 \quad \text{при } -\infty < \omega < +\infty. \quad (6.60)$$

2. Интеграл от $\tilde{S}(\omega)$ в бесконечных пределах равен дисперсии случайного процесса:

$$\int_{-\infty}^{+\infty} \tilde{S}_x(\omega) d\omega = D_x. \quad (6.61)$$

3. $\tilde{S}(\omega)$ – четная функция аргумента ω :

$$\tilde{S}_x(\omega) = \tilde{S}_x(-\omega). \quad (6.62)$$

Пара преобразований Фурье для корреляционной функции и спектральной плотности в комплексной форме имеет вид:

$$k_x(\tau) = \int_{-\infty}^{+\infty} \tilde{S}_x(\omega) e^{i\omega\tau} d\omega; \quad (6.63)$$

$$\tilde{S}_x(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} k_x(\tau) e^{-i\omega\tau} d\tau. \quad (6.64)$$

Следует отметить, что если спектральную плотность для действительных случайных процессов рассматривать как часть спектральной плотности $\tilde{S}_x(\omega)$, лежащую в области положительных частот, то мы получим определение $S(\omega)$, приведенное в строке 2 табл. 6.1. Однако в этом случае интеграл от нормированной спектральной плотности будет равен 0,5, а не единице, что с формальной точки зрения не очень хорошо.

6.9. Случайные процессы с независимыми сечениями

В гидрологической практике гипотеза о том, что ряд соответствует модели случайной величины, принимается, если отсутствует значимая корреляция между членами этого ряда при любом сдвиге τ . Однако в этом случае можно применить и модель случайного процесса с независимыми сечениями.

Случайный процесс называется *случайным процессом с независимыми сечениями*, если для любых значений t и t' ($t \neq t'$) корреляционная функция $K_x(t, t')$ равна нулю.

Этот процесс имеет следующие характеристики:

$$m_x(t) = m_x, \quad (6.65)$$

$$D_x(t) = D_x, \quad (6.66)$$

$$K_x(t, t') = k_x(\tau) = \begin{cases} D_x & \text{при } \tau = 0, \\ 0 & \text{при } \tau \neq 0. \end{cases} \quad (6.67)$$

Таким образом, этот процесс является стационарным и обладает эргодическим свойством. Для таких процессов характеристики одномерного закона распределения можно оценить как по любому сечению, так и по любой (достаточно продолжительной) реализации. У таких процессов отсутствует корреляция как между сечениями, так и между членами внутри любой реализации.

Принимая эту модель, мы рассматриваем ряд гидрологических величин не как выборку из генеральной совокупности, а как одну реализацию случайного процесса.

Ситуация, когда приходится исследовать всего одну реализацию случайного процесса, встречается довольно часто. В этом случае реализацию иногда называют временным рядом, а при независимости сечений — временным рядом с независимыми значениями.

Очевидно, что если речь идет об описании статистической структуры одного ряда, эта модель мало отличается от модели обычной случайной величины. Поэтому в гидрологической литературе модель ряда в виде случайной величины и в виде случайного процесса с независимыми сечениями иногда рассматриваются как тождественные. При этом следует понимать, что для случайного процесса с независимыми сечениями корректно говорить о таких

характеристиках, как корреляционная функция и спектр, а для случайной величины – нет.

Корреляционная функция и спектральная плотность случайного процесса с независимыми сечениями имеют очень специфический вид, и являются мало информативными. В частности, спектральная плотность такого случайного процесса является постоянной для любого конечного интервала разложения:

$$S_x(\omega) = \text{const} = c/2\pi. \quad (6.68)$$

В то же время эмпирическая оценка спектральной плотности по любой произвольно взятой реализации будет представлять собой беспорядочные всплески, случайным образом разбросанные по всему спектру частот со средним (по множеству реализаций) значением $c/2\pi$.

Случайный процесс с независимыми сечениями иногда называют "*белым шумом*" по аналогии с белым светом, у которого спектральный состав примерно однороден. При этом величину c называют интенсивностью белого шума.

6.10. Марковские процессы и цепи Маркова

Случайный процесс, протекающий в некоторой системе с дискретными состояниями, называется *марковским*, если для любого момента времени t_0 вероятность каждого из состояний системы в будущем (при $t > t_0$) зависит только от ее состояния в настоящем (при $t = t_0$) и не зависит от того, когда и как она пришла в это состояние, т.е. не зависит от ее состояния в прошлом (при $t < t_0$).

Марковский процесс с дискретными состояниями и дискретным временем называется *марковской цепью* или *простой марковской цепью*.

Марковский случайный процесс исчерпывающим образом описывается двумерным законом распределения.

Если Марковский случайный процесс является стационарным и обладает эргодическим свойством, его характеристики можно оценить по одной достаточно продолжительной реализации.

В настоящее время при описании гидрологических процессов широко используется гипотеза об их "марковости" совместно с гипотезой о стационарности и эргодичности.

Вводя в состав параметров, характеризующих настоящее состояние системы, те параметры из прошлого, от которых зависит будущее, можно, как говорится, "марковизовать" многие немарковские случайные процессы. В этом случае используют термин "сложная цепь Маркова".

Цепь, в которой условные вероятности состояний в будущем зависят от ее состояния на нескольких предыдущих шагах, называют *сложной цепью Маркова*.

6.11. Нормальные (Гауссовские) случайные процессы

Нормальным (гауссовским) случайным процессом $X(t)$ называется случайный процесс, у которого во всех сечениях случайная величина $X(t_i)$ имеет нормальное распределение.

Нормальный случайный процесс обладает рядом замечательных свойств.

– Двумерный закон распределения нормального случайного процесса $X(t)$ является его исчерпывающей характеристикой, так как все характеристики его n -мерного закона распределения зависят от двух функций – математического ожидания $m_x(t)$ и корреляционной функции $K_x(t, t')$.

– Если для нормального случайного процесса выполняется условие

$$\begin{cases} m_x(t) = \text{const}, \\ K_x(t, t') = K_x(t - t'), \end{cases} \quad (6.69)$$

т.е. если он является стационарным в широком смысле, то он будет стационарным и в узком смысле; кроме того, он будет эргодическим.

6.12. Периодически нестационарные случайные процессы

При описании вероятностной структуры гидрометеорологических рядов с шагом дискретности менее года (месяц, декада, сутки) необходимо учитывать ритмику колебаний, связанную с годовой цикличностью. А при шаге дискретности менее суток следует учитывать суточный ход гидрометеорологических элементов. В качестве математической модели в этом случае можно использовать модель *периодически нестационарного случайного процесса* (ПНСП).

Случайный процесс называют периодически нестационарным, если его вероятностные характеристики инвариантны относительно сдвигов на положительное число T . Например, при шаге дискретности один месяц инвариантность должна сохраняться при сдвигах 12, 24, 36 и т.д.

Процесс будет периодически нестационарным в узком смысле, если инвариантны его конечные распределения

$$F(x_1, \dots, x_n; t_1 + T, \dots, t_n + T) = F(x_1, \dots, x_n; t_1, \dots, t_n). \quad (6.70)$$

Процесс будет периодически нестационарным в широком смысле, если инвариантны математическое ожидание

$$m(t + T) = m(t) \quad (6.71)$$

и корреляционная функция

$$K(t_1 + T, t_2 + T) = K(t_1, t_2). \quad (6.72)$$

Случайный процесс, периодически нестационарный в широком смысле, называют *периодически коррелированным случайным процессом* (ПКСП).

Вводя обозначение $\tau = t_2 - t_1$, свойство (6.72) для ПКСП можно записать в виде

$$K(t + T, \tau) = K(t, \tau) \quad (6.73)$$

Для ПНСП корреляционная функция не симметрична относительно своего аргумента τ :

$$K(t, \tau) \neq K(t, -\tau). \quad (6.74)$$

но обладает свойством

$$K(t - \tau, \tau) = K(t, -\tau). \quad (6.75)$$

Если рассматривать гидрологический процесс как ПКСП с шагом дискретности один месяц, то математическое ожидание $m(t)$ представляет собой регулярную (периодическую) составляющую случайного процесса $X(t)$ и позволяет найти средний повторяющийся образ сезонного хода (рис. 6.9).

Дисперсия $D(t)$ характеризует разброс относительно регулярной составляющей. Функция $K(t, \tau)$ характеризует взаимосвязь значений гидрологических элементов в различные месяцы года и в аналогичные месяцы разных лет.

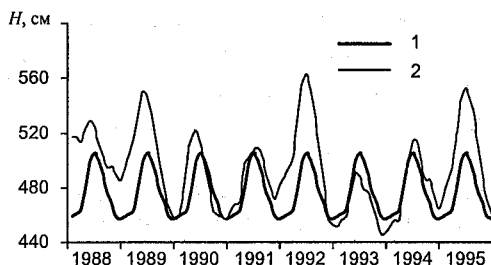


Рис. 6.9. Средние многолетние месячные уровни Ладожского озера (1) и среднемесячные уровни за 1988 – 1995 гг. (2).

Так же как и для стационарных процессов, для периодически нестационарных процессов можно определить свойство эргодичности. Периодически нестационарные процессы обладают свойством эргодичности, если их вероятностные характеристики, вычисленные по одной реализации и их ансамблю, совпадают.

6.13. Параметрические модели временных рядов

Методы описания случайных процессов с помощью корреляционных функций и спектров называют *непараметрическими*. Однако наряду с ними при исследовании гидрологических процессов применяются и *параметрические* методы.

Для описания стационарных случайных процессов в настоящее время довольно часто используются модели авторегрессии и их комбинация.

6.13.1. Модель авторегрессии (АР)

В этой модели текущие значения случайного процесса выражаются как конечная линейная комбинация предыдущих его значений и белого шума:

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + \dots + \phi_p \dot{X}_{t-p} + a_t, \quad (6.76)$$

где $\dot{X}_t$ — центрированный случайный процесс, $\dot{X}_t = X_t - m_x$; a_t — белый шум с нулевым математическим ожиданием и среднеквадратическим отклонением σ_a .

Модель (6.76) содержит $p + 2$ параметров: $\phi_1, \phi_2 \dots \phi_p, m_x, \sigma_a$, где m_x — математическое ожидание случайного процесса $X(t)$; $\phi_1, \phi_2 \dots \phi_p$ — коэффициенты модели (константы).

Модель (6.76) называют моделью авторегрессии p -того порядка и обозначают $AP(p)$.

При $p = 1$ получаем:

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + a_t. \quad (6.77)$$

Эту модель называют моделью авторегрессии первого порядка $AP(1)$ или моделью марковского процесса. Для процесса $AP(1)$ коэффициент ϕ_1 и ординаты автокорреляционной функции связаны соотношением

$$r_k = \phi_1 r_{k-1}, \quad \text{где } k > 0. \quad (6.78)$$

Учитывая, что $r_0 = 1$, получаем:

$$r_k = \phi_1^k. \quad (6.79)$$

Таким образом, для $AP(1)$ автокорреляционная функция полностью определяется своей первой ординатой. При этом $\phi_1 = r_1$.

При $p = 2$ получаем:

$$\dot{X}_t = \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + a_t. \quad (6.80)$$

Эту модель называют моделью авторегрессии второго порядка $AP(2)$ или процессом Юла.

Для процесса Юла коэффициенты ϕ_1 и ϕ_2 связаны с ординатами автокорреляционной функции следующими соотношениями:

$$\phi_1 = \frac{r_1(1-r_2)}{1-r_1^2}; \quad (6.81)$$

$$\phi_2 = \frac{r_2 - r_1^2}{1-r_1^2}. \quad (6.82)$$

$$r_1 = \frac{\phi_1}{1-\phi_2}; \quad (6.83)$$

$$r_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_2}. \quad (6.84)$$

И вообще для любого значения $k > 0$ процесса $AP(2)$:

$$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2}. \quad (6.85)$$

Таким образом, для $AP(2)$ автокорреляционная функция полностью определяется первыми двумя ординатами.

При моделировании гидрологических рядов по модели авторегрессии необходимо также знать среднеквадратическое отклонение шума. В общем случае СКО σ_a определяется формулой

$$\sigma_a = \sigma_x \sqrt{(1 - \phi_1 r_1 - \dots - \phi_k r_k)}. \quad (6.86)$$

Спектральная плотность процессов $AP(1)$ и $AP(2)$ может быть выражена через параметры модели.

Для процесса $AP(1)$ нормированная спектральная плотность определяется выражением

$$S(\omega) = \frac{1 - r_1^2}{\pi(1 - 2r_1 \cos \omega + r_1^2)}, \quad (6.87)$$

где $(0 \leq \omega \leq \pi)$; π — число «пи». При этом период и частота связаны соотношением $T = 2\pi/\omega$.

Для процесса $AP(2)$ нормированная спектральная плотность может быть вычислена по формуле

$$S(\omega) = \frac{1 - \phi_1 r_1 - \phi_2 r_2}{\pi(1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2) \cos \omega - 2\phi_2 \cos 2\omega)}. \quad (6.88)$$

6.13.1. Модель скользящего среднего (СС)

Модель скользящего среднего может быть получена из так называемой общей линейной модели (ОЛМ), если предположить, что ОЛМ содержит конечное число членов. В этой модели текущие значения процесса $X(t)$ выражаются через предыдущие значения белого шума $a_{t-1}, a_{t-2}, \dots, a_{t-q}$:

$$\dot{X}(t) = a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots - \psi_q a_{t-q}. \quad (6.89)$$

Модель содержит $q + 2$ параметров: $\psi_1, \psi_2, \dots, \psi_q, m_x, \sigma_a$, где $\dot{X}_t$ – центрированный случайный процесс, $\dot{X}_t = X_t - m_x$; a_t – белый шум с нулевым математическим ожиданием и среднеквадратическим отклонением σ_a ; $\psi_1, \psi_2, \dots, \psi_q$ – коэффициенты модели (константы).

Выражение (6.89) называют моделью скользящего среднего q -го порядка и обозначают $CC(q)$.

Автокорреляционная функция для процесса $CC(q)$ при $k \leq q$ определяется формулой

$$r_k = \frac{-\psi_k + \sum_{i=1}^{q-k} \psi_i \psi_{i+k}}{1 + \psi_1^2 + \dots + \psi_q^2}. \quad (6.90)$$

Среднеквадратическое отклонение белого шума равно:

$$\sigma_a = \sigma_x \frac{1}{\sqrt{1 + \psi_1^2 + \dots + \psi_q^2}}. \quad (6.91)$$

Если $r_1 \dots r_q$ известны, то из выражения (6.90) можно найти параметры модели $\psi_1 \dots \psi_q$. Однако следует иметь в виду, что при замене теоретических значений $r_1 \dots r_q$ на их эмпирические оценки мы получим не идеальные оценки параметров модели. Эти предварительные оценки параметров $\psi_1 \dots \psi_q$ можно использовать в качестве первого приближения для получения более эффективных оценок.

6.13.3. Смешанные модели (АРСС)

Иногда целесообразно объединить модели АР и СС. В этом случае мы получаем смешанную модель АРСС(p, q), где p – порядок авторегрессии, q – порядок скользящего среднего. Выражение для АРСС(p, q) имеет вид:

$$\begin{aligned} \dot{X}_t = & \phi_1 \dot{X}_{t-1} + \phi_2 \dot{X}_{t-2} + \dots + \phi_p \dot{X}_{t-p} + \\ & + a_t - \psi_1 a_{t-1} - \psi_2 a_{t-2} - \dots - \psi_q a_{t-q}. \end{aligned} \quad (6.92)$$

Такая модель может оказаться подходящей, например, когда наблюдаемый временной ряд является суммой двух или более неза-

висимых составляющих, каждая из которых описывается либо моделью АР, либо моделью СС, но которые непосредственно не изменяются.

Для описания гидрологических процессов в большинстве случаев достаточно использовать АРСС(p, q) при $p = 1$ и $q = 1$. В этом случае модель имеет вид:

$$\overset{\circ}{X}_t = \phi \overset{\circ}{X}_{t-1} + a_t - \psi a_{t-1}. \quad (6.93)$$

Первую ординату автокорреляционной функции процесса АРСС(1, 1) можно рассчитать по формуле

$$r_1 = \frac{(1 - \phi\psi)(\phi - \psi)}{1 - 2\phi\psi + \psi^2}. \quad (6.94)$$

При $k > 1$ автокорреляционная функция процесса АРСС(1, 1) определяется выражением

$$r_k = \phi^{k-1} r_1. \quad (6.95)$$

Дисперсия процесса X_t и дисперсия белого шума для АРСС(1, 1) связаны соотношением

$$\sigma_x^2 = \frac{1 - 2\phi\psi + \psi^2}{1 - \phi^2} \sigma_a^2. \quad (6.96)$$

6.13.4. Модель авторегрессии и проинтегрированного скользящего среднего (АРПСС)

Рассмотренные выше модели АР, СС и АРСС относятся к классу стационарных моделей, которые описывают процессы, для которых математическое ожидание и дисперсия являются константами. Однако в природе существует неограниченное число различных проявлений нестационарности.

Исследования Дж. Бокса и Г. Дженкина [2] показали, что во многих случаях от нестационарности можно избавиться, заменяя исходный ряд на ряд разностей:

$$Y_t = X_t - X_{t-1}. \quad (6.97)$$

Если от нестационарности избавиться не удалось можно взять разность повторно:

$$Z_t = Y_t - Y_{t-1}. \quad (6.98)$$

После проведенных преобразований к исходному ряду можно применить модель АРСС.

Полученную на основе таких рассуждений модель называют моделью Бокса-Дженкинса или моделью авторегрессии и проинтегрированного скользящего среднего АРПСС (p, d, q). В этой модели: p – порядок авторегрессии; d – порядок разности; q – порядок скользящего среднего.

Таким образом, модель АРПСС является обобщением всех рассмотренных в п. 6.12 моделей.

В настоящее время для оценки параметров моделей АРПСС разработаны специальные алгоритмы и программы. Одной из лучших программ является программа "Statistica" фирмы StatSoft.

6.14. Сглаживание и фильтрация

Если исходный временной ряд содержит некоторые частоты или периоды, которые в данный момент не представляют интереса для исследования, амплитуда этих волн может быть уменьшена с помощью статистической фильтрации. При этом изменяется спектр исходного временного ряда.

Одной из форм статистической фильтрации является сглаживание. В результате сглаживания создается временной ряд, в котором спектральные компоненты с высокой частотой уменьшены. Такой фильтр называется *низкочастотным*, так как сглаживание слабо влияет на волны с низкой частотой, т.е. на длиннопериодные волны.

Существуют и другие типы фильтров. Есть фильтры, которые позволяют отфильтровать низкие частоты, оставив в ряде только волны высокой частоты. Этот тип фильтра называется *высокочастотным*.

Можно отфильтровать как низкие, так и высокие частоты, оставив во временном ряде только средние частоты. Такой фильтр называется *фильтром пропускания полос*. Наконец, можно разработать статистические фильтры, которые будут усиливать во временном ряде волны высокой частоты таким образом, чтобы частично компенсировать эффект предыдущего сглаживания того же ряда. Этот процесс называется «разглаживанием», или «обратным сглаживанием».

Простейшим статистическим фильтром, или, как его называют, фильтрующей функцией, является скользящая средняя с равными весами. Скользящее среднее рассчитывается путем суммирования n последовательных величин временного ряда и делением полученной суммы на n . При расчете этих средних используются данные с обеих сторон сглаживаемых значений ряда, так что величины, использованные для расчетов соседних скользящих средних, значительно перекрываются. Поэтому такой тип скользящих средних называют также перекрывающимися средними. В данном случае все веса фильтра одинаковы и равны $1/n$.

Чтобы сохранить симметрию обычно используют нечетное значение n : 3, 5, 7... и т.д., а сглаженное значение относят к центру интервала осреднения.

Так, например, отфильтрованное значение 14,60 в табл. 6.2 получено как среднее из первых 5 значений: $(16 + 15 + 12 + 13 + 17)/5 = 14,60$.

Таблица 6.2

Сглаживание временного ряда скользящим средним
с равными весами при $n = 5$

Временной ряд	Отфильтрованные величины
16	14,60
15	
12	
13	
17	
20	15,40

В процессе анализа желательно знать, как фильтрация влияет на амплитуду волн временного ряда. Иными словами, как меняется спектр исходного временного ряда в результате фильтрации. Ответ на этот вопрос можно получить, если рассчитать частотную характеристику фильтра, которая представляет собой отношение амплитуды колебания данной частоты после фильтрации к исходной амплитуде до фильтрации. Это отношение меняется с частотой. Например, частотная характеристика сглаживающей функции меняется примерно от единицы (для низких частот) до нуля (для высоких).

Частотная характеристика любой дискретной симметричной сглаживающей или фильтрующей функции выражается следующим равенством:

$$H(\omega) = w_0 + 2 \sum_{k=1}^n w_k \cos(2k\pi\omega\Delta t), \quad (6.99)$$

где $H(\omega)$ – частотная характеристика; ω – частота; w_k – k -й вес, причем число k отсчитывается от главного весового коэффициента w_0 ; Δt – интервал времени между последовательными наблюдениями во временном ряде.

Для скользящего среднего с одинаковыми весами формула принимает вид:

$$H(\omega) = \frac{\sin(\pi\omega T)}{\pi\omega T}, \quad (6.100)$$

где T – интервал, за который осредняются значения временного ряда. На рис. 6.10 приведено графическое изображение этой функции.

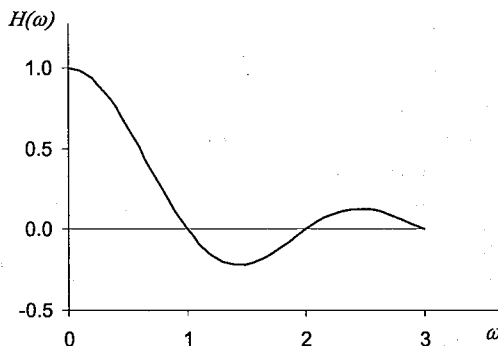


Рис. 6.10. График частотной характеристики скользящего среднего с равными весами.

Следует отметить, что для некоторых частот частотная характеристика отрицательна: не только амплитуды волн в этом интервале уменьшаются, но и их экстремумы становятся обратными, т.е. максимумы превращаются в минимумы и наоборот.

Этот недостаток скользящих средних с одинаковым весом можно устранить, применив сглаживающие функции, которые имеют веса, убывающие при удалении от центрального веса в каждом направлении пропорционально ординатам кривой нормального распределения вероятностей.

Можно рассчитать сглаживающие функции, которые приблизительно будут иметь форму нормального распределения, путем выбора весов биномиальных коэффициентов по формуле

$$c_m = \frac{N!}{m!(N-m)!}, \quad (6.101)$$

где m – порядковый номер весового коэффициента, который изменяется от 0 до N . Чтобы фильтрующая функция оставалась симметричной N должно быть четным.

Поскольку сумма весовых коэффициентов сглаживающей функции должна равняться единице, весовые коэффициенты получают по формуле:

$$w_m = c_m / \sum_{i=0}^{i=N} c_i. \quad (6.102)$$

По мере возрастания N распределение этих коэффициентов приближается по форме к нормальной кривой. При $N = 4$ биномиальные коэффициенты для $m = 0, 1, 2, 3, 4$ равны 1, 4, 6, 4, 1, веса сглаживающей функции будут равны соответственно 1/16, 4/16, 6/16, 4/16, 1/16, т.е. 0,0625; 0,250; 0,375; 0,250 и 0,0625.

Фильтр, весовые коэффициенты которого определяются с использованием формул (6.101), (6.102), называется *биномиальным фильтром*. Пример использования биномиального фильтра представлен в табл. 6.3.

Отфильтрованное значение 13,56 в табл. 6.2 получено как сумма произведений: $(0,0625 \cdot 16) + (0,250 \cdot 15) + (0,375 \cdot 12) + (0,250 \cdot 13) + (0,0625 \cdot 17)$.

Таблица 6.3

Сглаживание временного ряда биномиальным фильтром с коэффициентами $w_m = 0,0625; 0,250; 0,375; 0,250; 0,0625$

Временной ряд	Отфильтрованные величины
16	13,56 14,31
15	
12	
13	
17	
20	

Еще один вид сглаживания — *экспоненциальное сглаживание*. Оно называется экспоненциальным, так как вклад различных величин временного ряда в полученную на выходе сглаженную величину экспоненциально убывает в зависимости от интервала времени, прошедшего до момента, к которому отнесена осредненная величина. Будущие значения временного ряда не вносят никакого вклада.

Все физические приборы с постоянными коэффициентами инерции (постоянными времени) осредняют измеряемую величину по экспоненте. Ртутный термометр имеет почти постоянный коэффициент запаздывания и является примером прибора, который производит экспоненциальное сглаживание.

Поскольку весовые коэффициенты этого типа фильтра не являются симметричными относительно момента времени, к которому относится фильтрованная переменная, как это было в случае других, ранее описанных фильтров, этот фильтр сдвигает фазу волн временного ряда, полученного на выходе, относительно фазы волн той же частоты в исходном ряде. Этот сдвиг фаз является функцией частоты и выражается следующим образом:

$$\varphi = \arctg(-2\pi\omega\lambda), \quad (6.103)$$

где λ — постоянная времени. Угол φ всегда отрицателен и лежит в пределах от 0 до 90° . Таким образом, при экспоненциальном сглаживании фильтрованные волны всегда запаздывают по сравнению с исходными.

6.15. Оценка характеристик случайных процессов по эмпирическим данным

При исследовании гидрологических процессов во многих случаях требуется оценить их основные характеристики — математическое ожидание и корреляционную функцию.

Успешность решения этой задачи зависит от имеющейся информации. Наиболее благоприятной является ситуация, когда имеется достаточно большая совокупность реализаций случайного процесса. Однако довольно часто в распоряжении исследователя есть только одна реализация. В этом случае приходится вводить ряд допущений, в частности о стационарности и эргодичности исследуемого процесса.

В зависимости от количества эмпирической и априорной информации используются различные методы оценки характеристик случайных процессов.

6.15.1. Определение характеристик случайного процесса по множеству реализаций

Пусть имеется k реализаций ($k \geq 10$) случайного процесса $X(t)$ и пусть измерения этого процесса производятся в моменты времени $t_1, t_2, t_3, \dots, t_n$, тогда оценка математического ожидания случайного процесса $X(t)$ выражается формулой

$$m_x^*(t_i) = \frac{1}{k} \sum_{j=1}^k x_j(t_i). \quad (6.104)$$

Таким образом, для того чтобы получить оценку математического ожидания, необходимо рассчитать среднее значение для каждого сечения.

Аналогичным образом, для того чтобы получить оценку дисперсии случайного процесса $X(t)$, необходимо рассчитать оценку дисперсии для каждого сечения:

$$D_x^*(t_i) = \frac{1}{k-1} \sum_{j=1}^{j=k} [x_j(t_i) - m_x^*(t_i)]^2. \quad (6.105)$$

Оценка среднеквадратического отклонения случайного процесса $X(t)$ равна

$$S_x(t_i) = \sqrt{D_x^*(t_i)}. \quad (6.106)$$

Оценка корреляционной функции случайного процесса $X(t)$ в общем случае определяется выражением

$$K_x^*(t_i, t_q) = \frac{1}{k-1} \sum_{j=1}^{j=k} [x_j(t_i) - m_x^*(t_i)][x_j(t_q) - m_x^*(t_q)]. \quad (6.107)$$

Заметим, что при $i = q$ формула (6.107) превращается в формулу (6.105).

Оценка нормированной корреляционной функции случайного процесса $X(t)$ определяется выражением:

$$r_x^*(t_i, t_q) = \frac{K_x^*(t_i, t_q)}{S_x(t_i)S_x(t_q)}. \quad (6.108)$$

Как следует из формулы (6.108) эмпирической оценкой нормированной корреляционной функции процесса $X(t)$ будет корреляционная матрица вида:

$$\| r_{i,q} \| = \begin{vmatrix} r_{11} & r_{12} & r_{13} \cdots r_{1n} \\ & r_{22} & r_{23} \cdots r_{2n} \\ & & r_{33} \cdots r_{3n} \\ & & & \cdots \\ & & & & r_{nn} \end{vmatrix}, \quad (6.109)$$

где $r_{i,q}$ — коэффициенты парной корреляции между сечениями; n — число сечений.

Для стационарного случайного процесса математическое ожидание, дисперсия и СКО являются константами и их можно оценить по любому сечению. Однако более надежный результат будет получен, если сначала рассчитать оценку искомой характеристики для каждого сечения, а затем найти ее среднее значение по всем сечениям.

Корреляционная функция стационарного процесса не зависит от того, для каких моментов времени рассматриваются сечения, а зависит только от расстояния между сечениями τ . В этом случае оценка корреляционной функции записывается следующим образом:

$$k_x^*(\tau) = \frac{1}{k-1} \sum_{j=1}^{j=k} [x_j(t) - m_x^*][x_j(t+\tau) - m_x^*]. \quad (6.110)$$

Оценка нормированной корреляционной функции стационарного процесса будет иметь вид:

$$r_x^*(\tau) = \frac{K_x^*(\tau)}{S_x^2}. \quad (6.111)$$

Напомним, что эту функцию обычно называют автокорреляционной функцией.

Из формул (6.110) и (6.111) видно, что, меняя t , можно получить несколько оценок $r_x^*(\tau)$ для одного и того же τ .

Наилучшей эмпирической оценкой нормированной корреляционной функции стационарного случайного процесса будет столбец парных коэффициентов корреляции полученный осреднением диагональных элементов матрицы (6.109):

$$\begin{aligned}\bar{r}_x(0) &= \frac{r_{11} + r_{22} + r_{33} + \dots + r_{nn}}{n} \equiv 1; \\ \bar{r}_x(1) &= \frac{r_{12} + r_{23} + r_{34} + \dots + r_{(n-1)n}}{n-1}; \\ \bar{r}_x(2) &= \frac{r_{13} + r_{24} + r_{35} + \dots + r_{(n-2)n}}{n-2} \text{ и т.д.}\end{aligned}$$

Таким образом, если мы имеем дело с реализациями конечной длины, то при увеличении τ число коэффициентов парной корреляции, по которым производится осреднение, уменьшается, и точность оценки ординат корреляционной функции снижается.

Погрешность расчета выборочных ординат автокорреляционной функции стационарного случайного процесса в этом случае можно оценить по формуле

$$\varepsilon_r(\tau) = \frac{1 - [\bar{r}(\tau)]^2}{\sqrt{(k-1)(n-\tau)}}, \quad (6.112)$$

где k – количество реализаций ($k \geq 10$); n – длина реализаций (количество сечений); τ – расстояние между сечениями ($\tau < n$).

6.15.2. Определение характеристик стационарного эргодического процесса по одной реализации

Если для стационарного процесса $X(t)$ принимается гипотеза о том, что он обладает эргодическим свойством, то его математическое ожидание, дисперсию и СКО можно оценить по одной достаточно продолжительной реализации:

$$m_x^*(t) = m_x^* = \frac{1}{n} \sum_{i=1}^{i=n} x_i; \quad (6.113)$$

$$D_x^*(t) = D_x^* = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - m_x^*)^2; \quad (6.114)$$

$$\sigma_x^*(t) = \sigma_x^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - m_x^*)^2}, \quad (6.115)$$

где n — длина реализации (число членов ряда).

Отметим, что формулы (6.113) — (6.115) практически совпадают с формулами, которые используются при оценке этих параметров в рамках модели случайной величины. Разница состоит в том, что при использовании модели случайной величины имеющийся ряд рассматривается в качестве выборки из генеральной совокупности. Числовые характеристики, полученные по выборке, принимаются в качестве числовых характеристик случайной величины.

При использовании модели стационарного эргодического случайного процесса имеющийся ряд рассматривается как одна реализация случайного процесса. Числовые характеристики, полученные по этой реализации, принимаются в качестве числовых характеристик случайного процесса.

Корреляционная функция и автокорреляционная функция стационарного эргодического случайного процесса оцениваются по формулам:

$$k_x^*(\tau) = \frac{1}{n - \tau - 1} \sum_{i=1}^{n-\tau} (x_i - m_x^*)(x_{i+\tau} - m_x^*); \quad (6.116)$$

$$r_x^*(\tau) = \frac{k_x^*(\tau)}{D_x^*}, \quad (6.117)$$

где τ — расстояние между сечениями (используются также термин "сдвиг", а в зарубежной литературе — "лаг").

В практике гидрологических расчетов наиболее часто используется первая ордината автокорреляционной функции, которая называется *коэффициентом автокорреляции*. Как следует из формул (6.116) — (6.117), коэффициент автокорреляции определяется выражением

$$r_x^*(1) = \frac{\sum_{i=1}^{i=n-1} (x_i - m_x^*)(x_{i+1} - m_x^*)}{(n-2)D_x^*}. \quad (6.118)$$

Поясним методику оценки ординат автокорреляционной функции на следующем примере.

Даны среднегодовые расходы воды реки Невы в створе деревни Новосаратовки (табл. 6.4), длина ряда $n = 15$.

Таблица 6.4

Расчет ординат автокорреляционной функции ряда среднегодовых расходов воды; р. Нева – д. Новосаратовка

№ п/п	Год	Расход воды y , м ³ /с	$\tau = 0$ $m = 15$	$\tau = 1$ $m = 14$	$\tau = 2$ $m = 13$	$\tau = 3$ $m = 12$
			x_0	x_1	x_2	x_3
1	1971	2318	2318			
2	1972	2018	2018	2318		
3	1973	1579	1579	2018	2318	
4	1974	2012	2012	1579	2018	2318
5	1975	2508	2508	2012	1579	2018
6	1976	2188	2188	2508	2012	1579
7	1977	2230	2230	2188	2508	2012
8	1978	2300	2300	2230	2188	2508
9	1979	2200	2200	2300	2230	2188
10	1980	2100	2100	2200	2300	2230
11	1981	2450	2450	2100	2200	2300
12	1982	3100	3100	2450	2100	2200
13	1983	2610	2610	3100	2450	2100
14	1984	2690	2690	2610	3100	2450
15	1985	2380	2380	2690	2610	3100
				2380	2690	2610
					2380	2690
						2380
Коэффициент парной корреляции, $r(y, x_i)$			1,0	0,49	0,10	,08
Среднеквадратическая погрешность коэффициента парной корреляции			$\pm 0,0$	$\pm 0,14$	$\pm 0,26$	$\pm 0,28$

Для того чтобы рассчитать первые три ординаты автокорреляционной функции, сдвигаю ряд относительно самого себя на величину $\tau = 0, 1, 2, 3$. Исходный ряд обозначим y ; ряд, полученный при сдвиге $\tau = 0$, обозначим x_0 ; ряд, полученный при сдвиге $\tau = 1$, обозначим x_1 ; ряд, полученный при сдвиге $\tau = 2$, обозначим x_2 и ряд, полученный при сдвиге $\tau = 3$, обозначим x_3 .

Рассчитаем коэффициенты парной корреляции между y и каждым из вновь сформированных рядов: $r(y, x_0)$, $r(y, x_1)$, $r(y, x_2)$, $r(y, x_3)$. При этом в принципе рассчитывать коэффициент корреляции при $\tau = 0$ нет необходимости, так как $r(y, x_0) \equiv 1$.

Полученные коэффициенты парной корреляции $r(y, x_i)$ представляют собой оценки первых трех ординат искомой автокорреляционной функции.

Следует отметить, что при увеличении τ совместный период, по которому оценивается коэффициент парной корреляции m , уменьшается. В данном случае при $\tau = 0$ получаем $m = 15$, при $\tau = 1$ $m = 14$, при $\tau = 2$ $m = 13$; при $\tau = 3$ $m = 12$.

При уменьшении m неизбежно возрастает погрешность выборочных ординат автокорреляционной функции.

В классической статистике принято рассчитывать первые m ординат автокорреляционной функции:

$$\tau_m = \frac{1}{10} n. \quad (6.119)$$

В практике гидрологических расчетов нередко применяют более слабое ограничение:

$$\tau_m = \frac{1}{3} n. \quad (6.120)$$

Однако нужно помнить, что при малых n недостоверными могут оказаться не только значения выборочных ординат корреляционной функции, но даже их знак уже при $\tau \geq 2$ (см. табл. 6.2).

Погрешность ординат автокорреляционной функции для стационарного эргодического процесса оценивается по формуле

$$\varepsilon_r(\tau) = \frac{1 - r_\tau^2}{\sqrt{n - 1 - \tau}}. \quad (1.121)$$

6.15.3. Определение характеристик периодически коррелированного случайного процесса по одной реализации

Если периодически коррелированный случайный процесс $X(t)$ (ПКСП) с периодом коррелированности T обладает эргодическим свойством, то его характеристики можно определить по одной дос-

точно продолжительной реализации. Математическое ожидание и дисперсия для ПКСП являются периодическими функциями, поэтому достаточно получить значения искомой характеристики для моментов времени $t = 1, 2, 3, \dots T$.

Математическое ожидание и дисперсия оцениваются по формулам:

$$m^*(t) = \frac{1}{k} \sum_{i=0}^{i=k-1} X(t+iT), \quad (6.122)$$

$$D^*(t) = \frac{1}{k-1} \sum_{i=1}^{i=k-1} [X(t+iT) - m^*(t+iT)]^2, \quad (6.123)$$

где k – целое число, $kT = N$; N – длина реализации.

Так, например, если рассматривать в рамках модели ПКСП расходы воды за 30 лет с шагом дискретности один месяц, то мы будем иметь ряд из 360 среднемесячных расходов при периоде коррелированности $T = 12$ месяцев. В этом случае для того, чтобы оценить математическое ожидание и дисперсию случайного процесса, нужно рассчитать эти характеристики для каждого из двенадцати месяцев (рис. 6.11).

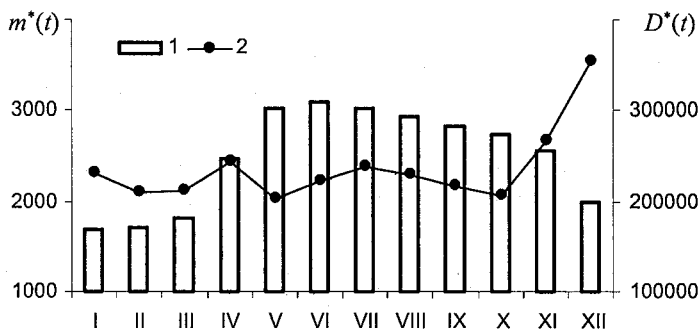


Рис. 6.11. Оценки математического ожидания (1) и дисперсии (2) среднемесячных расходов воды $\text{м}^3/\text{с}$ на интервале 12 месяцев; р. Нева – д. Новосаратовка.

Если полученные для 12 месяцев функции $m^*(t)$ и $D^*(t)$ "размножить" и "склеить" в цепочку, то мы получим искомые оценки математического ожидания и дисперсии, соответствующие эргодическому ПКСП (рис. 6.12).

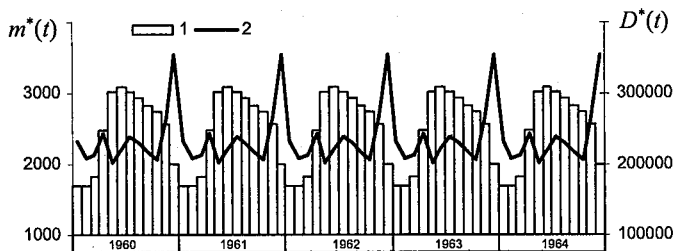


Рис. 6.12. Оценки математического ожидания (1) и дисперсии (2) среднеемесячных расходов воды м³/с на 5-летнем интервале времени в рамках модели ПКСП; р. Нева – д. Новосаратовка.

Автокорреляционная функция эргодического ПКСП может быть определена по формуле

$$r^*(t, \tau) = \frac{\sum_{i=0}^{k-1-\tau} \left[\overset{0}{X}(t+iT) \overset{0}{X}(t+iT+\tau) \right]}{(k-1-\tau) \sqrt{D^*(t+iT)} \sqrt{D^*(t+iT+\tau)}}, \quad (6.124)$$

где $\overset{0}{X}(t)$ – центрированный ПКСП; $\overset{0}{X}(t) = X(t) - m^*(t)$.

Рассмотрим методику оценки автокорреляционной функции эргодического ПКСП на примере ряда среднеемесячных расходов воды реки Невы в створе д. Новосаратовка за 1859 – 1986 гг. (табл. 6.5).

Таблица 6.5

Среднеемесячные расходы воды, м³/с; р. Нева – д. Новосаратовка

№ п/п	Год	Месяцы без сдвижки					Месяцы со сдвигом вверх на 1 год				
		1	2	...	11	12	1*	2*	...	11*	12*
1	1859	1080	1080	...	2320	1400	1450	1470	...	2430	1440
2	1860	1450	1470	...	2430	1440	1410	1470	...	2420	1520
3	1861	1410	1470	...	2420	1520	1520	1510	...	2300	1500
4	1862	1520	1510	...	2300	1500	1500	1460	...	2590	2200
5	1863	1500	1460	...	2590	2200	1690	1760	...	2550	2360
6	1864	1690	1760	...	2550	2360	2400	2410	...	2900	1950
7	1865	2400	2410	...	2900	1950	1900	1900	...	2340	1800
8	1866	1900	1900	...	2340	1800	1860	1900	...	3120	2660
	...	...	...	...	...	...	...	...	...	...	...
124	1982	1780	2000	...	3110	3070	2040	1790	...	2590	1450
125	1983	2040	1790	...	2590	1450	1360	1310	...	2970	2360
126	1984	1360	1310	...	2970	2360	1550	1820	...	2550	1370
127	1985	1550	1820	...	2550	1370	1400	1350	...	2670	1740
	1986	1400	1350	...	2670	1740					

Будем рассчитывать оценки первых 12 ординат. Для этого сдвинем исходную таблицу расходов на один год вверх (см. табл. 6.5). В результате получим 24 столбца расходов: (1, 2, 3, ...12) + (1*, 2*, 3*, ..., 12*). Каждый столбец содержит 127 значений.

Используя эти данные, рассчитываем корреляционную матрицу (табл. 6.6). Полученная корреляционная матрица будет иметь размер 24×24 (по числу столбцов в табл. 6.5).

Таблица 6.6

Корреляционная матрица среднемесячных расходов воды;
р. Нева – д. Новосаратовка

Месяц	1	2	3	4	...	12	1*	2*	3*	...	12*
1	1										
2	0,95	1									
3	0,89	0,92	1								
...	...	...	...	...	...						
12	0,45	0,48	0,46	0,41	...	1					
1*	0,52	0,58	0,53	0,41	...	0,85	1				
2*	0,45	0,53	0,48	0,37	...	0,81	0,95	1			
3*	0,38	0,45	0,41	0,29	...	0,78	0,88	0,92	1		
...	...	...	...	...	...	...	...	...	...	...	
12*	0,07	0,07	0,09	0,04	...	0,29	0,44	0,48	0,46	...	1

Поскольку в практике гидрологических расчетов интерес представляет зависимость настоящего от прошлого, а не от будущего, выбираем из таблицы соответствующие строки (в табл. 6.6 они выделены рамками).

Так, например, для месяца, обозначенного символом 3*, при сдвиге $\tau = 0$ получаем $r(t, \tau) = r(3^*, 0) = 1$, при $\tau = 1$ получаем $r(3^*, 1) = 0,92$, при $\tau = 2$ получаем $r(3^*, 2) = 0,88$, при $\tau = 3$ получаем $r(3^*, 3) = 0,78$ и т. д.

При сдвиге $\tau = 12$ мы получим коэффициент корреляции между мартом текущего года и мартом предшествующего года: $r(3^*, 12) = 0,41$.

Выбранные из табл. 6.6 ординаты автокорреляционной функции заносим в табл. 6.7. Таким образом, в данном случае автокорреляционная функция ПКСП представляет собой набор из 12 автокорреляционных функций – для января, февраля, ..., декабря. На рис. 6.13 в качестве примера представлены две из 12 функций.

Таблица 6.7

Ординаты автокорреляционной функции среднемесячных расходов воды,
 $r(t, \tau)$; р. Нева – д. Новосаратовка

τ	Месяц											
	1	2	3	4	5	6	7	8	9	10	11	12
0	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
1	0,85	0,95	0,92	0,81	0,83	0,95	0,97	0,99	0,99	0,98	0,91	0,84
2	0,85	0,81	0,88	0,75	0,86	0,74	0,94	0,95	0,96	0,95	0,87	0,76
3	0,86	0,86	0,78	0,75	0,88	0,78	0,69	0,91	0,92	0,92	0,81	0,72
4	0,82	0,87	0,81	0,68	0,83	0,81	0,78	0,65	0,87	0,87	0,76	0,67
5	0,78	0,82	0,78	0,75	0,69	0,75	0,82	0,74	0,61	0,81	0,71	0,62
6	0,73	0,75	0,73	0,69	0,78	0,58	0,76	0,78	0,70	0,56	0,67	0,57
7	0,67	0,69	0,66	0,64	0,77	0,68	0,58	0,72	0,74	0,63	0,47	0,53
8	0,61	0,63	0,61	0,59	0,71	0,69	0,67	0,54	0,68	0,69	0,51	0,40
9	0,41	0,56	0,54	0,54	0,65	0,63	0,69	0,62	0,49	0,62	0,55	0,46
10	0,53	0,37	0,48	0,47	0,59	0,58	0,63	0,64	0,56	0,42	0,48	0,48
11	0,58	0,48	0,29	0,43	0,53	0,52	0,57	0,58	0,59	0,50	0,32	0,44
12	0,52	0,53	0,41	0,30	0,48	0,45	0,52	0,52	0,53	0,52	0,35	0,29

В рассмотренном примере рассчитывались ординаты автокорреляционной функции ПКСП до максимального сдвига $\tau_{\max} = 12$. Если $12 < \tau_{\max} \leq 24$, то на первом этапе формируется таблица не из 24 столбцов, а из 36 столбцов. Первые 12 столбцов – таблица исходных данных; следующие 12 столбцов получают сдвигом исходной таблицы вверх на 1 год; еще 12 столбцов получают сдвигом исходной таблицы вверх на 2 года.

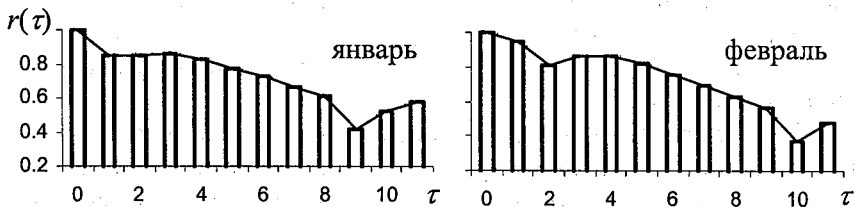


Рис. 6.13. Эмпирические автокорреляционные функции среднемесячных расходов воды для января и февраля; р. Нева – д. Новосаратовка.

6.15.4. Выборочный спектр стационарного эргодического процесса

Выборочный нормированный спектр стационарного случайного процесса $X(t)$ на конечном интервале разложения можно определить через выборочную автокорреляционную функцию:

$$s^*(\omega) = \frac{1}{\pi} \left[1 + 2 \sum_{\tau=1}^{\tau=\tau_m} r(\tau) \cos(\omega\tau) \right], \quad (6.125)$$

где ω – угловая частота ($0 \leq \omega \leq \pi$); $r(\tau)$ – выборочная автокорреляционная функция; τ_m – максимальный сдвиг при оценке автокорреляционной функции. При назначении τ_m можно ориентироваться на следующее неравенство: $(n - \tau_m) \geq 10$ (где n – длина ряда).

Период и угловая частота в формуле (6.125) связаны соотношением:

$$T = \frac{2\pi}{\omega} \quad (6.126)$$

Заметим, что в различных ситуациях спектр приходится нормировать с помощью различных множителей, например, если используется абсолютная частота $f = \omega/2\pi$, пробегающая интервал $[0, 0,5]$, то множитель $(1/\pi)$ нужно заменить на 2. Период и частота в этом случае будут связаны соотношением:

$$T = \frac{1}{f} \quad (6.127)$$

Формула (6.125) получена переходом от теоретического спектра к конечным разностям, однако эта оценка не является состоятельной. К лучшим оценкам спектральной плотности приводят методы, основанные на предварительном сглаживании автокорреляционной функции.

Процедура сглаживания состоит в том, что каждая ордината автокорреляционной функции умножается на весовой коэффициент. При этом учитывается тот факт, что с увеличением сдвига увеличивается ошибка выборочных ординат автокорреляционной функции. Поэтому с увеличением сдвига весовые коэффициенты убывают.

Функцию, сглаживающую автокорреляционную функцию, принято называть *корреляционным окном*, обычно ее обозначают $\lambda(\tau)$. В этом случае формула (6.125) принимает вид:

$$s^*(\omega) = \frac{1}{\pi} \left[1 + 2 \sum_{\tau=1}^{\tau=\tau_m} \lambda(\tau) r(\tau) \cos(\omega\tau) \right]. \quad (6.128)$$

В настоящее время разработано достаточно большое количество корреляционных окон, ниже представлены наиболее известные из них.

1. Прямоугольное окно:

$$\lambda(\tau) = \begin{cases} 1 & \text{при } |\tau| \leq \tau_m, \\ 0 & \text{при } |\tau| > \tau_m. \end{cases} \quad (6.129)$$

2. Функция Барлетта (треугольное окно):

$$\lambda(\tau) = \begin{cases} 1 - \frac{|\tau|}{\tau_m} & \text{при } |\tau| \leq \tau_m, \\ 0 & \text{при } |\tau| > \tau_m. \end{cases} \quad (6.130)$$

3. Функция Хемминга

$$\lambda(\tau) = \begin{cases} 0,54 + 0,46 \cos(\pi\tau/\tau_m) & \text{при } |\tau| \leq \tau_m, \\ 0 & \text{при } |\tau| > \tau_m. \end{cases} \quad (6.131)$$

4. Функция Ханна

$$\lambda(\tau) = \begin{cases} 0,50 + 0,5 \cos(\pi\tau/\tau_m) & \text{при } |\tau| \leq \tau_m, \\ 0 & \text{при } |\tau| > \tau_m. \end{cases} \quad (6.132)$$

5. Функция Парзена

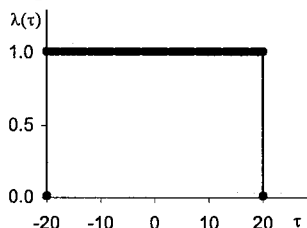
$$\lambda(\tau) = \begin{cases} 1 - \left(\frac{\tau}{\tau_m} \right)^2 & \text{при } |\tau| \leq \tau_m, \\ 0 & \text{при } |\tau| > \tau_m. \end{cases} \quad (6.133)$$

5. Функция Наттолла

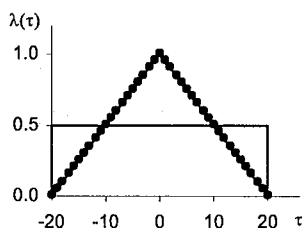
$$\lambda(\tau) = \begin{cases} \sum_{k=0}^{k=3} a_k \cos[(\pi k \tau)/\tau_m] & \text{при } |\tau| \leq \tau_m, \\ 0 & \text{при } |\tau| > \tau_m. \end{cases} \quad (6.134)$$

Здесь $a_0 = 0,3635819$; $a_1 = 0,4891775$; $a_2 = 0,1365995$; $a_3 = 0,0106441$.

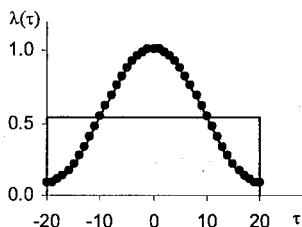
На рис. 6.14 даны графики перечисленных спектральных окон. Как видно на рисунке, прямоугольное окно дает одинаковые весовые коэффициенты и только усекает автокорреляционную функцию. У треугольного окна весовые коэффициенты убывают линейно. Эти окна не являются идеальными и в настоящее время используются редко.



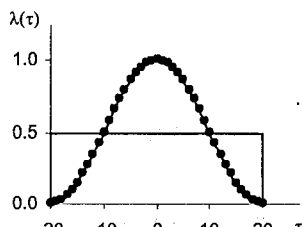
Прямоугольное окно



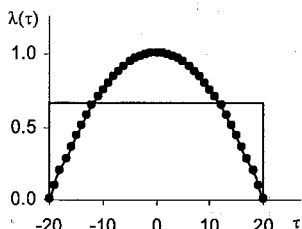
Функция Барлетта (треугольное окно)



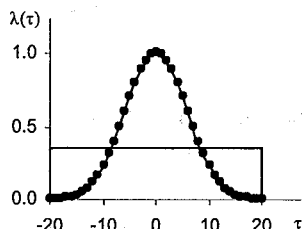
Функция Хемминга



Функция Ханна



Функция Парзена



Функция Наттолла

Рис. 6.14. Различные корреляционные окна, при максимальном сдвиге $\tau_m = 20$.

Наиболее часто используются окна Хемминга, Ханна и Парзена.

Окно Наттолла в гидрологии используется довольно редко, однако оно имеет определенные преимущества перед остальными окнами. В частности, этот метод максимально снижает уровень второстепенных шумов. Правда, в качестве расплаты за это, мы получаем более сглаженный спектр.

Доверительный интервал для выборочного спектра приближенно описывается неравенством

$$\frac{\chi_{\alpha}^2}{\nu} S_0(\omega) \leq S(\omega) \leq \frac{\chi_{1-\alpha}^2}{\nu} S_0(\omega), \quad (6.135)$$

где χ_{α}^2 – квантиль распределения хи-квадрат при уровне значимости 2α ; $S_0(\omega)$ – среднее значение спектра; ν – число степеней свободы.

Число степеней свободы определяется по формуле

$$\nu = 2n/\tau_m - 0,5. \quad (6.136)$$

Если в качестве нулевой гипотезы мы принимаем модель ряда в виде случайного процесса с независимыми сечениями, то средний спектр является константой. Для нормированного спектра при любом значении частоты ω средний спектр равен

$$S_0(\omega) = \frac{1}{\pi} = 0,318. \quad (6.137)$$

Если в качестве нулевой гипотезы принята модель авторегрессии первого или второго порядка (см. п. 6.12.1), то средний спектр можно определить по формулам (6.87), (6.88).

6.15.5. Использование методов декомпозиции при оценке характеристик случайных процессов

Если при анализе гидрологических рядов используется интервал дискретности по времени месяц и менее, то описать статистическую структуру ряда простой моделью довольно сложно. В этой ситуации можно воспользоваться моделью ПКСП (см. п. 6.11) или моделью Бокса – Дженкинса (см. п. 6.12.4). Однако существует более общий подход – декомпозиция случайного процесса.

В рамках этого подхода предполагается, что случайный процесс можно представить в виде нескольких процессов, каждый из которых описывается достаточно простой моделью.

Формы разложения (декомпозиции) временного ряда на отдельные составляющие могут различаться. В гидрологической практике обычно используются либо аддитивная модель, либо мультипликативная модель, причем аддитивная модель используется чаще.

Аддитивной моделью временного ряда называется представление исходного процесса в виде суммы более простых случайных процессов.

Мультипликативной моделью временного ряда называется представление исходного процесса в виде произведения более простых случайных процессов.

В настоящее время разработано несколько статистических пакетов, где реализованы процедуры декомпозиции:

- универсальный статистический пакет STADIA разработан НПО "Информатика и компьютеры" при активном участии МГУ им. М. В. Ломоносова;

- универсальный статистический пакет SPSS фирмы SPSS, inc. (США).

- универсальный статистический пакет STATISTICA фирмы StatSoft, inc. (США).

Указанные пакеты позволяют выделить из исходного ряда следующие компоненты: линейный тренд, периодическую (сезонную) составляющую, тренд-циклическую составляющую и случайную составляющую (шум).

Выделять линейный тренд имеет смысл, если на всем анализируемом интервале времени ряд имеет тенденцию к повышению или понижению. Линейный тренд представляет собой детерминированную компоненту, которая описывается линейной функцией времени.

Учет периодической составляющей позволяет выделить из ряда детерминированную циклическую компоненту, обусловленную регулярными внутригодовыми или внутрисуточными изменениями (рис. 6.15).

Тренд-циклическая составляющая отражает низкочастотные колебания, например, многолетние колебания водности реки или озера. Эта составляющая представляет собой случайный процесс, который может быть описан в рамках модели АРСС.

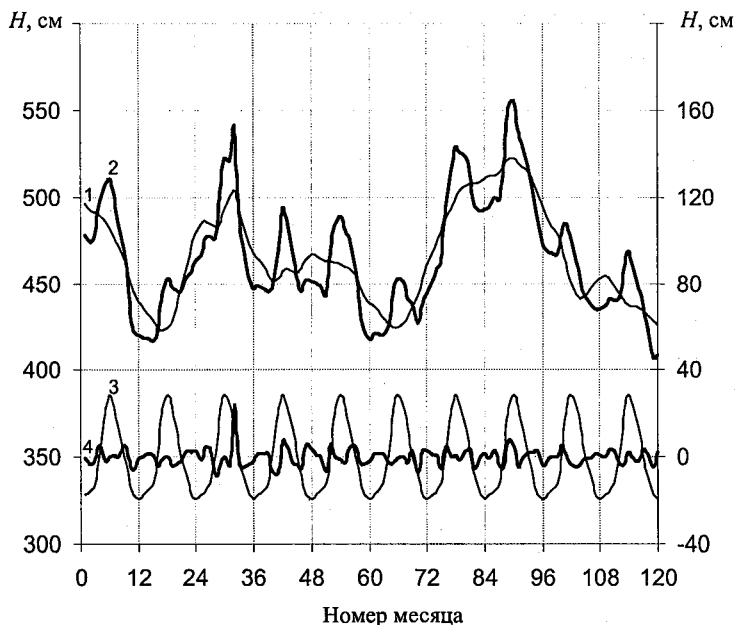


Рис. 6.15. Среднемесячные уровни Ладожского озера за 1882 – 1891 гг.
Основная шкала: 1 – тренд-циклическая компонента; 2 – исходный ряд уровней.
Вспомогательная шкала: 3 – периодическая составляющая;
4 – случайная составляющая.

Если три первых компоненты выделены корректно, то в идеале остаток должен представлять собой случайную составляющую в виде "белого гауссовского шума".

6.16. Моделирование искусственных гидрологических рядов

При решении некоторых гидрологических задач необходимо иметь ряды продолжительностью в несколько сотен и даже тысяч лет. Поскольку ряды такой продолжительности отсутствуют, были разработаны методы моделирования искусственных гидрологических рядов. При использовании этих методов вначале производится оценка параметров распределения имеющегося эмпирического ряда, а затем моделируется искусственный ряд с такой же статистической структурой, но гораздо большей продолжительности.

В основе моделирования искусственных гидрологических рядов лежит метод Монте-Карло. *Метод Монте-Карло* – это метод решения математических задач при помощи моделирования случайных величин.

Процесс моделирования включает несколько этапов. На первом этапе требуется получить последовательность случайных чисел, равномерно распределенных на интервале $[0 - 1]$ (см. п. 2.2).

Различают три способа получения случайных чисел: таблицы случайных чисел, генераторы случайных чисел и метод псевдослучайных чисел.

Простейшим генератором случайных чисел является рулетка. С помощью устройств типа рулетки легко сгенерировать случайные последовательности любой продолжительности. Полученные таким образом случайные последовательности можно оформить в виде таблиц. Самая большая из опубликованных таблиц случайных чисел содержит один миллион цифр. Однако при работе на компьютере удобнее пользоваться так называемыми *псевдослучайными* числами.

Первый алгоритм для получения псевдослучайных чисел был предложен Дж. Нейманом. Он называется *методом середины квадратов*. Поясним его на примере.

Пусть задано 4-значное число $\gamma_0 = 0,9876$. Возведем его в квадрат. Получим 8-значное число ($\gamma_0^2 = 0,97535376$). Выберем четыре средние цифры этого числа и положим $\gamma_1 = 0,5353$. Затем возведем γ_1 в квадрат ($\gamma_1^2 = 0,28654609$) и снова извлечем четыре средние цифры. Получим $\gamma_2 = 0,6546$. Далее, $\gamma_2^2 = 0,42850116$, $\gamma_3 = 0,8501$; $\gamma_3^2 = 0,72267001$, $\gamma_4 = 2670$; $\gamma_4^2 = 0,07128900$, $\gamma_5 = 0,1289$ и т. д.

Но этот алгоритм не оправдал себя: получалось больше чем нужно малых значений. Позже разными исследователями были разработаны другие, более совершенные алгоритмы.

В настоящее время на любом персональном компьютере имеется возможность генерации случайных чисел, равномерно распределенных в интервале $[0 - k]$, где k – положительное целое число.

В качестве примера рассмотрим программу на языке Pascal, которая генерирует 20 значений случайной величины, равномерно распределенных на интервале $[1 - 50]$:

```

var i: integer;
begin
Randomize; {смена базы генератора псевдослучайных чисел}
For i := 1 to 20 do
begin
writeln(Random(50)) {генерация случайного числа}
end
end.

```

Процедура Randomize устанавливает начальное значение датчика случайных чисел в зависимости от показания системного таймера. Поэтому при каждом запуске программы будет генерироваться новая выборка.

Вместо процедуры Randomize можно использовать системную переменную RandSeed, которая позволяет загрузить конкретную базу для датчика случайных чисел. Если переменной RandSeed присваивать одно и то же значение, то будут генерироваться одинаковые выборки. RandSeed может принимать любые целые значения из диапазона $-2147483648 \dots 2147483647$. Значит имеется возможность сгенерировать более 4 миллиардов различных выборок.

На следующем этапе моделирования каждое значение случайной величины, равномерно распределенной на интервале $[0-1]$, рассматривается как вероятность непревышения и по нему рассчитывается соответствующий квантиль заданного закона распределения (рис. 6.16).

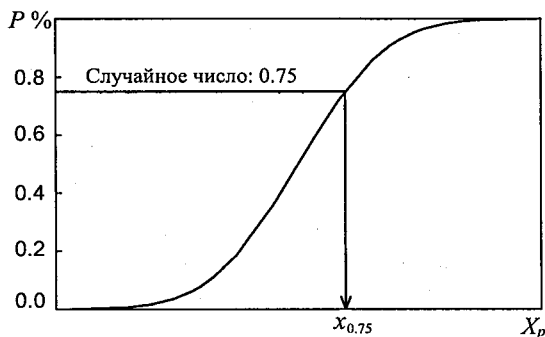


Рис. 6.16. Схема моделирования случайной величины с заданным законом распределения.

Переход от случайных чисел равномерно распределенных на отрезке $[0 - 1]$ к случайным числам с заданным законом распределения выполняется аналитически с использованием обратной функции от интегрального закона распределения. Если аналитического решения обратной задачи не существует, то задача решается с использованием численных методов.

Применяя изложенную схему, можно написать компьютерную программу на любом языке программирования. Однако довольно часто достаточно воспользоваться уже готовыми программами. Например, "Пакет анализа", который входит в состав Microsoft Excel, позволяет моделировать ряды для 7 законов распределения. Еще большими возможностями обладают специальные статистические пакеты.

Если вероятностная структура ряда соответствует модели случайного процесса, то алгоритм несколько сложнее.

Так для моделирования ряда соответствующего модели автокорреляции первого порядка (см. п. 6.13) можно воспользоваться рекуррентной формулой

$$x_{i+1}^0 = r x_i^0 + a_i, \quad (6.138)$$

где $x_i^0 = (x_i - \bar{x})$ – центрированный случайный процесс; r – коэффициент автокорреляции; a_i – случайная составляющая (белый шум).

Соответствующая блок-схема представлена на рис. 6.17.

Для моделирования необходимо иметь следующие параметры: среднее значение $\bar{x}$, среднеквадратическое отклонение σ_x , отношение C_s/C_v , коэффициент автокорреляции r и начальное значение моделируемой гидрологической характеристики x_1 . В качестве начального значения можно в большинстве случаев принять среднее значение (в этом случае $x_1 = 0$). Кроме того, требуется указать тип распределения, например Крицкого – Менкеля или Пирсона III типа.

При моделировании по формуле (6.138) на каждом шаге генерируется значение случайной величины a_i по ранее изложенной схеме.

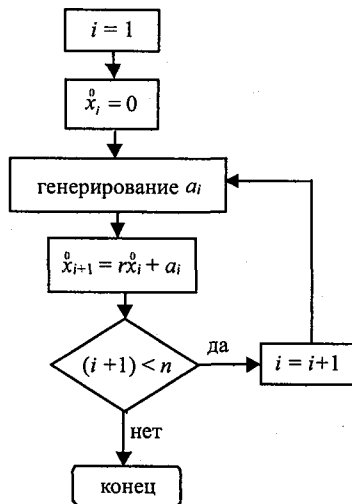


Рис. 6.17. Блок-схема для генерации рядов по модели авторегрессии первого порядка.

Тип распределения для шумовой составляющей a_i и отношение C_s/C_v принимаются такими же, как для исходного ряда, среднее значение может быть принято равным нулю, а СКО в соответствии с формулой (6.86) определяется выражением

$$\sigma_a = \sigma_x \sqrt{1 - r^2}. \quad (6.139)$$

Для моделирования рядов соответствующих модели авторегрессии второго порядка используется формула (6.80).

В п. 6.13 приводятся также формулы для моделирования по моделям скользящего среднего (СС) и АРСС.

7. ТРЕБОВАНИЯ НОРМАТИВНЫХ ДОКУМЕНТОВ К СТАТИСТИЧЕСКОЙ ОБРАБОТКЕ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

В настоящее время основным нормативным документом при выполнении гидрологических расчетов является СП 33-101-2003. Рассмотрим, как в рамках этого документа производится расчет гидрологических характеристик при наличии длинного ряда гидрометрических наблюдений.

В гидрологии ряд считается длинным, если погрешность расчета среднего значения ряда не превышает 5 – 10 %, а погрешность расчета коэффициента вариации не превышает 10 – 15 %. Хотя расчет каждой гидрологической характеристики имеет свои особенности, методический подход во всех случаях един и базируется на использовании методов теории вероятностей и математической статистики. Общая схема расчета включает следующие этапы.

1. Предварительный анализ исходных данных. На этом этапе производится анализ надежности экстраполяции кривой $Q = f(H)$; выполняется проверка полноты учета стока воды на поймах и в протоках; оценивается точность расчета стока за различные интервалы времени и т. д. Цель такой проверки – оценить надежность исходной информации и выявить грубые ошибки и опечатки.

2. Проверка ряда на случайность. Цель данной проверки – выбор наиболее адекватной математической модели для описания вероятностной структуры гидрологического ряда. В качестве нулевой принимается гипотеза о том, что ряд соответствует модели случайной величины. Если нулевая гипотеза опровергается, то принимается альтернативная гипотеза. В большинстве случаев – это модель авторегрессии первого порядка (простая цепь Маркова).

Для проверки случайности используются различные статистические критерии, описание которых дано в п. 4.8. Однако в практике гидрологических расчетов наиболее часто используется критерий значимости коэффициента автокорреляции:

$$r < \sigma_r t_{2\alpha}, \quad (7.1)$$

где r – коэффициент автокорреляции, определяемый по формуле (6.118); σ_r – его абсолютная погрешность; $t_{2\alpha}$ – нормированная ор-

дината нормального закона распределения (табл. 2.1) при уровне значимости 2α (обычно $2\alpha = 5\%$ или $2\alpha = 10\%$).

Значение σ_r рассчитывается по формуле

$$\sigma_r = \frac{1-r^2}{\sqrt{n-2}}. \quad (7.2)$$

Гипотеза о случайности ряда не опровергается, если выполняется неравенство (7.1).

Следует учитывать, что погрешность выборочного коэффициента автокорреляции достаточно велика, и делать вывод о принятии той или иной вероятностной модели по результатам обработки только одного ряда очень рискованно. Поэтому рекомендуется проводить проверку на случайность по группе рек в пределах гидрологически однородного района.

Гипотеза о соответствии гидрологических рядов модели случайной величины опровергается, если процент опровержений гипотезы по рекам района превысит принятый уровень значимости.

В этом случае принимается модель авторегрессии первого порядка, а в качестве r для всех рек устанавливается средний по району коэффициент автокорреляции.

3. Проверка ряда на однородность. Гидрологический ряд считается однородным, если в течение всего рассматриваемого периода условия формирования стока оставались неизменными. Антропогенное воздействие и природные катаклизмы могут привести к изменению условий формирования стока и нарушить однородность ряда. Например, в результате землетрясения могут измениться границы водосбора, а работа водохранилища сезонного регулирования приводит к перераспределению стока внутри года и увеличению суммарного испарения. В этом случае без предварительной корректировки ряда нельзя применять стандартные методы статистической обработки.

В рамках модели случайной величины для проверки гидрологических рядов на однородность используются стандартные статистические критерии Фишера и Стьюдента (см. п. 4.6.1, 4.6.2).

Если для описания вероятностной структуры ряда используется модель авторегрессии первого порядка, то при проверке однородности следует применять таблицы Фишера и Стьюдента, обобщенные

на случай автокоррелированных рядов. Такие таблицы даны в СП 33-101-2003.

Вместо указанных таблиц при определении критического значения статистики Стьюдента можно использовать формулу

$$t'_{2\alpha} = C_t t_{2\alpha}, \quad (7.3)$$

где $t'_{2\alpha}$ – критическое значение статистики Стьюдента с учетом автокорреляции; $t_{2\alpha}$ – критическое значение статистики Стьюдента без учета автокорреляции; C_t – переходный коэффициент, определяемый по табл. 7.1.

Таблица 7.1

Коэффициенты $C_t = f(r)$, предназначенные для пересчета критических значений статистики Стьюдента

r	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	1,00	1,00	1,01	1,02	1,02	1,03	1,04	1,05	1,06	1,07
0,1	1,07	1,08	1,09	1,10	1,11	1,12	1,13	1,14	1,16	1,17
0,2	1,18	1,19	1,20	1,22	1,23	1,24	1,26	1,27	1,29	1,31
0,3	1,33	1,34	1,35	1,37	1,39	1,41	1,42	1,44	1,46	1,49
0,4	1,51	1,52	1,54	1,56	1,59	1,67	1,63	1,65	1,68	1,70
0,5	1,72	1,75	1,78	1,81	1,84	1,88	1,92	1,95	1,99	2,03
0,6	2,06	2,07	2,13	2,17	2,21	2,24	2,28	2,32	2,36	2,40

При проверке однородности по критерию Фишера можно также воспользоваться формулами перехода. При объемах выборок $n_1 \geq 25$ и $n_2 \geq 25$ допустимо использовать классическое F -распределение с новыми степенями свободы, которые зависят от автокорреляции и асимметрии и определяются по формулам:

$$v_1 = \frac{n_1 g}{1 + \frac{2r^2}{1-r^2} \left[1 - \frac{1-r^{2n_1}}{n_1(1-r^2)} \right]}; \quad (7.4)$$

$$v_2 = \frac{n_2 g}{1 + \frac{2r^2}{1-r^2} \left[1 - \frac{1-r^{2n_2}}{n_2(1-r^2)} \right]}, \quad (7.5)$$

где n_1 и n_2 – соответственно длина первой и второй выборки; r – коэффициент автокорреляции; g – параметр, определяемый по табл. 7.2.

Зависимость коэффициента g от C_s при определении новых степеней свободы для критерия Фишера

C_s	0	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0
g	1,0	0,82	0,62	0,45	0,30	0,24	0,17	0,14	0,10

Критерии однородности в классической статистике применяются для выборок большой продолжительности, как правило, из нескольких сотен членов. В гидрологии приходится работать в условиях коротких выборок, поэтому оценку однородности рекомендуется проводить не для одного ряда, а по группе рек в пределах однородного (по условиям формирования стока) района. Кроме того, необходимо использовать дополнительную информацию о видах и интенсивности антропогенной деятельности на водосборе.

4. Расчет параметров распределения. Обычно рассчитываются среднее значение ряда, коэффициент вариации и коэффициент асимметрии. Расчет производится методом моментов по формулам (3.13), (3.17), (3.19).

Если коэффициент вариации больше 0,6, следует произвести уточнение указанных характеристик, используя метод наибольшего правдоподобия (см. п. 3.2.2). При этом необходимо заранее принять решение о выборе аналитической кривой обеспеченностей, так как решение в рамках этого метода зависит от типа распределения (в России основной расчетной кривой является кривая обеспеченностей Крицкого – Менкеля).

Учитывая то, что коэффициент асимметрии и отношение C_s/C_v рассчитываются с большой погрешностью, рекомендуется использовать районное отношение C_s/C_v . Районное отношение C_s/C_v получают путем осреднения этого показателя по группе рек в пределах гидрологически однородного района. Исключение составляют крупные реки, пересекающие несколько климатических зон. Для таких рек соотношение C_s/C_v следует подбирать индивидуально по наилучшему соответствию эмпирической и аналитической кривых обеспеченностей.

Если для описания вероятностной структуры гидрологических рядов использовалась модель авторегрессии первого порядка и при этом $C_v > 0,6$, то значение коэффициента вариации корректируется по формуле

$$C_v = \left(a_1 + \frac{a_2}{n}\right) + \left(a_3 + \frac{a_4}{n}\right) \tilde{C}_v + \left(a_5 + \frac{a_6}{n}\right) \tilde{C}_v^2, \quad (7.3)$$

где n – длина ряда; $\tilde{C}_v$ – смещенная оценка C_v (т.е. без учета автокорреляции.); $a_1 \dots a_6$ – коэффициенты, определяемые в зависимости от соотношения C_s / C_v и коэффициента автокорреляции r по табл. 7.3.

Таблица 7.3

Значения параметров a_i в формуле (7.3)

C_s / C_v	r	Коэффициенты					
		a_1	a_2	a_3	a_4	a_5	a_6
2	0	0	0,19	0,99	-0,88	0,01	1,54
	0,3	0	0,22	0,99	-0,41	0,01	1,51
	0,5	0	0,18	0,98	0,41	0,02	1,47
3	0	0	0,69	0,98	-4,34	0,01	6,78
	0,3	0	1,15	1,02	-7,53	-0,04	12,38
	0,5	0	1,75	1,00	-11,79	-0,05	21,13
4	0	0	1,36	1,02	-9,68	-0,05	15,55
	0,3	-0,02	2,61	1,13	-19,85	-0,22	34,15
	0,5	-0,02	3,47	1,18	-29,71	-0,41	58,08

5. Расчет погрешностей параметров распределения. Если расчет параметров распределения производился в рамках модели случайной величины (см. п. 3.2.1, 3.2.2), то оценка их погрешностей производится в соответствии с методикой, изложенной в п. 3.3.

При использовании модели авторегрессии первого порядка относительные погрешности среднего значения и коэффициента вариации определяются по приближенным формулам:

$$\sigma_{\bar{x}} = \frac{C_v}{\sqrt{n}} \sqrt{\frac{1+r}{1-r}} 100\%; \quad (7.4)$$

$$\sigma_{C_v} = \frac{1}{n + 4C_v^2} \sqrt{\frac{n(1 + C_v^2)}{2} \left(1 + \frac{3C_v r^2}{1+r}\right)} 100\%, \quad (7.5)$$

где C_v – выборочный коэффициент вариации; r – коэффициент автокорреляции; n – длина выборки.

Следует иметь ввиду, что формулы (7.4), (7.5) разработаны для условий: $r \leq 0,5$ и $C_s / C_v = 2$.

Расчет считается надежным, если ошибки среднего значения и коэффициента вариации не превышают 5 – 10 %. При расчетах максимального стока удовлетворительной считается ошибка до 20 %.

Если ошибки превышают допустимые значения, то необходимо удлинить ряд, используя более продолжительные ряды по рекам-аналогам и аппарат линейной корреляции (см. раздел 5).

6. Построение эмпирической и аналитической кривых обеспеченностей. Для построения эмпирической кривой обеспеченностей значения исходного ряда ранжируются (располагаются в убывающем порядке). Затем для каждого члена ранжированного ряда (x_m) рассчитывается эмпирическая обеспеченность по формуле (3.6).

Эмпирическая кривая обеспеченностей строится на специальной клетчатке вероятностей в виде последовательности точек с координатами (p_m, x_m). Кривую обеспеченностей можно строить как в абсолютных значениях гидрологической характеристики, так и в модульных коэффициентах (см. пример 3.1).

Для объективного сглаживания эмпирической кривой обеспеченностей и экстраполяции ее в область больших и малых значений используются специальные аналитические кривые. В России для этой цели используется кривая Крицкого – Менкеля (трехпараметрическое гамма-распределение).

Построение эмпирической и аналитической кривых обеспеченностей производится для визуальной проверки их соответствия. Более строго такую проверку можно провести с использованием критериев согласия (см. п. 4.7).

Построение эмпирической и аналитической кривых обеспеченностей позволяют также выявить точки, резко отклоняющиеся от аналитической кривой в области больших или малых значений. В этой ситуации необходимо выяснить, относится ли данная точка к той же генеральной совокупности, что и остальные члены выборки или нет.

В первом случае отклонение можно объяснить тем, что данный гидрологический ряд содержит значение очень редкой повторяемости, во втором – ряд следует признать неоднородным.

Для анализа таких ситуаций в СП 33-101-2003 включены два дополнительных критерия однородности – критерий Диксона и критерий Смирнова – Граббса, обобщенных на случай автокоррелированных рядов.

При использовании указанных критериев исходный ряд ранжируется в возрастающем порядке: $x_1 < x_2 < \dots < x_n$.

Статистики критерия Диксона для максимального и минимального членов выборки имеют вид:

$$D_{\max}^* = \frac{x_n - x_{n-1}}{x_n - x_1}; \quad (7.6)$$

$$D_{\min}^* = \frac{x_1 - x_2}{x_1 - x_n}. \quad (7.7)$$

Гипотеза об однородности ряда по критерию Диксона не опровергается, если

$$D^* < D_\alpha, \quad (7.8)$$

где D^* – эмпирическое значение статистики Диксона; D_α – теоретическое значение статистики Диксона, определяемое в зависимости от уровня значимости α %, коэффициента автокорреляции, коэффициента асимметрии и длины выборки по приложению 10.

Статистики критерия Смирнова-Граббса для максимального и минимального членов выборки имеют вид:

$$G_{\max}^* = \frac{x_n - \bar{x}}{S_x}; \quad (7.9)$$

$$G_{\min}^* = \frac{\bar{x} - x_1}{S_x}, \quad (7.10)$$

где $\bar{x}$ – среднее значение; S_x – выборочное среднеквадратическое отклонение.

Гипотеза об однородности ряда по критерию Смирнова-Граббса не опровергается, если

$$G^* < G_\alpha, \quad (7.11)$$

где G^* – эмпирическое значение статистики Смирнова – Граббса; G_α – теоретическое значение статистики Смирнова – Граббса, определяемое в зависимости от уровня значимости α %, коэффициента

автокорреляции, коэффициента асимметрии и длины выборки по приложению 11.

Если аналитическая кривая хорошо согласуется с эмпирическими точками, она используется для определения расходов расчетной обеспеченности.

Следует иметь в виду, что изложенная схема является достаточно общей, и расчет каждой гидрометеорологической характеристики имеет свои особенности. Так, например, при расчетах максимального стока для расходов, обеспеченностью $p = 0,01 \%$, рекомендуется вводить гарантийную поправку. А при расчетах минимального стока нередко приходится пользоваться сглаженными эмпирическими кривыми из-за невозможности подобрать аналитическую кривую обеспеченностей. Эти и другие вопросы инженерной гидрологической практики рассматриваются в курсе «Гидрологические расчеты».

СПИСОК ЛИТЕРАТУРЫ

1. Бендат Дж., Пирсол А. Применение корреляционного и спектрального анализа. – М.: Мир, 1983. – 312 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. – М.: Мир, 1974. – 406 с.
3. Боровиков В.П. Популярное введение в программу STATISTICA. – М.: «Компьютер Пресс», 1998. – 267 с.
4. Боровиков В.П., Боровиков И.П. STATISTICA. Статистический анализ и обработка данных в среде Windows. – М.: Информационно-издательский дом «Филин», 1998. – 608 с.
5. Боровиков В.П., Ивченко В.П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере: Учебное пособие – М.: Финансы и статистика, 1999. – 384 с.
6. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. – СПб.: «Питер», 2003. – 688 с.
7. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. – М.: Наука, 1988. – 480 с.
8. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения. – М.: Наука, 1991. – 384 с.
9. Закс Л. Статистическое оценивание. М.: «Статистика», 1976. – 600 с.
10. Казакевич Д.И. Основы теории случайных функций в задачах гидрометеорологии. – Л.: Гидрометиздат, 1989. – 230 с.
11. Кайсл Ч. Анализ временных рядов гидрологических данных. Л.: Гидрометиздат, 1972. – 138 с.
12. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
13. Международное руководство по методам расчета основных гидрологических характеристик. – Л.: Гидрометеоздат, 1984. – 247 с.
14. Митропольский А.К. Техника статистических вычислений. – М.: Наука, 1971. – 576 с.
15. Определение основных расчетных гидрологических характеристик. СП 33-101-2003. – М.: Стройиздат, 2004. – 72 с.
16. Пособие по определению расчетных гидрологических характеристик. – Л.: Гидрометеоздат, 1984. – 447 с.
17. Расчеты паводочного стока. Методика расчета на основе мирового опыта / Под ред. А.А. Соколова, С.Е. Рантца и др. – Л.: Гидрометеоздат, 1978. – 304 с.
18. Рождественский А.В., Ежов А.В., Сахарюк А.В. Оценка точности гидрологических расчетов. – Л.: Гидрометеоздат, 1990. – 273 с.
19. Рождественский А.В. Оценка точности кривых распределения гидрологических характеристик. – Л.: Гидрометеоздат, 1977. – 268 с.
20. Рождественский А.В., Чеботарев А.И. Статистические методы в гидрологии. – Л.: Гидрометеоздат, 1974. – 424 с.
21. Сванидзе Г.Г. Математическое моделирование гидрологических рядов. – Л.: Гидрометеоздат, 1977. – 296 с.
22. Справочник по прикладной статистике. Т. 1, 2 / Под ред. Э. Ллойда, У. Линдермана. – М.: Финансы и статистика, 1990.

23. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003, – 544 с.
24. *Химмельблау Д.* Анализ процессов статистическими методами. – М.: Наука, 1973. – 958 с.
25. *Шелутко В.А.* Численные методы в гидрологии. – Л.: Гидрометеиздат, 1991. – 238 с.

ПРИЛОЖЕНИЯ

Нормированные ординаты кривой обеспеченностей Пирсона III типа,

$$t_p = f(C_s, p); t_p = (x_p - \bar{x}) / \sigma = (k_p - 1) / C_v$$

C_s	Обеспеченность, $p\%$																	S
	0,01	0,1	1,0	5,0	10	20	25	30	50	70	75	80	90	95	97	99	99,9	
-1,8	1,11	1,11	1,09	1,02	0,94	0,80	0,72	0,64	0,28	-0,24	-0,42	-0,64	-1,32	-1,99	-2,46	-3,50	-5,64	-0,51
-1,6	1,26	1,24	1,20	1,10	0,99	0,81	0,73	0,64	0,25	-0,28	-0,46	-0,68	-1,33	-1,97	-2,42	-3,39	-5,37	-0,45
-1,4	1,41	1,39	1,32	1,17	1,04	0,83	0,73	0,64	0,22	-0,31	-0,49	-0,71	-1,34	-1,95	-2,37	-3,27	-5,09	-0,39
-1,2	1,68	1,58	1,45	1,24	1,08	0,84	0,74	0,63	0,19	-0,35	-0,52	-0,73	-1,34	-1,92	-2,31	-3,15	-4,81	-0,34
-1,0	1,92	1,79	1,59	1,32	1,13	0,85	0,73	0,62	0,16	-0,38	-0,55	-0,76	-1,34	-1,88	-2,25	-3,02	-4,53	-0,27
-0,8	2,23	2,02	1,74	1,38	1,17	0,86	0,73	0,60	0,13	-0,41	-0,58	-0,79	-1,34	-1,84	-2,18	-2,89	-4,24	-0,22
-0,6	2,57	2,27	1,88	1,45	1,20	0,85	0,72	0,59	0,10	-0,44	-0,61	-0,80	-1,33	-1,80	-2,12	-2,75	-3,96	-0,17
-0,4	2,98	2,54	2,03	1,52	1,23	0,85	0,71	0,57	0,07	-0,47	-0,63	-0,82	-1,32	-1,75	-2,04	-2,61	-3,66	-0,11
-0,2	3,37	2,81	2,18	1,58	1,26	0,85	0,69	0,55	0,03	-0,50	-0,65	-0,83	-1,30	-1,70	-1,96	-2,47	-3,38	-0,05
0,0	3,72	3,09	2,33	1,64	1,28	0,84	0,67	0,52	0,00	-0,52	-0,67	-0,84	-1,28	-1,64	-1,88	-2,33	-3,09	0,00
0,2	4,16	3,38	2,47	1,70	1,30	0,83	0,65	0,50	-0,03	-0,55	-0,69	-0,85	-1,26	-1,58	-1,79	-2,81	-2,81	0,06
0,4	4,61	3,66	2,61	1,75	1,32	0,82	0,63	0,47	-0,07	-0,57	-0,71	-0,85	-1,23	-1,52	-1,70	-2,03	-2,54	0,11
0,6	5,05	3,96	2,75	1,80	1,33	0,80	0,61	0,44	-0,10	-0,59	-0,72	-0,85	-1,20	-1,45	-1,61	-1,88	-2,27	0,17
0,8	5,50	4,24	2,89	1,84	1,34	0,78	0,58	0,41	-0,13	-0,60	-0,73	-0,86	-1,17	-1,38	-1,52	-1,74	-2,02	0,22
1,0	5,96	4,53	3,02	1,88	1,34	0,76	0,55	0,38	-0,16	-0,62	-0,73	-0,85	-1,13	-1,32	-1,42	-1,59	-1,79	0,28
1,2	6,41	4,81	3,15	1,92	1,34	0,73	0,52	0,35	-0,19	-0,63	-0,74	-0,84	-1,08	-1,24	-1,33	-1,45	-1,58	0,34
1,4	6,87	5,09	3,27	1,95	1,34	0,71	0,49	0,31	-0,22	-0,64	-0,73	-0,83	-1,04	-1,17	-1,23	-1,32	-1,39	0,39
1,6	7,31	5,37	3,39	1,97	1,33	0,68	0,46	0,28	-0,25	-0,64	-0,73	-0,81	-0,99	-1,10	-1,14	-1,20	-1,24	0,45
1,8	7,76	5,64	3,50	1,99	1,32	0,64	0,42	0,24	-0,28	-0,64	-0,72	-0,80	-0,94	-1,02	-1,06	-1,09	-1,11	0,51
2,0	8,21	5,91	3,60	2,00	1,30	0,61	0,39	0,20	-0,31	-0,64	-0,71	-0,78	-0,90	-0,95	-0,97	-0,99	-1,00	0,57
2,2	8,63	6,14	3,68	2,02	1,27	0,57	0,35	0,16	-0,33	-0,64	-0,69	-0,75	-0,84	-0,88	-0,90	-0,90	-0,91	0,62
2,4	9,00	6,37	3,78	2,00	1,25	0,52	0,29	0,12	-0,35	-0,62	-0,67	-0,72	-0,79	-0,82	-0,83	-0,83	-0,83	0,67
2,6	9,39	6,54	3,86	2,00	1,21	0,48	0,25	0,085	-0,37	-0,61	-0,66	-0,70	-0,75	-0,76	-0,77	-0,77	-0,77	0,72
2,8	9,77	6,86	3,96	2,00	1,18	0,44	0,22	0,057	-0,39	-0,60	-0,64	-0,67	-0,70	-0,71	-0,71	-0,72	-0,72	0,76
3,0	10,16	7,10	4,05	1,97	1,13	0,39	0,19	0,027	-0,40	-0,59	-0,62	-0,64	-0,66	-0,66	-0,67	-0,67	-0,67	0,80

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 0,5$**

$p \%$	Коэффициент вариации, C_v							
	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7
0,01	1,00	1,38	1,76	2,13	2,47	2,74	2,91	2,89
0,1	1,00	1,31	1,63	1,95	2,25	2,50	2,69	2,74
0,3	1,00	1,28	1,56	1,85	2,12	2,36	2,55	2,64
0,5	1,00	1,26	1,53	1,79	2,05	2,28	2,48	2,59
1,0	1,00	1,24	1,48	1,72	1,95	2,17	2,37	2,50
3	1,00	1,19	1,38	1,58	1,78	1,97	2,16	2,33
5	1,00	1,17	1,33	1,51	1,68	1,86	2,03	2,22
10	1,00	1,13	1,26	1,39	1,53	1,67	1,83	2,01
20	1,00	1,08	1,17	1,25	1,35	1,44	1,56	1,70
25	1,00	1,07	1,13	1,20	1,27	1,35	1,45	1,56
30	1,00	1,05	1,10	1,16	1,21	1,27	1,34	1,42
40	1,00	1,02	1,05	1,07	1,10	1,12	1,15	1,16
50	1,00	0,999	0,997	0,993	0,988	0,980	0,962	0,920
60	1,00	0,974	0,946	0,915	0,881	0,839	0,780	0,690
70	1,00	0,947	0,882	0,834	0,769	0,693	0,596	0,476
75	1,00	0,932	0,862	0,789	0,709	0,615	0,503	0,376
80	1,00	0,915	0,829	0,740	0,643	0,533	0,409	0,282
90	1,00	0,872	0,744	0,615	0,480	0,343	0,215	0,115
95	1,00	0,837	0,676	0,517	0,362	0,221	0,113	0,047
97	1,00	0,814	0,633	0,458	0,295	0,160	0,070	0,024
99	1,00	0,772	0,554	0,354	0,189	0,080	0,025	0,006
99,5	1,00	0,748	0,511	0,302	0,144	0,051	0,013	0,002
99,7	1,00	0,732	0,482	0,269	0,117	0,037	0,008	0,001
99,9	1,00	0,700	0,428	0,210	0,076	0,019	0,003	0,000

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 1,0$**

$p \%$	● Коэффициент вариации, C_v									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,01	1,38	1,81	2,26	2,70	3,15	3,57	3,95	4,31	4,64	4,92
0,1	1,32	1,67	2,03	2,40	2,77	3,13	3,48	3,82	4,13	4,42
0,3	1,28	1,59	1,91	2,23	2,56	2,89	3,21	3,53	3,84	4,14
0,5	1,27	1,55	1,84	2,15	2,46	2,77	3,08	3,38	3,69	3,99
1,0	1,24	1,49	1,76	2,03	2,30	2,59	2,88	3,16	3,46	3,75
3	1,19	1,39	1,60	1,82	2,04	2,27	2,50	2,75	3,01	3,29
5	1,17	1,34	1,52	1,70	1,90	2,10	2,30	2,53	2,76	3,02
10	1,13	1,26	1,40	1,54	1,68	1,83	1,99	2,16	2,35	2,55
20	1,08	1,17	1,25	1,34	1,42	1,51	1,60	1,70	1,80	1,90
25	1,07	1,13	1,20	1,26	1,33	1,39	1,46	1,52	1,59	1,64
30	1,05	1,10	1,15	1,20	1,24	1,29	1,33	1,37	1,39	1,40
40	1,02	1,04	1,06	1,08	1,09	1,10	1,10	1,08	1,05	0,995
50	0,998	0,993	0,985	0,972	0,954	0,928	0,891	0,836	0,76	0,665
60	0,973	0,943	0,909	0,870	0,824	0,768	0,698	0,613	0,512	0,406
70	0,946	0,890	0,830	0,764	0,692	0,609	0,515	0,413	0,309	0,215
75	0,932	0,861	0,787	0,708	0,622	0,528	0,426	0,321	0,224	0,144
80	0,915	0,829	0,740	0,648	0,549	0,445	0,338	0,237	0,151	0,088
90	0,873	0,748	0,623	0,500	0,378	0,264	0,165	0,092	0,045	0,019
95	0,838	0,683	0,533	0,392	0,263	0,157	0,081	0,036	0,013	0,004
97	0,816	0,642	0,478	0,329	0,202	0,107	0,048	0,018	0,005	0,001
99	0,775	0,568	0,383	0,229	0,115	0,047	0,015	0,004	0,001	0,000
99,5	0,752	0,528	0,335	0,182	0,081	0,028	0,008	0,002	0,000	0,000
99,7	0,737	0,502	0,303	0,154	0,062	0,019	0,004	0,001	0,000	0,000
99,9	0,707	0,451	0,247	0,108	0,036	0,008	0,001	0,000	0,000	0,000

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 1,0$**

$p \%$	Коэффициент вариации, C_v									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0,01	5,16	5,34	5,46	5,58	5,68	5,76	5,82	5,88	5,92	5,96
0,1	4,69	4,92	5,06	5,18	5,29	5,37	5,44	5,49	5,54	5,58
0,3	4,44	4,74	4,92	5,06	5,16	5,24	5,31	5,36	5,42	5,46
0,5	4,29	4,58	4,75	4,91	5,02	5,11	5,18	5,24	5,28	5,32
1,0	4,06	4,36	4,55	4,72	4,84	4,94	5,00	5,07	5,12	5,16
3	3,59	3,92	4,14	4,33	4,46	4,58	4,68	4,76	4,84	4,92
5	3,31	3,63	3,84	4,02	4,16	4,28	4,40	4,50	4,60	4,69
10	2,78	3,03	3,26	3,46	3,64	3,81	3,94	4,05	4,15	4,25
20	2,00	2,10	2,20	2,32	2,44	2,56	2,67	2,80	2,92	3,03
25	1,68	1,69	1,70	1,70	1,68	1,66	1,61	1,56	1,51	1,46
30	1,39	1,34	1,26	1,17	1,07	0,96	0,84	0,72	0,60	0,45
40	0,916	0,808	0,72	0,60	0,50	0,38	0,28	0,20	0,11	0,04
50	0,559	0,446	0,34	0,26	0,20	0,15	0,105	0,07	0,04	0,01
60	0,306	0,216	0,19	0,13	0,10	0,075	0,055	0,035	0,015	0,00
70	0,141	0,085	0,06	0,045	0,035	0,025	0,015	0,010	0,000	0,00
75	0,086	0,046	0,025	0,020	0,010	0,005	0,000	0,000	0,000	0,00
80	0,047	0,023	0,015	0,005	0,000	0,000	0,000	0,000	0,000	0,00
90	0,007	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00
95	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00
97	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00
99	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00
99,5	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00
99,7	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00
99,9	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 2,0$**

p %	0,28 Коэффициент вариации, C_v									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,01	1,42	1,92	2,52	3,20	3,98	4,85	5,81	6,85	7,98	9,21
0,1	1,34	1,73	2,19	2,70	3,27	3,87	4,56	5,30	6,08	6,91
0,3	1,30	1,64	2,02	2,45	2,91	3,42	3,96	4,55	5,16	5,81
0,5	1,28	1,59	1,94	2,32	2,74	3,20	3,68	4,19	4,74	5,30
1,0	1,25	1,52	1,82	2,16	2,51	2,89	3,29	3,71	4,15	4,60
3	1,20	1,41	1,64	1,87	2,13	2,39	2,66	2,94	3,21	3,51
5	1,17	1,35	1,54	1,74	1,94	2,15	2,36	2,57	2,78	3,00
10	1,13	1,26	1,40	1,54	1,67	1,80	1,94	2,06	2,19	2,30
20	1,08	1,16	1,24	1,31	1,38	1,44	1,50	1,54	1,58	1,61
25	1,06	1,13	1,18	1,23	1,28	1,31	1,34	1,37	1,38	1,39
30	1,05	1,09	1,13	1,16	1,19	1,21	1,22	1,22	1,22	1,20
40	1,02	1,04	1,05	1,05	1,04	1,03	1,01	0,984	0,955	0,916
50	0,997	0,986	0,970	0,948	0,918	0,886	0,846	0,800	0,748	0,693
60	0,972	0,938	0,898	0,852	0,803	0,748	0,692	0,632	0,568	0,511
70	0,945	0,886	0,823	0,760	0,691	0,622	0,552	0,488	0,424	0,357
75	0,931	0,858	0,784	0,708	0,634	0,556	0,489	0,416	0,352	0,288
80	0,915	0,830	0,745	0,656	0,574	0,496	0,419	0,352	0,280	0,223
90	0,873	0,754	0,640	0,532	0,436	0,352	0,272	0,208	0,154	0,105
95	0,842	0,696	0,565	0,448	0,342	0,256	0,181	0,120	0,082	0,051
97	0,821	0,660	0,517	0,392	0,288	0,202	0,139	0,088	0,046	0,030
99	0,782	0,594	0,436	0,304	0,206	0,130	0,076	0,040	0,019	0,010
99,5	0,761	0,560	0,394	0,269	0,166	0,099	0,054	0,027	0,012	0,005
99,7	0,748	0,537	0,374	0,240	0,144	0,082	0,042	0,019	0,008	0,003
99,9	0,719	0,492	0,319	0,192	0,107	0,052	0,027	0,008	0,004	0,001

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 2,0$**

p %	Коэффициент вариации, C_v									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0,01	10,5	11,8	13,2	14,7	16,4	18,2	20,2	22,2	24,4	26,6
0,1	7,75	8,65	9,60	10,6	11,6	12,5	13,5	14,6	15,8	17,0
0,3	6,47	7,10	7,98	8,70	9,50	10,5	11,0	11,9	12,7	13,6
0,5	5,90	6,50	7,13	7,80	8,42	9,0	9,5	10,1	10,8	11,4
1,0	5,05	5,53	6,02	6,55	7,08	7,5	8,0	8,6	9,2	9,8
3	3,80	4,12	4,42	4,71	4,98	5,2	5,5	5,8	6,2	6,5
5	3,22	3,40	3,60	3,80	3,96	4,0	4,3	4,5	4,7	5,0
10	2,40	2,50	2,57	2,64	2,70	2,7	2,6	2,6	2,6	2,6
20	1,62	1,63	1,62	1,61	1,59	1,6	1,6	1,56	1,5	1,5
25	1,39	1,35	1,33	1,31	1,28	1,26	1,24	1,22	1,2	1,18
30	1,18	1,14	1,11	1,08	1,04	1,015	0,98	0,95	0,92	0,89
40	0,870	0,830	0,770	0,725	0,670	0,625	0,58	0,53	0,48	0,44
50	0,640	0,580	0,520	0,460	0,405	0,355	0,310	0,265	0,23	0,20
60	0,450	0,390	0,334	0,283	0,234	0,190	0,160	0,130	0,105	0,085
70	0,300	0,250	0,203	0,155	0,120	0,090	0,070	0,060	0,05	0,045
75	0,241	0,193	0,146	0,106	0,077	0,060	0,050	0,040	0,03	0,025
80	0,175	0,130	0,094	0,065	0,046	0,035	0,027	0,020	0,015	0,010
90	0,074	0,049	0,030	0,016	0,009	0,005	0,004	0,003	0,002	0,001
95	0,030	0,016	0,009	0,004	0,002	0,001	0,000	0,000	0,000	0,000
97	0,016	0,008	0,004	0,002	0,001	0,000	0,000	0,000	0,000	0,000
99	0,005	0,002	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000
99,5	0,002	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
99,7	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
99,9	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 3,0$**

<i>p</i> %	Коэффициент вариации, C_v									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,01	1,46	2,05	2,83	3,80	4,94	6,26	7,70	9,30	11,0	12,8
0,1	1,36	1,81	2,35	3,01	3,74	4,56	5,44	6,38	7,37	8,41
0,3	1,31	1,69	2,12	2,65	3,21	3,82	4,48	5,17	5,88	6,61
0,5	1,28	1,63	2,03	2,48	2,97	3,50	4,06	4,64	5,24	5,84
1,0	1,25	1,55	1,90	2,26	2,66	3,07	3,50	3,96	4,41	4,87
3	1,20	1,42	1,66	1,91	2,17	2,43	2,69	2,95	3,21	3,47
5	1,17	1,36	1,55	1,75	1,95	2,14	2,34	2,52	2,70	2,88
10	1,13	1,26	1,40	1,52	1,65	1,76	1,87	1,97	2,06	2,15
20	1,08	1,16	1,23	1,29	1,34	1,38	1,42	1,45	1,47	1,49
25	1,07	1,12	1,17	1,21	1,24	1,26	1,28	1,28	1,29	1,29
30	1,05	1,09	1,12	1,14	1,15	1,16	1,16	1,15	1,14	1,13
40	1,02	1,03	1,03	1,03	1,01	0,995	0,972	0,946	0,915	0,883
50	0,997	0,981	0,959	0,930	0,898	0,862	0,823	0,783	0,741	0,699
60	0,972	0,933	0,890	0,843	0,794	0,745	0,695	0,646	0,597	0,549
70	0,945	0,884	0,822	0,758	0,696	0,636	0,578	0,523	0,471	0,422
75	0,931	0,858	0,786	0,715	0,647	0,583	0,522	0,465	0,412	0,363
80	0,915	0,830	0,748	0,669	0,596	0,528	0,465	0,407	0,354	0,306
90	0,876	0,761	0,656	0,563	0,479	0,406	0,341	0,284	0,235	0,193
95	0,844	0,708	0,588	0,487	0,400	0,326	0,263	0,210	0,166	0,129
97	0,825	0,675	0,548	0,443	0,355	0,282	0,221	0,171	0,131	0,099
99	0,786	0,618	0,484	0,369	0,283	0,213	0,158	0,116	0,083	0,058
99,5	0,769	0,588	0,446	0,334	0,249	0,182	0,131	0,092	0,064	0,043
99,7	0,756	0,568	0,422	0,312	0,228	0,163	0,114	0,079	0,053	0,034
99,9	0,732	0,531	0,381	0,273	0,192	0,131	0,088	0,057	0,036	0,022

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 3,0$**

$p \%$	Коэффициент вариации, C_v									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0,01	14,8	16,8	19,0	21,2	23,5	25,9	28,4	31,0	33,7	36,5
0,1	9,49	10,6	11,8	13,0	14,2	15,4	16,7	18,0	19,4	20,8
0,3	7,37	8,15	8,94	9,75	10,6	11,4	12,3	13,1	14,0	14,8
0,5	6,47	7,10	7,75	8,41	9,07	9,74	10,4	11,1	11,8	12,4
1,0	5,33	5,79	6,26	6,74	7,21	7,68	8,14	8,61	9,07	9,53
3	3,73	3,98	4,20	4,44	4,67	4,89	5,10	5,31	5,51	5,70
5	3,05	3,22	3,37	3,52	3,66	3,80	3,92	4,04	4,15	4,26
10	2,23	2,30	2,36	2,42	2,47	2,51	2,55	2,58	2,60	2,62
20	1,50	1,50	1,50	1,49	1,48	1,46	1,45	1,42	1,40	1,37
25	1,28	1,27	1,25	1,23	1,20	1,18	1,15	1,12	1,08	1,05
30	1,11	1,08	1,06	1,03	0,997	0,964	0,929	0,892	0,855	0,818
40	0,848	0,812	0,775	0,736	0,697	0,659	0,620	0,581	0,544	0,507
50	0,656	0,614	0,572	0,531	0,491	0,452	0,415	0,379	0,345	0,313
60	0,503	0,459	0,417	0,377	0,339	0,304	0,271	0,240	0,212	0,186
70	0,375	0,333	0,293	0,257	0,224	0,194	0,166	0,142	0,121	0,102
75	0,318	0,277	0,239	0,206	0,176	0,149	0,125	0,105	0,087	0,071
80	0,263	0,224	0,190	0,160	0,133	0,110	0,090	0,073	0,059	0,047
90	0,156	0,126	0,100	0,078	0,061	0,047	0,035	0,026	0,019	0,014
95	0,100	0,076	0,057	0,042	0,030	0,022	0,015	0,010	0,007	0,004
97	0,073	0,054	0,038	0,027	0,018	0,012	0,008	0,005	0,003	0,002
99	0,040	0,027	0,017	0,011	0,007	0,004	0,002	0,001	0,001	0,000
99,5	0,028	0,018	0,011	0,006	0,004	0,002	0,001	0,000	0,000	0,000
99,7	0,022	0,014	0,008	0,004	0,002	0,001	0,000	0,000	0,000	0,000
99,9	0,013	0,007	0,004	0,002	0,001	0,000	0,000	0,000	0,000	0,000

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 4,0$**

<i>p</i> %	Коэффициент вариации, <i>C_v</i>									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0,01	1,50	2,18	3,17	4,43	5,91	7,58	9,41	11,4	13,4	15,5
0,1	1,38	1,88	2,53	3,29	4,15	5,07	6,05	7,08	8,15	9,26
0,3	1,32	1,74	2,24	2,82	3,44	4,09	4,79	5,50	6,22	6,96
0,5	1,29	1,67	2,12	2,61	3,13	3,68	4,26	4,85	5,43	6,03
1,0	1,25	1,58	1,94	2,31	2,75	3,17	3,59	4,03	4,47	4,91
3	1,20	1,44	1,68	1,93	2,18	2,43	2,68	2,92	3,16	3,39
5	1,17	1,36	1,56	1,75	1,94	2,12	2,29	2,46	2,62	2,78
10	1,13	1,26	1,39	1,51	1,62	1,72	1,81	1,90	1,98	2,05
20	1,08	1,15	1,22	1,27	1,31	1,34	1,37	1,40	1,41	1,42
25	1,07	1,12	1,16	1,19	1,21	1,23	1,24	1,24	1,24	1,24
30	1,05	1,08	1,11	1,12	1,13	1,13	1,13	1,12	1,11	1,10
40	1,02	1,02	1,02	1,01	0,996	0,976	0,954	0,929	0,902	0,873
50	0,997	0,976	0,950	0,920	0,888	0,853	0,818	0,781	0,744	0,707
60	0,972	0,929	0,885	0,839	0,793	0,747	0,702	0,658	0,614	0,572
70	0,945	0,883	0,821	0,761	0,704	0,649	0,597	0,548	0,501	0,457
75	0,931	0,858	0,788	0,722	0,660	0,601	0,546	0,495	0,448	0,403
80	0,915	0,832	0,754	0,681	0,614	0,553	0,496	0,443	0,395	0,351
90	0,877	0,767	0,671	0,586	0,511	0,444	0,384	0,331	0,284	0,243
95	0,846	0,719	0,611	0,519	0,440	0,372	0,312	0,261	0,217	0,180
97	0,829	0,690	0,576	0,481	0,400	0,332	0,274	0,224	0,182	0,147
99	0,790	0,638	0,516	0,417	0,336	0,269	0,214	0,168	0,132	0,102
99,5	0,776	0,612	0,485	0,386	0,305	0,239	0,186	0,144	0,110	0,083
99,7	0,762	0,594	0,466	0,366	0,286	0,221	0,170	0,129	0,097	0,072
99,9	0,742	0,561	0,430	0,331	0,252	0,189	0,141	0,104	0,075	0,054

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 4,0$**

$p \%$	Коэффициент вариации, C_v									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0,01	17,9	20,3	22,8	25,4	28,0	30,8	33,6	36,5	39,4	42,4
0,1	10,4	11,6	12,8	14,0	15,3	16,6	17,9	19,2	20,6	21,9
0,3	7,73	8,53	9,31	10,1	10,9	11,7	12,5	13,3	14,2	15,0
0,5	6,65	7,29	7,91	8,53	9,16	9,79	10,4	11,0	11,7	12,3
1,0	5,34	5,79	6,22	6,66	7,09	7,52	7,95	8,73	8,78	9,19
3	3,62	3,83	4,04	4,25	4,45	4,64	4,83	5,01	5,18	5,34
5	2,93	3,07	3,21	3,34	3,46	3,57	3,68	3,78	3,87	3,96
10	2,12	2,18	2,24	2,28	2,32	2,36	2,39	2,42	2,44	2,45
20	1,43	1,44	1,43	1,43	1,42	1,41	1,39	1,38	1,36	1,33
25	1,23	1,22	1,21	1,19	1,17	1,15	1,13	1,10	1,08	1,05
30	1,08	1,06	1,04	1,01	0,985	0,958	0,929	0,90	0,871	0,841
40	0,843	0,812	0,781	0,748	0,716	0,684	0,652	0,62	0,588	0,558
50	0,670	0,634	0,598	0,562	0,529	0,495	0,464	0,433	0,403	0,375
60	0,532	0,494	0,457	0,421	0,388	0,356	0,327	0,299	0,273	0,249
70	0,416	0,377	0,341	0,308	0,227	0,248	0,223	0,199	0,177	0,157
75	0,362	0,325	0,290	0,258	0,230	0,203	0,179	0,158	0,139	0,121
80	0,311	0,274	0,242	0,212	0,185	0,162	0,140	0,122	0,105	0,090
90	0,207	0,176	0,148	0,125	0,104	0,087	0,072	0,060	0,049	0,040
95	0,148	0,121	0,098	0,080	0,064	0,051	0,041	0,032	0,025	0,019
97	0,119	0,095	0,075	0,059	0,046	0,036	0,028	0,021	0,016	0,012
99	0,078	0,060	0,045	0,034	0,025	0,018	0,013	0,009	0,006	0,004
99,5	0,062	0,046	0,034	0,024	0,017	0,012	0,008	0,006	0,004	0,003
99,7	0,053	0,038	0,027	0,019	0,013	0,009	0,006	0,004	0,003	0,002
99,9	0,038	0,026	0,018	0,012	0,008	0,005	0,003	0,002	0,001	0,001

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 5,0$**

<i>p</i> %	Коэффициент вариации, C_v								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,01	1,54	2,34	3,43	4,91	6,65	8,70	10,70	12,71	15,05
0,1	1,40	1,95	2,66	3,51	4,44	5,40	6,43	7,54	8,64
0,3	1,34	1,78	2,31	2,92	3,52	4,22	4,91	5,69	6,41
0,5	1,31	1,70	2,16	2,69	3,21	3,77	4,34	4,94	5,52
1,0	1,27	1,61	1,98	2,38	2,79	3,21	3,65	4,06	4,50
3	1,20	1,44	1,67	1,93	2,17	2,42	2,62	2,88	3,10
5	1,17	1,36	1,55	1,74	1,90	2,08	2,22	2,41	2,54
10	1,13	1,26	1,37	1,49	1,60	1,70	1,79	1,86	1,94
20	1,08	1,15	1,21	1,25	1,30	1,32	1,34	1,36	1,36
25	1,06	1,11	1,15	1,17	1,20	1,20	1,20	1,22	1,22
30	1,05	1,08	1,09	1,10	1,10	1,11	1,10	1,10	1,09
40	1,02	1,02	1,01	1,00	0,98	0,97	0,94	0,92	0,90
50	0,99	0,97	0,94	0,92	0,88	0,85	0,82	0,78	0,75
60	0,97	0,93	0,88	0,84	0,79	0,75	0,71	0,67	0,63
70	0,94	0,88	0,82	0,77	0,71	0,66	0,61	0,56	0,52
75	0,93	0,86	0,79	0,73	0,67	0,62	0,56	0,51	0,47
80	0,91	0,83	0,75	0,69	0,63	0,57	0,52	0,47	0,42
90	0,88	0,77	0,68	0,61	0,53	0,47	0,41	0,36	0,32
95	0,84	0,73	0,63	0,55	0,47	0,40	0,34	0,29	0,25
97	0,82	0,70	0,60	0,51	0,43	0,36	0,31	0,26	0,22
99	0,78	0,66	0,55	0,45	0,37	0,31	0,25	0,20	0,16
99,5	0,76	0,63	0,52	0,42	0,34	0,28	0,23	0,18	0,14
99,7	0,75	0,62	0,51	0,41	0,32	0,26	0,21	0,16	0,12
99,9	0,73	0,59	0,47	0,37	0,29	0,23	0,18	0,14	0,10

**Ординаты кривой обеспеченностей Крицкого – Менкеля
(трехпараметрического гамма-распределения)
в модульных коэффициентах, $C_s/C_v = 6,0$**

$p \%$	Коэффициент вариации, C_v								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,01	1,60	2,48	3,75	5,48	7,30	9,39	11,55	13,80	16,40
0,1	1,41	2,02	2,80	3,68	4,58	5,54	6,57	7,63	8,79
0,3	1,35	1,83	2,38	2,98	3,64	4,31	5,00	5,66	6,38
0,5	1,32	1,74	2,22	2,73	3,26	3,82	4,38	4,93	5,51
1,0	1,29	1,63	2,01	2,40	2,81	3,22	3,63	4,03	4,44
3	1,21	1,45	1,68	1,92	2,14	2,38	2,60	2,82	3,04
5	1,18	1,37	1,55	1,73	1,89	2,05	2,20	2,36	2,51
10	1,14	1,26	1,37	1,47	1,56	1,66	1,73	1,82	1,90
20	1,08	1,14	1,19	1,23	1,27	1,30	1,32	1,34	1,36
25	1,07	1,10	1,13	1,16	1,18	1,19	1,20	1,21	1,20
30	1,04	1,07	1,08	1,10	1,10	1,10	1,10	1,09	1,08
40	1,02	1,02	1,01	0,99	0,98	0,96	0,94	0,92	0,89
50	0,99	0,97	0,94	0,91	0,88	0,85	0,82	0,79	0,75
60	0,96	0,92	0,88	0,84	0,80	0,76	0,72	0,68	0,64
70	0,94	0,88	0,83	0,77	0,72	0,67	0,63	0,58	0,54
75	0,93	0,86	0,80	0,74	0,68	0,63	0,58	0,53	0,49
80	0,91	0,84	0,77	0,70	0,64	0,58	0,53	0,48	0,44
90	0,88	0,78	0,70	0,62	0,55	0,49	0,43	0,38	0,33
95	0,85	0,74	0,65	0,56	0,49	0,43	0,37	0,32	0,27
97	0,83	0,72	0,62	0,53	0,46	0,39	0,33	0,28	0,24
99	0,80	0,67	0,57	0,48	0,40	0,33	0,28	0,23	0,19
99,5	0,78	0,65	0,55	0,45	0,37	0,31	0,25	0,20	0,17
99,7	0,76	0,64	0,53	0,43	0,36	0,29	0,24	0,19	0,15
99,9	0,75	0,61	0,50	0,40	0,33	0,26	0,21	0,16	0,12

**Ординаты кривых обеспеченностей H -распределения
(А.В. Ежова) в модульных коэффициентах**

$Cs/Cv = 0,3$

$p \%$	Коэффициент вариации, Cv			
	0,1	0,2	0,3	0,4
0,01	1,38	1,76	2,15	2,52
0,05	1,34	1,67	2,01	2,35
0,1	1,31	1,63	1,95	2,26
0,5	1,26	1,52	1,79	2,05
1	1,24	1,47	1,71	1,95
3	1,19	1,38	1,57	1,77
5	1,17	1,33	1,50	1,67
10	1,13	1,26	1,39	1,52
20	1,08	1,17	1,25	1,34
30	1,05	1,10	1,16	1,21
40	1,02	1,05	1,07	1,10
50	1,00	1,00	1,00	0,99
60	0,97	0,95	0,92	0,89
70	0,95	0,89	0,84	0,78
80	0,92	0,83	0,74	0,65
90	0,87	0,74	0,62	0,48
95	0,84	0,67	0,51	0,34
97	0,81	0,63	0,45	0,26
99	0,77	0,54	0,33	0,13
99,5	0,74	0,50	0,26	0,08
99,9	0,70	0,41	0,15	0,02

$Cs/Cv = 0,5$

$p \%$	Коэффициент вариации, Cv				
	0,1	0,2	0,3	0,4	0,5
0,01	1,38	1,78	2,17	2,57	2,76
0,05	1,34	1,68	2,04	2,38	2,55
0,1	1,32	1,64	1,97	2,30	2,46
0,5	1,26	1,53	1,80	2,08	2,21
1	1,24	1,48	1,72	1,97	2,09
3	1,19	1,38	1,58	1,78	1,88
5	1,17	1,34	1,51	1,68	1,77
10	1,13	1,26	1,39	1,52	1,59
20	1,08	1,17	1,25	1,34	1,38
30	1,05	1,10	1,15	1,21	1,23
40	1,02	1,05	1,07	1,09	1,10
50	1,00	1,00	0,99	0,99	0,99
60	0,97	0,95	0,92	0,88	0,87
70	0,95	0,89	0,84	0,78	0,74
80	0,92	0,83	0,74	0,65	0,60
90	0,87	0,74	0,62	0,48	0,42
95	0,84	0,68	0,52	0,36	0,28
97	0,81	0,63	0,46	0,28	0,20
99	0,77	0,55	0,34	0,16	0,09
99,5	0,75	0,51	0,29	0,11	0,05
99,9	0,70	0,42	0,18	0,04	0,01

**Ординаты кривой обеспеченностей H -распределения
(А.В. Ежова) в модульных коэффициентах, $C_s/C_v = 1,0$**

$P \%$	Коэффициент вариации, C_v					
	0,1	0,2	0,3	0,4	0,5	0,6
0,01	1,39	1,81	2,26	2,72	3,18	3,65
0,05	1,34	1,71	2,10	2,50	2,91	3,32
0,1	1,32	1,67	2,03	2,40	2,79	3,17
0,5	1,27	1,55	1,84	2,15	2,46	2,78
1	1,24	1,49	1,76	2,03	2,31	2,60
3	1,19	1,39	1,60	1,82	2,04	2,26
5	1,17	1,34	1,52	1,70	1,89	2,09
10	1,13	1,26	1,40	1,53	1,68	1,82
20	1,08	1,17	1,25	1,33	1,42	1,51
30	1,05	1,10	1,15	1,19	1,24	1,28
50	1,00	0,99	0,98	0,97	0,96	0,93
70	0,95	0,89	0,83	0,76	0,69	0,62
80	0,92	0,83	0,74	0,65	0,55	0,45
90	0,87	0,75	0,62	0,50	0,38	0,26
95	0,84	0,68	0,53	0,39	0,26	0,15
97	0,82	0,64	0,48	0,33	0,20	0,10
99	0,77	0,57	0,38	0,22	0,11	0,04
99,5	0,75	0,53	0,33	0,18	0,07	0,02
99,9	0,70	0,45	0,24	0,10	0,03	0,0004

**Ординаты кривой обеспеченностей H -распределения
(А.В. Ежова) в модульных коэффициентах, $C_s/C_v = 1,5$**

$p \%$	Коэффициент вариации, C_v							
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
0,01	1,40	1,86	2,37	2,92	3,51	4,14	4,80	5,50
0,05	1,35	1,75	2,18	2,66	3,16	3,37	4,27	4,86
0,1	1,33	1,70	2,10	2,54	3,00	3,50	4,02	4,57
0,5	1,27	1,57	1,89	2,23	2,60	2,99	3,40	3,83
1	1,24	1,51	1,79	2,09	2,41	2,75	3,11	3,48
3	1,20	1,40	1,62	1,85	2,09	2,34	2,61	2,88
5	1,17	1,35	1,53	1,72	1,92	2,13	2,35	2,57
10	1,13	1,26	1,40	1,54	1,68	1,82	1,97	2,12
20	1,08	1,17	1,25	1,32	1,40	1,48	1,55	1,61
30	1,05	1,10	1,14	1,18	1,21	1,24	1,26	1,28
50	1,00	0,99	0,98	0,96	0,93	0,90	0,86	0,81
70	0,95	0,89	0,83	0,76	0,69	0,61	0,54	0,46
80	0,92	0,83	0,74	0,65	0,56	0,47	0,38	0,30
90	0,87	0,75	0,63	0,52	0,41	0,31	0,22	0,15
95	0,84	0,69	0,54	0,42	0,31	0,21	0,13	0,08
97	0,82	0,65	0,50	0,36	0,25	0,16	0,09	0,05
99	0,78	0,58	0,41	0,27	0,16	0,09	0,045	0,019
99,5	0,76	0,54	0,37	0,23	0,13	0,06	0,028	0,011
99,9	0,71	0,47	0,29	0,15	0,07	0,03	0,006	0,002

χ^2 -распределение.

Значения квантилей даны в зависимости от числа степеней свободы ν
и уровня значимости α %

$F(x)$		0,5	2,5	5	10	90	95	97,5	99,5
α		99,5	97,5	95	90	10	5	2,5	0,5
Число степеней свободы, ν	3	0,071	0,216	0,352	0,584	6,251	7,815	9,348	12,838
	4	0,207	0,484	0,711	1,064	7,779	9,488	11,143	14,830
	5	0,412	0,831	1,145	1,610	9,236	11,070	12,832	16,750
	6	0,676	1,237	1,635	2,204	10,645	12,592	14,449	18,475
	7	0,989	1,690	2,167	2,833	12,017	14,067	16,013	20,278
	8	1,314	2,180	2,733	3,490	13,362	15,507	17,535	21,955
	9	1,735	2,700	3,325	4,168	14,684	16,919	19,023	23,589
	10	2,156	3,247	3,940	4,865	15,987	18,307	20,484	25,188
	11	2,603	3,816	4,575	5,578	17,275	19,675	21,920	26,757
	12	3,074	4,404	5,226	6,304	18,549	21,026	23,336	28,300
	13	3,565	5,009	5,892	7,042	19,812	22,362	24,736	29,819
	14	4,075	5,629	6,571	7,790	21,064	23,685	26,119	31,319
	15	4,601	6,262	7,261	8,547	22,307	24,996	27,488	32,804
	16	5,142	6,908	7,962	9,312	23,542	26,296	28,845	34,267
	17	5,697	7,564	8,672	10,085	24,769	27,587	30,191	35,713
	18	6,265	8,231	9,390	10,865	25,989	28,869	31,526	37,156
	19	6,844	8,907	10,117	11,651	27,204	30,144	32,852	38,582
	20	7,434	9,591	10,851	12,443	28,412	31,410	34,170	39,897
	21	8,034	10,283	11,591	13,240	29,615	32,671	35,479	41,401
	22	8,643	10,982	12,338	14,041	30,813	33,924	36,781	42,796
	23	9,260	11,688	13,091	14,848	32,007	35,172	38,076	44,181
	24	9,886	12,401	13,848	15,659	33,196	36,415	39,364	45,558
	25	10,520	13,120	14,671	16,473	34,382	37,652	40,646	46,928
	26	11,160	13,844	15,379	17,292	35,563	38,885	41,923	48,290
	27	11,808	14,573	16,151	18,114	36,741	40,113	42,194	49,645
	28	12,461	15,308	16,928	18,939	37,916	41,337	44,461	50,993
	29	13,121	16,047	17,708	19,768	39,087	42,557	45,722	52,336
	30	13,787	16,791	18,498	20,599	40,256	43,773	46,979	53,672

Распределение Стьюдента.

Абсолютные значения квантилей даны в зависимости от числа степеней свободы ν для одностороннего (α %) и двухстороннего (2α %) уровней значимости

$F(t)$		90	95	97,5	99,0	99,5
2α		20	10	5	2	1
α		10	5	2,5	1	0,5
Число степеней свободы ν	1	3,077	6,313	12,71	31,82	63,66
	2	1,885	2,920	4,302	6,964	9,924
	3	1,638	2,353	3,182	4,540	5,840
	4	1,533	2,132	2,776	3,746	4,604
	5	1,476	2,015	2,571	3,365	4,032
	6	1,439	1,943	2,446	3,142	3,707
	7	1,415	1,895	2,365	2,998	3,500
	8	1,397	1,859	2,306	2,897	3,355
	9	1,383	1,833	2,262	2,821	3,250
	10	1,372	1,813	2,228	2,764	3,169
	11	1,363	1,796	2,201	2,718	3,105
	12	1,356	1,782	2,179	2,681	3,055
	13	1,350	1,771	2,160	2,650	3,012
	14	1,345	1,761	2,145	2,625	2,976
	15	1,341	1,753	2,131	2,603	2,947
	16	1,336	1,746	2,119	2,583	2,920
	17	1,333	1,740	2,110	2,567	2,898
	18	1,330	1,734	2,101	2,551	2,878
	19	1,328	1,729	2,093	2,540	2,861
	20	1,325	1,725	2,086	2,528	2,845
	21	1,323	1,721	2,079	2,517	2,831
	22	1,321	1,717	2,074	2,508	2,819
	24	1,318	1,711	2,064	2,492	2,797
	26	1,315	1,706	2,055	2,478	2,778
	28	1,313	1,701	2,048	2,467	2,763
	30	1,310	1,697	2,042	2,457	2,750
	32	1,308	1,693	2,036	2,448	2,738
	34	1,307	1,691	2,032	2,441	2,728
	36	1,305	1,688	2,028	2,435	2,720
	38	1,304	1,686	2,024	2,429	2,712
	40	1,303	1,684	2,021	2,423	2,705
	50	1,299	1,676	2,009	2,403	2,678
	60	1,296	1,671	2,000	2,390	2,660
	80	1,292	1,664	1,990	2,373	2,638
	100	1,290	1,660	1,984	2,364	2,626
	200	1,286	1,653	1,972	2,345	2,601
	300	1,284	1,650	1,968	2,339	2,592
	500	1,283	1,647	1,964	2,333	2,585

F-распределение (Фишера).

Значения квантилей даны в зависимости от числа степеней свободы ν_1
(для большей дисперсии) и ν_2 (для меньшей дисперсии)
при уровне значимости $2\alpha = 10\%$

ν_2	Число степеней свободы ν_1										
	1	4	6	8	10	15	20	30	60	120	∞
1	161	225	234	239	242	246	248	250	252	253	254
2	18,5	19,2	19,3	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5
3	10,1	9,12	8,94	8,84	8,78	8,70	8,66	8,62	8,57	8,56	8,53
4	7,71	6,39	6,16	6,04	5,96	5,86	5,80	5,74	5,68	5,66	5,63
5	6,61	5,19	4,95	4,82	4,47	4,62	4,56	4,50	4,42	4,40	4,63
6	5,99	4,53	4,28	4,15	4,06	3,94	3,87	3,81	3,72	3,71	3,67
7	5,59	4,12	3,87	3,73	3,63	3,51	3,44	3,38	3,28	3,28	3,23
8	5,32	3,84	3,58	3,44	3,34	3,22	3,15	3,08	3,00	2,98	2,93
9	5,12	3,63	3,37	3,23	3,13	3,01	2,93	2,86	2,77	2,76	2,71
10	4,96	3,48	3,22	3,07	2,97	2,85	2,77	2,70	2,61	2,59	2,54
11	4,84	3,36	3,09	2,95	2,86	2,72	2,65	2,57	2,47	2,45	2,40
12	4,75	3,26	3,00	2,85	2,76	2,62	2,54	2,46	2,36	2,35	2,30
13	4,67	3,18	2,92	2,77	2,67	2,53	2,46	2,38	2,28	2,26	2,21
14	4,60	3,11	2,85	2,70	2,60	2,46	2,39	2,31	2,21	2,19	2,13
15	4,54	3,06	2,79	2,64	2,55	2,40	2,33	2,25	2,15	2,12	2,07
16	4,49	3,01	2,74	2,59	2,49	2,35	2,28	2,20	2,09	2,07	2,01
17	4,45	2,96	2,70	2,55	2,45	2,31	2,23	2,15	2,04	2,02	1,96
18	4,41	2,93	2,66	2,51	2,41	2,27	2,19	2,11	2,00	1,98	1,92
19	4,38	2,90	2,63	2,48	2,38	2,23	2,15	2,07	1,96	1,94	1,88
20	4,35	2,87	2,60	2,45	2,35	2,20	2,12	2,04	1,92	1,90	1,84
21	4,32	2,84	2,57	2,42	2,32	2,18	2,09	2,00	1,89	1,87	1,81
22	4,30	2,82	2,55	2,40	2,30	2,15	2,07	1,98	1,87	1,84	1,78
23	4,28	2,80	2,53	2,38	2,28	2,13	2,04	1,96	1,84	1,82	1,76
24	4,26	2,78	2,51	2,36	2,26	2,11	2,02	1,94	1,82	1,80	1,73
25	4,24	2,76	2,49	2,34	2,24	2,09	2,00	1,92	1,80	1,77	1,71
26	4,23	2,74	2,47	2,32	2,22	2,07	1,99	1,90	1,78	1,76	1,69
27	4,21	2,73	2,46	2,30	2,20	2,06	1,97	1,88	1,76	1,74	1,67
28	4,20	2,71	2,44	2,29	2,19	2,04	1,96	1,87	1,75	1,72	1,65
29	4,18	2,70	2,43	2,28	2,18	2,03	1,94	1,85	1,73	1,71	1,64
30	4,17	2,69	2,42	2,27	2,16	2,01	1,93	1,84	1,72	1,69	1,62
40	4,08	2,61	2,34	2,18	2,07	1,92	1,84	1,74	1,61	1,59	1,51
60	4,00	2,52	2,25	2,10	1,99	1,84	1,75	1,65	1,50	1,48	1,39
120	3,92	2,44	2,17	2,01	1,90	1,75	1,65	1,65	1,49	1,36	1,25
∞	3,84	2,37	2,09	1,94	1,83	1,67	1,57	1,46	1,28	1,24	1,00

F-распределение (Фишера).

Значения квантилей даны в зависимости от числа степеней свободы ν_1
(для большей дисперсии) и ν_2 (для меньшей дисперсии)
при уровне значимости $2\alpha = 5\%$

ν_2	Число степеней свободы ν_1										
	1	4	6	8	10	15	20	30	60	120	∞
1	648	899	937	957	969	985	993	1001	1010	1014	1018
2	38,5	39,2	39,3	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5
3	17,4	15,1	14,7	14,6	14,4	14,3	14,2	14,1	14,0	13,9	13,9
4	12,2	9,60	9,20	8,98	8,84	8,66	8,56	8,46	8,36	8,31	8,26
5	10,0	7,39	6,98	6,76	6,62	6,42	6,33	6,23	6,12	6,07	6,02
6	8,81	6,23	5,82	5,60	5,46	5,27	5,17	5,07	4,96	4,90	4,85
7	8,07	5,52	5,12	4,90	4,76	4,57	4,47	4,36	4,25	4,20	4,14
8	7,57	5,05	4,65	4,43	4,30	4,10	4,00	3,89	3,78	3,73	3,67
9	7,21	4,72	4,32	4,10	3,96	3,77	3,67	3,56	3,45	3,39	3,33
10	6,94	4,47	4,07	3,85	3,72	3,52	3,42	3,31	3,20	3,14	3,08
11	6,72	4,28	3,88	3,66	3,53	3,33	3,23	3,12	3,00	2,94	2,88
12	6,55	4,12	3,73	3,51	3,37	3,18	3,07	2,96	2,85	2,79	2,72
13	6,41	4,00	3,61	3,39	3,25	3,05	2,95	2,84	2,72	2,66	2,60
14	6,30	3,89	3,50	3,29	3,15	2,95	2,84	2,73	2,61	2,55	2,49
15	6,20	3,80	3,41	3,20	3,05	2,86	2,76	2,64	2,52	2,46	2,40
16	6,16	3,73	3,34	3,12	2,99	2,79	2,68	2,57	2,45	2,38	2,32
17	6,04	3,66	3,28	3,06	2,92	2,72	2,62	2,50	2,38	2,32	2,25
18	5,98	3,61	3,22	3,01	2,87	2,67	2,56	2,44	2,32	2,26	2,19
19	5,92	3,56	3,17	2,96	2,82	2,62	2,51	2,38	2,27	2,20	2,13
20	5,87	3,51	3,13	2,91	2,77	2,57	2,46	2,35	2,22	2,16	2,09
21	5,83	3,47	3,09	2,87	2,73	2,53	2,42	2,31	2,18	2,11	2,04
22	5,79	3,44	3,05	2,84	2,70	2,50	2,39	2,27	2,14	2,08	2,03
23	5,75	3,41	3,02	2,81	2,67	2,47	2,36	2,24	2,11	2,04	1,97
24	5,72	3,38	2,99	2,78	2,64	2,44	2,33	2,21	2,08	2,01	1,94
25	5,69	3,35	2,97	2,75	2,61	2,41	2,30	2,18	2,05	1,98	1,91
26	5,66	3,33	2,94	2,72	2,59	2,39	2,28	2,16	2,03	1,95	1,88
27	5,63	3,31	2,92	2,71	2,57	2,34	2,25	2,13	2,00	1,93	1,85
28	5,61	3,29	2,90	2,69	2,55	2,34	2,23	2,11	1,98	1,91	1,83
29	5,59	3,27	2,88	2,67	2,53	2,32	2,21	2,09	1,96	1,89	1,81
30	5,57	3,25	2,87	2,65	2,51	2,31	2,20	2,07	1,94	1,87	1,77
40	5,42	3,13	2,74	2,53	2,39	2,18	2,07	1,94	1,80	1,72	1,74
60	5,28	3,01	2,63	2,41	2,27	2,06	1,94	1,82	1,67	1,58	1,48
120	5,15	2,89	2,52	2,30	2,16	1,95	1,82	1,69	1,53	1,43	1,31
∞	5,02	2,79	2,41	2,19	2,05	1,83	1,71	1,57	1,39	1,27	1,00

F-распределение (Фишера).

Значения квантилей даны в зависимости от числа степеней свободы ν_1 (для большей дисперсии) и ν_2 (для меньшей дисперсии)
при уровне значимости $2\alpha = 1\%$

$\square \nu_2$	Число степеней свободы ν_1										
	1	4	6	8	10	15	20	30	60	120	∞
2	198	199	199	199	199	199	199	199	199	199	199
3	55,6	46,2	44,8	44,1	43,7	43,1	42,8	42,5	42,1	42,0	41,8
4	31,3	23,2	22,0	21,4	21,0	20,4	20,2	19,9	19,6	19,5	19,3
5	22,8	15,6	14,5	14,0	13,6	13,1	12,9	12,7	12,4	12,3	12,1
6	18,6	12,1	11,1	10,6	10,3	9,81	9,59	9,36	9,12	9,00	8,88
7	16,2	10,1	9,16	8,68	8,38	7,97	7,75	7,53	7,31	7,19	7,08
8	14,7	8,81	7,95	7,50	7,21	6,81	6,61	6,40	6,18	6,06	5,95
9	13,6	7,96	7,13	6,69	6,42	6,03	5,83	5,52	5,41	5,30	5,19
10	12,8	7,34	6,51	6,12	5,85	5,47	5,27	5,07	4,86	4,75	4,64
11	12,2	6,88	6,10	5,68	5,42	5,05	4,86	4,65	4,45	4,34	4,23
12	11,7	6,52	5,76	5,35	5,09	4,72	4,53	4,33	4,12	4,01	3,90
13	11,4	6,23	5,48	5,08	4,82	4,46	4,27	3,97	3,76	3,76	3,65
14	11,1	6,00	5,26	4,86	4,60	4,25	4,06	3,86	3,66	3,55	3,44
15	10,8	6,80	5,08	4,67	4,42	4,07	3,88	3,69	3,48	3,37	3,26
16	10,6	5,54	4,91	4,52	4,27	3,96	3,73	3,54	3,33	3,22	3,11
17	10,4	5,50	4,78	4,39	4,14	3,79	3,61	3,41	3,21	3,10	2,98
18	10,2	5,37	4,66	4,28	4,03	3,68	3,50	3,30	3,10	2,99	2,87
19	10,1	5,27	4,56	4,18	3,93	3,59	3,40	3,21	3,00	2,89	2,78
20	9,94	5,17	4,47	4,09	3,85	3,50	3,32	3,12	2,92	2,81	2,69
21	9,83	5,09	4,39	4,01	3,77	3,43	3,24	3,05	2,84	2,73	2,61
22	9,73	5,02	4,32	3,94	3,70	3,36	3,18	2,98	2,77	2,66	2,55
23	9,63	4,95	4,26	3,88	3,64	3,30	3,12	2,92	2,71	2,60	2,48
24	9,55	4,89	4,20	3,83	3,59	3,25	3,06	2,87	2,66	2,55	2,43
25	9,48	4,84	4,15	3,76	3,54	3,20	3,01	2,82	2,61	2,50	2,38
26	9,41	4,79	4,10	3,73	3,49	3,15	2,97	2,77	2,56	2,45	2,33
27	9,34	4,74	4,06	3,69	3,44	3,11	2,93	2,73	2,52	2,41	2,29
28	9,28	4,70	4,02	3,65	3,41	3,07	2,89	2,69	2,48	2,37	2,25
29	9,23	4,66	3,98	3,61	3,38	3,04	2,86	2,66	2,45	2,33	2,21
30	9,18	4,62	3,95	3,58	3,34	3,01	2,82	2,63	2,42	2,30	2,18
40	8,83	4,37	3,71	3,35	3,12	2,78	2,59	2,40	2,18	2,06	1,93
60	8,48	4,14	3,49	3,13	2,90	2,57	2,39	2,19	1,96	1,83	1,69
120	8,18	3,92	3,25	2,93	2,71	2,37	2,19	1,98	1,75	1,61	1,43
∞	7,88	3,72	3,09	2,74	2,52	2,19	2,00	1,79	1,53	1,36	1,00

Распределение Колмогорова.
Вероятности превышения даны в долях единицы.

$$P(\lambda) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2\lambda^2}$$

λ	0	1	2	3	4	5	6	7	8	9
0,3				0,9999	0,9998	0,9997	0,9995	0,9992	0,9987	0,9981
0,4	0,9972	0,9960	0,9945	0,9926	0,9903	0,9874	0,9840	0,9800	0,9753	0,9700
0,5	0,9639	0,9572	0,9497	0,9415	0,9325	0,9228	0,9124	0,9013	0,8896	0,8772
0,6	0,8643	0,8508	0,8368	0,8222	0,8073	0,7920	0,7764	0,7604	0,7442	0,7278
0,7	0,7112	0,6945	0,6777	0,6609	0,6440	0,6272	0,6104	0,5936	0,5770	0,5605
0,8	0,5441	0,5280	0,5120	0,4962	0,4806	0,4653	0,4503	0,4355	0,4209	0,4067
0,9	0,3927	0,3791	0,3657	0,3527	0,3399	0,3275	0,3154	0,3036	0,2921	0,2809
1,0	0,2700	0,2594	0,2492	0,2392	0,2296	0,2202	0,2111	0,2024	0,1939	0,1857
1,1	0,1777	0,1700	0,1626	0,1555	0,1486	0,1420	0,1356	0,1294	0,1235	0,1177
1,2	0,1122	0,1070	0,1019	0,0970	0,0924	0,0879	0,0836	0,0794	0,0755	0,0717
1,3	0,0681	0,0646	0,0613	0,0582	0,0551	0,0522	0,0495	0,0469	0,0444	0,0420
1,4	0,0397	0,0375	0,0354	0,0335	0,0316	0,0298	0,0282	0,0266	0,0250	0,0236
1,5	0,0222	0,0209	0,0197	0,0185	0,0174	0,0164	0,0154	0,0145	0,0136	0,0127
1,6	0,0120	0,0112	0,0105	0,0098	0,0092	0,0086	0,0081	0,0076	0,0071	0,0066
1,7	0,0062	0,0058	0,0054	0,0050	0,0047	0,0044	0,0041	0,0038	0,0035	0,0033
1,8	0,0031	0,0029	0,0027	0,0025	0,0023	0,0021	0,0020	0,0019	0,0017	0,0016
1,9	0,0015	0,0014	0,0013	0,0012	0,0011	0,0010	0,0009	0,0009	0,0008	0,0007
2,0	0,0007	0,0006	0,0006	0,0005	0,0005	0,0004	0,0004	0,0004	0,0003	0,0003
2,1	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
2,2	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
2,3	0,0001	0,0000								

**Эффект ложной корреляции для зависимости $y = f(z)$
(по Бенсону)**

Переменная		Условие	$r_{y,z}$ (аппроксимация)	Коэффициент вариации, C	$r_{y,z}$ (значение)
y	z				
$x_1 x_2$	x_2	$r_{1,2} = 0$	$\frac{C_2}{(C_1^2 + C_2^2)^{1/2}}$	$C_1 = C_2$ $C_1 > C_2$	0,71 > 0,71
$x_1 x_2$	$x_3 x_2$	$r_{1,2} = 0$ $r_{1,3} = 0$ $r_{2,3} = 0$	$\frac{C_2^2}{(C_1^2 + C_2^2)^{1/2} (C_2^2 + C_3^2)^{1/2}}$	$C_2 > C_1 C_3$ $C_2 = 2C_1 = 2C_3$ $C_2 = 2C_1 = 3C_3$ $C_2 = 3C_1 = 3C_3$	> 0,50 0,80 0,85 0,90
x_1/x_2	x_2	$r_{1,2} = 0$	$\frac{-C_2}{(C_1^2 + C_2^2)^{1/2}}$	$C_2 > C_1$ $C_2 = 2C_1$ $C_2 = 3C_1$	< -0,70 -0,89 -0,95
x_1/x_2	x_3/x_2	$r_{1,2} = 0$ $r_{1,3} = 0$ $r_{2,3} = 0$	$\frac{C_2^2}{(C_1^2 + C_2^2)^{1/2} (C_2^2 + C_3^2)^{1/2}}$	$C_1 > C_2 C_3$ $C_2 = 2C_1 = 2C_3$ $C_2 = 3C_1 = 3C_3$	> 0,50 0,80 0,90

Приложение 9.1

**Таблица для определения коэффициента вариации C_v
методом приближенного наибольшего правдоподобия
в зависимости от статистики $-\lambda_2$
(для кривой обеспеченностей Пирсона III типа) $C_s/C_v = 1$**

C_v	0	1	2	3	4	5	6	7	8	9
0,1	0,0022	0,0026	0,0032	0,0038	0,0044	0,0050	0,0058	0,0066	0,0074	0,0082
0,2	0,0090	0,0100	0,0111	0,0122	0,0132	0,0143	0,0157	0,0171	0,0184	0,0199
0,3	0,0212	0,0229	0,0246	0,0263	0,0280	0,0297	0,0318	0,0339	0,0360	0,0381
0,4	0,0402	0,0428	0,0454	0,0479	0,0505	0,0530	0,0562	0,0593	0,0624	0,0655
0,5	0,0686	0,0723	0,0760	0,0797	0,0835	0,0872	0,0916	0,0961	0,101	0,105
0,6	0,109	0,115	0,120	0,125	0,130	0,136	0,142	0,148	0,154	0,160
0,7	0,166	0,173	0,181	0,188	0,195	0,202	0,210	0,218	0,227	0,235
0,8	0,243	0,252	0,262	0,271	0,281	0,290	0,300	0,311	0,321	0,332
0,9	0,342	0,354	0,366	0,377	0,389	0,401	0,414	0,427	0,440	0,435
1,0	0,466	0,480	0,495	0,509	0,523	0,538	0,553	0,569	0,585	0,600
1,1	0,616	0,633	0,650	0,668	0,684	0,701	0,719	0,737	0,755	0,774
1,2	0,792	—	—	—	—	—	—	—	—	—

**Таблица для определения коэффициента вариации C_v
методом приближенного наибольшего правдоподобия
в зависимости от статистики $-\lambda_2$
(для кривой обеспеченностей Пирсона III типа) $C_s/C_v = 2$**

C_v	0	1	2	3	4	5	6	7	8	9
0,1	0,0022	0,0026	0,0031	0,0037	0,0043	0,0049	0,0056	0,0063	0,0071	0,0079
0,2	0,0087	0,0096	0,0106	0,0116	0,0126	0,0137	0,0148	0,0160	0,0172	0,0185
0,3	0,0198	0,0212	0,0296	0,0241	0,0256	0,0271	0,0287	0,0304	0,0321	0,0339
0,4	0,0357	0,0375	0,0394	0,0414	0,0434	0,0454	0,0476	0,0497	0,0519	0,0542
0,5	0,0565	0,0589	0,0613	0,0638	0,0664	0,0690	0,0716	0,0743	0,0771	0,0799
0,6	0,0828	0,0858	0,0887	0,0918	0,0949	0,0981	0,101	0,105	0,108	0,111
0,7	0,115	0,118	0,122	0,126	0,129	0,133	0,137	0,141	0,145	0,149
0,8	0,153	0,157	0,162	0,166	0,170	0,175	0,180	0,184	0,189	0,194
0,9	0,198	0,203	0,208	0,213	0,218	0,224	0,229	0,234	0,240	0,245
1,0	0,251	0,256	0,262	0,268	0,274	0,280	0,286	0,292	0,298	0,304
1,1	0,311	0,317	0,323	0,330	0,337	0,343	0,350	0,357	0,364	0,371
1,2	0,378	0,386	0,393	0,400	0,408	0,415	0,423	0,431	0,438	0,446
1,3	0,454	0,462	0,471	0,479	0,487	0,496	0,504	0,513	0,521	0,530
1,4	0,539	0,548	0,557	0,566	0,575	0,584	0,594	0,603	0,613	0,623
1,5	0,632	0,642	0,652	0,662	0,672	0,682	0,692	0,703	0,713	0,724
1,6	0,734	0,745	0,756	0,767	0,778	0,789	0,800	0,811	0,823	0,824
1,7	0,846	0,857	0,860	0,881	0,893	0,905	0,917	0,929	0,911	0,953
1,8	0,966	0,978	0,991	1,004	1,016	1,029	1,042	1,055	1,069	1,082
1,9	1,095	1,109	1,122	1,136	1,150	1,163	1,177	1,191	1,205	1,220

**Таблица для определения коэффициента вариации C_v
методом приближенного наибольшего правдоподобия
в зависимости от статистики – λ_2
(для кривой обеспеченностей Пирсона III типа) $C_s/C_v = 3$**

C_v	0	1	2	3	4	5	6	7	8	9
0,1	0,0022	0,0026	0,0031	0,0037	0,0043	0,0048	0,0055	0,0062	0,0070	0,0078
0,2	0,0085	0,0094	0,0104	0,0113	0,0122	0,0132	0,0143	0,0154	0,0165	0,0176
0,3	0,0188	0,0201	0,0214	0,0227	0,0240	0,0253	0,0267	0,0282	0,0297	0,0311
0,4	0,0326	0,0342	0,0359	0,0375	0,0391	0,0408	0,0425	0,0443	0,0461	0,0479
0,5	0,0497	0,0526	0,0535	0,0555	0,0574	0,0594	0,0614	0,0635	0,0656	0,0677
0,6	0,0698	0,0720	0,0742	0,0764	0,0786	0,0808	0,0832	0,0856	0,0879	0,0903
0,7	0,0926	0,0951	0,0976	0,100	0,103	0,105	0,108	0,110	0,113	0,116
0,8	0,118	0,121	0,124	0,127	0,129	0,132	0,135	0,138	0,141	0,144
0,9	0,147	0,150	0,153	0,156	0,159	0,162	0,165	0,168	0,171	0,174
1,0	0,177	0,181	0,184	0,187	0,191	0,194	0,197	0,201	0,204	0,208
1,1	0,211	0,215	0,218	0,222	0,225	0,229	0,232	0,236	0,240	0,243
1,2	0,247	0,251	0,255	0,258	0,262	0,266	0,270	0,274	0,277	0,281
1,3	0,286	0,289	0,294	0,298	0,302	0,306	0,311	0,315	0,319	0,323
1,4	0,327	0,332	0,336	0,340	0,345	0,349	0,354	0,358	0,363	0,367
1,5	0,372	0,376	0,381	0,386	0,390	0,395	0,400	0,404	0,409	0,414
1,6	0,419	0,424	0,428	0,433	0,438	0,443	0,448	0,453	0,458	0,463
1,7	0,468	0,474	0,479	0,484	0,489	0,495	0,500	0,505	0,511	0,516
1,8	0,521	0,527	0,532	0,538	0,543	0,549	0,555	0,560	0,566	0,572
1,9	0,578	0,584	0,589	0,595	0,600	0,606	0,612	0,618	0,624	0,630

**Таблица для определения коэффициента вариации C_v
методом приближенного наибольшего правдоподобия**

в зависимости от статистики - λ_2

(для кривой обеспеченностей Пирсона III типа) $C_v/C_v = 4$

C_v	0	1	2	3	4	5	6	7	8	9
0,1	0,0022	0,0026	0,0031	0,0037	0,0043	0,0048	0,0055	0,0062	0,0069	0,0076
0,2	0,0083	0,0092	0,0101	0,0110	0,0119	0,0128	0,0138	0,0148	0,0159	0,0169
0,3	0,0180	0,0192	0,0203	0,0215	0,0227	0,0239	0,0252	0,0265	0,0279	0,0292
0,4	0,0305	0,0319	0,0334	0,0348	0,0363	0,0377	0,0392	0,0408	0,0423	0,0439
0,5	0,0454	0,0471	0,0488	0,0505	0,0522	0,0538	0,0556	0,0574	0,0592	0,0609
0,6	0,0627	0,0646	0,0665	0,0683	0,0702	0,0721	0,0741	0,0760	0,0780	0,0800
0,7	0,0820	0,0840	0,0861	0,0882	0,0903	0,0923	0,0946	0,0966	0,0989	0,101
0,8	0,103	0,105	0,108	0,110	0,112	0,114	0,117	0,119	0,121	0,124
0,9	0,126	0,128	0,131	0,133	0,136	0,138	0,141	0,143	0,146	0,148
1,0	0,151	0,153	0,156	0,158	0,161	0,163	0,166	0,169	0,171	0,174
1,1	0,177	0,179	0,182	0,185	0,188	0,190	0,193	0,196	0,199	0,201
1,2	0,204	0,207	0,210	0,213	0,216	0,219	0,222	0,224	0,227	0,230
1,3	0,233	0,236	0,239	0,242	0,245	0,248	0,251	0,254	0,258	0,261
1,4	0,264	0,267	0,270	0,273	0,276	0,279	0,283	0,286	0,289	0,292
1,5	0,296	0,299	0,302	0,305	0,309	0,312	0,315	0,319	0,322	0,325
1,6	0,329	0,332	0,336	0,339	0,342	0,346	0,349	0,353	0,356	0,360
1,7	0,363	0,367	0,370	0,374	0,377	0,381	0,384	0,388	0,392	0,395
1,8	0,399	0,403	0,406	0,410	0,414	0,417	0,421	0,425	0,429	0,432
1,9	0,436	0,440	0,444	0,447	0,451	0,455	0,459	0,463	0,467	0,471

**Критические значения статистики Диксона
для максимального члена выборки $D_{\max}$
в зависимости от уровня значимости α %, r , n , C_s**

α, %	Коэффициент автокорреляции, <i>r</i>	Объем выборки, <i>n</i>						
		6	10	20	30	50	70	100
<i>C_s</i> = 0								
1	0	0,70	0,53	0,39	0,34	0,31	0,28	0,25
	0,5	0,70	0,52	0,38	0,32	0,28	0,26	0,23
	0,9	0,66	0,47	0,32	0,27	0,22	0,20	0,17
5	0	0,56	0,41	0,30	0,26	0,22	0,20	0,18
	0,5	0,56	0,40	0,29	0,25	0,21	0,19	0,17
	0,9	0,54	0,38	0,24	0,20	0,16	0,14	0,12
10	0	0,48	0,35	0,25	0,22	0,19	0,17	0,15
	0,5	0,48	0,34	0,24	0,21	0,18	0,16	0,14
	0,9	0,46	0,32	0,19	0,16	0,13	0,12	0,10
<i>C_s</i> = 0,5								
1	0	0,74	0,58	0,48	0,43	0,39	0,37	0,33
	0,5	0,74	0,57	0,44	0,39	0,36	0,34	0,31
	0,9	0,68	0,54	0,36	0,30	0,26	0,24	0,21
5	0	0,63	0,48	0,38	0,33	0,29	0,26	0,25
	0,5	0,62	0,47	0,36	0,31	0,27	0,26	0,23
	0,9	0,56	0,41	0,28	0,24	0,19	0,18	0,16
10	0	0,55	0,42	0,32	0,28	0,24	0,23	0,21
	0,5	0,54	0,39	0,30	0,26	0,22	0,21	0,19
	0,9	0,48	0,35	0,22	0,19	0,16	0,15	0,13

**Критические значения статистики Диксона
для максимального члена выборки $D_{\max}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 1,0$								
1	0	0,79	0,65	0,55	0,50	0,47	0,45	0,41
	0,5	0,78	0,63	0,52	0,47	0,43	0,41	0,38
	0,9	0,73	0,58	0,42	0,36	0,32	0,30	0,27
5	0	0,68	0,55	0,45	0,40	0,36	0,34	0,31
	0,5	0,67	0,53	0,43	0,38	0,33	0,31	0,28
	0,9	0,59	0,45	0,31	0,37	0,23	0,22	0,19
10	0	0,60	0,48	0,39	0,34	0,30	0,28	0,26
	0,5	0,59	0,45	0,36	0,31	0,27	0,26	0,24
	0,9	0,51	0,38	0,26	0,23	0,19	0,18	0,16
$C_s = 1,5$								
1	0	0,84	0,71	0,62	0,57	0,54	0,51	0,47
	0,5	0,82	0,69	0,58	0,53	0,50	0,48	0,45
	0,9	0,75	0,62	0,46	0,41	0,37	0,35	0,31
5	0	0,73	0,60	0,51	0,46	0,42	0,40	0,37
	0,5	0,72	0,58	0,49	0,44	0,39	0,38	0,34
	0,9	0,62	0,49	0,37	0,32	0,28	0,26	0,24
10	0	0,67	0,54	0,45	0,40	0,36	0,34	0,31
	0,5	0,64	0,51	0,42	0,38	0,33	0,32	0,29
	0,9	0,53	0,42	0,31	0,27	0,23	0,22	0,18

**Критические значения статистики Диксона
для максимального члена выборки $D_{\max}$
в зависимости от уровня значимости α %, r , n , C_s**

α, %	Коэффициент автокорреляции, <i>r</i>	Объем выборки, <i>n</i>						
		6	10	20	30	50	70	100
<i>C_s</i> = 2,0								
1	0	0,88	0,77	0,68	0,64	0,61	0,58	0,53
	0,5	0,88	0,76	0,65	0,61	0,57	0,54	0,50
	0,9	0,79	0,61	0,52	0,46	0,41	0,39	0,36
5	0	0,79	0,67	0,58	0,52	0,48	0,45	0,42
	0,5	0,77	0,65	0,55	0,49	0,44	0,42	0,39
	0,9	0,67	0,56	0,41	0,36	0,32	0,30	0,27
10	0	0,72	0,60	0,51	0,46	0,41	0,39	0,36
	0,5	0,69	0,57	0,49	0,44	0,38	0,36	0,33
	0,9	0,57	0,46	0,34	0,30	0,26	0,25	0,23
<i>C_s</i> = 3,0								
1	0	0,96	0,87	0,78	0,74	0,70	0,67	0,62
	0,5	0,96	0,86	0,76	0,71	0,67	0,63	0,58
	0,9	0,91	0,83	0,69	0,61	0,55	0,53	0,49
5	0	0,88	0,79	0,69	0,63	0,58	0,55	0,50
	0,5	0,88	0,78	0,67	0,61	0,55	0,52	0,47
	0,9	0,79	0,69	0,54	0,48	0,42	0,39	0,36
10	0	0,83	0,72	0,62	0,56	0,50	0,47	0,44
	0,5	0,81	0,70	0,60	0,54	0,48	0,45	0,42
	0,9	0,69	0,59	0,46	0,40	0,35	0,33	0,30

**Критические значения статистики Диксона
для минимального члена выборки $D_{\min}$
в зависимости от уровня значимости α %, r , n , C_s**

α, %	Коэффициент автокорреляции, <i>r</i>	Объем выборки, <i>n</i>						
		6	10	20	30	50	70	100
<i>C_s</i> = 0								
1	0	0,70	0,53	0,39	0,34	0,28	0,24	0,21
	0,5	0,69	0,51	0,38	0,29	0,26	0,22	0,19
	0,9	0,69	0,49	0,35	0,25	0,20	0,19	0,17
5	0	0,56	0,41	0,30	0,26	0,21	0,19	0,16
	0,5	0,56	0,40	0,29	0,25	0,20	0,17	0,15
	0,9	0,56	0,36	0,24	0,19	0,15	0,13	0,12
10	0	0,48	0,35	0,25	0,21	0,17	0,15	0,13
	0,5	0,48	0,34	0,24	0,18	0,16	0,14	0,12
	0,9	0,47	0,30	0,20	0,15	0,12	0,11	0,09
<i>C_s</i> = 0,5								
1	0	0,65	0,45	0,31	0,25	0,19	0,16	0,14
	0,5	0,65	0,45	0,31	0,25	0,19	0,16	0,14
	0,9	0,65	0,45	0,29	0,22	0,16	0,15	0,13
5	0	0,51	0,34	0,22	0,18	0,14	0,12	0,10
	0,5	0,51	0,34	0,22	0,18	0,14	0,12	0,10
	0,9	0,51	0,33	0,21	0,16	0,12	0,10	0,09
10	0	0,44	0,29	0,18	0,14	0,11	0,10	0,09
	0,5	0,44	0,29	0,18	0,14	0,11	0,10	0,09
	0,9	0,44	0,28	0,17	0,12	0,10	0,08	0,07

**Критические значения статистики Диксона
для минимального члена выборки $D_{\min}$
в зависимости от уровня значимости α %, r , n , C_s**

α, %	Коэффициент автокорреляции, <i>r</i>	Объем выборки, <i>n</i>						
		6	10	20	30	50	70	100
<i>C_s</i> = 1,0								
1	0	0,58	0,38	0,22	0,17	0,12	0,09	0,08
	0,5	0,62	0,38	0,22	0,17	0,12	0,09	0,08
	0,9	0,66	0,43	0,26	0,18	0,12	0,09	0,08
5	0	0,44	0,27	0,16	0,12	0,08	0,07	0,06
	0,5	0,46	0,27	0,16	0,12	0,08	0,07	0,06
	0,9	0,52	0,30	0,17	0,13	0,09	0,07	0,06
10	0	0,37	0,23	0,12	0,09	0,07	0,06	0,05
	0,5	0,39	0,24	0,13	0,09	0,07	0,06	0,05
	0,9	0,42	0,25	0,14	0,10	0,07	0,06	0,05
<i>C_s</i> = 1,5								
1	0	0,51	0,30	0,14	0,10	0,06	0,04	0,03
	0,5	0,57	0,32	0,17	0,11	0,07	0,05	0,04
	0,9	0,63	0,39	0,22	0,14	0,09	0,06	0,05
5	0	0,38	0,20	0,10	0,07	0,04	0,03	0,02
	0,5	0,41	0,22	0,11	0,07	0,04	0,03	0,02
	0,9	0,48	0,27	0,14	0,10	0,06	0,04	0,03
10	0	0,31	0,17	0,08	0,05	0,03	0,025	0,02
	0,5	0,34	0,18	0,09	0,06	0,04	0,025	0,02
	0,9	0,39	0,22	0,11	0,07	0,045	0,03	0,02

**Критические значения статистики Диксона
для минимального члена выборки $D_{\min}$
в зависимости от уровня значимости α %, r , n , C_s**

α, %	Коэффициент автокорреляции, <i>r</i>	Объем выборки, <i>n</i>						
		6	10	20	30	50	70	100
<i>C_s</i> = 2,0								
1	0	0,45	0,22	0,09	0,05	0,025	0,015	0,01
	0,5	0,52	0,25	0,11	0,06	0,03	0,02	0,01
	0,9	0,60	0,35	0,18	0,11	0,06	0,04	0,03
5	0	0,31	0,14	0,06	0,03	0,015	0,01	0,007
	0,5	0,36	0,17	0,08	0,04	0,02	0,01	0,007
	0,9	0,45	0,24	0,11	0,07	0,04	0,02	0,015
10	0	0,24	0,11	0,04	0,02	0,01	0,008	0,005
	0,5	0,29	0,13	0,05	0,03	0,01	0,008	0,005
	0,9	0,36	0,19	0,09	0,05	0,03	0,02	0,01
<i>C_s</i> = 3,0								
1	0	0,32	0,11	0,02	0,008	0,002	0,001	0,000
	0,5	0,43	0,15	0,04	0,01	0,004	0,002	0,001
	0,9	0,56	0,29	0,12	0,07	0,03	0,015	0,007
5	0	0,20	0,06	0,009	0,003	0,001	0,000	0,000
	0,5	0,27	0,09	0,02	0,006	0,002	0,001	0,000
	0,9	0,39	0,18	0,07	0,03	0,01	0,005	0,002
10	0	0,14	0,04	0,006	0,002	0,001	0,000	0,000
	0,5	0,18	0,06	0,01	0,003	0,001	0,000	0,000
	0,9	0,30	0,13	0,05	0,02	0,006	0,003	0,001

**Критические значения статистики Смирнова – Граббса
для максимального члена выборки $G_{\max}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 0$								
1	0	1,95	2,43	2,94	3,17	3,38	3,52	3,54
	0,5	1,94	2,39	2,84	3,11	3,35	3,50	3,51
	0,9	1,92	2,30	2,67	2,86	3,08	3,10	3,18
5	0	1,84	2,17	2,60	2,79	3,00	3,10	3,21
	0,5	1,83	2,15	2,52	2,74	2,93	3,02	3,15
	0,9	1,80	2,10	2,37	2,51	2,67	2,79	2,90
10	0	1,75	2,04	2,44	2,60	2,851	2,91	3,02
	0,5	1,73	2,03	2,36	2,54	2,78	2,88	2,97
	0,9	1,70	1,99	2,22	2,34	2,52	2,62	2,72
$C_s = 0,5$								
1	0	1,98	2,54	3,26	3,60	3,97	4,26	4,33
	0,5	1,98	2,50	3,17	3,52	3,90	4,21	4,30
	0,9	1,93	2,47	2,88	3,12	3,44	3,67	3,71
5	0	1,89	2,32	2,89	3,18	3,52	3,69	3,85
	0,5	1,88	2,29	2,80	3,10	3,43	3,61	3,74
	0,9	1,82	2,22	2,53	2,71	2,97	3,12	3,32
10	0	1,81	2,19	2,70	2,99	3,28	3,46	3,62
	0,5	1,79	2,16	2,64	2,88	3,20	3,39	3,52
	0,9	1,73	2,06	2,35	2,54	2,78	2,94	3,09

**Критические значения статистики Смирнова – Граббса
для максимального члена выборки $G_{\max}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 1,0$								
1	0	2,00	2,64	3,53	3,96	4,59	4,96	5,16
	0,5	2,00	2,59	3,42	3,87	4,42	4,83	5,04
	0,9	1,96	2,52	3,06	3,36	3,72	4,10	4,18
5	0	1,93	2,46	3,18	3,54	4,00	4,24	4,47
	0,5	1,92	2,43	3,05	3,45	3,93	4,19	4,36
	0,9	1,85	2,29	2,68	2,92	3,25	3,50	3,74
10	0	1,87	2,34	2,98	3,34	3,74	3,98	4,22
	0,5	1,85	2,29	2,88	3,20	3,61	3,84	4,08
	0,9	1,77	2,13	2,49	2,74	3,05	3,26	3,47
$C_s = 1,5$								
1	0	2,02	2,71	3,73	4,29	5,06	5,54	5,86
	0,5	2,02	2,68	3,58	4,19	4,90	5,40	5,76
	0,9	1,98	2,57	3,25	3,70	4,18	4,54	4,70
5	0	1,97	2,57	3,42	3,87	4,43	4,81	5,06
	0,5	1,96	2,53	3,30	3,74	4,28	4,70	4,94
	0,9	1,89	2,38	2,86	3,15	3,57	3,92	4,17
10	0	1,92	2,47	3,23	3,66	4,14	4,46	4,77
	0,5	1,90	2,40	3,11	3,52	4,00	4,32	4,60
	0,9	1,80	2,22	2,66	2,96	3,35	3,59	3,87

**Критические значения статистики Смирнова – Граббса
для максимального члена выборки $G_{\max}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 2,0$								
1	0	2,03	2,76	3,91	4,57	5,49	6,08	6,51
	0,5	2,03	2,75	3,85	4,48	5,31	5,97	6,39
	0,9	1,99	2,65	3,44	4,04	4,62	5,01	5,24
5	0	2,00	2,65	3,62	4,14	4,86	5,30	5,62
	0,5	1,99	2,62	3,51	4,04	4,67	5,14	5,50
	0,9	1,92	2,48	3,06	3,42	3,92	4,30	4,58
10	0	1,96	2,57	3,44	3,92	4,49	4,91	5,28
	0,5	1,95	2,52	3,33	3,79	4,35	4,79	5,08
	0,9	1,85	2,33	2,85	3,17	3,66	3,94	4,26
$C_s = 3,0$								
1	0	2,04	2,82	4,10	4,93	6,06	6,88	7,63
	0,5	2,04	2,82	4,04	4,87	5,94	6,68	7,47
	0,9	2,03	2,80	3,94	4,51	5,23	5,95	6,37
5	0	2,03	2,77	3,91	4,59	5,51	6,11	6,63
	0,5	2,03	2,75	3,84	4,51	5,34	6,00	6,41
	0,9	2,00	2,67	3,56	4,00	4,61	5,10	5,50
10	0	2,01	2,72	3,76	4,38	5,15	5,67	6,23
	0,5	2,01	2,68	3,68	4,28	5,02	5,56	5,95
	0,9	1,94	2,54	3,30	3,69	4,25	4,70	5,12

**Критические значения статистики Смирнова – Граббса
для минимального члена выборки $G_{\min}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 0$								
1	0	1,95	2,43	2,94	3,17	3,38	3,52	3,54
	0,5	1,94	2,39	2,84	3,11	3,35	3,50	3,51
	0,9	1,92	2,30	2,67	2,86	3,08	3,10	3,18
5	0	1,84	2,17	2,60	2,79	3,00	3,10	3,21
	0,5	1,83	2,15	2,52	2,74	2,93	3,02	3,15
	0,9	1,80	2,10	2,37	2,51	2,67	2,79	2,90
10	0	1,75	2,04	2,44	2,60	2,851	2,91	3,02
	0,5	1,73	2,03	2,36	2,54	2,78	2,88	2,97
	0,9	1,70	1,99	2,22	2,34	2,52	2,62	2,72
$C_s = 0,5$								
1	0	1,88	2,27	2,50	2,51	2,53	2,55	2,57
	0,5	1,90	2,28	2,53	2,60	2,63	2,64	2,65
	0,9	1,93	2,29	2,56	2,64	2,72	2,74	2,86
5	0	1,74	2,00	2,22	2,28	2,36	2,39	2,42
	0,5	1,76	2,02	2,24	2,30	2,38	2,41	2,44
	0,9	1,78	2,04	2,26	2,32	2,40	2,44	2,48
10	0	1,66	1,88	2,09	2,17	2,27	2,29	2,32
	0,5	1,65	1,87	2,07	2,16	2,26	2,28	2,31
	0,9	1,64	1,86	2,06	2,15	2,25	2,27	2,30

**Критические значения статистики Смирнова – Граббса
для минимального члена выборки $G_{\min}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 1,0$								
1	0	1,83	2,09	2,12	2,12	2,05	2,03	2,01
	0,5	1,87	2,10	2,17	2,13	2,11	2,11	2,04
	0,9	1,92	2,22	2,40	2,40	2,40	2,40	2,40
5	0	1,66	1,80	1,91	1,91	1,91	1,91	1,91
	0,5	1,70	1,86	1,93	1,93	1,93	1,93	1,93
	0,9	1,74	1,94	2,09	2,09	2,10	2,10	2,13
10	0	1,56	1,70	1,78	1,80	1,83	1,83	1,83
	0,5	1,60	1,72	1,82	1,83	1,85	1,85	1,85
	0,9	1,65	1,81	1,95	1,95	1,96	1,98	1,99
$C_s = 1,5$								
1	0	1,78	1,91	1,82	1,73	1,66	1,60	1,56
	0,5	1,83	1,93	1,87	1,81	1,70	1,66	1,61
	0,9	1,89	2,11	2,26	2,22	2,10	2,01	2,00
5	0	1,56	1,61	1,60	1,57	1,53	1,50	1,48
	0,5	1,62	1,69	1,68	1,62	1,58	1,54	1,51
	0,9	1,71	1,86	1,94	1,87	1,84	1,81	1,79
10	0	1,45	1,51	1,50	1,48	1,46	1,44	1,44
	0,5	1,52	1,57	1,56	1,53	1,51	1,48	1,47
	0,9	1,60	1,72	1,79	1,75	1,71	1,70	1,69

**Критические значения статистики Смирнова – Граббса
для минимального члена выборки $G_{\min}$
в зависимости от уровня значимости α %, r , n , C_s**

α , %	Коэффициент автокорреляции, r	Объем выборки, n						
		6	10	20	30	50	70	100
$C_s = 2,0$								
1	0	1,70	1,71	1,53	1,44	1,35	1,28	1,24
	0,5	1,78	1,77	1,62	1,53	1,43	1,35	1,31
	0,9	1,88	2,04	2,10	1,97	1,84	1,73	1,71
5	0	1,46	1,46	1,35	1,30	1,25	1,20	1,17
	0,5	1,54	1,54	1,44	1,36	1,29	1,25	1,22
	0,9	1,66	1,79	1,78	1,69	1,61	1,56	1,52
10	0	1,34	1,34	1,27	1,23	1,18	1,15	1,15
	0,5	1,43	1,43	1,35	1,29	1,23	1,20	1,17
	0,9	1,55	1,63	1,63	1,59	1,50	1,46	1,43
$C_s = 3,0$								
1	0	1,53	1,38	1,12	1,04	0,97	0,92	0,87
	0,5	1,65	1,52	1,28	1,15	1,04	0,96	0,92
	0,9	1,83	1,87	1,84	1,64	1,46	1,35	1,31
5	0	1,26	1,17	1,02	0,94	0,89	0,85	0,82
	0,5	1,38	1,28	1,10	1,03	0,94	0,89	0,85
	0,9	1,76	1,63	1,52	1,41	1,28	1,21	1,12
10	0	1,15	1,08	0,95	0,89	0,84	0,81	0,80
	0,5	1,25	1,18	1,02	0,95	0,88	0,85	0,82
	0,9	1,50	1,46	1,38	1,30	1,17	1,10	1,04

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Выборка 53
- Генеральная совокупность 52
- Дисперсия 17, 175
- Доверительный интервал 100, 104
- Закон распределения
 - случайной величины 8
 - случайного процесса 172
- Интервальная оценка 100
- Квантиль 21
- Клетчатка вероятности 46
- Ковариация 140
- Коэффициент
 - асимметрии 18
 - вариации 17
 - детерминации 145
 - корреляции
 - парной 139
 - множественной 163
 - ранговой Спирмэна 152
 - частный 166
 - модульный 21
 - регрессии 139, 161
- Критерии
 - однородности 107
 - случайности 129
 - согласия 119
- Критерий
 - Диксона 229
 - Зигеля -Тьюки 116
 - Колмогорова 125
 - наибольшей длины серий 132
 - общего числа серий 130
 - Смирнова-Граббса 229
 - Стьюдента 109
 - Уилкоксона-Манна-Уитни 112
 - Фишера 111
 - хи-квадрат (Пирсона) 120
 - числа повышений-понижений 133
 - числа экстремумов 134
 - эн-омега-квадрат 127
- Критическая область 104
- Ложная корреляция 167
- Математическое ожидание 15, 174
- Медиана 15
- Метод расчета параметров распределения
 - графический 83
 - графоаналитический 74
 - моментов 65
 - наибольшего правдоподобия 69
- Мода 14
- Моменты случайной величины 16
- Несмещенность 54
- Погрешности оценок параметров
 - распределения 88
 - уравнения линейной регрессии 145
- Распределение
 - Гумбеля 33
 - Джонсона 43
 - Колмогорова 125
 - Крицкого-Менкеля 41
 - логнормальное 28
 - нормальное 24
 - Пирсона III типа 36
 - равномерной плотности 27
 - Стьюдента (t-распределение) 97
 - Фишера (F-распределение) 99
 - хи-квадрат 96
- Реализация случайного процесса 171
- Сечение случайного процесса 169
- Случайная величина 8
- Случайный процесс 170
 - стационарный 178
 - эргодический 180
- Состоятельность 53
- Среднеквадратическое отклонение 17, 175
- Средняя линия 141
- Стандартная нормированная СВ 21
- Статистическая гипотеза 103
- Уравнение регрессии 137, 161
- Уровень значимости 106
- Функция
 - автокорреляционная 179
 - корреляционная 176
 - обеспеченностей 9
 - плотности вероятности 10
 - распределения (интегральная) 8
- Число степеней свободы 146
- Числовые характеристики СВ 14
- Экспесс 18
- Эмпирическая
 - кривая обеспеченностей 54
 - обеспеченность (формулы) 59

СОДЕРЖАНИЕ

Предисловие	3
Список условных обозначений и сокращений	5
Введение	7
1. Некоторые сведения из теории вероятностей	8
1.1. Случайные величины и функции распределения	8
1.2. Дискретные и непрерывные случайные величины	11
1.3. Случайная величина и время	13
1.4. Числовые характеристики случайных величин	14
1.4.1. Мода. Медиана. Математическое ожидание	14
1.4.2. Моменты случайной величины	16
1.4.3. Дисперсия. Среднеквадратическое отклонение. Коэффициент вариации	17
1.4.4. Асимметрия и эксцесс	18
1.4.5. Свойства числовых характеристик случайных величин	19
1.5. Стандартные преобразования случайной величины	21
1.6. Квантили распределения	21
2. Аналитические функции распределения, используемые в гидрологии	23
2.1. Нормальное распределение	24
2.2. Закон равномерной плотности	27
2.3. Логарифмически нормальное распределение	28
2.4. Закон распределения крайних членов выборки (распределение Гумбеля)	33
2.5. Распределение Пирсона III типа (непрерывное биномиальное распределение)	36
2.6. Распределение Крицкого – Менкеля (трехпараметрическое гамма-распределение)	41
2.7. Распределение Джонсона	43
2.8. Графическое представление функций распределения на клетчатке вероятности	46
2.9. Комментарии к разделу	48
3. Построение кривых обеспеченностей и оценка параметров распределения по эмпирическим данным	52
3.1. Эмпирические кривые обеспеченностей	54
3.2. Методы расчета оценок параметров распределения	65
3.2.1. Метод моментов	65
3.2.2. Метод наибольшего правдоподобия	69
3.2.3. Графоаналитический метод (метод квантилей)	74
3.2.3.1. Графоаналитический метод на основе кривой Пирсона III типа	74
3.2.3.2. Графоаналитический метод на основе трехпараметрической логнормальной кривой	77
3.2.3.3. Графоаналитический метод на основе кривой Крицкого- Менкеля	79
3.2.3.4. Комментарии	82
3.2.4. Графический метод	83
3.2.5. Оценка параметров распределения Джонсона	85
3.3. Оценка погрешностей выборочных параметров распределения	88
3.4. Экономические аспекты определения расчетных вероятностей	91
3.4.1. Цели определения расчетных вероятностей	91
3.4.2. Определение оптимальных расчетных вероятностей	92
4. Интервальное оценивание параметров и проверка статистических гипотез	96
4.1. Распределение χ^2 (хи-квадрат)	96

TABLE OF CONTENTS

Foreword	3
Notation and abbreviations	5
Introduction	7
1 Some Information From the Theory of Probabilities	8
1.1. Random variables and distribution functions	8
1.2. Discrete and continuous random variables.	11
1.3. Random variable and time	13
1.4. Numerical characteristics of random variables	14
1.4.1. Modal value. Median. Expectation	14
1.4.2. The moments of a random variable	16
1.4.3. Variance. Standard deviation. Spread coefficient	17
1.4.4. Skewness and kurtosis	18
1.4.5. Properties of numerical characteristics of random variables.	19
1.5. Standard transformations of a random variable	21
1.6. Quantiles of distribution	21
2. Analytical Distribution Functions Used in Hydrology	23
2.1. Normal distribution	24
2.2. The law of uniform frequency	27
2.3. Log-Normal distribution	28
2.4. The law of distribution of extreme terms of a sample (Gumbel distribution) ...	33
2.5. Pearson distribution of the third type (continuous binomial distribution)	36
2.6. Distribution of Kritsky – Menkel (three-parametric F -distribution)	41
2.7. Distribution of Johnson	43
2.8. Graphic performance of distribution functions on standard probability paper ..	46
2.9. Comments to the chapter	48
3. Construction of Cumulative Distribution Functions... Estimate of Distribution Parameters Based on the Empirical Data	52
3.1. Empirical cumulative distribution functions.	54
3.2. Methods of calculating of distribution parameters on empirical data	65
3.2.1. Method of the moments	65
3.2.2. Method of the maximum likelihood	69
3.2.3. Method of quantiles	74
3.2.3.1. Method of quantiles on the basis of the Pearson distribution of the third type	74
3.2.3.2. Method of quantiles on the basis of the three-parametric logarithmic normal distribution	77
3.2.3.3. Method of quantiles on the basis of the Kritsky – Menkel distribution	79
3.2.3.4. Comments	82
3.2.4. Graphic method.	83
3.2.5. Estimate of Johnson's distribution parameters	85
3.3. Estimate of errors of empirical parameters of distribution	88
3.4. Economical aspects of selection of calculated probabilities	91
3.4.1. The purposes of selection of calculated probabilities	91
3.4.2. Selection of optimum probabilities	92
4. Parameters Interval Estimation and Check of Statistical Hypotheses	96
4.1. Chi-squared distribution	96

4.2. t-distribution (Student's distribution)	97
4.3. F-distribution (Fisher's distribution)	99
4.4. Interval estimations of distribution parameters	100
4.4.1. Interval estimation of expectation	101
4.4.2. Interval estimation of variance	102
4.5. Statistical hypothesis. Test of a statistical hypothesis	103
4.6. Tests used for check of a homogeneity of hydrological series	107
4.6.1. Student's test for check of a significance of distinction of mean values of two samples	109
4.6.2. Test of equality of two variances (Fisher's test)	111
4.6.3. Rank-summarized test of Wilcoxon – Mann – Whitney	112
4.6.4. Rank test of scattering of Siegel – Tukey	116
4.7. Tests of consent	119
4.7.1. χ^2 -test (Pearson's test)	120
4.7.2. Kolmogorov's test	125
4.7.3. $n\omega^2$ -test (test of Cramer – Mises – Smirnov)	127
4.8. Tests for conformity to random variable	129
4.8.1. Criterion of sequences	130
4.8.1.1. Test of common number of sequences	130
4.8.1.2. Test of greatest length of sequence	132
4.8.2. Test of number of raises and lowerings	133
4.8.3. Test of number of extremums	134
5. Statistical Analysis of Associations Between Hydrological Variables	136
5.1. Least square method. The equation of a linear regression for two variables ...	137
5.2. Linearization of nonlinear associations	143
5.3. Estimate of an exactitude of the equation of a linear regression for two variables	145
5.4. Interval estimation and estimation of a significance of parameters of a linear regression for two variables	150
5.4.1. Interval estimation of coefficient of correlation Z-transformation of Fisher	150
5.4.2. Коэффициент ранговой корреляции Спирмэна	152
5.4.3. Interval estimation of a slope	156
5.4.4. Interval estimation of an intercept	157
5.4.5. F-test of significance of regression	158
5.4.6. Construction of a confidence interval for the equation of a linear regression	159
5.5. Multiple linear correlation	160
5.5.1. Estimate of an exactitude of the equation of a multiple linear regression	163
5.5.2. Selection of the "best" equation of a regression	164
5.5.3. Nonsense correlation	167
6. Random processes	168
6.1. Basic concepts	168
6.2. Classification of random processes "according to time" and "according to the state"	171
6.3. Distribution law and the basic characteristics of random processes.	172
6.4. Stationary random processes	178

6.5. Ergodic random processes	180
6.6. Elementary random processes	182
6.7. Canonical expansion of random processes	183
6.8. A spectral resolution of a stationary random process	184
6.9. Random processes with independent cuts	190
6.10. Markovian processes and Markov chains	191
6.11. Normal (Gauss) random processes	192
6.12. Periodically non-stationary random processes	192
6.13. Parametric models of time series	194
6.13.1. Autoregression model (AR)	194
6.13.2. Moving average model (MA)	196
6.13.3. The mixed models (ARMA)	197
6.13.4. Model of autoregression and integrated moving average (ARIMA)	198
6.14. Flattening and filtration	199
6.15. Determination of characteristics of random processes based on empirical datas	203
6.15.1. Determination of characteristics of random processes based on a set of realizations	204
6.15.2. Determination of characteristics of stationary ergodic process based on one realization	206
6.15.3. Determination of characteristics of periodically correlated random process based on one realization	209
6.15.4. A selective spectrum of a stationary ergodic process	214
6.15.5. Employment of methods of decomposition in estimating characteristics of random process	217
6.16. Modelling of simulated hydrological series	219
7. Guidelines from normative documents for statistical processing of hydrometeorological information	224
References	232
Appendices	234
1. The normalized ordinates of Pearson distribution of the third type (continuous binomial distribution)	235
2. Ordinates of distribution of Kritsky – Menkel	236
3. Ordinates of H-distribution (of A. Ezhov)	247
4. χ^2 -distribution	250
5. t -distribution (of Student)	251
6. F -distribution (of Fisher)	252
7. Distribution of Kolmogorov	255
8. Effect of a nonsense correlation for association $y=f(z)$ (according to Benson)	256
9. The table for finding of coefficient of a spread C_v by a method of the maximum likelihood depending on a parameter λ_2 and relation C_v/C_v (for distribution of Pearson of the third type)	256
10. Critical values of Dickson's statistics	260
11. Critical values of the Smirnov-Grabbs statistics	266
Subject index	272

Учебное издание

А.В. Сикан

МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ
ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Учебник

Редакторы: И.Г. Максимова, Т.В. Иващенко

ЛР № 020309 от 30.12.96.

Подписано в печать 28.03.07. Формат 60×90 1/16. Гарнитура Times New Roman.
Бумага офсетная. Печать офсетная. Усл. печ.л. 17,52. Уч.-изд. л. 17,52. Тираж 700 экз. Заказ № 20/07
РГГМУ, 195196, Санкт-Петербург, Малоохтинский пр., 98.
ЗАО «НПП «Система», 195112, Санкт-Петербург, Малоохтинский пр., 80/2.

the first of these is the fact that the system is not a simple one, and the second is the fact that the system is not a simple one.

the first of these is the fact that the system is not a simple one, and the second is the fact that the system is not a simple one.

the first of these is the fact that the system is not a simple one, and the second is the fact that the system is not a simple one.

the first of these is the fact that the system is not a simple one, and the second is the fact that the system is not a simple one.

the first of these is the fact that the system is not a simple one, and the second is the fact that the system is not a simple one.