

Ю. С. Харин, Н. М. Зуев, Е. Е. Жук

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ,
МАТЕМАТИЧЕСКАЯ
И ПРИКЛАДНАЯ СТАТИСТИКА**

Утверждено
Министерством образования Республики Беларусь
в качестве учебника для студентов
учреждений высшего образования
по специальностям «Прикладная математика», «Информатика»,
«Экономическая кибернетика», «Актuarная математика»,
«Компьютерная безопасность», «Статистика»

Минск
БГУ
2011

УДК
ББК
Х

Р е ц е н з е н т:

кафедра экономической кибернетики и теории вероятностей Гомельского государственного университета им. Ф. Скорины (зав. кафедрой доктор физико-математических наук, профессор Ю. В. Малинковский)

Харин Ю. С.

Х Теория вероятностей, математическая и прикладная статистика : учебник / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. – Минск: БГУ, 2011. – 464 с. – (Классическое университетское издание)

ISBN 985-485-440-.

Учебник содержит лекционный материал и упражнения по всем разделам теории вероятностей, теории случайных процессов, математической и прикладной статистики.

Для студентов высших учебных заведений по математическим специальностям.

УДК 519.2(075.8)

ББК

© Харин Ю. С.,
Зуев Н. М.,
Жук Е. Е.,
© БГУ, 2011

ПРЕДИСЛОВИЕ

Математическое моделирование разнообразных явлений окружающей действительности невозможно без учета случайных (стохастических) факторов. Теория вероятностей и математическая статистика – единственная математическая дисциплина, разрабатывающая и использующая методы математического моделирования случайных явлений. В связи с этим в вузовские учебные планы большинства современных учебных специальностей неизбежно входит дисциплина «Теория вероятностей и математическая статистика», содержание которой зависит от типа вуза, специальности и объема учебных часов. В современных классических и профильных университетах этот курс включает три приблизительно равновеликих (по объему учебного материала) модуля, иногда выделяемых в самостоятельные учебные дисциплины: «Теория вероятностей», «Математическая статистика» и «Теория случайных процессов».

Учебник содержит лекционный материал и упражнения по всем разделам этих модулей и предназначен для студентов, обучающихся по специальностям «Прикладная математика», «Информатика», «Экономическая кибернетика», «Актuarная математика», «Компьютерная безопасность», «Статистика», а также может быть полезен студентам других математических и статистических специальностей вузов.

Учебник состоит из 21 главы, которые сгруппированы в 5 частей.

В части 1, включающей пять глав, изложены базовые понятия по теории вероятностей и необходимый учебный материал по случайным событиям и их вероятностям, случайным величинам и их распределениям вероятностей, случайным последовательностям и предельным теоремам.

В части 2 рассматриваются основы теории случайных процессов, включая общие свойства случайных функций, стационарные процессы, случайные процессы с независимыми приращениями, марковские процессы и мартингалы.

В части 3 представлен учебный материал по теоретическим основам математической статистики.

Части 4, 5 содержат дополнительные главы по математической статистике.

В учебном процессе материал данного учебника рекомендуется использовать следующим образом. Части 1–3 изучаются как трехсеместровый базовый курс. Части 4, 5 следует использовать при организации спецкурса «Прикладная статистика» для студентов старших курсов либо как лекционный курс «Дополнительные главы математической и прикладной статистики» в магистратуре.

В конце глав приводятся упражнения, позволяющие студенту работать самостоятельно и оценить степень усвоения лекционного курса. Расширенный перечень задач, упражнений и тестовых заданий опубликован в [31].

В приложении представлены справочная информация по двадцати основным распределениям вероятностей одномерных и многомерных случайных величин, а также краткое описание «Электронного учебно-методического комплекса по теории вероятностей и математической статистике».

Следует отметить особенности, выделяющие данный учебник среди существующей отечественной учебной литературы по теории вероятностей и математической статистике:

- использование зарубежных литературных источников для подготовки упражнений;

- наличие «Электронного учебно-методического комплекса», подготовленного в БГУ (под руководством Ю. С. Харина), который представлен в приложении и предназначен для интенсивной самостоятельной работы студентов;
- включение в учебник разделов по прикладной статистике, что необходимо при практическом использовании методов теории вероятностей и математической статистики.

Содержание учебника полностью охватывает университетскую программу учебной дисциплины «Теория вероятностей и математическая статистика». Книга написана на основе 25-летнего опыта преподавания курса «Теория вероятностей и математическая статистика» на факультете прикладной математики и информатики Белорусского государственного университета.

Авторство в учебнике распределено следующим образом: Ю. С. Харин – главы 1–16, 18–21; Н. М. Зуев – главы 1–10; Е. Е. Жук – главы 13–21.

Авторы признательны рецензентам: доктору физико-математических наук профессору Ю. В. Малинковскому, доктору физико-математических наук профессору А. Д. Егорову за замечания и рекомендации по улучшению книги. Слова благодарности авторы адресуют доктору физико-математических наук профессору Г. А. Медведеву – основателю кафедры теории вероятностей и математической статистики БГУ за поддержку и методические советы при подготовке учебника, а также О. Г. Кадуриной и О. А. Куцепаловой за техническую помощь в оформлении рукописи в издательской системе \LaTeX .

Предложения и замечания просьба направлять авторам по адресу: факультет прикладной математики и информатики, Белорусский государственный университет, пр. Независимости, 4, Минск, 220050, Республика Беларусь; тел. +375 17 2095530; e-mail: kharin@bsu.by.

ВВЕДЕНИЕ

В истории развития теории вероятностей и математической статистики можно выделить три этапа.

Этап I. Возникновение теории вероятностей

Возникновение теории вероятностей как науки принято относить к середине XVII в. и связывать с именами Паскаля, Ферма, Гюйгенса. На этом этапе сформировался некоторый класс задач подсчета шансов в азартных играх, исходы которых к началу игры не могут быть достоверно вычислены. Методика решения таких необычных для математики задач с неопределенностью, случайностью была наиболее полно изложена Гюйгенсом в 1657 г. в книге «О расчетах в азартной игре».

Этап II. Применение аналитических методов в теории вероятностей

Опубликованная переписка Паскаля и Ферма по комбинаторным вычислениям шансов, а также указанная выше книга Гюйгенса привлекли внимание математиков. Началась разработка математических моделей, аналитических методов для решения новых задач о случайности. Наибольший вклад в создание «аналитики» теории вероятностей в этот период внесли: Я. Бернулли, который развил теорию схемы независимых испытаний; Лаплас, давший объяснение «чуду Лапласа» (стремление к нормальному закону распределения вероятностей суммы независимых, равномерно малых случайных слагаемых); Гаусс и Пуассон, получившие асимптотические распределения сумм случайных слагаемых. На этом этапе значительный вклад в развитие теории вероятностей внесла русская математическая школа: П. Л. Чебышев, А. А. Марков, А. М. Ляпунов, создавшие аппарат асимптотического анализа сумм случайных величин.

Этап III. Развитие теории вероятностей на аксиоматической основе

В 1933 г. академик А. Н. Колмогоров в книге «Основные понятия теории вероятностей» предложил аксиоматику, которая теперь носит его имя и которая не только систематизировала все ранее полученные результаты, но и явилась основой для развития теории случайных процессов, случайных полей, случайных множеств, а также математической статистики. С этого момента теория вероятностей и математическая статистика становятся математической наукой.

Математическая статистика занимается задачами, обратными задачам теории вероятностей: если в теории вероятностей для заданной вероятностной модели вычисляются вероятности некоторых событий, то в математической статистике на основе результатов наблюдения за некоторыми событиями необходимо принять решение (осуществить статистический вывод) о вероятностной модели исследуемого явления. Приведем пример Ж. Бертрана [16], показывающий, что хотя математическая статистика и является молодой математической наукой, но имеет глубокие исторические корни: «Однажды в Неаполе преподобный Галиани увидел человека из Базиликаты, который, встряхивая 3 игральные кости в чашке, держал пари, что выбросит 3 шестерки... Конечно, такая удача возможна. Однако человеку из Базиликаты это удалось и во второй раз, и пари повторилось. Он клал кости назад в чашку 3, 4, 5 раз и каждый раз выбрасывал 3 шестерки. “Черт возьми, – вскричал преподобный, – кости налиты свинцом!” И так оно и было». Это типичный пример статистического вывода, т. е. вывода, основанного на наблюдениях. Если бы кости были симметричны и не были бы «налиты свинцом», то наблюдаемое событие «5 раз подряд выпали шестерки на трех костях» имело бы ни-

чтожно малую вероятность $((1/6)^3)^5 = 6^{-15} \approx 4,7 \cdot 10^{-11}$. Поскольку это событие все-таки наступило, преподаватель сделал вывод, что модель симметричной кости не верна.

Математическая статистика наиболее интенсивно начала развиваться начиная с 30-х гг. XX в. Наибольший вклад в ее развитие внесли: Р. Фишер, К. Пирсон, Г. Крамер, А. Вальд, Э. Леман, С. Рао, М. Кендалл, А. Стьюарт, С. Уилкс, А. Н. Колмогоров, Б. В. Гнеденко, Н. В. Смирнов, Л. Н. Большев, Ю. В. Прохоров.

В настоящее время математическая наука «Теория вероятностей и математическая статистика» интенсивно развивается в следующих направлениях: случайные множества, случайные поля, стохастический анализ, случайные операторы, байесовская статистика, робастная статистика, непараметрическая статистика, квантовая статистика, статистический анализ данных, последовательный статистический анализ, анализ изображений. В наши дни теория вероятностей и математическая статистика, являясь математической основой моделирования объектов и явлений внешнего мира при наличии неопределенных и стохастических факторов, находит многочисленные и весьма плодотворные применения в технике, информатике, экономике и финансах, биологии и генетике, медицине, физике и химии, социальных науках и других областях. Издается более 300 научных журналов по теории вероятностей и математической статистике, ежегодно проводятся сотни международных конференций, во многих странах мира созданы статистические ассоциации. В нашей стране организатором Белорусской статистической ассоциации в 1998 г. явился Белорусский государственный университет.

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

Множества и пространства

\mathbb{N}	–	множество натуральных чисел
\mathbb{Z}	–	множество целых чисел
\mathbb{R}	–	множество действительных чисел
\mathbb{R}^n	–	n -мерное евклидово пространство
\mathbb{Q}	–	множество рациональных чисел
\mathbb{C}	–	множество комплексных чисел
$\{x_1, x_2, \dots\}$	–	множество, состоящее из элементов x_1, x_2, \dots
$\{x : x \in A\}$	–	множество элементов x , удовлетворяющих условию A
\emptyset	–	пустое множество
\cap	–	знак пересечения множеств
\cup	–	знак объединения множеств
\setminus	–	знак разности множеств
\subset	–	знак включения
$\Omega = \{\omega\}$	–	пространство элементарных событий
\mathcal{A}	–	алгебра событий
\mathcal{P}	–	семейство распределений
\mathcal{S}	–	алфавит классов
\mathcal{D}	–	множество решающих правил
$\mathcal{L}\{\xi\}$	–	закон распределения вероятностей ξ
\mathcal{F}	–	алгебра или σ -алгебра событий
(Ω, \mathcal{F})	–	измеримое пространство
(Ω, \mathcal{F}, P)	–	вероятностное пространство
$\overline{\lim} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$	–	верхний предел последовательности случайных событий (наступает бесконечно много событий среди $\{A_n\}$)
$\underline{\lim} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$	–	нижний предел последовательности случайных событий (наступают все $\{A_n\}$, за исключением их конечного числа)
$\mathcal{B} = \mathcal{B}(\mathbb{R})$	–	борелевская σ -алгебра подмножеств числовой прямой
$\mathcal{B}^m = \mathcal{B}(\mathbb{R}^m)$	–	борелевская σ -алгебра подмножеств \mathbb{R}^m
$(\mathbb{R}, \mathcal{B}, P_{\xi})$	–	вероятностное пространство, порожденное случайной величиной ξ

Функции

$m!$	–	факториал числа $m \in \mathbb{N}$, $0! = 1$
$C_n^m = \frac{n!}{m!(n-m)!}$, $0 \leq m \leq n$	–	биномиальный коэффициент
$ A $	–	модуль числа A , или длина вектора A , или определитель матрицы A , или мощность множества A
δ_{ij}	–	символ Кронекера ($\delta_{ij} = \{1, i = j; 0, i \neq j\}$)
$\mathbb{I}_A(x)$	–	индикаторная функция множества A ($\mathbb{I}_A(x) = \{1, x \in A; 0, x \notin A\}$)
$\mu(A) = \text{mes}_m(A)$	–	мера Лебега в \mathbb{R}^m ($A \subseteq \mathbb{R}^m$)
$P\{A\}$	–	вероятность случайного события A , вероятностная мера
$P\{A B\}$	–	условная вероятность события A при условии события B
$\xi = \xi(\omega)$	–	случайная величина (\mathcal{F} -измеримое отображение $\Omega \rightarrow \mathbb{R}$)
$P_\xi(B)$	–	распределение вероятностей случайной величины ξ
$F_\xi(x) = P\{\xi < x\}$	–	функция распределения случайной величины ξ
$\mathcal{L}\{\xi\}$	–	закон распределения случайной величины ξ
$p_\xi(x) = F'_\xi(x) = \frac{d}{dx}F_\xi(x)$	–	плотность распределения вероятностей ξ
$F_\xi(x C)$	–	условная функция распределения ξ при условии события C
$F_{\xi \eta}(x y)$	–	условная функция распределения случайной величины ξ при условии $\eta = y$
$p_{\xi_1} * p_{\xi_2}$	–	свертка плотностей распределения вероятностей ξ_1, ξ_2
$E\{\xi\}$	–	математическое ожидание случайной величины ξ
$E\{\xi; \xi \in B\}$	–	усеченное математическое ожидание
$E\{\xi B\}$	–	условное математическое ожидание ξ при условии B
$E\{\xi \eta = y\}$	–	условное математическое ожидание ξ при условии $\eta = y$
$E\{\xi \eta\}$	–	условное математическое ожидание ξ относительно η
$E\{\xi \mathcal{F}\}$	–	условное математическое ожидание ξ относительно σ -алгебры \mathcal{F}
$\overset{\circ}{\xi} = \xi - E\{\xi\}$	–	центрированная случайная величина
$\nu_k = E\{\xi^k\}$	–	начальный момент k -го порядка
$\mu_k = E\left\{\overset{\circ}{\xi}^k\right\}$	–	центральный момент k -го порядка
$D\{\xi\} = \sigma_\xi^2 = \mu_2$	–	дисперсия случайной величины
$\sigma_\xi = \sqrt{D\{\xi\}}$	–	среднеквадратичное (стандартное) отклонение
$\text{cov}\{\xi_i, \xi_j\} = \sigma_{ij} = E\left\{\overset{\circ}{\xi}_i \overset{\circ}{\xi}_j\right\}$	–	ковариация случайных величин
$\Sigma = \text{cov}\{\xi, \xi\} = E\left\{\overset{\circ}{\xi} \overset{\circ T}{\xi}\right\}$	–	ковариационная матрица случайного вектора $\xi = (\xi_i)$
$\text{corr}\{\xi_i, \xi_j\} = \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$	–	коэффициент корреляции случайных величин
$H\{\xi\}$	–	энтропия случайной величины ξ

$I\{\xi, \eta\}$	–	количество информации по Шеннону, содержащейся в η о ξ
M	–	мода распределения вероятностей
m	–	медиана распределения вероятностей
$F_{\xi}^{-1}(p)$	–	квантиль уровня p для случайной величины ξ
$f_{\xi}(t) = E\{e^{it\xi}\}$	–	характеристическая функция случайной величины ξ
$\psi_{\xi}(z) = E\{z^{\xi}\}$	–	производящая функция случайной величины ξ
$\Phi_{\xi}(t) = \ln f_{\xi}(t)$	–	кумулянтная функция случайной величины ξ
$\xi_n \xrightarrow{P} \xi$	–	сходимость по вероятности
$\xi_n \xrightarrow{\text{п. н.}} \xi$	–	сходимость почти наверное
$\xi_n \xrightarrow{L_r} \xi$	–	сходимость в среднем порядка r (L_r -сходимость)
$\xi_n \xrightarrow{\text{ср. кв.}} \xi$ ($l.i.m. \xi_n = \xi$)	–	сходимость в среднеквадратическом (L_2)
$\xi_n \xrightarrow{D} \xi$	–	сходимость по распределению
$F_n(\cdot) \Rightarrow F(\cdot)$	–	слабая сходимость последовательности функций распределения
$w(t)$	–	стандартный винеровский процесс
$F(\lambda)$	–	спектральная функция случайного процесса
$f(\lambda)$	–	спектральная плотность случайного процесса
$R(t, u)$	–	корреляционная функция случайного процесса
$r(t, u)$	–	ковариационная функция случайного процесса
$N_N(\mu, \Sigma)$	–	N -мерный нормальный (гауссовский) закон распределения вероятностей с вектором математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицей Σ
$n_N(x \mu, \Sigma)$	–	плотность N -мерного нормального распределения $N_N(\mu, \Sigma)$, вычисленная в точке $x \in \mathbb{R}^N$: $n_N(x \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^N \Sigma }} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$
$\Phi(\cdot), \Phi^{-1}(\cdot)$	–	функция и квантиль распределения вероятностей стандартного нормального (гауссовского) закона $N_1(0, 1)$
$\phi(\cdot)$	–	плотность распределения вероятностей стандартного нормального закона: $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$, $z \in \mathbb{R}$
χ_m^2	–	закон χ^2 -распределения с m степенями свободы
$F_{\chi_m^2}(\cdot), F_{\chi_m^2}^{-1}(\cdot)$	–	функция и квантиль χ^2 -распределения с m степенями свободы
t_m	–	закон распределения Стьюдента с m степенями свободы
$F_{t_m}(\cdot), F_{t_m}^{-1}(\cdot)$	–	функция и квантиль t -распределения Стьюдента с m степенями свободы
$F_{m,n}$	–	закон распределения Фишера с m и n степенями свободы
$F_{m,n}(\cdot), F_{m,n}^{-1}(\cdot)$	–	функция и квантиль F -распределения Фишера с m и n степенями свободы
$W_N(\Sigma, m)$	–	закон распределения Уишарта с параметрами: $(N \times N)$ -матрицей Σ ($ \Sigma \neq 0$) и $m \geq 1$
$O(\cdot), o(\cdot)$	–	символы Ландау

Матрицы

$I_n = (\delta_{ij})$	–	единичная матрица порядка n
$\mathbf{1}_M$	–	M -вектор, все компоненты которого равны 1
$\mathbf{0}_M, \mathbf{0}_{M \times m}$	–	нулевые M -вектор и $(M \times m)$ -матрица
B^T	–	транспонированная матрица B
$\text{tr}(B)$	–	след матрицы B
$\text{rank}(B)$	–	ранг матрицы B
B^{-1}	–	обращение квадратной матрицы B

Сокращения

АР(p)	–	авторегрессия порядка p
АРПСС(p, d, q)	–	авторегрессия и проинтегрированное скользящее среднее порядков p, d и q
АРСС(p, q)	–	авторегрессия и скользящее среднее порядков p и q
БРП	–	байесовское решающее правило
ВП	–	винеровский процесс
ВР	–	временной ряд
ДСВ	–	дискретная случайная величина
ЗБЧ	–	закон больших чисел
КОП	–	критерий отношения правдоподобия
ММП	–	метод максимального правдоподобия
МНК	–	метод наименьших квадратов
н. о. р.	–	независимые одинаково распределенные
ОМК	–	оценка минимального контраста (МК-оценка)
ОММ	–	оценка по методу моментов
ОМП	–	оценка максимального правдоподобия (МП-оценка)
ПРВ	–	плотность распределения вероятностей
РВ	–	распределение вероятностей
РИ	–	равномерная интегрируемость
РП	–	решающее правило
СВ	–	случайная величина
СЛАУ	–	система линейных алгебраических уравнений
СП	–	случайный процесс
СС(q)	–	последовательность скользящего среднего порядка q
УЗБЧ	–	усиленный закон больших чисел
ФР	–	функция распределения
ЦДИ	–	центральный доверительный интервал
ЦПТ	–	центральная предельная теорема
□	–	конец доказательства

ЧАСТЬ 1

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Глава 1

СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

1.1. ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Определение 1.1. Теория вероятностей – математическая наука, изучающая математические модели случайных экспериментов.

Любой эксперимент S характеризуется двумя факторами:

- 1) U (комплекс условий, при котором S происходит);
- 2) результаты эксперимента, т. е. некоторые события A_1, A_2, \dots , наступление или ненаступление которых регистрируется в ходе проведения S с помощью некоторых приборов.

Для любого эксперимента комплекс условий U влияет на результат: $U \rightarrow A_i$. В зависимости от типа этого влияния существует следующая классификация типов экспериментов. Всевозможные эксперименты делятся на детерминированные и недетерминированные, среди последних, в свою очередь, различают случайные эксперименты и прочие.

Определение 1.2. Эксперимент S называется детерминированным, если комплекс условий U однозначно определяет результат (исход) эксперимента. В противном случае эксперимент называется недетерминированным.

Приведем примеры S_1 – S_5 и дадим их классификацию.

S_1 : $U = \{\text{дистиллированная вода, объем 1 л, } p = 760 \text{ мм рт. ст., } T = 20^\circ \text{C}\}$, $A_1 = \{\text{вода в жидком состоянии}\}$, $A_2 = \{\text{вода в твердом состоянии}\}$, $A_3 = \{\text{вода в газообразном состоянии}\}$. Очевидно: $U \rightarrow A_1, A_2, A_3$ – однозначно. Таким образом, S_1 – детерминированный эксперимент.

Определение 1.3. Событие A , которое при данном комплексе условий U неизбежно наступает, называется достоверным событием. Событие B , которое неизбежно не наступает, называется невозможным событием.

События A_2, A_3 (в данном примере) – невозможные события, а событие A_1 – достоверное.

В детерминированных экспериментах события либо достоверны, либо невозможны.

Приведем теперь примеры недетерминированных экспериментов.

S_2 : $U = \{\text{экзаменатор перемешивает 40 экзаменационных билетов, студент наудачу извлекает 1 билет}\}$, $A = \{\text{номер извлеченного билета четный}\}$. $U \rightarrow A$ или \bar{A} ;

S_3 : $U = \{\text{над плоской поверхностью наудачу бросают симметричную монету (герб – решетка)}\}$, $A = \{\text{выпал герб}\}$. $U \rightarrow A$ или \bar{A} ;

S_4 : $U = \{\text{рассматривается рабочий день филиала акционерного коммерческого банка «Prior»}\}$, $A = \{\text{число посетителей} = k\}$, где k – некоторое целое число. $U \rightarrow A$ или \bar{A} ;

S_5 : $U = \{\text{аудитория 513 главного корпуса БГУ, 2056 г., 7 февраля, 9.00 утра}\}$, $A = \{\text{в аудитории находится 10 студентов}\}$. $U \rightarrow A$ или \bar{A} .

Определение 1.4. Недетерминированный эксперимент S называется случайным экспериментом, если выполняются следующие два свойства:

- 1) он допускает массовое, n -кратное ($n \rightarrow \infty$) повторение;
- 2) выполняется свойство статистической устойчивости частот.

Свойства статистической устойчивости частот

Пусть эксперимент S осуществлен n -кратно (серия из n независимых экспериментов). Обозначим:

$m_n(A)$ – число наступлений события A в эксперименте (абсолютная частота A);

$\nu_n(A) = \frac{m_n(A)}{n}$ – относительная частота A , $0 \leq \nu_n(A) \leq 1$.

Свойство статистической устойчивости состоит в том, что для любого события A при возрастании n последовательность относительных частот сходится к некоторому пределу:

$\nu_n(A) \rightarrow P = P(A)$, где $P(A) \in [0, 1]$ – вероятность случайного события A .

Для экспериментов S_2, S_3, S_4 свойство n -кратного повторения выполняется, а для S_5 – нет, поэтому S_5 не является случайным экспериментом.

Проверим свойство статистической устойчивости для S_2, S_3, S_4 .

S_2 : $\nu_n(A) \rightarrow 1/2$ (т. е. приблизительно в половине случаев выпадает четный номер билета).

S_3 : $\nu_n(A) \rightarrow 1/2$.

S_4 : $\nu_n(A) \rightarrow P(A) = (\lambda^k e^{-\lambda})/k!$, $\lambda > 0$, $k \in \mathbb{N}$, где λ – некоторый параметр, который характеризует определенный филиал данного банка. Это установлено статистической обработкой архивных наблюдений.

В случайном эксперименте приходится иметь дело не только с достоверными и невозможными событиями, но и со случайными событиями.

Задание 1.1. Прочитать историю возникновения и развития теории вероятностей [13, 26], Введение в данном учебнике.

Задание 1.2. Проверить свойства статистической устойчивости частот в следующем компьютерном эксперименте: наудачу бросается точка в отрезок $[0; 1]$. $A = \{\text{точка попала на } [0; 0,5]\}$. Вычислить $\nu_n(A)$, задав n , и построить график зависимости ν_n от n . Использовать компьютерные датчики случайных чисел.

1.2. СЛУЧАЙНЫЕ СОБЫТИЯ И СООТНОШЕНИЯ МЕЖДУ НИМИ

Существует конструктивная идея для построения математической модели случайных событий: определить некоторый полный набор всевозможных простейших (элементарных) исходов случайных экспериментов S и построить интересующее нас событие из этих элементарных исходов.

Определение 1.5. События, составляющие множество $\{\omega\}$ простейших исходов случайного эксперимента S , называются элементарными событиями (ЭС), если:

- 1) все эти элементарные события различны;
- 2) наступление одного из исходов исключает наступление всех остальных;
- 3) в ходе эксперимента одно из элементарных событий неизбежно наступает.

При этом множество $\Omega = \{\omega\}$, составленное из всех этих элементарных событий, называется пространством элементарных событий.

Определение 1.6. Объединение A некоторых элементарных событий из Ω называется случайным событием. Иначе говоря, случайное событие – подмножество пространства элементарных событий: $A \subseteq \Omega$. При этом $A = \Omega$ – достоверное событие (включает все элементарные события); $A = \emptyset$ – невозможное событие (не содержит ни одного случайного события).

Замечание 1.1. A наступает всякий раз, когда наступает некоторое элементарное событие $\omega' \in A$. Если $\omega' \in A$, то говорят, что элементарное событие (исход) ω' благоприятствует наступлению случайного события A .

Замечание 1.2. Элементарные события и пространство элементарных событий строятся неоднозначно.

Примеры пространств элементарных событий

1. Бросание наудачу одной монеты (см. с. 17, пример S_3): $\omega_1 = \{\Gamma\} = \{\text{монета выпала гербом кверху}\}$, $\omega_2 = \{P\} = \{\text{монета выпала решеткой кверху}\}$, $\Omega = \{\omega_1, \omega_2\}$, $|\Omega| = 2$.
2. Бросание наудачу двух монет: $\omega_1 = \{(\Gamma, \Gamma)\}$, $\omega_2 = \{(\Gamma, P)\}$, $\omega_3 = \{(P, \Gamma)\}$, $\omega_4 = \{(P, P)\}$, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $|\Omega| = 4$.
3. Экзаменатор – студент – $2N$ билетов (см. с. 17, пример S_2): $\omega_i = \{\text{номер извлеченного билета} = i\}$, $i = 1, \dots, 2N$, $\Omega = \{\omega_1, \omega_2, \dots, \omega_{2N}\}$, $A = \{\omega_2, \omega_4, \dots, \omega_{2N}\} \subset \Omega$, $|A| = N$.
4. АКБ «Prior» (см. с. 17, пример S_2): $\omega_i = \{\text{число посетителей} = i\}$, $i = 0, 1, 2, \dots$, $\Omega = \{\omega_0, \omega_1, \dots\}$ – счетное множество, $A = \{\text{число посетителей больше } k\} = \{\omega_{k+1}, \omega_{k+2}, \dots\}$ – счетное подмножество.
5. Имеется некоторая производственная коммерческая фирма. Нас интересует ее ежедневный доход: $\omega_x = \{\text{доход за один день составил величину } x\}$, где $x \in \mathbb{R}$, $\Omega = \{\omega_x : x \in \mathbb{R}\}$; $A = \{\text{доход меньше числа } z\} = \{\omega_x : x < z\}$.

Зададим соотношение между событиями с помощью табл. 1.1, используя язык теории множеств. Пусть $A, B, C \subseteq \Omega$.

Таблица 1.1

Обозначение	Название	Пояснение
$B \subset C$	Событие B влечет C	C наступает всякий раз, когда наступает B
$B = C$	События B и C эквивалентны: $B \subset C$ и $C \subset B$	B наступает тогда и только тогда, когда наступает C
$A = \overline{B}$	A есть событие, противоположное B	A наступает тогда и только тогда, когда не наступает B
$A = B \cap C$	A есть произведение событий B и C	A наступает тогда и только тогда, когда B и C наступают вместе
$B \cap C = \emptyset$	B и C несовместны	Совместное наступление B и C невозможно
$A = B \cup C$	A есть сумма B и C	A наступает тогда и только тогда, когда наступает B или C или оба вместе
$A = B \setminus C = B \cap \overline{C}$	A есть разность событий B минус C	A наступает тогда и только тогда, когда наступает B и не наступает событие C
$A = B \Delta C$	A есть симметричная разность B и C	A наступает тогда и только тогда, когда наступает только одно из событий: либо B , либо C

Правило де Моргана

Пусть событие A есть результат применения к событиям B, C, D, \dots действий: \cap, \cup, \subset . Тогда, чтобы получить \overline{A} , достаточно все события поменять на противоположные: $\overline{B}, \overline{C}, \overline{D}, \dots$, а действия поменять по схеме $\cup \rightarrow \cap, \cap \rightarrow \cup, \subset \rightarrow \supset$.

1.3. ПОНЯТИЕ ВЕРОЯТНОСТИ. ПРОСТЕЙШИЕ ВЕРОЯТНОСТНЫЕ МОДЕЛИ

Рассмотрим три типа простейших вероятностных моделей: классическую, дискретную, геометрическую.

Классическая вероятностная модель – это математическая модель простейших случайных экспериментов.

Пример 1.1. Экзаменатор – студент – N билетов: $\Omega = \{\omega_1, \dots, \omega_N\}$, $\omega_i = \{\text{номер извлеченного билета} = i\}$, $i = 1, \dots, N$.

Свойства классической вероятностной модели

1. Число элементарных событий конечно: $N < \infty$.
2. Предположим, данный эксперимент осуществлен n -кратно и каждый раз регистрировалось наступление или ненаступление события $A \subseteq \Omega$, следовательно, для любого события $A \subseteq \Omega$ мы можем поставить в соответствие число $\nu_n = \nu_n(A) = m_n(A)/n$ – относительная частота наступления события A в этой серии экспериментов. Заметим, что

$$\nu_n(\emptyset) = 0, \nu_n(\Omega) = 1, 0 \leq \nu_n(A) \leq 1, \nu_n(A) \rightarrow P = P(A).$$

3. Для любых несовместных случайных событий $A, B \subseteq \Omega$ ($A \cap B = \emptyset$) функция $\nu_n(A)$ аддитивна, т. е.

$$\nu_n(A \cup B) = \frac{m_n(A \cup B)}{n} = \frac{m_n(A) + m_n(B)}{n} = \nu_n(A) + \nu_n(B).$$

4. Все N элементарных событий равновозможны, т. е. $\nu_n(\omega_i) \rightarrow 1/N$, $i = 1, \dots, N$.

Эти четыре свойства лежат в основе определения *классической вероятностной модели*, задаваемой следующими четырьмя аксиомами.

A 1. Пространство элементарных событий конечно: $\Omega = \{\omega_1, \dots, \omega_N\}$, $N < \infty$.

A 2. Каждому случайному событию $A \in \Omega$ поставлено в соответствие такое число P , что

$$P = P(A), 0 \leq P(A) \leq 1, P(\emptyset) = 0, P(\Omega) = 1.$$

При этом число P называется *вероятностью события A* , а функция $P(\cdot)$, заданная на подмножествах Ω , – *вероятностной функцией*.

A 3 (аксиома конечной аддитивности). Для любых несовместных случайных событий $A, B \in \Omega$ ($A \cap B = \emptyset$) вероятность суммы событий равна сумме вероятностей этих событий:

$$P(A \cup B) = P(A) + P(B).$$

A 4 (аксиома равновероятности). Все N элементарных событий имеют равную вероятность: $P(\omega_1) = P(\omega_2) = \dots = P(\omega_N) = \text{const} = p$, при этом число $p \in (0; 1)$ называется *элементарной вероятностью*.

Замечание 1.3. $P = P(A)$ – функция, заданная на подмножествах $A \subseteq \Omega$. Покажем, что данная система аксиом является полной и однозначно определяет вероятностную функцию $P = P(A)$.

Теорема 1.1. Для классической вероятностной модели, определяемой аксиомами A1–A4, $p = 1/N$ и для любого случайного события $A \subseteq \Omega$ вероятность определяется соотношением

$$P = P(A) = \frac{|A|}{|\Omega|} = \frac{M}{N}, \quad (1.1)$$

где $N = |\Omega|$ – полное число элементарных событий; $M = |A|$ – число элементарных событий, благоприятствующих наступлению A .

Доказательство. Выберем произвольное случайное событие $A \subseteq \Omega$, состоящее из M элементарных событий ($0 \leq M \leq N$):

$$A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_M}\} = \bigcup_{j=1}^M \omega_{i_j}, \quad (1.2)$$

где $1 \leq i_1 < i_2 < \dots < i_M \leq N$ – номера элементарных событий, образующих A .

Рассмотрим $\bigcup_{j=1}^M \omega_{i_j}$ как сумму одноточечных множеств. Применим к (1.2) аксиомы А3, А4:

$$P(A) = P\left(\bigcup_{j=1}^M \omega_{i_j}\right) = \sum_{j=1}^M P(\omega_{i_j}) = Mp. \quad (1.3)$$

Положим в (1.3) $A = \Omega$, тогда в силу (1.2) $M = N$, $P(\Omega) = Np = 1 \Rightarrow p = 1/N$. Подставляя полученный результат в (1.3), получим соотношение (1.1). \square

Следствие 1.1. Вероятность события A в классической вероятностной модели равна отношению числа M элементарных событий, благоприятствующих наступлению события A , к общему числу N элементарных событий.

Замечание 1.4. Для вычислений $|A|, |\Omega|$ в формуле (1.1) целесообразно использовать правила комбинаторики.

Пример 1.2 (гипергеометрическое распределение). В одном непрозрачном сосуде тщательно перемешано K однотипных шаров, среди которых k ($k \leq K$) красных и $(K - k)$ белых. Наудачу извлекается комплект из L шаров (без учета порядка извлечения). Вычислить вероятность события $A_l = \{\text{среди } L \text{ извлеченных шаров имеется } l \text{ красных}\}$ для $l = 0, 1, 2, \dots$.

Шаг 1. Построим пространство элементарных событий Ω и вероятностную меру. Мысленно пронумеруем все шары: $\omega_i = \{i\text{-й вариант извлечения комплекта шаров из } K \text{ имеющихся без учета порядка извлечения}\}$, $i = 1, \dots, N$, $N = C_K^L$.

Очевидно, что аксиомы А1–А3 выполняются. Учитывая постановку задачи, все исходы равновозможны, значит, выполняется А4. Поскольку все аксиомы выполняются, то используем классическую вероятностную модель.

Шаг 2. Вычислим, согласно (1.1), $P_l = P(A_l) = |A_l|/N = M_l/N = M_l/C_K^L$, где $M_l = C_k^l C_{K-k}^{L-l}$; $0 \leq l \leq \min\{k, L\}$; $l > \min\{k, L\} \Rightarrow A_l = \emptyset \Rightarrow M_l = 0$. Получаем

$$P_l = \begin{cases} P(A_l) = \frac{C_k^l C_{K-k}^{L-l}}{C_K^L}, & l = 0, 1, \dots, \min\{k, L\}, \\ 0, & l > \min\{k, L\}. \end{cases} \quad (1.4)$$

Определение 1.7. Набор вероятностей $\{P_l\}$, определенный формулой (1.4), называется гипергеометрическим распределением вероятностей.

Задание 1.3. Показать, что

$$\sum_{l=0}^{\min\{k, L\}} P_l = 1.$$

Пример 1.3. Студент ФПМИ заполнил карточку спортлото 6 из 49. Вычислить вероятности угадывания l номеров в очередном тираже: $K = 49$, $L = 6$, $k = 6$, $l = 0, 1, 2, \dots, 6$.

Задание 1.4. По формуле (1.4) вычислить шансы студента в примере 1.3.

Задание 1.5. Рассмотреть обобщение предыдущего упражнения на случай, когда для одного и того же тиража заполнено несколько карточек. Вычислить искомые вероятности методом имитационного моделирования при заполнении нескольких карточек.

Дискретная вероятностная модель есть обобщение классической вероятностной модели для случаев, когда пространство элементарных событий счетно или элементарные события не равновероятны. *Дискретная вероятностная модель* определяется следующей системой трех аксиом (штрих обозначает обобщение аксиомы):

A1'. Ω – дискретное множество: $\Omega = \{\omega_1, \dots, \omega_N\}$, $N \leq \infty$.

A2. Без изменений.

A3' (аксиома счетной аддитивности). Для любой последовательности $A_1, A_2, \dots \subseteq \Omega$ попарно несовместных случайных событий ($A_i \cap A_j = \emptyset, i \neq j$) выполняется соотношение

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Заметим, что аксиома A4 здесь снимается, так как при $N = \infty$ она не имеет смысла: $p = 0$. Вероятность i -го события – $P(\omega_i) = p_i$ ($i = 1, \dots, N$) – называется *i -й элементарной вероятностью*.

Теорема 1.2. В рамках дискретной вероятностной модели, определяемой аксиомами A1', A2, A3', вероятность случайного события A равна сумме ряда, составленного из тех элементарных вероятностей, которые соответствуют элементарным событиям, входящим в A (благоприятствующим A):

$$P = P(A) = \sum_{k: \omega_k \in A} p_k, \quad A \subseteq \Omega. \quad (1.5)$$

Доказательство. Выберем произвольное случайное событие $A \subseteq \Omega$, состоящее из M элементарных событий, $M \leq \infty$:

$$A = \{\omega_{i_1}, \dots, \omega_{i_M}\} = \bigcup_{j=1}^M \omega_{i_j}.$$

Пользуясь аксиомой A3', имеем

$$P(A) = \sum_{j=1}^M P(\omega_{i_j}) = \sum_{j=1}^M p_{i_j} = \sum_{k: \omega_k \in A} p_k,$$

что совпадает с (1.5). \square

Следствие 1.2. Набор всех элементарных вероятностей удовлетворяет условию нормировки:

$$\sum_{k=1}^N p_k = 1.$$

Доказательство. Положим в соотношении (1.5) $A = \Omega$: $P(\Omega) = \sum_{k=1}^N p_k = 1$. \square

Пример 1.4. (Продолжение примера S_4 из п. 1.1.) Эта модель согласуется со следующей моделью Пуассона:

$$p_k = P(\omega_k) = \frac{\lambda e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots,$$

где $\lambda > 0$ – среднее число посетителей.

Задание 1.6. Вычислить набор вероятностей $\{p_k\}$ при $\lambda = 10$ и проверить условие нормировки.

Геометрическая вероятностная модель – это обобщение классической вероятностной модели на случай, когда пространство элементарных событий Ω есть ограниченное подмножество m -мерного евклидова пространства \mathbb{R}^m , $m = 1, 2, \dots$.

Здесь также предполагается, что все $\omega \in \Omega$ равновозможны. Но воспользоваться формулой (1.1) нельзя, так как нет возможности сосчитать, сколько точек на отрезке ($m = 1$) или на плоской фигуре ($m = 2$) и т. д., т. е. нельзя вычислить мощность $A \subseteq \Omega$.

Конструктивная идея: вместо понятия $|A|$ ввести в этой ситуации меру множества A : $|A| = \mu(A)$.

Определение 1.8. Числовая функция $\mu = \mu(A)$, определенная на подмножествах $A \subseteq \Omega$, называется мерой, если выполняются следующие свойства:

- 1) неотрицательность: $\mu(A) \geq 0$;
- 2) ограниченность: $\mu(\Omega) < \infty$;
- 3) счетная аддитивность: $\mu(\cdot)$ удовлетворяет свойству, аналогичному АЗ'.

При $m = 1$ $\mu(A)$ – длина отрезка A , при $m = 2$ $\mu(A)$ – площадь плоской фигуры A , при $m = 3$ $\mu(A)$ – объем тела A и т. д.

В \mathbb{R}^m будем использовать меру Лебега: $\mu(A) = \text{mes}_m(A)$. В рамках геометрической вероятностной модели вероятность случайного события $A \subseteq \Omega$ определяется следующим соотношением:

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}. \quad (1.6)$$

Замечание 1.5. Не для всех подмножеств A существует понятие длины, площади, объема и т. д. Здесь требуется, чтобы $A \subseteq \Omega$ было измеримо по Лебегу.

Пример 1.5 (жребий в игре «Что? Где? Когда?»). Отсчитываем угловое положение остановившейся стрелки – ω ; $\Omega = [0; 2\pi)$. Случайное событие $A = \{\text{стрелка остановилась в круговом секторе } [\psi_1, \psi_2)\} = [\psi_1, \psi_2)$. Это иллюстрирует рис. 1.1. При этом согласно формуле (1.6)

$$P(A) = \frac{\psi_2 - \psi_1}{2\pi}.$$

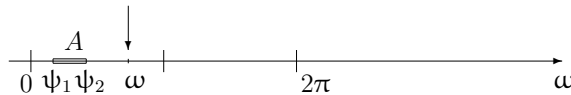


Рис. 1.1. Пример геометрической вероятностной модели

1.4. АЛГЕБРА, σ -АЛГЕБРА И ИХ СВОЙСТВА. ИЗМЕРИМОЕ ПРОСТРАНСТВО

Определение 1.9. Пусть Ω – произвольное пространство элементарных событий, тогда некоторая система \mathcal{F} подмножеств из Ω называется алгеброй случайных событий, если выполняются следующие свойства:

- 1) $\Omega \in \mathcal{F}$;
- 2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$;
- 3) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$.

Очевидно, справедливы еще два свойства:

- 4) $\emptyset \in \mathcal{F}$;
- 5) $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.

Доказательство. По правилу де Моргана: противоположность произведению – сумма, а она принадлежит \mathcal{F} . \square

Следствие 1.3. Алгебра замкнута относительно конечного числа операций \cap , \cup , $\bar{\cdot}$.

Доказательство. Из определения 1.9 и следствия 1.3: эти операции не выведут нас за пределы \mathcal{F} . \square

Пусть теперь Ω – бесконечное множество.

Определение 1.10. Алгебра \mathcal{F} подмножеств из Ω называется σ -алгеброй, если свойства 3) и 5) выполняются в обобщенном виде для счетного множества событий:

$$\forall A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}, \quad \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Следствие 1.4. σ -Алгебра замкнута относительно счетного множества операций \cap , \cup , $\bar{\cdot}$.

Определение 1.11. Измеримым пространством называется пара математических объектов (Ω, \mathcal{F}) , где Ω – пространство элементарных событий, \mathcal{F} – алгебра или σ -алгебра подмножеств из Ω .

Приведем примеры измеримых пространств в порядке возрастания их сложности.

1. **Измеримое пространство $(\Omega, 2^\Omega)$.** Здесь Ω – дискретное множество (конечное либо счетное), $\mathcal{F} = 2^\Omega$ – множество всех подмножеств из Ω . В частности, если $|\Omega| = N < \infty$, то $|\mathcal{F}| = 2^N$. Отсюда и происходит обозначение этого измеримого пространства.
2. **Измеримое пространство $(\mathbb{R}, \mathcal{B})$.** Здесь $\Omega = \mathbb{R}$ – числовая прямая. Для построения σ -алгебры на числовой прямой выберем $\forall x \in \mathbb{R}$ и обозначим полуограниченный числовой промежуток: $A_x = (-\infty, x) \subset \Omega$.

Определение 1.12. Назовем базовой системой множеств на числовой прямой следующую бесконечную систему интервалов: $\mathcal{F}_0 = \{\emptyset, A_x = (-\infty, x) : x \in \mathbb{R}\}$.

Заметим: \mathcal{F}_0 не является даже алгеброй, так как не выполняется свойство 2 алгебры событий: $\overline{A_x} = [x, +\infty) \notin \mathcal{F}_0$.

Определение 1.13. σ -Алгеброй $\sigma(\mathcal{F}_0)$, порожденной некоторой системой множеств \mathcal{F}_0 , называется наименьшая σ -алгебра, содержащая \mathcal{F}_0 :

$$\sigma(\mathcal{F}_0) = \bigcap_{\alpha} \mathcal{F}^{(\alpha)},$$

где $\mathcal{F}^{(\alpha)}$ – σ -алгебра, содержащая \mathcal{F}_0 .

Определение 1.14. Борелевской (в честь французского математика Э. Бореля) σ -алгеброй на числовой прямой называется σ -алгебра, порожденная системой интервалов $\mathcal{F}_0 = \{\emptyset, (-\infty, x) : x \in \mathbb{R}\}$: $\mathcal{B} = \sigma(\mathcal{F}_0)$.

Подмножества числовой прямой, которые принадлежат борелевской σ -алгебре \mathcal{B} , называются борелевскими множествами.

Свойства борелевской σ -алгебры \mathcal{B}

C1. Для $\forall x \in \mathbb{R}$ одноточечное множество $\{x\}$ является борелевским множеством: $\{x\} \in \mathcal{B}$.

Доказательство. Представим множество $\{x\}$ следующим образом:

$$\{x\} = \bigcap_{n=1}^{\infty} \left[x, x + \frac{1}{n} \right) = \bigcap_{n=1}^{\infty} \left(A_{x+\frac{1}{n}} \setminus A_x \right) = \bigcap_{n=1}^{\infty} \left(A_{x+\frac{1}{n}} \cap \overline{A_x} \right).$$

По построению полученное выражение принадлежит \mathcal{B} . \square

C2. Борелевскими множествами являются произвольные числовые промежутки вида (a, b) , $[a, b)$, $(a, b]$, $[a, b]$, где $a, b \in \mathbb{R}$, $a < b$.

C3. Множество рациональных чисел является борелевским: $\mathbb{Q} \in \mathcal{B}$.

Доказательство. $\mathbb{Q} = \bigcup_{n=1}^{\infty} \{q_n\}$ – счетная сумма одноточечных множеств, следовательно, $\mathbb{Q} \in \mathcal{B}$. □

C4. Множество иррациональных чисел – борелевское: $\overline{\mathbb{Q}} \in \mathcal{B}$.

Доказательство. Множество иррациональных чисел принадлежит \mathcal{B} как дополнение к \mathbb{Q} . □

C5. В выборе базовой системы \mathcal{F}_0 имеется произвол:

$$\begin{aligned} \mathcal{B} &= \sigma((a, b) : a, b \in \mathbb{R}) = \sigma([a, b) : a, b \in \mathbb{R}) = \\ &= \sigma((a, b] : a, b \in \mathbb{R}) = \sigma([a, b] : a, b \in \mathbb{R}). \end{aligned}$$

Таким образом, борелевская σ -алгебра достаточно богата, хотя беднее множества всех подмножеств числовой прямой.

Замечание 1.6. Существует обобщение измеримого пространства $(\mathbb{R}, \mathcal{B})$ на многомерный случай: измеримое пространство $(\mathbb{R}^m, \mathcal{B}^m)$, т. е. в m -мерном пространстве строится своя борелевская σ -алгебра \mathcal{B}^m . Строится она по аналогичной схеме, в качестве базовой системы \mathcal{F}_0 берется система параллелепипедов в \mathbb{R}^m :

$$\mathcal{F}_0 = \{\emptyset, A_x = A_{x_1} \times A_{x_2} \times \dots \times A_{x_m} : x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{R}^m\}.$$

1.5. АКСИОМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ. ВЕРОЯТНОСТНОЕ ПРОСТРАНСТВО

Определение 1.15 (определение А. Н. Колмогорова). Пусть Ω – любое пространство элементарных событий, $\mathcal{F} - \forall$ система подмножеств из Ω . Числовая функция $P = P(A) : \mathcal{F} \rightarrow \mathbb{R}$ называется вероятностной мерой, любое подмножество $A \in \mathcal{F}$ называется случайным событием, а число $P = P(A)$ – вероятностью случайного события A , если выполняются следующие аксиомы Колмогорова.

A 1. \mathcal{F} есть алгебра подмножеств из Ω .

A 2. $0 \leq P(A) \leq 1$, $A \in \mathcal{F}$.

A 3 (аксиома нормировки). $P(\Omega) = 1$.

A 4 (аксиома конечной аддитивности). Для любых несовместных случайных событий A, B ($A \cap B = \emptyset$) выполняется $P(A \cup B) = P(A) + P(B)$.

При этом если Ω – бесконечное множество, то аксиомы A1, A4 расширяются (обобщаются) следующим образом:

A 1'. \mathcal{F} – σ -алгебра подмножеств из Ω .

A 4' (аксиома счетной аддитивности). Для любой последовательности попарно несовместных случайных событий $A_1, A_2, \dots \in \mathcal{F}$ ($A_i \cap A_j = \emptyset, i \neq j$) справедливо равенство

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Определение 1.16. Вероятностным пространством называется тройка математических объектов (Ω, \mathcal{F}, P) :

Ω – пространство элементарных событий;

\mathcal{F} – алгебра или σ -алгебра подмножеств из Ω ;

P – вероятностная мера, определенная на \mathcal{F} .

В теории вероятностей есть еще один вариант расширения (обобщения) аксиом на случай бесконечного пространства элементарных событий Ω . Для определения этого варианта введем понятия последовательностей случайных событий и их пределов.

Определение 1.17. Пусть на измеримом пространстве (Ω, \mathcal{F}) определена произвольная последовательность случайных событий $A_1, A_2, \dots \in \mathcal{F}$. Тогда случайные события

$$\begin{aligned} A^+ &= \overline{\lim_{n \rightarrow \infty}} A_n ::= \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \in \mathcal{F}, \\ A^- &= \lim_{n \rightarrow \infty} A_n ::= \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \in \mathcal{F} \end{aligned} \quad (1.7)$$

называются соответственно верхним и нижним пределами последовательности случайных событий.

Выясним содержательный смысл событий (1.7) на языке «наступление – ненаступление». Проведем рассуждения для A^+ , для A^- они проводятся аналогично. Имеем: A^+ наступает \Leftrightarrow для $\forall k \in \mathbb{N}$ наступает $B_k = \bigcup_{n=k}^{\infty} A_n \Leftrightarrow$ для $\forall k = 1, 2, \dots \exists$ хотя бы один номер $n_k \geq k$, что наступает $A_{n_k} \Leftrightarrow \{ \text{наступает бесконечно много случайных событий среди } A_1, A_2, \dots \}$.

Упражнение 1.1. Провести аналогичные рассуждения для A^- самостоятельно.

Итак, из (1.7) следует:

$$\begin{aligned} A^+ &= \{ \text{наступает бесконечно много случайных событий среди } \{A_n\} \}; \\ A^- &= \{ \text{наступают все случайные события } \{A_n\}, \text{ за исключением лишь их конечного числа} \}. \end{aligned} \quad (1.8)$$

Из (1.8) очевидно:

$$A^- \subseteq A^+. \quad (1.9)$$

Определение 1.18. Если в (1.9) $A^- = A^+ = A \in \mathcal{F}$, то случайное событие A называется пределом последовательности случайных событий $\{A_n\}$:

$$A = \lim_{n \rightarrow \infty} A_n.$$

Определение 1.19. Последовательность случайных событий $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, называется монотонно убывающей и обозначается $A_n \downarrow$, если $A_1 \supset A_2 \supset \dots$. Последовательность случайных событий $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, называется монотонно возрастающей и обозначается $A_n \uparrow$, если $A_1 \subset A_2 \subset \dots$.

Теорема 1.3. Любая монотонная последовательность случайных событий $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, имеет предел при $n \rightarrow \infty$, который равен

$$A = \lim_{n \rightarrow \infty} A_n = \begin{cases} \bigcap_{n=1}^{\infty} A_n, & \text{если } A_n \downarrow, \\ \bigcup_{n=1}^{\infty} A_n, & \text{если } A_n \uparrow. \end{cases}$$

Доказательство. Вычислим нижний и верхний пределы и сравним их.

По формулам (1.7) с учетом монотонности $A_n \downarrow$ имеем

$$A^+ = \bigcap_{k=1}^{\infty} \left(\bigcup_{n=k}^{\infty} A_n \right) \equiv \bigcap_{k=1}^{\infty} A_k, \quad A^- = \bigcup_{k=1}^{\infty} \left(\bigcap_{n=k}^{\infty} A_n \right) \equiv \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} A_n,$$

так как $\bigcup_{n=k}^{\infty} A_n \equiv A_k$ и $\bigcup_{n=k}^{\infty} A_n \equiv \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

Значит, $A^- = A^+ = A$. Аналогично исследуется случай $A_n \uparrow$. \square

Итак, наряду с вариантом расширения аксиом Колмогорова $K = \{A1', A2, A3, A4'\}$ существует еще один вариант: $\tilde{K} = \{A1', A2, A3, A4, A5\}$, где $A5$ – это дополнительная аксиома.

А 5 (аксиома непрерывности меры). Для любой монотонно убывающей последовательности случайных событий $B_n \downarrow$, имеющей по теореме 1.3 предел $\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n$, допускается предельный переход под знаком вероятностной меры:

$$\lim_{n \rightarrow \infty} P(B_n) = P\left(\lim_{n \rightarrow \infty} B_n\right) = P(B).$$

Теорема 1.4. Если справедливы $A1', A2, A3$, то $A4'$ равносильна паре аксиом $\{A4, A5\}$.

Доказательство. [9, с. 30]. \square

Следствие 1.5. Эквивалентны следующие два варианта расширения аксиом: $K \Leftrightarrow \tilde{K}$.

1.6. СВОЙСТВА ВЕРОЯТНОСТНОЙ МЕРЫ (ВЕРОЯТНОСТИ)

Пусть (Ω, \mathcal{F}, P) – вероятностное пространство. Исследуем свойства вероятностной меры (вероятности) $P = P(A)$, вытекающие из аксиом Колмогорова.

Свойства вероятности

C1. $\forall A \in \mathcal{F} \quad P(\bar{A}) = 1 - P(A)$.

Доказательство. Разложим Ω : $\Omega = A \cup \bar{A}$, $A \cap \bar{A} = \emptyset$, тогда по аксиоме A4: $P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$. В силу A3: $P(\Omega) = 1 \Rightarrow P(\bar{A}) = 1 - P(A)$. \square

C2. Вероятность невозможного события равна нулю: $P(\emptyset) = 0$.

Доказательство следует из свойства C1, если положить $\Omega = A$. \square

C3. (Монотонность вероятностной меры.) Если $A \subset B$, то $P(A) \leq P(B)$.

Доказательство. Разложим B на два несовместных: $B = A \cup (B \setminus A)$, и по A2, A4 имеем $P(B) = P(A) + P(B \setminus A) \geq P(A) \Rightarrow P(B) \geq P(A)$. \square

Следствие 1.6. Если $A \subset B$, то $P(B \setminus A) = P(B) - P(A)$.

C4. (Формула сложения вероятностей.) Пусть (Ω, \mathcal{F}, P) – произвольное вероятностное пространство. Тогда для любых случайных событий $A, B \in \mathcal{F}$ справедлива формула сложения вероятностей:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (1.10)$$

Доказательство. Проиллюстрируем доказательство диаграммой Вьенна на рис. 1.2.

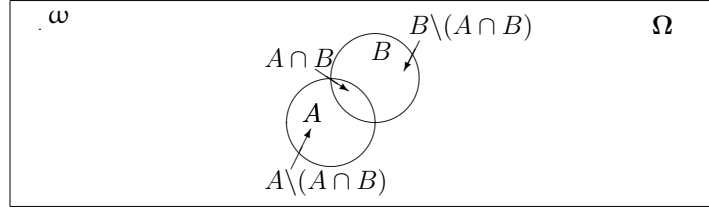


Рис. 1.2. Диаграмма Венна для суммы событий

Запишем разложения случайных событий на несовместные:

$$A \cup B = A \cup (B \setminus (A \cap B)); \quad (1.11)$$

$$B = (B \setminus (A \cap B)) \cup (A \cap B). \quad (1.12)$$

Применим к (1.11) и (1.12) аксиому А3 и получим

$$P(A \cup B) = P(A) + P(B \setminus (A \cap B)); \quad (1.13)$$

$$P(B) = P(A \cap B) + P(B \setminus (A \cap B)). \quad (1.14)$$

Почленно вычтем (1.14) из (1.13) и получим (1.10). \square

Следствие 1.7. Если A и B несовместны, $A \cap B = \emptyset$, то формула сложения вероятностей (1.10) превращается в А3.

Следствие 1.8. Для любых случайных событий A и B вероятность $P(A \cup B) \leq P(A) + P(B)$.

С5. (Обобщенная формула сложения вероятностей.) Пусть $N \geq 2$, тогда для любых случайных событий $A_1, A_2, \dots, A_N \in \mathcal{F}$ справедливо следующее обобщение формулы (1.10):

$$\begin{aligned} P\left(\bigcup_{i=1}^N A_i\right) &= \sum_{i=1}^N P(A_i) - \sum_{i=1}^{N-1} \sum_{j=i+1}^N P(A_i \cap A_j) + \\ &+ \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N P(A_i \cap A_j \cap A_k) - \dots + (-1)^{N-1} P\left(\bigcap_{i=1}^N A_i\right). \end{aligned} \quad (1.15)$$

Доказательство. Соотношение (1.15) доказывается по математической индукции по N ; для $N = 2$ результат следует из (1.10). \square

С6. Для любой последовательности $N (2 \leq N \leq \infty)$ случайных событий $A_1, \dots, A_N \in \mathcal{F}$ справедливо неравенство

$$P\left(\bigcup_{i=1}^N A_i\right) \leq \sum_{i=1}^N P(A_i).$$

Доказательство. Прежде всего отметим, что случай $N = 2$ рассмотрен в следствии 1.8. Построим вспомогательную последовательность случайных событий (используя «принцип новизны»):

$$B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2), \dots, B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j, i = 1, \dots, N.$$

Построенные события попарно несовместны: $B_i \cap B_j = \emptyset, i \neq j$. Очевидно, что

$$B_i \subseteq A_i, \bigcup_{i=1}^N B_i = \bigcup_{i=1}^N A_i.$$

Тогда в силу A4' и свойства C3

$$P\left(\bigcup_{i=1}^N A_i\right) = P\left(\bigcup_{i=1}^N B_i\right) = \sum_{i=1}^N P(B_i) \leq \sum_{i=1}^N P(A_i).$$

□

C7. (Эквивалент A5.) Для любой монотонной возрастающей последовательности случайных событий $A_n \uparrow$, имеющей предел

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n,$$

допустим предельный переход под знаком вероятностной меры:

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\lim_{n \rightarrow \infty} A_n\right).$$

Доказательство. Введем вспомогательную последовательность случайных событий $B_n = \overline{A_n}$. Очевидно, что $B_n \downarrow$, тогда по теореме 1.3 (см. п. 1.5)

$$\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n.$$

Поэтому с учетом свойства C1 и правила де Моргана имеем

$$\begin{aligned} \lim_{n \rightarrow \infty} P(A_n) &= \lim_{n \rightarrow \infty} (1 - P(B_n)) = 1 - P\left(\lim_{n \rightarrow \infty} B_n\right) = \\ &= 1 - P\left(\bigcap_{n=1}^{\infty} \overline{A_n}\right) = P\left(\overline{\bigcap_{n=1}^{\infty} \overline{A_n}}\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\lim_{n \rightarrow \infty} A_n\right). \end{aligned}$$

□

1.7. УСЛОВНАЯ ВЕРОЯТНОСТЬ И ЕЕ СВОЙСТВА. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ. ФОРМУЛА БАЙЕСА

Рассмотрим события $B = \{\text{студент сдал первый экзамен на 3}\}$; $A = \{\text{он сдал второй экзамен на 10}\}$. Как показывают экзаменационные статистические данные, вероятность события A зависит от наступления или ненаступления B . Удобной математической моделью для зависимости таких событий является условная вероятность.

Определение 1.20. Пусть (Ω, \mathcal{F}, P) – произвольное вероятностное пространство, на котором задано произвольное случайное событие $B \in \mathcal{F}$: $P(B) > 0$. Тогда условной вероятностью случайного события A при условии события B называется величина

$$P(A | B) = \frac{P(A \cap B)}{P(B)}; \quad (1.16)$$

при этом вероятность $P(A)$, изученная ранее, называется безусловной вероятностью.

Замечание 1.7. Символ « $|$ » читается: «при условии» (или «относительно»).

Замечание 1.8. $P(A | B) \in [0; 1]$. Это следует из (1.16), так как вероятность неотрицательна, $(A \cap B) \subseteq B$ и вероятностная мера монотонна.

Выясним содержательный смысл соотношения (1.16). Рассмотрим серию n независимых случайных экспериментов: $n = 1, 2, \dots$ и представим результаты в виде табл. 1.2.

Таблица 1.2

Случайное событие	A	B	$A \cap B$
Частота случайного события	$m_n(A)$	$m_n(B)$	$m_n(A \cap B)$
Относительная частота	$\nu_n(A) = m_n(A)/n$	$\nu_n(B) = m_n(B)/n$	$\nu_n(A \cap B) = m_n(A \cap B)/n$

По свойству статистической устойчивости

$$\nu_n(A) \xrightarrow{n \rightarrow \infty} P(A), \nu_n(B) \xrightarrow{n \rightarrow \infty} P(B), \nu_n(A \cap B) \xrightarrow{n \rightarrow \infty} P(A \cap B).$$

Вычислим относительную частоту наступления события A при условии, что B наступило, и предел при $n \rightarrow \infty$ с учетом свойства статистической устойчивости и формулы (1.16):

$$\nu_n(A|B) = \frac{m_n(A \cap B)}{m_n(B)} \cdot \frac{1/n}{1/n} = \frac{\nu_n(A \cap B)}{\nu_n(B)} \xrightarrow{n \rightarrow \infty} \frac{P(A \cap B)}{P(B)} = P(A|B).$$

Таким образом, условная вероятность (1.16) характеризует относительную частоту наступления события A при условии события B в бесконечной серии экспериментов.

Свойства условной вероятности

С1. При фиксированном случайном событии $B \in \mathcal{F} : P(B) > 0$ числовая функция $P_B(A) ::= P(A|B)$, где $A \in \mathcal{F}$, удовлетворяет всем аксиомам теории вероятностей А1–А5.

Следствие 1.9. $P_B(A)$ является некоторой вероятностной мерой.

С2. Условная вероятность удовлетворяет всем шести свойствам безусловной вероятности (см. п. 1.6), в частности

$$P(\bar{A} | B) = 1 - P(A | B),$$

$$P(A \cup B | C) = P(A | C) + P(B | C) - P(A \cap B | C).$$

С3. (Формула умножения вероятностей.) Если $P(B) > 0$, то справедлива формула

$$P(A \cap B) = P(B)P(A|B). \quad (1.17)$$

Доказательство. Формула получается из (1.16), если произвести умножение на $P(B)$. \square

Следствие 1.10. (Симметричная формула умножения вероятностей.) Если $P(A) > 0$ и $P(B) > 0$, то справедлива симметричная формула умножения вероятностей

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).$$

С4. (Обобщенная формула умножения вероятностей.) Для любого конечного числа N и любых N случайных событий $A_1, \dots, A_N \in \mathcal{F}$ таких, что $P\left(\bigcup_{i=1}^{N-1} A_i\right) > 0$, справедлива формула

$$P\left(\bigcap_{i=1}^N A_i\right) = P(A_1) \prod_{i=1}^{N-1} P\left(A_{i+1} \mid \bigcap_{j=1}^i A_j\right).$$

Доказательство. Требуемый результат можно получить, $N - 1$ раз применив формулу (1.17), последовательно отщепляя по одному событию. \square

Определение 1.21. Пусть (Ω, \mathcal{F}, P) – произвольное вероятностное пространство, тогда любая конечная или бесконечная система из N ($2 \leq N < \infty$) случайных событий $H_1, H_2, \dots, H_N \in \mathcal{F}$ называется *полной системой случайных событий (полной группой гипотез)*, если выполнены следующие три свойства:

С1. $\bigcup_{i=1}^N H_i = \Omega$;

С2. $H_i \cap H_j = \emptyset, i \neq j$;

С3. $P(H_i) > 0, i = 1, 2, \dots, N$.

Теорема 1.5 (формула полной вероятности). Пусть на произвольном вероятностном пространстве (Ω, \mathcal{F}, P) определена полная система случайных событий $\{H_i\}$. Тогда для любого случайного события $A \in \mathcal{F}$ его безусловная вероятность допускает разложение

$$P(A) = \sum_{i=1}^N P(H_i)P(A \mid H_i).$$

Доказательство. Построим разложение A на несовместные события:

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^N H_i\right) = \bigcup_{i=1}^N (A \cap H_i).$$

Преобразовав вероятность события A с помощью формулы умножения вероятностей, получим

$$P(A) = \sum_{i=1}^N P(A \cap H_i) = \sum_{i=1}^N P(H_i)P(A \mid H_i).$$

\square

Замечание 1.9. Формула полной вероятности допускает следующую более общую формулировку. Если случайное событие A может наступать лишь с одним из событий G_1, \dots, G_N ($N \leq +\infty$), то справедлива формула

$$P(A) = \sum_{i=1}^N P(G_i)P(A \mid G_i).$$

Теорема 1.6 (формула Байеса). Пусть выполнены условия теоремы 1.5, тогда, если $A \in \mathcal{F}$ – такое случайное событие, что $P(A) > 0$, то справедлива формула

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^N P(H_j)P(A|H_j)}, \quad i = 1, 2, \dots, N.$$

Доказательство. Воспользуемся симметричной формулой умножения вероятностей:

$$P(A \cap H_i) = P(H_i)P(A|H_i) = P(A)P(H_i|A),$$

и, применив формулу полной вероятности, разрешим полученное уравнение относительно $P(H_i|A)$:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)} = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^N P(H_j)P(A|H_j)}.$$

□

Замечание 1.10. $P(H_i)$ называется априорной вероятностью (от лат. «a priori», что означает «до опыта»), так как она известна до проведения и наблюдения результатов эксперимента.

Замечание 1.11. $P(H_i|A)$ называется апостериорной вероятностью (от лат. «a posteriori», что означает «после опыта»).

Замечание 1.12. Формула Байеса позволяет пересчитывать априорные вероятности в апостериорные:

$$\{P(H_i)\} \rightarrow \{P(H_i|A)\},$$

и в этом смысле ее называют математической моделью обучения.

1.8. НЕЗАВИСИМЫЕ СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ СВОЙСТВА

Основатель теории вероятностей А. Н. Колмогоров считал понятие независимости случайных событий важнейшим в теории вероятностей. Понятия вероятностной независимости и причинной независимости различны.

Определение 1.22. Пусть (Ω, \mathcal{F}, P) – произвольное вероятностное пространство. Любые события $A, B \in \mathcal{F}$ называются независимыми случайными событиями на (Ω, \mathcal{F}, P) , если вероятность их совместного наступления равна произведению вероятностей:

$$P(A \cap B) = P(A)P(B). \quad (1.18)$$

В противном случае события A и B зависимы.

Свойства независимых случайных событий

C1. Если $P(B) > 0$, то A и B независимы тогда и только тогда, когда

$$P(A|B) = P(A). \quad (1.19)$$

Доказательство. По правилу умножения вероятностей

$$P(A \cap B) = P(B)P(A|B),$$

тогда из (1.18) получаем (1.19). □

Следствие 1.11. Свойство (1.19) может рассматриваться как критерий независимости случайных событий A и B .

Доказательство. Возможны два случая:

- а) если $P(B) > 0$, то согласно С1 (1.18) и (1.19) эквивалентны;
- б) если $P(B) = 0$, тогда в (1.18) правая и левая части равны нулю:

$$A \cap B \subseteq B \Rightarrow P(A \cap B) \leq P(B) = 0.$$

□

Замечание 1.13. Из (1.19) виден содержательный смысл определения 1.22. События A и B независимы, если наступление одного из этих событий не влияет на вероятность наступления другого.

- С2. Свойство независимости переносится на противоположные события. В частности, если A, B – независимые случайные события, то \bar{A}, \bar{B} – также независимые случайные события.

Доказательство. Проверим критерий (1.19):

$$P(\bar{A}|B) = 1 - P(A|B) = 1 - P(A) = P(\bar{A}).$$

□

- С3. Пусть $A, B, C \in \mathcal{F}$ – любые случайные события на (Ω, \mathcal{F}, P) такие, что A, C – независимы, B, C – независимы и A, B – несовместны, тогда независимы $A \cup B$ и C .

Доказательство. Проверим (1.19):

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C) = P(A) + P(B) - 0 = P(A \cup B).$$

□

Определение 1.23. Случайные события $A_1, A_2, \dots, A_n \in \mathcal{F}$ называются независимыми в совокупности на (Ω, \mathcal{F}, P) , если для $\forall m \in \{2, 3, \dots, n\}$ и любых упорядоченных значений m индексов $1 \leq i_1 < i_2 < \dots < i_m \leq n$ выполняется обобщение (1.18):

$$P\left(\bigcap_{j=1}^m A_{i_j}\right) = \prod_{j=1}^m P(A_{i_j}). \quad (1.20)$$

Если же (1.20) выполняется лишь для $m = 2$, то случайные события A_1, \dots, A_n называются попарно независимыми.

- С4. Из независимости в совокупности следует попарная независимость случайных событий. Обратное, вообще говоря, неверно.

Доказательство. Прямое утверждение очевидно. Приведем контрпример Бернштейна, показывающий, что из попарной независимости событий не следует их независимости в совокупности. Над плоской поверхностью бросается симметрический тетраэдр, грани которого раскрашены следующим образом: одна – в красный цвет, вторая – в синий, третья – в зеленый, четвертая имеет полосы всех трех цветов. Определим случайные события: $A_1 = \{\text{на выпавшей грани есть красный цвет}\}$; $A_2 = \{\text{на выпавшей грани есть синий цвет}\}$; $A_3 = \{\text{на выпавшей грани есть зеленый цвет}\}$. Проверим попарную независимость и независимость в совокупности этих событий: $P(A_i) = 2/4 = 1/2$; $P(A_i \cap A_j) = 1/4$; $P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \Rightarrow$ события попарно независимы; $P(A_1 \cap A_2 \cap A_3) = 1/4 \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = 1/8$, т. е. A_1, A_2, A_3 – зависимы. □

- С5. (Обобщенная формула сложения вероятностей для независимых в совокупности случайных событий.) Если случайные события A_1, A_2, \dots, A_N независимы в совокупности, то вероятность наступления хотя бы одного из этих событий вычисляется по формуле

$$P\left(\bigcup_{i=1}^N A_i\right) = 1 - \prod_{i=1}^N (1 - P(A_i)).$$

Доказательство провести самостоятельно. \square

Задание 1.7. Указать все события, которые не зависят от любых событий $A \in \mathcal{F}$.

Задание 1.8. Доказать или опровергнуть: если случайные события A и B имеют ненулевую вероятность и несовместны, то они не могут быть независимыми.

1.9. СХЕМА НЕЗАВИСИМЫХ ИСПЫТАНИЙ БЕРНУЛЛИ. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

Рассмотрим случайный эксперимент S , который является составным и состоит из $N \geq 2$ случайных экспериментов S_1, S_2, \dots, S_N . Например, студент сдает в сессию 5 экзаменов: $N = 5$.

Рассмотрим S_n – n -й эксперимент, $n = 1, \dots, N$. Его математическая модель – это вероятностное пространство (Ω, \mathcal{F}, P) , где Ω_n – пространство элементарных событий n -го эксперимента; $\omega_n \in \Omega_n$ – исход n -го эксперимента; \mathcal{F}_n – σ -алгебра подмножеств из Ω_n ; P_n – вероятностная мера, действующая в n -м эксперименте. Обозначим: $\omega = (\omega_1, \dots, \omega_N)^T$ – составной вектор исходов (протокол всей сессии); $\omega \in \Omega = \Omega_1 \times \dots \times \Omega_N$; \mathcal{F} – σ -алгебра, порожденная произведением $\mathcal{F}_1 \times \dots \times \mathcal{F}_N$. $P(A)$, $A \in \mathcal{F}$ – вероятностная мера составного эксперимента.

Определение 1.24. Указанная выше схема N случайных испытаний называется схемой независимых испытаний, если произвольные N случайных событий A_1, \dots, A_N , относящиеся к различным случайным экспериментам, независимы в совокупности, т. е. $\forall A = A_1 \times \dots \times A_N \in \mathcal{F}$, $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2, \dots, A_N \in \mathcal{F}_N$, выполняется

$$P(A) = \prod_{n=1}^N P_n(A_n).$$

Имеет место схема независимых испытаний Бернулли, если вдобавок к предыдущему определению выполняются:

- 1) $\Omega_n = \{0, 1\}$, т. е. состоит из двух исходов, $|\Omega_n| = 2$: $\omega_n = 1$ – «успех», $\omega_n = 0$ – «неуспех»;
- 2) $\mathcal{F}_n = 2^{\Omega_n} = \{\emptyset, \{0\}, \{1\}, \Omega_n\}$;
- 3) $P_n(\omega_n) = \begin{cases} p, & \omega_n = 1, \\ 1 - p, & \omega_n = 0, \end{cases} = p^{\omega_n}(1-p)^{1-\omega_n}$, $\omega_n \in \Omega_n$, где $p \in (0, 1)$ является параметром и называется вероятностью успеха;
- 4) $\Omega = \{\omega = (\omega_1, \dots, \omega_N)^T : \omega_n \in (0, 1), n = 1, 2, \dots, N\}$ – множество всевозможных двоичных векторов, $\mathcal{F} = 2^\Omega$ – множество всех подмножеств из Ω :

$$P(\omega) = \prod_{n=1}^N P_n(\omega_n) = p^{\sum_{n=1}^N \omega_n} (1-p)^{N - \sum_{n=1}^N \omega_n},$$

где $\sum_{n=1}^N \omega_n$ – число единиц в двоичном векторе ω .

Рассмотрим следующее случайное событие в этой модели: $B_m = \{\text{в указанной схеме независимых испытаний Бернулли наступило } m \text{ «успехов»}\}$, $m = 0, 1, \dots, N$. Вычислим его вероятность:

$$P(B_m) = \sum_{\omega: \sum_{n=1}^N \omega_n = m} P(\omega) = \sum_{\omega: \sum_{n=1}^N \omega_n = m} p^m (1-p)^{N-m} = C_N^m p^m (1-p)^{N-m}.$$

Определение 1.25. *Набор вероятностей*

$$p_m = C_N^m p^m (1-p)^{N-m}, \quad m = 0, 1, \dots, N, \quad (1.21)$$

называется *биномиальным распределением вероятностей* и обозначается $Bi(N, p)$.

Замечание 1.14. Биномиальное распределение вероятностей (1.21) можно получить на основе следующей формулы бинома Ньютона:

$$(p + (1-p))^N = \sum_{n=0}^N p_n \equiv 1.$$

Замечание 1.15. При больших значениях N, m и для вычисления (1.21) используют приближенные формулы Муавра – Лапласа, которые получаются из формулы Стирлинга (см. п. 5.6).

1.10. УПРАЖНЕНИЯ

1. Пусть \mathcal{F} – σ -алгебра подмножеств из Ω , а $B \in \mathcal{F}$ – некоторое случайное событие. Показать, что $G_B = \{g = A \cap B : A \in \mathcal{F}\}$ есть σ -алгебра подмножеств из B .
2. Какие из нижеприведенных соотношений тождественно верны? Для тех соотношений, которые не верны, указать условия, при которых они выполняются:
 - а) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
 - б) $A \cap (B \cap C) = (A \cap B) \cap C$;
 - в) $(A \cup B) \cap C = A \cup (B \cap C)$;
 - г) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.
3. Доказать справедливость неравенств: $\max(P(A), P(B)) \leq P(A \cup B) \leq P(A) + P(B)$.
4. Доказать справедливость неравенств: $\max(0, P(A) - P(\overline{B}), P(B) - P(\overline{A})) \leq P(A \cap B) \leq \min(P(A), P(B))$.
5. Пусть $A_n, \quad n \in N$, – последовательность случайных событий, каждое из которых имеет вероятность $P(A_n) = 1$. Показать, что $P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$.
6. Пусть $n \geq 3$ и A_1, \dots, A_n – случайные события, о которых известно, что по крайней мере одно из этих событий в эксперименте определенно наступает, но более двух наступить не может. Если $P(A_i) = p_1, \quad P(A_i \cap A_j) = p_2$, показать, что $p_1 \geq 1/n, \quad p_2 \leq 2/n$.
7. Показать, что если $P(A|B) > P(A)$, то $P(B|A) > P(B)$.
8. Пусть p – простое число, $\Omega = \{1, 2, \dots, p\}$, $\mathcal{F} = 2^\Omega$, и для любого подмножества $A \in \mathcal{F}$ вероятностная мера задана следующим образом: $P(A) = |A|/p$. Показать, что если $A, B \in \mathcal{F}$ независимы, то A и B совпадают с \emptyset или Ω .
9. Показать, что в общем случае условная независимость A, B при условии C не влечет, не влечется безусловной независимостью A, B . Для каких событий C – произвольные события A, B независимы тогда и только тогда, когда они условно независимы при условии C ?
10. Пусть F, G – некоторые σ -алгебры подмножеств из Ω . Показать, что $F \cap G$ – тоже σ -алгебра, а $F \cup G$, вообще говоря, не является σ -алгеброй.
11. Показать, что вероятность того, что в точности одно из событий A, B наступает, равна $P(A) + P(B) - 2P(A \cap B)$.

12. Доказать, что $P(A_1 \cup A_2 \cup A_3) = 1 - P(\bar{A}_1|\bar{A}_2 \cap \bar{A}_3) \cdot P(\bar{A}_2|\bar{A}_3) \cdot P(\bar{A}_3)$.

13. Доказать:

- а) если A не зависит от самого себя, то $P(A) = 0 \vee 1$;
 б) если $P(A) = 0 \vee 1$, то A не зависит от $\forall B \in \mathcal{F}$.

14. Доказать неравенства (иногда называемые неравенствами Буля):

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i), P\left(\bigcup_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n P(\bar{A}_i).$$

15. Пусть события A и B_1 независимы и события A и B_2 независимы. Показать, что события A и $B_1 \cup B_2$ независимы тогда и только тогда, когда события A и $B_1 \cap B_2$ независимы.

16. Доказать:

$$P\left(\bigcap_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cup A_j) + \\ + \sum_{i < j < k} P(A_i \cup A_j \cup A_k) - \dots - (-1)^n P(A_1 \cup A_2 \cup \dots \cup A_n).$$

17. Доказать следующее утверждение (иногда называемое теоремой Варинга). Пусть $A_1, A_2, \dots, A_n \in \mathcal{F}$ – некоторые события, $N_k \in \mathcal{F}$ – случайное событие, состоящее в том, что в точности k событий среди $\{A_i\}$ наступает. Тогда

$$P(N_k) = \sum_{i=0}^{n-k} (-1)^i C_{k+i}^k S_{k+i},$$

$$\text{где } S_j = \sum_{i_1 < i_2 < \dots < i_j} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_j}).$$

18. («Парадокс дней рождения».) Предполагая, что дни рождения присутствующих на лекции по теории вероятностей и математической статистике n студентов, рожденных в одном и том же году, чисто случайны и независимы, показать, что вероятность хотя бы одного совпадения дней рождения

$$p_n = 1 - \frac{365!}{(365-n)!365^n}.$$

Показать, что начиная с $n = 23$ эта вероятность $p_n > 1/2$.

Глава 2

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

2.1. ПОНЯТИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ И ЕЕ СВОЙСТВА

В гл. 1 был рассмотрен случайный эксперимент S , в котором каждому $\omega \in \Omega$ соответствует некоторое случайное событие A . Но существуют более сложные по результатам случайные эксперименты, в которых измеряются числовые величины, т. е. элементарному исходу ω ставится в соответствие действительное число:

$$\xi = \xi(\omega) \in \mathbb{R}.$$

Удобной математической моделью для таких случайных экспериментов является случайная величина.

Определение 2.1. Пусть (Ω, \mathcal{F}, P) – произвольное вероятностное пространство, $(\mathbb{R}, \mathcal{B})$ – измеримое пространство, где $\mathcal{B} = \mathcal{B}(\mathbb{R})$ – борелевская σ -алгебра на числовой прямой. Числовая функция $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$ называется \mathcal{F} -измеримой функцией, если прообраз любого борелевского множества $B \in \mathcal{B}$ является случайным событием, т. е. выполняется

$$\xi^{-1}(B) ::= \{\omega \in \Omega : \xi(\omega) \in B\} = \{\xi \in B\} \in \mathcal{F}. \quad (2.1)$$

Определение 2.2. \mathcal{F} -измеримая числовая функция $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$, удовлетворяющая (2.1), называется случайной величиной, заданной на (Ω, \mathcal{F}, P) .

Обозначим для произвольного борелевского множества $B \in \mathcal{B}$ его ξ -прообраз:

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{F}.$$

Теорема 2.1. Совокупность случайных событий

$$\mathcal{F}_\xi ::= \{\xi^{-1}(B) : B \in \mathcal{B}\}$$

является σ -алгеброй, причем $\mathcal{F}_\xi \subseteq \mathcal{F}$.

Доказательство. Проверим свойства, определяющие σ -алгебру.

Во-первых, так как $\xi^{-1}(\mathbb{R}) = \Omega$, то $\Omega \in \mathcal{F}_\xi$.

Во-вторых, выберем произвольные борелевские множества $B_1, B_2 \in \mathcal{B}$ и построим случайные события:

$$A_1 = \xi^{-1}(B_1) \in \mathcal{F}_\xi, A_2 = \xi^{-1}(B_2) \in \mathcal{F}_\xi.$$

Из определения борелевской σ -алгебры \mathcal{B} имеем

$$B_1 \cup B_2 \in \mathcal{B}, B_1 \cap B_2 \in \mathcal{B}, \overline{B_1} \in \mathcal{B}, \overline{B_2} \in \mathcal{B}.$$

Отсюда следует:

$$A_1 \cup A_2 = \{\xi \in B_1\} \cup \{\xi \in B_2\} = \{\xi \in B_1 \cup B_2\} \in \mathcal{F}_\xi;$$

$$A_1 \cap A_2 = \{\xi \in B_1\} \cap \{\xi \in B_2\} = \{\xi \in B_1 \cap B_2\} \in \mathcal{F}_\xi;$$

$$\overline{A_i} = \overline{\{\xi \in B_i\}} = \{\xi \in \overline{B_i}\} \in \mathcal{F}_\xi, \quad i = 1, 2, \dots$$

В-третьих, выберем произвольно $B_1, B_2, \dots \in \mathcal{B}$. Тогда

$$\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}, \quad A_i = \xi^{-1}(B_i) \in \mathcal{F}_{\xi}, \quad i = 1, 2, \dots$$

Поэтому

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \xi^{-1}(B_i) = \xi^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) \in \mathcal{F}_{\xi}.$$

Таким образом, \mathcal{F}_{ξ} является σ -алгеброй.

Осталось показать, что $\mathcal{F}_{\xi} \subseteq \mathcal{F}$. Предположим противное: $\mathcal{F} \subset \mathcal{F}_{\xi}$. Это значит, что $\mathcal{F}_{\xi} \setminus \mathcal{F} \neq \emptyset$ и найдется $B' \in \mathcal{B}$ такое, что

$$\xi^{-1}(B') \in \mathcal{F}_{\xi}, \quad \text{но} \quad \xi^{-1}(B') \notin \mathcal{F}.$$

Последнее противоречит определению случайной величины ξ . Следовательно, $\mathcal{F}_{\xi} \subseteq \mathcal{F}$. \square

Определение 2.3. σ -Алгебра $\mathcal{F}_{\xi} = \{\xi^{-1}(B) : B \in \mathcal{B}\}$ называется σ -алгеброй, порожденной случайной величиной ξ .

С помощью случайной величины $\xi = \xi(\omega)$ и вероятностной меры P построим еще одну числовую функцию:

$$P_{\xi}(B) ::= P(\xi^{-1}(B)) = P\{\xi \in B\}, \quad B \in \mathcal{B}. \quad (2.2)$$

Пользуясь формулой (2.2), легко проверить, что эта функция удовлетворяет всем аксиомам Колмогорова и, следовательно, является некоторой вероятностной мерой.

Определение 2.4. Вероятностная мера $P_{\xi}(B)$, определяемая (2.2), называется распределением вероятностей случайной величины $\xi = \xi(\omega)$. Вероятностное пространство $(\mathbb{R}, \mathcal{B}, P_{\xi})$ называется вероятностным пространством, порожденным случайной величиной ξ (индуцированным случайной величиной ξ).

Замечание 2.1. Функция $P_{\xi}(B)$ определена на борелевских множествах $B \in \mathcal{B}$ и показывает, как распределена вероятность по этим множествам. Отсюда происходит ее название.

Замечание 2.2. Борелевская σ -алгебра \mathcal{B} была построена на основе базовой системы интервалов: $\mathcal{F}_0 = \{\emptyset, A_x = (-\infty, x) : x \in \mathbb{R}\}$. В связи с этим определение случайной величины можно упростить.

Определение 2.5 (эквивалентное понятие СВ). Числовая функция $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$ называется случайной величиной на вероятностном пространстве (Ω, \mathcal{F}, P) , если для любого $x \in \mathbb{R}$ прообраз интервала $A_x = (-\infty, x)$ является случайным событием, т. е.

$$\xi^{-1}(A_x) = \{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{F}.$$

Определение 2.6. Пусть $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$ – произвольная случайная величина на (Ω, \mathcal{F}, P) , имеющая распределение вероятностей $P_{\xi}(B) = P\{\xi \in B\}$, $B \in \mathcal{B}$. Тогда действительная функция

$$F_{\xi}(x) = P\{\xi < x\} = P_{\xi}(A_x), \quad x \in \mathbb{R}, \quad (2.3)$$

называется функцией распределения случайной величины ξ .

Распределение вероятностей $P_{\xi}(\cdot)$ однозначно определяет функцию распределения $F_{\xi}(\cdot)$. Обратное соответствие задается следующей теоремой.

Теорема 2.2. Для произвольного $B \in \mathcal{B}$ распределение вероятностей $P_{\xi}(B) = P\{\xi \in B\}$ однозначно выражается через функцию $F_{\xi}(\cdot)$. В частности, для $\forall x, x_1, x_2 \in \mathbb{R}, x_1 < x_2$, справедливы соотношения:

$$P\{\xi \geq x\} = 1 - F_\xi(x); \quad (2.4)$$

$$P\{x_1 \leq \xi < x_2\} = F_\xi(x_2) - F_\xi(x_1); \quad (2.5)$$

$$P\{\xi = x\} = F_\xi(x+0) - F_\xi(x). \quad (2.6)$$

Доказательство. Результат теоремы вытекает из свойств вероятностной меры P и из того факта, что \mathcal{B} построена на основе базовой системы интервалов: $F_0 = \{\emptyset, A_x : x \in \mathbb{R}\}$. Доказательство проведем для вышеприведенных частных случаев.

По свойствам вероятности из (2.3) для (2.4) имеем

$$P\{\xi \geq x\} = 1 - P\{\xi < x\} = 1 - F_\xi(x).$$

Теперь рассмотрим (2.5), учитывая, что $[x_1, x_2) = A_{x_2} \setminus A_{x_1}$:

$$\begin{aligned} P\{x_1 \leq \xi < x_2\} &= P_\xi\{\xi \in A_{x_2} \setminus A_{x_1}\} = \\ &= P\{\xi \in A_{x_2}\} - P\{\xi \in A_{x_1}\} = F_\xi(x_2) - F_\xi(x_1). \end{aligned}$$

Обратимся к (2.6), используя аксиому непрерывности А5:

$$\begin{aligned} P\{\xi = x\} &= P\left(\bigcap_{n=1}^{\infty} \left\{x \leq \xi < x + \frac{1}{n}\right\}\right) = \\ &= \lim_{n \rightarrow \infty} P\left\{x \leq \xi < x + \frac{1}{n}\right\} = \lim_{n \rightarrow \infty} F_\xi\left(x + \frac{1}{n}\right) - F_\xi(x) = F_\xi(x+0) - F_\xi(x). \end{aligned}$$

□

Следствие 2.1. Распределение вероятностей $P_\xi(B)$, $B \in \mathcal{B}$, и функция распределения $F_\xi(x)$, $x \in \mathbb{R}$, находятся во взаимно-однозначном соответствии.

Теорема 2.3 (о свойствах ФР). Пусть $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$ – произвольная случайная величина, определенная на вероятностном пространстве (Ω, \mathcal{F}, P) и имеющая функцию распределения $F_\xi(x)$, $x \in \mathbb{R}$, и распределение вероятностей $P_\xi(B)$, $B \in \mathcal{B}$. Тогда $F_\xi(\cdot)$ удовлетворяет следующим свойствам:

- 1) область ее значений есть $[0, 1]$: $0 \leq F_\xi(x) \leq 1$, $x \in \mathbb{R}$;
- 2) функция распределения F_ξ – неубывающая функция, т. е. $\forall x_1 < x_2 \Rightarrow F_\xi(x_1) \leq F_\xi(x_2)$;
- 3) полное колебание функции распределения $\omega(F_\xi) = 1$, при этом $F_\xi(-\infty) = 0$, $F_\xi(+\infty) = 1$;
- 4) функция распределения непрерывна слева в каждой точке области определения, т. е. $F_\xi(x-0) = F_\xi(x)$, $x \in \mathbb{R}$.

Доказательство. Для начала графически проиллюстрируем вид $F_\xi(\cdot)$ (рис. 2.1).

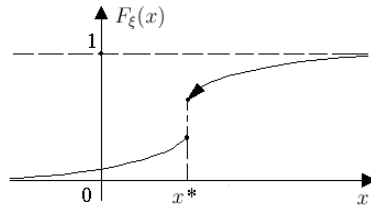


Рис. 2.1. График функции распределения

Свойства 1) и 2) вытекают из формулы (2.3) и свойств вероятности. Докажем свойство 3). Функция $F_\xi(\cdot)$ – неубывающая, поэтому полное колебание $\omega = F_\xi(+\infty) - F_\xi(-\infty)$. Покажем, что $F_\xi(-\infty) = 0$. Выберем вспомогательную монотонно убывающую не ограниченную снизу числовую последовательность $x_n \rightarrow -\infty$ и с ее помощью построим монотонно убывающую последовательность числовых промежутков $A_{x_n} \downarrow$.

Как известно, $\exists \lim_{n \rightarrow \infty} A_{x_n} = \bigcap_{n=1}^{\infty} A_{x_n} = \emptyset$. Поэтому с учетом (2.3) и аксиомы А5 имеем:

$$F_\xi(-\infty) = \lim_{n \rightarrow \infty} F_\xi(x_n) = \lim_{n \rightarrow \infty} P_\xi(A_{x_n}) = P_\xi(\lim_{n \rightarrow \infty} A_{x_n}) = P_\xi(\emptyset) = 0.$$

Самостоятельно доказать, что $F_\xi(+\infty) = 1$, выбрав вспомогательную не ограниченную сверху последовательность $y_n \rightarrow +\infty$.

Докажем свойство 4) с учетом (2.3) и теоремы о монотонной последовательности случайных событий: $F_\xi(x-0) = \lim_{n \rightarrow \infty} F_\xi(x-1/n) = \lim_{n \rightarrow \infty} P_\xi(A_{x-1/n}) = P_\xi(\lim_{n \rightarrow \infty} A_{x-1/n}) = P_\xi(A_x) = F_\xi(x)$. \square

Теорема 2.4. Пусть $F(x)$, $x \in \mathbb{R}$, – произвольная действительная функция, обладающая свойствами 1–4. Тогда найдется такое вероятностное пространство (Ω, \mathcal{F}, P) и на нем такая случайная величина $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$, что ее функция распределения в точности совпадет с данной функцией $F_\xi(\cdot) = F(\cdot)$.

Доказательство. Доказательство данной теоремы носит конструктивный характер. Его разобрать самостоятельно согласно работе [9]. \square

Следствие 2.2. Свойства 1–4 полностью характеризуют семейство функций распределения.

2.2. КЛАССИФИКАЦИЯ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ, РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ И СЛУЧАЙНЫХ ВЕЛИЧИН. ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Пусть $\xi = \xi(\omega)$ – произвольная случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) . Математической моделью этой случайной величины является вероятностное пространство $(\mathbb{R}, \mathcal{B}, P_\xi)$, где \mathbb{R} – числовая прямая, \mathcal{B} – борелевская σ -алгебра, $P_\xi(B)$, $B \in \mathcal{B}$, – распределение вероятностей.

Многообразие математических моделей случайных величин есть многообразие их функций распределений, поэтому проведем классификацию случайных величин и распределений вероятностей, классифицируя функции распределения. Существует следующая классификация функций распределения.

1. Дискретные функции распределения (дискретные распределения вероятностей, дискретные случайные величины).
2. Абсолютно непрерывные функции распределения (абсолютно непрерывные распределения вероятностей, непрерывные случайные величины).
3. Сингулярные функции распределения (сингулярные распределения вероятностей, сингулярные случайные величины).

2.2.1. Дискретные функции распределения

Определение 2.7. Принято говорить, что случайная величина $\xi = \xi(\omega)$, определенная на вероятностном пространстве (Ω, \mathcal{F}, P) , имеет дискретное распределение вероятностей, если множество X всевозможных ее значений дискретно (конечно или счетно), т. е. для $\forall \omega \in \Omega$ $\xi(\omega) \in X = \{x_1, \dots, x_M\}$, $M \leq +\infty$, причем $-\infty < x_1 < x_2 < \dots < x_M < +\infty$ и элементарные вероятности $p_i = P\{\xi = x_i\} > 0$, $i = 1, \dots, M$.

При этом случайная величина ξ называется *дискретной случайной величиной*, а ее функция распределения $F_\xi(x)$ – *дискретной функцией распределения*.

Дискретное распределение вероятностей удобно задавать с помощью следующей таблицы:

Значение x_i	x_1	...	x_M
Элементарная вероятность p_i	p_1	...	p_M

Элементарные вероятности удовлетворяют условию нормировки $\sum_{i=1}^M p_i = 1$, так как сумма слева определяет вероятность достоверного события.

Вычислим функцию распределения для дискретной случайной величины. Из определения следует:

$$F_\xi(x) = P\{\xi < x\} = \sum_{i: x_i < x} P\{\xi = x_i\}.$$

Таким образом, дискретная функция распределения имеет следующий общий вид (рис. 2.2):

$$F_\xi(x) = \sum_{i: x_i < x} p_i, \quad x \in \mathbb{R}.$$

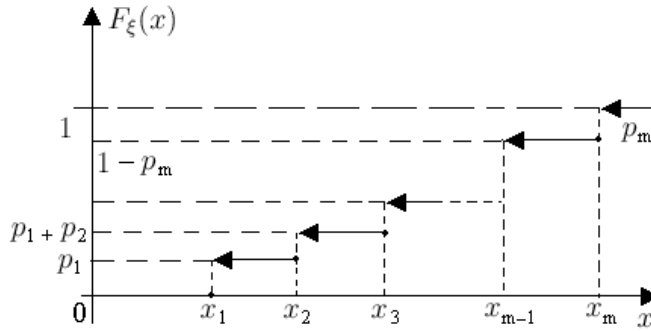


Рис. 2.2. График дискретной функции распределения

Существует бесконечное множество дискретных функций распределения, различающихся в зависимости от \mathcal{X} , $\{p_x : x \in \mathcal{X}\}$.

На практике встречаются некоторые типовые дискретные распределения вероятностей, которые принято называть законами дискретного распределения вероятностей.

Основные законы дискретного распределения вероятностей

1. *Закон распределения вероятностей Бернулли* с параметром $p \in [0, 1]$ (при этом $\mathcal{X} = \{0, 1\}$, $M = 2$):

$$p_x = P\{\xi = x\} = \begin{cases} p, & x = 1, \\ 1 - p, & x = 0, \end{cases} = p^x(1 - p)^{1-x}, \quad x \in \mathcal{X}.$$

Здесь $p \in [0, 1]$ – параметр, который принято называть элементарной вероятностью или вероятностью успеха. Такая случайная величина $\xi = \xi(\omega) \in \{0, 1\}$ характеризует случайный исход единичного испытания Бернулли:

$$\begin{aligned}\xi = 1 & \text{ — «успех»,} \\ \xi = 0 & \text{ — «неудача»}.\end{aligned}$$

Пример 2.1. Представим специалиста, который контролирует качество некоторой продукции (микросхем, дискет, ...). Ему для проверки предоставили огромный контейнер этих деталей. Пусть в исследуемой партии доля брака равна p . Контролер наудачу берет одну деталь и проверяет. С точки зрения контролера:

$$\xi = \begin{cases} 1, & \text{«успех» — бракованное изделие,} \\ 0, & \text{«неуспех» — годное изделие.} \end{cases}$$

Таким образом, введенная случайная величина ξ является случайной величиной Бернулли.

2. *Биномиальный закон распределения вероятностей* с параметрами $p \in [0, 1]$, $N \in \mathbb{N}$. Он задается следующими соотношениями:

$$\begin{aligned}\mathcal{X} &= \{0, 1, \dots, N\}, \quad M = N + 1; \\ p_x &= P\{\xi = x\} = C_N^x p^x (1 - p)^{N-x}, \quad x \in \mathcal{X}.\end{aligned}$$

Пример 2.2. $\xi(\omega)$ – случайное число бракованных деталей, обнаруженных контролером при извлечении N деталей наудачу, $\xi = \xi(\omega)$ – случайное число успехов в схеме N независимых испытаний Бернулли.

В теории вероятностей для краткости записи законов используют следующее обозначение (Binomial Law):

$$\mathcal{L}\{\xi\} = Bi(N, p)$$

(читается: «закон распределения величины ξ есть биномиальный закон с параметрами N и p »).

При $N = 1$ имеем

$$\mathcal{L}\{\xi\} = Bi(1, p) -$$

закон распределения вероятностей Бернулли с параметром $p \in [0, 1]$.

3. *Дискретный равномерный на множестве \mathcal{X} закон распределения вероятностей:*

$$\mathcal{X} = \{x_1, \dots, x_M\}, \quad M < +\infty, \quad p_x = \frac{1}{M}, \quad x \in \mathcal{X}.$$

Обозначение:

$$\mathcal{L}\{\xi\} = DR(\mathcal{X}) - Discrete Rectangular.$$

4. *Гипергеометрический закон распределения вероятностей* с параметрами K, k, L :

$$\mathcal{L}\{\xi\} = H(K, k, L) - Hypergeometric, \quad L < K.$$

При этом множество возможных значений $\mathcal{X} = \{N_1, N_1 + 1, \dots, N_2\}$, где $N_1 = \max(0, L - (K - k))$, $N_2 = \min\{k, L\}$:

$$p_x = \frac{C_k^x C_{K-k}^{L-x}}{C_K^L}, \quad x \in \mathcal{X}.$$

Пример 2.3. В урне перемешано K шаров: k – красных, $K - k$ – белых; извлекаем L шаров, $\xi = \xi(\omega)$ – случайное число извлеченных красных шаров, $p_x = P\{\xi = x\}$.

Пример 2.4. Контроль качества продукции: K микросхем, среди которых k бракованных, наудачу извлекаем L микросхем и проверяем их, ξ – случайное число обнаруженных бракованных микросхем.

5. *Геометрическое распределение вероятностей* с параметром $p \in [0, 1]$:

$$\mathcal{L}\{\xi\} = G(p) - \text{Geometric}.$$

Здесь p – элементарная вероятность, $\mathcal{X} = \{0, 1, 2, \dots\}$, $M = |\mathcal{X}| = \infty$, $p_x = P\{\xi = x\} = (1-p)^x p$, $x \in \mathcal{X}$.

Пример 2.5. Контролер проверяет первую микросхему, вторую и т. д. – до первой бракованной; $\xi = \xi(\omega)$ – случайное число проверок до обнаружения первой бракованной микросхемы, p – доля брака, $1-p$ – вероятность извлечь годную микросхему, $(1-p)^x$ – вероятность извлечь годную деталь x раз. Поскольку события независимы, то вероятность после x годных деталей извлечь бракованную равна $(1-p)^x p$.

6. *Закон распределения Пуассона* с параметром $\lambda > 0$:

$$\mathcal{L}\{\xi\} = \Pi(\lambda),$$

$$\mathcal{X} = \{0, 1, 2, \dots\}, \quad M = +\infty, \quad p_x = P\{\xi = x\} = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \mathcal{X}.$$

Задание 2.1. Проверить условие нормировки.

Пример 2.6. $\xi = \xi(\omega)$ – случайное число посетителей акционерно-коммерческого банка в течение рабочего дня, $\lambda > 0$ – среднее число посетителей.

2.2.2. Абсолютно непрерывные функции распределения вероятностей

Пусть на некотором вероятностном пространстве (Ω, \mathcal{F}, P) определена некоторая случайная величина $\xi = \xi(\omega)$, имеющая распределение вероятностей $P_\xi(B)$, $B \in \mathcal{B}$, и функцию распределения $F_\xi(x) = P\{\xi < x\}$, $x \in \mathbb{R}$.

Определение 2.8. Функция распределения $F_\xi(\cdot)$ называется абсолютно непрерывной функцией распределения вероятностей, если существует неотрицательная интегрируемая функция $p_\xi(x) \geq 0$ такая, что справедливо следующее интегральное представление:

$$F_\xi(x) = \int_{-\infty}^x p_\xi(y) dy, \quad x \in \mathbb{R}; \quad (2.7)$$

при этом распределение вероятностей $P_\xi(\cdot)$ называется абсолютно непрерывным распределением вероятностей, случайная величина ξ – непрерывной случайной величиной, а функция $p_\xi(\cdot)$ – плотностью распределения вероятностей.

Свойства плотности распределения вероятностей

C1. $p_\xi(x) \geq 0$, $x \in \mathbb{R}$.

C2. Условие нормировки:

$$\int_{-\infty}^{+\infty} p_\xi(y) dy = 1.$$

Доказательство. Из интегрального представления (2.7) и свойства функции распределения при $x \rightarrow \infty$ имеем

$$F_{\xi}(+\infty) = \int_{-\infty}^{+\infty} p_{\xi}(y) dy,$$

но по свойствам функций распределения $F_{\xi}(+\infty) = 1$. □

С3. Для $\forall B \in \mathcal{B}$ справедливо

$$P\{\xi \in B\} = \int_B p_{\xi}(y) dy, \quad (2.8)$$

и, в частности,

$$P\{a \leq \xi < b\} = \int_a^b p_{\xi}(y) dy. \quad (2.9)$$

Доказательство. Это свойство является следствием трех фактов:

- а) \mathcal{B} построена на основе базовой системы интервалов $\mathcal{F}_0 = \{\emptyset, A_x = (-\infty, x)\}$;
- б) для $\forall B \in \mathcal{F}_0$ формула (2.8) верна, так как она превращается в представление (2.7);
- в) интеграл обладает свойством аддитивности, следовательно, так как для базовых множеств (2.8) выполняется, а \mathcal{B} порождено их объединениями, то (2.8) выполняется $\forall B \in \mathcal{B}$.

Проиллюстрируем эту схему доказательства на формуле (2.9), используя свойства функции распределения:

$$\begin{aligned} P\{a \leq \xi < b\} &= F_{\xi}(b) - F_{\xi}(a) = \int_{A_b \setminus A_a} p_{\xi}(y) dy = \\ &= \int_{-\infty}^b p_{\xi}(y) dy - \int_{-\infty}^a p_{\xi}(y) dy = \int_a^b p_{\xi}(y) dy. \end{aligned}$$

□

Соотношение (2.8) имеет наглядную графическую иллюстрацию (рис. 2.3). Плотность распределения вероятностей неотрицательна и площадь фигуры под кривой равна единице.

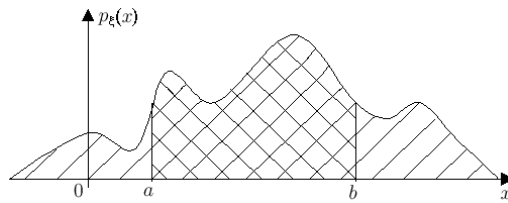


Рис. 2.3. График плотности распределения вероятностей

Таким образом, вероятность $P\{a \leq \xi < b\}$ есть площадь криволинейной трапеции, выделенной двойной штриховкой.

- С4. Каким бы ни было фиксированное значение $x \in \mathbb{R}$, для непрерывной случайной величины $\xi = \xi(\omega)$ справедливо соотношение

$$P\{\xi = x\} = 0.$$

Доказательство. Поскольку функция распределения $F_\xi(\cdot)$ непрерывна, то $P\{\xi = x\} = F_\xi(x+0) - F_\xi(x) = 0$. \square

Замечание 2.3. Это свойство приводит к парадоксу. Например, наудачу бросается точка на отрезок $[0,1]$. Одно из событий вида $\{\xi = x\}$ произойдет, но в то же время $P\{\xi = x\} = 0$.

- С5. Если x – точка непрерывности $p_\xi(\cdot)$, то справедливо

$$p_\xi(x) = F'_\xi(x). \quad (2.10)$$

Доказательство. Дифференцируя $F_\xi(x) = \int_{-\infty}^x p_\xi(y)dy$, получаем требуемое равенство. \square

Следствие 2.3. Справедливы соотношения:

$$F_\xi(x) = \int_{-\infty}^x p_\xi(y)dy, \quad x \in \mathbb{R}, \quad p_\xi(x) = F'_\xi(x).$$

- С6. В условиях С5 справедливо предельное соотношение

$$p_\xi(x) = \lim_{\Delta \rightarrow 0} \frac{P\{x \leq \xi < x + \Delta\}}{\Delta}. \quad (2.11)$$

Доказательство. В силу свойств функции распределения имеем

$$p_\xi(x) = \lim_{\Delta \rightarrow 0} \frac{F_\xi(x + \Delta) - F_\xi(x)}{\Delta} = \lim_{\Delta \rightarrow 0} \frac{P\{x \leq \xi < x + \Delta\}}{\Delta}.$$

\square

Замечание 2.4. Из соотношения (2.11) происходит название «плотность распределения». В физике плотность распределения массы на числовой прямой

$$\rho = \frac{m}{V} = \lim_{\Delta \rightarrow 0} \frac{m([x, x + \Delta))}{\Delta},$$

т. е. ρ – это плотность массы, сосредоточенной в точке x . Аналогично в теории вероятностей: плотность распределения СВ ξ характеризует плотность вероятности в точке x .

Задание 2.2. Показать справедливость формулы

$$P\{a < \xi < b\} = P\{a \leq \xi \leq b\} = P\{a \leq \xi < b\} = P\{a < \xi \leq b\},$$

если ξ имеет абсолютно непрерывное распределение вероятностей.

Основные абсолютно непрерывные законы распределения вероятностей

1. *Равномерный (прямоугольный) закон распределения вероятностей* на $[a, b]$, $a < b$:

$$\mathcal{L}\{\xi\} = R[a, b] - \text{Rectangular},$$

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases} \quad F_{\xi}(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & x > b. \end{cases}$$

Вероятность попадания зависит лишь от длины интервала, а не от его расположения, т. е. вероятностная масса распределена равномерно (рис. 2.4).

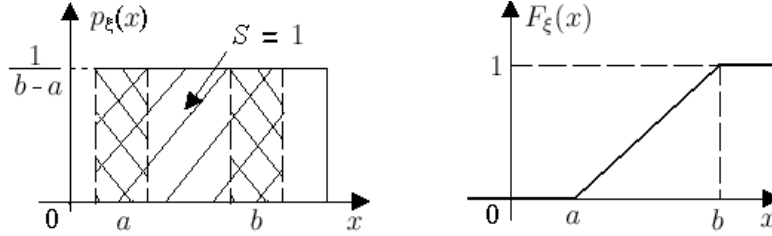


Рис. 2.4. Иллюстрация равномерного закона

Пример 2.7 (геометрическая вероятностная модель). В ранее рассмотренном примере (1.5) игры «Что? Где? Когда?» $\xi = \xi(\omega) = \omega$ – случайный угол остановившейся стрелки. Шансы остановиться в любой точке, принадлежащей $[0, 2\pi)$, одинаковы; $\mathcal{L}\{\xi\} = R[0, 2\pi)$.

Определение 2.9. Если $a = 0$, $b = 1$, то закон $R[0, 1]$ называется стандартным равномерным законом распределения вероятностей.

Пример 2.8. $\xi = \xi(\omega)$ – выходное значение программного датчика случайных чисел, например *Random* в языке *Pascal*.

2. **Нормальный (гауссовский) закон распределения вероятностей** с параметрами $a \in \mathbb{R}$, $D > 0$:

$$\mathcal{L}\{\xi\} = N(a, D),$$

где N происходит от англ. «Normal»:

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi D}} e^{-\frac{(x-a)^2}{2D}} =: n(x|a, D).$$

Графики этой плотности распределения для различных a, D приведены на рис. 2.5.

Функция распределения для нормального закона

$$F_{\xi}(x) = \int_{-\infty}^x n(y|a, D) dy, \quad x \in \mathbb{R}.$$

Пример 2.9. Допустим, на некотором производстве выпускают стержни для шариковых ручек. Пусть a – стандартная длина стержня. С конвейера наудачу берем один из изготовленных стержней и измеряем его длину: $\xi = \xi(\omega)$. Она отличается от a и является случайной величиной по следующим причинам:

- а) случайные погрешности настройки станка;
- б) случайные ошибки персонала;
- в) случайная неоднородность материала.

Если строить гистограмму, она будет иметь вид гауссоиды. При этом a – точка максимума, стандарт, который хотелось бы иметь; D характеризует точность изготовления. Заметим, что $P\{\xi = a\} = 0$, т. е. в точности стержень длины a появится в выборке с вероятностью 0.

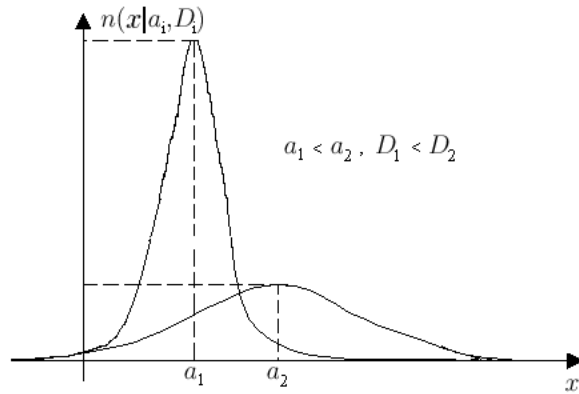


Рис. 2.5. Иллюстрация нормального закона

Определение 2.10. Если $a = 0, D = 1$, то закон распределения $N(0, 1)$ с плотностью $n(x|0, 1)$ называется стандартным нормальным законом распределения вероятностей;

$$\phi(x) = n(x|0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}, -$$

стандартная нормальная плотность распределения вероятностей;

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy, \quad x \in \mathbb{R}, -$$

стандартная нормальная функция распределения (функция Лапласа, или интеграл вероятностей).

Замечание 2.5. Функции ϕ, Φ имеются в таблицах, и существуют стандартные компьютерные программы для их вычисления.

Задание 2.3. Построить с помощью математических пакетов графики $\phi(x), \Phi(x)$.

Свойства ϕ, Φ

- C1. Функция $\phi(\cdot)$ – четная: $\phi(-x) = \phi(x), \quad x \in \mathbb{R}; \quad \Phi(-x) = 1 - \Phi(x)$.
 C2. Любой нормальный закон $\mathcal{L}\{\xi\} = N(a, D)$ выражается через стандартный нормальный закон:

$$p_{\xi}(x) = \frac{1}{\sqrt{D}} \phi\left(\frac{x-a}{\sqrt{D}}\right); \quad F_{\xi}(x) = \Phi\left(\frac{x-a}{\sqrt{D}}\right).$$

3. Закон распределения вероятностей Коши (обозначение происходит от англ. Cauchy) с параметрами $a \in \mathbb{R}, b > 0$:

$$\mathcal{L}\{\xi\} = C(a, b);$$

$$p_{\xi}(x) = \frac{1}{\pi b \left(1 + \left(\frac{x-a}{b} \right)^2 \right)}, \quad x \in \mathbb{R};$$

$$F_{\xi}(x) = \int_{-\infty}^x \frac{1}{\pi b \left(1 + \left(\frac{y-a}{b} \right)^2 \right)} dy.$$

Задание 2.4. Построить с помощью математических пакетов графики $p_{\xi}(\cdot)$ и $F_{\xi}(\cdot)$.
 4. Экспоненциальный закон распределения вероятностей с параметром $\lambda > 0$:

$$\mathcal{L}\{\xi\} = E\{\lambda\},$$

$$p_{\xi}(x) = \lambda e^{-\lambda x}, \quad x \geq 0; \quad F_{\xi}(x) = 1 - e^{-\lambda x}.$$

Задание 2.5. Построить с помощью математических пакетов графики $p_{\xi}(\cdot)$ и $F_{\xi}(\cdot)$.

2.2.3. Сингулярные функции распределения вероятностей

Определение 2.11. Точка $g \in \mathbb{R}$ называется точкой роста для $F_{\xi}(x)$, если для $\forall \epsilon > 0$ приращение $F_{\xi}(g + \epsilon) - F_{\xi}(g - \epsilon) > 0$. При этом $G(F_{\xi}) ::= \{g\} \subseteq \mathbb{R}$ – множество всех точек роста.

Приведем классификацию функций распределения в табл. 2.1.

Таблица 2.1

Класс функций распределения	Дискретные	Абсолютно непрерывные	Сингулярные
Непрерывна ли $F_{\xi}(\cdot)$?	нет	да	да
Мера Лебега $\mu(G(F_{\xi}))$	0	$\mu > 0$	0

Определение 2.12. Функция распределения $F_{\xi}(x)$ называется сингулярной функцией распределения, если она непрерывна, и в то же время множество точек роста имеет нулевую меру Лебега.

Примером сингулярной функции распределения может служить «лестница Кантора» [9, с. 51].

Замечание 2.6. Существует эквивалентное определение сингулярной функции распределения: $F_{\xi}(\cdot)$ непрерывна и существует борелевское множество (носитель) $S \in \mathcal{B}$ нулевой меры Лебега $\mu(S) = \text{mes}(S) = 0$ такое, что $P\{\xi \in S\} = 1$.

Теорема 2.5 (Лебега). Любая функция распределения $F(x)$ представима в виде линейной комбинации некоторой дискретной $F_{\partial}(x)$, некоторой абсолютно непрерывной $F_{an}(x)$ и некоторой сингулярной $F_c(x)$ функций распределения:

$$F(x) = p_1 \cdot F_{\partial}(x) + p_2 \cdot F_{an}(x) + p_3 \cdot F_c(x), \quad x \in \mathbb{R}, \quad (2.12)$$

где $p_1, p_2, p_3 \in [0, 1]$, $p_1 + p_2 + p_3 = 1$; (2.12) – смесь функций распределения; p_1, p_2, p_3 – коэффициенты этой смеси.

Следствие 2.4. Произведенная классификация функций распределения в три класса вместе со всевозможными их смесями является полной.

2.3. n -МЕРНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА И СЛУЧАЙНЫЙ n -ВЕКТОР. n -МЕРНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, n -МЕРНАЯ ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ И ИХ СВОЙСТВА

Эти понятия являются обобщением понятий одномерных случайных величин на многомерный случай. Раньше рассматривались случаи, когда каждому исходу ставилось в соответствие некоторое число: $\omega \rightarrow \xi = \xi(\omega) \in \mathbb{R}$ – случайная величина. Но часто в экспериментах измеряется сразу $n > 1$ случайных величин, т. е.

$$\omega \rightarrow \xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \in \mathbb{R}^n -$$

многомерная случайная величина (случайный вектор).

Определение 2.13. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определено n случайных величин: $\xi_i = \xi_i(\omega) : \Omega \rightarrow \mathbb{R}, i = 1, \dots, n$. Тогда упорядоченный по индексу i набор случайных величин

$$\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} = \xi(\omega) = \begin{pmatrix} \xi_1(\omega) \\ \vdots \\ \xi_n(\omega) \end{pmatrix}$$

называется n -мерной случайной величиной или случайным n -вектором на (Ω, \mathcal{F}, P) . При этом $\xi_i = \xi_i(\omega)$ – i -я компонента случайного вектора.

Рассмотрим измеримое пространство $(\mathbb{R}^n, \mathcal{B}^n)$: \mathbb{R}^n – евклидово n -мерное пространство; \mathcal{B}^n – борелевская σ -алгебра подмножеств из \mathbb{R}^n . \mathcal{B}^n строится на основе базовой системы параллелепипедов:

$$F_0 = \{\emptyset, A_x = A_{x_1} \times A_{x_2} \times \dots \times A_{x_n} : x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\}, \quad A_{x_i} = (-\infty, x_i).$$

Поскольку $\xi_i = \xi_i(\omega)$ \mathcal{F} -измерима, $\forall i = 1, \dots, n$, то векторная функция $\xi = \xi(\omega)$ тоже \mathcal{F} -измерима, т. е. $\xi^{-1}(B) \in \mathcal{F}, \forall B \in \mathcal{B}^n$.

Определение 2.14. Распределением вероятностей случайного вектора $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}^n$, заданного на вероятностном пространстве (Ω, \mathcal{F}, P) , называется вероятностная мера

$$P_\xi(B) ::= P(\xi^{-1}(B)) = P\{\xi \in B\}, \quad B \in \mathcal{B}^n.$$

Определение 2.15. Функция n действительных переменных

$$F_\xi(x) = F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P\{\xi_1 < x_1, \dots, \xi_n < x_n\} = P_\xi(A_x),$$

$$x = (x_i) \in \mathbb{R}^n,$$

называется n -мерной функцией распределения или совместной функцией распределения случайных величин ξ_1, \dots, ξ_n .

Свойства n -мерной функции распределения

C1. Распределение вероятностей случайного вектора $\mathbf{P}_\xi(\cdot)$ и n -мерная функция распределения $F_\xi(\cdot)$ находятся во взаимно-однозначном соответствии:

$$\mathbf{P}_\xi(\cdot) \Leftrightarrow F_\xi(\cdot).$$

C2. Для $\forall x \in \mathbb{R}^n$ $F_\xi(x) \in [0, 1]$.

C3. $F_\xi(\cdot)$ – неубывающая функция по каждому из своих аргументов.

C4. Справедливы предельные соотношения ($i = 1, \dots, n$):

$$F_\xi(x) \big|_{x_i = -\infty} = 0$$

(здесь достаточно устремить к $-\infty$ по крайней мере одну координату);

$$F_\xi(x) \big|_{x_1 = +\infty, \dots, x_n = +\infty} = 1$$

(в данном случае необходимо устремить к $+\infty$ все координаты вектора x).

C5. Функция распределения $F_\xi(\cdot)$ непрерывна слева по совокупности своих переменных. Доказательства C1–C5 проводятся аналогично случаю $n = 1$.

C6. (Свойство согласованности.) n - и $(n - 1)$ -мерные функции распределения согласованы следующим образом:

$$\begin{aligned} F_\xi(x) \big|_{x_i = +\infty} &= F_{\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), \\ i &= 1, \dots, n, \quad x_1, \dots, x_n \in \mathbb{R}. \end{aligned} \quad (2.13)$$

Доказательство. В определении 2.15 устремим координату $x_i \rightarrow +\infty$.

Левая часть (2.13)

$$\begin{aligned} F_\xi(x) \big|_{x_i = +\infty} &= \mathbf{P}\{\xi_1 < x_1, \dots, \xi_{i-1} < x_{i-1}, \xi_i < +\infty, \xi_{i+1} < x_{i+1}, \\ &\quad \xi_{i+2} < x_{i+2}, \dots, \xi_n < x_n\}. \end{aligned}$$

Учитывая, что $\{\xi_i < +\infty\}$ есть достоверное событие, удаляя его из правой части и используя определение 2.15, получаем правую часть формулы (2.13). \square

Определение 2.16. Пусть k ($1 \leq k \leq n - 1$) – произвольное натуральное число и задан произвольный набор k индексов: $1 \leq i_1 < i_2 < \dots < i_k \leq n$. Тогда упорядоченный набор k компонент исходного случайного n -вектора $\xi \in \mathbb{R}^n$

$$\xi' = \begin{pmatrix} \xi_{i_1} \\ \vdots \\ \xi_{i_k} \end{pmatrix} \in \mathbb{R}^k$$

называется случайным подвектором, а его k -мерная функция распределения – k -мерной маргинальной (от англ. *marginal*) (частной) функцией распределения.

Обозначим k -мерную маргинальную ФР:

$$F_{\xi'}(x'), \quad x' = \begin{pmatrix} x_{i_1} \\ \vdots \\ x_{i_k} \end{pmatrix} \in \mathbb{R}^k.$$

Следствие 2.5. Чтобы получить k -мерную маргинальную функцию распределения $F_{\xi'}(x')$, достаточно в исходной n -мерной функции распределения $F_\xi(x)$ устремить к $+\infty$ все $n - k$ переменных, за исключением x_{i_1}, \dots, x_{i_k} .

Примем обозначения:

$\xi = (\xi_i) \in \mathbb{R}^n$ – случайный n -вектор с n -мерной функцией распределения $F_\xi(x)$;
 $C_i = [a_i, b_i]$ – i -й числовой промежуток, где a_i, b_i ($a_i < b_i$) – действительные числа;
 $C = C_1 \times C_2 \times \dots \times C_n \in \mathbb{R}^n$ – прямоугольный параллелепипед;

$$\Delta_i F_\xi(x) = F_\xi(x) \big|_{x_i=b_i} - F_\xi(x) \big|_{x_i=a_i} -$$

приращение функции распределения по i -й переменной на промежутке C_i .

C7. Вероятность попадания вектора ξ в прямоугольный параллелепипед C равна n -кратному приращению функции распределения:

$$P\{\xi \in C\} = \Delta_1 \Delta_2 \dots \Delta_n F_\xi(x) \geq 0. \quad (2.14)$$

Доказательство. Введем в рассмотрение вспомогательную функцию:

$$P\{\xi \in C\} =: Q(C_1, \dots, C_n) = P\{\xi_1 \in C_1, \dots, \xi_n \in C_n\},$$

$$C_i = A_{b_i} \setminus A_{a_i},$$

где $A_y = (-\infty, y)$. Тогда по свойствам вероятности имеем цепочку равенств:

$$Q(C_1, \dots, C_n) = Q(A_{b_1} \setminus A_{a_1}, C_2, \dots, C_n) = P\{\xi_1 \in A_{b_1}, \xi_2 \in C_2,$$

$$\xi_3 \in C_3, \dots, \xi_n \in C_n\} - P\{\xi_1 \in A_{a_1}, \xi_2 \in C_2, \dots, \xi_n \in C_n\} =$$

$$= Q(A_{b_1}, C_2, \dots, C_n) - Q(A_{a_1}, C_2, \dots, C_n) =$$

$$= \Delta_1 Q(A_{x_1}, C_2, \dots, C_n) = \dots = \Delta_1 \Delta_2 \dots \Delta_n Q(A_{x_1}, \dots, A_{x_n}) =$$

$$= \Delta_1 \Delta_2 \dots \Delta_n F_\xi(x) \geq 0.$$

□

Задание 2.6. Записать соотношение (2.14) в явном виде при $n = 2$.

C8. Справедливо обобщение теоремы Лебега о классификации функций распределения.

Определение 2.17. n -Мерная функция распределения $F_\xi(x)$, $x \in \mathbb{R}^n$, называется абсолютно непрерывной функцией распределения, если существует такая неотрицательная интегрируемая функция n действительных переменных $p_\xi(y)$, $y = (y_i) \in \mathbb{R}^n$, что справедливо интегральное представление

$$F_\xi(x) = F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p_\xi(y_1, \dots, y_n) dy_1, \dots, dy_n;$$

при этом функция $p_\xi(y)$ называется плотностью распределения вероятностей случайного вектора ξ или совместной плотностью распределения случайных величин ξ_1, \dots, ξ_n .

Свойства n -мерной плотности распределения вероятностей

C1. $p_\xi(x) \geq 0$, $x \in \mathbb{R}^n$.

C2. Выполняется условие нормировки: $\int_{\mathbb{R}^n} p_\xi(x) dx = 1$.

C3. Для $\forall B \in \mathcal{B}^n$ вероятность $P\{\xi \in B\} = \int_B p_\xi(x) dx$.

С4. Если x – точка непрерывности $p_\xi(\cdot)$, то справедлива формула

$$p_\xi(x) = \frac{\partial^n F_\xi(x)}{\partial x_1 \dots \partial x_n}.$$

Свойства С1–С4 доказываются аналогично случаю с $n = 1$.

С5. (Свойство согласованности плотностей распределения вероятностей.) n -Мерная и $(n - 1)$ -мерная плотности распределения вероятностей согласованы:

$$\begin{aligned} & \int_{-\infty}^{+\infty} p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_i = \\ & = p_{\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), \\ & \forall x_1, \dots, x_n \in \mathbb{R}^n, \forall i = 1, \dots, n. \end{aligned}$$

Доказательство. Достаточно продифференцировать почленно по переменным $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ соотношение (2.13) и воспользоваться С4. \square

Определение 2.18. k -Мерной маргинальной плотностью распределения вероятностей случайного подвектора $\xi', \xi' = (\xi_{i_1}, \dots, \xi_{i_k})^T \in \mathbb{R}^k$ называется совместная плотность распределения вероятностей случайных величин $\xi_{i_1}, \dots, \xi_{i_k}$.

Следствие 2.6. Чтобы получить k -Мерную маргинальную ПРВ, достаточно исходную n -мерную плотность распределения вероятностей $p_\xi(x)$, $x \in \mathbb{R}^n$, проинтегрировать в бесконечных пределах $n - k$ раз по всем переменным, за исключением x_{i_1}, \dots, x_{i_k} .

Определение 2.19. Случайный m -вектор $\xi = (\xi_1, \dots, \xi_m)^T \in \mathbb{R}^m$ называется гауссовским случайным вектором, или говорят, что многомерный случайный вектор ξ имеет n -мерное нормальное распределение вероятностей с параметрами $a = (a_1, \dots, a_m) \in \mathbb{R}^m$ и $B = (b_{ij}) = (m \times m)$ -матрицей, и это кратко обозначается $\mathcal{L}(\xi) = N_m(a, B)$, если плотность распределения ξ имеет вид

$$p_\xi(x) = (2\pi)^{-\frac{m}{2}} |B|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (x - a)^T B^{-1} (x - a) \right) =: n_m(x|a, B),$$

$$x = (x_i) \in \mathbb{R}^m$$

(здесь $B = B^T \succ 0$ – положительно определенная симметрическая матрица).

Задание 2.7. Построить график этой плотности распределения при $m = 2$.

Следствие 2.7. Для гауссовского случайного вектора все маргинальные распределения вероятностей являются гауссовскими. В частности, одномерное маргинальное распределение i -й компоненты имеет вид

$$\mathcal{L}\{\xi_i\} = N_1(a_i, b_{ii}), \quad i = 1, \dots, m.$$

Задание 2.8. Используя С5, доказать формулу для одномерного маргинального распределения i -й компоненты при $m = 2$.

Определение 2.20. Принято говорить, что случайный вектор $\xi = (\xi_1, \dots, \xi_m)^T \in \mathbb{R}^m$ имеет полиномиальное распределение вероятностей с параметрами N, m, p_1, \dots, p_m : $\mathcal{L}\{\xi\} = \text{Pol}(N, m, p_1, \dots, p_m)$, если дискретное распределение вероятностей этого вектора имеет вид

$$P\{\xi_1 = x_1, \dots, \xi_m = x_m\} = \frac{N!}{x_1! \dots x_m!} \cdot p_1^{x_1} \cdot \dots \cdot p_m^{x_m},$$

где

$$x_1, \dots, x_m \in \{0, 1, \dots, N\}, \sum_{i=1}^N x_i \equiv N, 0 \leq p_i \leq 1, \sum_{i=1}^m p_i = 1.$$

Это распределение возникает в схеме независимых испытаний, когда в каждом независимом испытании имеется m исходов, а N – число испытаний, p_i – вероятность i -го исхода в единичном испытании.

Пример 2.10. Студент сдает сессию, состоящую из $N = 5$ экзаменов, $|\{3, 4, \dots, 10\}| = m = 8$ – число исходов, p_i – вероятность получить соответствующую оценку ($p_1 = 3, p_2 = 4, \dots, p_8 = 10$), ξ_i – число оценок типа i , которое получил студент за всю сессию (случайное число оценок i -го типа).

Следствие 2.1. Допустим, $m = 2$ («зачет – незачет») и в сессии N зачетов. Получаем схему независимых испытаний Бернулли, и полиномиальное распределение вероятностей превращается в биномиальное.

Следствие 2.2. Одномерное маргинальное распределение является биномиальным:

$$\mathcal{L}\{\xi_i\} = B_i(N, p_i), i = 1, \dots, m.$$

2.4. УСЛОВНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, УСЛОВНАЯ ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ И ИХ СВОЙСТВА

На практике часто приходится находить распределение вероятностей случайной величины ξ при условии, что некоторая другая случайная величина приняла фиксированное значение $\eta = y$. Например:

1) η – оценка, полученная на первом экзамене;

ξ – оценка, полученная на втором экзамене.

Как влияет η на распределение вероятностей ξ ?

2) η – расход фирмы на рекламу в течение месяца;

ξ – доход фирмы.

Если затраты на рекламу составили $\eta = y$, то каким будет распределение дохода?

Обозначим: $\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \in \mathbb{R}^n$ – случайный вектор; $x = (x_i) \in \mathbb{R}^n$; $A_{x_i} = (-\infty, x_i)$; $A_x = A_{x_1} \times \dots \times A_{x_n} \in \mathcal{B}^n$; \mathcal{B}^n – борелевская σ -алгебра подмножеств в \mathbb{R}^n .

Определение 2.21. Пусть на (Ω, \mathcal{F}, P) определен вектор $\xi = (\xi_i) \in \mathbb{R}^n$, $F_\xi(x) = F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) ::= P\{\xi \in A_x\}$, и на этом же пространстве определено случайное событие $C \in \mathcal{F}$, $P(C) > 0$; тогда условной функцией распределения случайного вектора ξ при условии случайного события C называется функция n действительных переменных

$$F_\xi(x|C) ::= P\{\xi \in A_x|C\} = \frac{P(\{\xi \in A_x\} \cap C)}{P(C)}, x \in \mathbb{R}^n. \quad (2.15)$$

При этом $F_\xi(x)$ называется безусловной функцией распределения.

Определение 2.22. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены случайный вектор $\xi = (\xi_i) \in \mathbb{R}^n$ и дискретный случайный вектор $\eta = (\eta_j) \in Y \subset \mathbb{R}^m$. Тогда, если вероятность $P\{\eta = y\} > 0 (y \in Y)$, то условной функцией распределения случайного вектора ξ при условии $\eta = y$ называется функция $n + m$ действительных переменных:

$$F_{\xi|\eta}(x|y) ::= P\{\xi \in A_x|\eta = y\}, x \in \mathbb{R}^n, y \in Y.$$

Пусть теперь составной случайный вектор $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbb{R}^{n+m}$ имеет абсолютно непрерывный закон распределения вероятностей с некоторой совместной плотностью распределения вероятностей $p_{\xi, \eta}(x, y)$; $p_{\eta}(y) = \int_{\mathbb{R}^n} p_{\xi, \eta}(x, y) dx$ – маргинальная плотность распределения подвектора η . Но воспользоваться определением 2.22 не представляется возможным, так как $P\{\eta = y\} = 0$.

Обозначим ϵ -кирпич: $C_y(\epsilon) ::= \{v = (v_j) \in \mathbb{R}^m : y_j \leq v_j < y_j + \epsilon, j = 1, \dots, m\} \subset \mathbb{R}^m$.

Определение 2.23. Пусть составной случайный вектор $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbb{R}^{n+m}$ имеет абсолютно непрерывное распределение вероятностей с непрерывной плотностью $p_{\xi, \eta}(x, y)$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Тогда если $p_{\eta}(y) > 0$, то условной функцией распределения ξ при условии $\eta = y$ называется функция $n + m$ действительных переменных, определяемая соотношением

$$F_{\xi, \eta}(x|y) = \lim_{\epsilon \rightarrow 0} P\{\xi \in A_x | \eta \in C_y(\epsilon)\}. \quad (2.16)$$

Покажем, что в указанных условиях предел в правой части (2.16) существует и, следовательно, определение корректно.

По определению условной вероятности имеем

$$\lim_{\epsilon \rightarrow 0} P\{\xi \in A_x | \eta \in C_y(\epsilon)\} = \lim_{\epsilon \rightarrow 0} \frac{\int_{A_x} du \int_{C_y(\epsilon)} p_{\xi, \eta}(u, v) dv}{\int_{C_y(v)} p_{\eta}(v) dv}.$$

Применим к интегралам по v в числителе и знаменателе теорему о среднем, обозначив $v', v'' \in C_y(\epsilon)$ – «промежуточные точки». С учетом этого и непрерывности $p_{\xi, \eta}(\cdot)$, $p_{\xi}(\cdot)$ имеем

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} P\{\xi \in A_x | \eta \in C_y(\epsilon)\} &= \lim_{\epsilon \rightarrow 0} \frac{\int_{A_x} p_{\xi, \eta}(u, v'') du \cdot \epsilon^m}{p_{\eta}(v') \epsilon^m} = \\ &= \int_{A_x} \frac{p_{\xi, \eta}(u, v)}{p_{\eta}(y)} du =: \int_{A_x} p_{\xi|\eta}(u|y) du, \end{aligned} \quad (2.17)$$

где обозначено

$$p_{\xi|\eta}(x|y) = \frac{p_{\xi, \eta}(x, y)}{p_{\eta}(y)} \geq 0, \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^m. \quad (2.18)$$

Проведенный анализ позволяет сделать выводы:

- 1) предел в (2.16) существует, и, следовательно, определение корректно;
- 2) из (2.17) видно: условная функция распределения $F_{\xi|\eta}(\cdot)$ представима в виде интеграла от неотрицательной функции $p_{\xi|\eta}(u|y) \geq 0$ по параллелепипеду A_x , а из этого следует, что эта условная функция распределения абсолютно непрерывна; функция (2.18) является условной плотностью распределения вероятностей.

Определение 2.24. Функция $n + m$ действительных переменных $p_{\xi|\eta}(x|y)$, определяемая (2.18), называется условной плотностью распределения вероятностей случайного вектора ξ при условии $\eta = y$, при этом $p_{\xi}(x)$ – безусловная плотность распределения вероятностей.

Свойства условной плотности распределения вероятностей

C1. Неотрицательность: $p_{\xi|\eta}(x|y) \geq 0$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$.

C2. Условие нормировки: $\int_{\mathbb{R}^n} p_{\xi|\eta}(x|y) dx = 1$, $\forall y \in \mathbb{R}^m$.

C3. $\forall B \in \mathcal{B}^n$ справедлива формула

$$P\{\xi \in B | \eta = y\} = \int_B p_{\xi|\eta}(x|y) dx, \quad y \in \mathbb{R}^m.$$

C4. Если x – точка непрерывности условной плотности распределения вероятностей $p_{\xi|\eta}(x|y)$, то справедливо следующее равенство:

$$p_{\xi|\eta}(x|y) = \frac{\partial^n F_{\xi|\eta}(x|y)}{\partial x_1 \dots \partial x_n}, \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^m.$$

Свойства C1–C4 доказываются аналогично свойствам безусловной плотности распределения вероятностей.

C5. (Симметрическая формула умножения плотностей.) Если $p_{\xi}(x) > 0$ и $p_{\eta}(y) > 0$, то

$$p_{\xi, \eta}(x, y) = p_{\eta}(y) \cdot p_{\xi|\eta}(x|y) = p_{\xi}(x) \cdot p_{\eta|\xi}(y|x).$$

Доказательство. Эта формула следует из (2.18). □

C6. Формула безусловной плотности распределения вероятностей:

$$p_{\xi}(x) = \int_{\mathbb{R}^m} p_{\xi|\eta}(x|y) p_{\eta}(y) dy.$$

Доказательство. Эту формулу можно получить из C5 и свойства маргинальной плотности распределения вероятностей. □

C7. (Формула Байеса для плотностей распределения вероятностей.)

$$p_{\eta|\xi}(y|x) = \frac{p_{\xi|\eta}(x|y) \cdot p_{\eta}(y)}{\int_{\mathbb{R}^m} p_{\xi|\eta}(x|v) p_{\eta}(v) dv}, \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^m.$$

Доказательство. Эту формулу можно получить из свойств C5 и C6. □

C8. Пусть $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \in \mathbb{R}^2$ – двухмерный гауссовский случайный вектор, $\mathcal{L}\{\xi\} = N_2(a, B)$, где $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. Тогда условное распределение вероятностей ξ_1 при условии $\xi_2 = x_2$ – тоже гауссово:

$$\mathcal{L}\{\xi_1 | \xi_2 = x_2\} = N_1(a_{1|2}, b_{11|2}),$$

$$a_{1|2} = a_1 + \frac{b_{12}}{b_{22}}(x_2 - a_2); \quad b_{11|2} = b_{11} - \frac{b_{12}^2}{b_{22}}.$$

Доказать самостоятельно.

2.5. НЕЗАВИСИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН

Определение 2.25. Функция $y = f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^N$ называется борелевской функцией, если прообраз любого борелевского множества является борелевским множеством: $\forall B \in \mathcal{B}^N \ f^{-1}(B) \in \mathcal{B}^n$.

Определение 2.26. Случайные величины ξ_1, \dots, ξ_n ($n \geq 2$) называются независимыми в совокупности на вероятностном пространстве (Ω, \mathcal{F}, P) , если для $\forall B_1, \dots, B_n \in \mathcal{B}^1$ независимы в совокупности случайные события: $\{\xi_1 \in B_1\}, \dots, \{\xi_n \in B_n\} \in \mathcal{F}$, так что выполняется соотношение

$$P\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \prod_{i=1}^n P\{\xi_i \in B_i\}. \quad (2.19)$$

В противном случае величины называются зависимыми. Если вышеуказанные случайные события попарно независимы, то и случайные величины называются попарно независимыми.

Свойства независимых случайных величин

C1. Свойство независимости сохраняется при борелевских функциональных преобразованиях.

Доказательство. Пусть $\xi = (\xi_i) \in \mathbb{R}^n$, $\eta = (\eta_j) \in \mathbb{R}^m$ — независимы и заданы борелевские функции: $y = f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^N$ и $v = g(u) : \mathbb{R}^m \rightarrow \mathbb{R}^M$. Покажем, что $\alpha = f(\xi) \in \mathbb{R}^N$ и $\beta = g(\eta) \in \mathbb{R}^M$ независимы. Проверим выполнение (2.19). Для любых $B_1 \in \mathcal{B}^N$, $B_2 \in \mathcal{B}^M$ имеем в силу независимости ξ и η

$$\begin{aligned} P\{\alpha \in B_1, \beta \in B_2\} &= P\{f(\xi) \in B_1, g(\eta) \in B_2\} = \\ &= P\{\xi \in f^{-1}(B_1), \eta \in g^{-1}(B_2)\} = P\{\xi \in f^{-1}(B_1)\} \times \\ &\times P\{\eta \in g^{-1}(B_2)\} = P\{\alpha \in B_1\} \times P\{\beta \in B_2\}. \end{aligned}$$

Поскольку (2.19) выполняется, то α и β независимы. \square

C2. (Общий критерий независимости случайных величин.) Случайные величины ξ_1, \dots, ξ_n независимы в совокупности на вероятностном пространстве (Ω, \mathcal{F}, P) тогда и только тогда, когда их совместная n -мерная функция распределения совпадает с произведением n одномерных маргинальных функций распределения:

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n), x_1, \dots, x_n \in \mathbb{R}. \quad (2.20)$$

Доказательство. Необходимость. Пусть ξ_1, \dots, ξ_n независимы в совокупности, тогда выполняется соотношение (2.19):

$$P\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = P\{\xi_1 \in B_1\} \cdot \dots \cdot P\{\xi_n \in B_n\},$$

$$\forall B_1, \dots, B_n \in \mathcal{B}.$$

Выберем $B_i = A_{x_i} = (-\infty, x_i)$, $i = 1, \dots, n$, тогда $\{\xi_i \in B_i\} = \{\xi_i < x_i\}$, и левая часть (2.19) получается равной левой части (2.20), и правые их части соответственно тоже равны.

Достаточность. Пусть выполняется соотношение (2.20). Тогда (2.19) вытекает из следующих двух фактов:

а) соотношение (2.19) выполняется для любого базового борелевского множества;

б) $\forall B \in \mathcal{B}$ порождается множествами из базовой системы \mathcal{F}_0 применением счетного числа объединений, пересечений и дополнений.

Проиллюстрируем это при $B_i = [a_i, b_i] = A_{b_i} \setminus A_{a_i}$, $i = 1, \dots, n$. Обозначим прямоугольный параллелепипед: $B = B_1 \times \dots \times B_n$. Тогда по свойствам n -мерной функции распределения вероятностей с учетом (2.20) имеем

$$\begin{aligned} P\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} &= P\{\xi \in B\} = \\ &= \Delta_1 \dots \Delta_n F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \\ &= \prod_{i=1}^n (F_{\xi_i}(b_i) - F_{\xi_i}(a_i)) = \prod_{i=1}^n P\{\xi_i \in B_i\}. \end{aligned}$$

□

С3. (Критерий независимости для абсолютно непрерывных случайных величин.) Непрерывные случайные величины ξ_1, \dots, ξ_n независимы в совокупности на вероятностном пространстве (Ω, \mathcal{F}, P) тогда и только тогда, когда их совместная n -мерная плотность распределения вероятностей равна произведению n одномерных маргинальных плотностей распределения вероятностей:

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = p_{\xi_1}(x_1) \cdot \dots \cdot p_{\xi_n}(x_n), \quad x_1, \dots, x_n \in \mathbb{R}. \quad (2.21)$$

Доказательство. Для доказательства этого свойства достаточно продифференцировать (2.20) n раз по x_1, \dots, x_n . □

Следствие 2.3. Случайные величины ξ, η (вообще говоря, многомерные) независимы тогда и только тогда, когда их условная плотность совпадает с безусловной:

$$p_{\xi|\eta}(x|y) = p_{\xi}(x), \quad \forall x, y. \quad (2.22)$$

Доказательство. Во-первых, из соотношения (2.21) имеем $p_{\xi, \eta}(x, y) = p_{\xi}(x)p_{\eta}(y)$, а во-вторых, $p_{\xi, \eta}(x, y) = p_{\eta}(y) \cdot p_{\xi|\eta}(x|y)$. Справедлив и симметричный результат. □

Следствие 2.4. Пусть $\xi = (\xi_1, \dots, \xi_n)^T \in \mathbb{R}^n$ – гауссовский (нормальный) случайный вектор, имеющий распределение $\mathcal{L}\{\xi\} = N_n(a, B)$. Тогда величины ξ_1, \dots, ξ_n независимы в совокупности тогда и только тогда, когда матрица B является диагональной: $B = \text{diag}\{b_{11}, \dots, b_{nn}\}$.

Доказательство. Достаточно проверить соотношение (2.21), используя вид гауссовской плотности. □

Следствие 2.5. Случайные величины ξ, η независимы тогда и только тогда, когда условная функция распределения совпадает с безусловной функцией распределения

$$F_{\xi|\eta}(x|y) = F_{\xi}(x), \quad \forall x, y.$$

Доказательство. Формула получается интегрированием соотношения (2.22). □

Следствие 2.6 (критерий независимости ДСВ). Случайные величины $\xi_1, \dots, \xi_n \in \mathcal{X}$ (\mathcal{X} – дискретное множество) независимы в совокупности на (Ω, \mathcal{F}, P) тогда и только тогда, когда выполняется соотношение

$$P\{\xi_1 = x_1, \dots, \xi_n = x_n\} = P\{\xi_1 = x_1\} \cdot \dots \cdot P\{\xi_n = x_n\},$$

$$\forall x_1, \dots, x_n \in \mathcal{X}.$$

2.6. ФУНКЦИОНАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определен произвольный случайный n -вектор ($n \geq 1$)

$$\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} = \xi(\omega) = \begin{pmatrix} \xi_1(\omega) \\ \vdots \\ \xi_n(\omega) \end{pmatrix} \in \mathbb{R}^n,$$

имеющий некоторую функцию распределения

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_\xi(x) = P\{\xi \in A_x\}, \quad (2.23)$$

$A_x = A_{x_1} \times \dots \times A_{x_n}$; $A_{x_j} = (-\infty, x_j]$; $x = (x_j) \in \mathbb{R}^n$, $j = 1, \dots, n$.

Пусть также задано некоторое борелевское функциональное преобразование $\mathbb{R}^n \rightarrow \mathbb{R}^m$:

$$y = f(x), \text{ где } y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m, \quad f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}. \quad (2.24)$$

С помощью соотношений (2.23) и (2.24) определим новый m -мерный случайный вектор – результат функционального преобразования:

$$\eta = f(\xi) \in \mathbb{R}^m, \text{ где } \eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_m \end{pmatrix} \in \mathbb{R}^m, \quad \eta_i = f_i(\xi_1, \dots, \xi_n),$$

$$i = 1, \dots, m. \quad (2.25)$$

Рассмотрим задачу отыскания закона распределения $\mathcal{L}\{\eta\}$ по заданным $f(\cdot)$ и $F_\xi(\cdot)$. Будем искать функцию распределения $F_\eta(\cdot)$. Используя (2.25), из определения получаем

$$F_\eta(y) = P\{\eta \in A_y\} = P\{f(\xi) \in A_y\} = P\{\xi \in f^{-1}(A_y)\},$$

$$A_y = A_{y_1} \times \dots \times A_{y_m}, \quad A_{y_i} = (-\infty, y_i], \quad i = 1, \dots, m.$$

Таким образом, справедлива общая формула

$$F_\eta(y) = P\{\xi \in f^{-1}(A_y)\}, \quad y = (y_i) \in \mathbb{R}^m. \quad (2.26)$$

Отметим, что в этой формуле $B = f^{-1}(A_y) \in \mathcal{B}^n$, поэтому $P\{\xi \in B\}$ можно выразить через $F_\xi(\cdot)$. К сожалению, в общем виде записать это выражение невозможно. Поэтому рассмотрим наиболее важные частные случаи.

Предположим, преобразование (2.25) непрерывно дифференцируемо, так что существует матрица Якоби:

$$J_f = \left(\frac{\partial f_i(x_1, \dots, x_n)}{\partial x_j} \right), \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

Кроме того, предположим, что вектор ξ имеет абсолютно непрерывное распределение вероятностей, так что существует плотность распределения вероятностей:

$$p_\xi(x) = \frac{\partial^n F_\xi(x)}{\partial x_1 \dots \partial x_n}.$$

Выделим два случая.

I. Пусть функциональное преобразование (2.25) – невырожденное, $\text{rank}(\mathcal{J}_f) = t = n$, функции $f_1(x), \dots, f_n(x)$ функционально независимы и

$$|\mathcal{J}_f| = \frac{Df(x)}{Dx} \neq 0, \exists ! x = f^{-1}(y), \frac{Df^{-1}(y)}{Dy} \neq 0. \quad (2.27)$$

Теорема 2.6. Пусть $\xi = (\xi_i) \in \mathbb{R}^n$ – случайный n -вектор с n -мерной непрерывной плотностью распределения вероятностей $p_\xi(x)$, а функциональное преобразование (2.25) – непрерывно дифференцируемо, невырождено, так что выполняется (2.27). Тогда случайный n -вектор $\eta = f(\xi) \in \mathbb{R}^n$ также имеет абсолютно непрерывное распределение вероятностей, и его плотность распределения вероятностей вычисляется по формуле

$$p_\eta(y) = p_\xi(f^{-1}(y)) \cdot \left| \frac{Df^{-1}(y)}{Dy} \right|, \quad y \in \mathbb{R}^n. \quad (2.28)$$

Доказательство. Воспользуемся формулой (2.26) и свойством многомерной плотности распределения вероятностей: $F_\eta(y) = \int_{f^{-1}(A_y)} p_\xi(x) dx$.

Сделаем замену переменных $x = f^{-1}(z)$, $z = f(x)$ в кратном интеграле:

$$F_\xi(y) = \int_{A_y} p_\xi(f^{-1}(z)) \cdot \left| \frac{Df^{-1}(z)}{Dz} \right| dz =: \int_{A_y} p_\eta(z) dz.$$

Отсюда получаем:

- 1) $F_\eta(y) = \int_{A_y} p_\eta(z) dz$, так как подынтегральная функция неотрицательна, то эта функция распределения абсолютно непрерывна;
- 2) подынтегральная функция $p_\eta(z)$ – плотность этого распределения вероятностей.

□

Следствие 2.7 (случай линейного преобразования). Пусть ξ – абсолютно непрерывная случайная величина с плотностью $p_\xi(x)$, $x \in \mathbb{R}^1$, $y = a + bx$ – невырожденное линейное преобразование ($a \in \mathbb{R}$, $b \neq 0$). Тогда случайная величина $\eta = a + b\xi \in \mathbb{R}^1$ также является абсолютно непрерывной случайной величиной и имеет плотность распределения вероятностей

$$p_\eta(y) = \frac{1}{|b|} p_\xi\left(\frac{y-a}{b}\right), \quad y \in \mathbb{R}. \quad (2.29)$$

Доказательство. Полагаем $t = n = 1$ и применяем (2.28). □

Следствие 2.8. Если в условиях следствия 2.7 ξ – гауссовская случайная величина с некоторым распределением $\mathcal{L}\{\xi\} = N_1(\mu, D)$, то при линейном преобразовании гауссовость сохраняется: $\mathcal{L}\{\eta\} = N_1(a + b\mu, b^2 D)$.

Доказательство. Достаточно воспользоваться (2.29). □

Следствие 2.9. Если функциональное преобразование таково, что обратная функция $x = f^{-1}(y)$ – не единственная, т. е. если уравнение $f(x) = y$ имеет $K > 1$ решений: $x = f_{(1)}^{-1}(y), \dots, x = f_{(K)}^{-1}(y)$, то формула (2.28) обобщается следующим образом:

$$p_\eta(y) = \sum_{k=1}^K p_\xi\left(f_{(k)}^{-1}(y)\right) \left| \frac{Df_{(k)}^{-1}(y)}{Dy} \right|, \quad y \in \mathbb{R}^n.$$

Доказательство повторяет доказательство теоремы, только прообраз $f^{-1}(A_y)$ является суммой K прообразов. \square

II. Пусть функциональное преобразование (2.25) – вырожденное, $\text{rank}(\mathcal{J}_f) = m < n$. При этом $f_1(x), \dots, f_m(x)$ – независимые функции и существует $n - m$ вспомогательных функций $f_{m+1}(x), \dots, f_n(x)$, которые в совокупности дают систему n функционально независимых элементов.

Сведем вырожденный случай к невырожденному, рассмотренному ранее. Введем расширенное функциональное преобразование $\mathbb{R}^n \rightarrow \mathbb{R}^n$. Обозначим его $\bar{y} = \bar{f}(x)$, где

$$\bar{y} = \begin{pmatrix} y \\ \dots \dots \\ y_{m+1} \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n, \quad \bar{f}(x) = \begin{pmatrix} f(x) \\ \dots \dots \\ f_{m+1}(x) \\ \vdots \\ f_n(x) \end{pmatrix} \in \mathbb{R}^n.$$

Тогда построим расширенный случайный вектор $\bar{\eta} = \bar{f}(\xi) = (\eta, \dots, \eta_{m+1}, \dots, \eta_n)^T$. Вспомогательные функции выбираем так, чтобы составное преобразование удовлетворяло условиям (2.27). Тогда по теореме 2.6 имеем

$$p_{\bar{\eta}}(\bar{y}) = p_{\xi}(\bar{f}^{-1}(\bar{y})) \cdot \left| \frac{D\bar{f}^{-1}(\bar{y})}{D\bar{y}} \right|, \quad \bar{y} \in \mathbb{R}^n. \quad (2.30)$$

Плотность распределения вероятностей подвектора η получим, используя свойство маргинальной плотности распределения вероятностей. Для этого проинтегрируем плотность (2.30) по всем вспомогательным переменным:

$$p_{\eta}(y) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p_{\bar{\eta}}(\bar{y}) dy_{m+1} \dots dy_n, \quad y \in \mathbb{R}^m. \quad (2.31)$$

Замечание 2.7. В выборе вспомогательных функций $f_{m+1}(x), \dots, f_n(x)$ имеется произвол. Им следует пользоваться в целях упрощения процесса решения задачи.

Теорема 2.7 (формула свертки). Пусть $\xi_1, \xi_2 \in \mathbb{R}^1$ – независимые случайные величины, имеющие плотности распределения вероятностей $p_{\xi_1}(x_1), p_{\xi_2}(x_2)$. Тогда плотность распределения вероятностей суммы $\eta = \xi_1 + \xi_2$ определяется по формуле свертки

$$p_{\eta}(y) = \int_{-\infty}^{+\infty} p_{\xi_1}(y - z) p_{\xi_2}(z) dz =: p_{\xi_1} * p_{\xi_2} = p_{\xi_2} * p_{\xi_1}.$$

Доказательство. Имеем вырожденное преобразование (т. е. случай II), в котором $n = 2, m = 1$. Построим расширенное функциональное преобразование: $y_1 = f_1(x_1, x_2) = x_1 + x_2, y_2 = f_2(x_1, x_2) = x_2$. Используя формулы (2.30), (2.31) и учитывая независимость, получаем доказываемое. \square

Следствие 2.10. Если в условии теоремы суммируются гауссовские случайные величины, т. е. $\mathcal{L}\{\xi_i\} = N(a_i, b_i), i = 1, 2$, то сумма – тоже гауссовская случайная величина: $\mathcal{L}\{\eta\} = N(a_1 + a_2, b_1 + b_2)$.

2.7. УПРАЖНЕНИЯ

1. Пусть случайная величина ξ имеет функцию распределения $F(x)$, $x \in \mathbb{R}$, $a, b \in \mathbb{R}$ – константы. Найти функцию распределения линейно преобразованной случайной величины $\eta = a + b\xi$.
2. Показать, что если $F(\cdot), G(\cdot)$ – некоторые функции распределения и $0 \leq \lambda \leq 1$, то смесь $H_1 = \lambda F + (1 - \lambda)G$ тоже является функцией распределения. Будут ли $H_2 = FG$, $H_3 = F^\lambda G^{1-\lambda}$ функциями распределения?
3. Пусть $F(\cdot)$ – некоторая функция распределения, а $n \in \mathbb{N}$ – натуральное число. Показать, что функциями распределения являются следующие функции:

$$\begin{aligned} H_1(x) &= F^n(x); \\ H_2(x) &= 1 - (1 - F(x))^n; \\ H_3(x) &= F(x) + (1 - F(x)) \log(1 - F(x)); \\ H_4(x) &= (F(x) - 1)e + e^{1-F(x)}. \end{aligned}$$

4. Выразить функции распределения каждой из следующих случайных величин:

$$-\xi, \quad \xi^+ = \max\{0, \xi\}, \quad \xi^- = -\min\{0, \xi\}, \quad |\xi| = \xi^+ + \xi^-,$$

через функцию распределения $F(x)$ случайной величины ξ .

5. Числовая функция $d : S \times S \rightarrow \mathbb{R}$ называется *метрикой* на S , если:

- а) $d(s, t) = d(t, s) \geq 0, \quad \forall s, t \in S$;
- б) $d(s, t) = 0 \iff s = t$;
- в) $d(s, t) \leq d(s, u) + d(u, t), \quad \forall s, t, u \in S$.

Если $F(\cdot), G(\cdot)$ – функции распределения, то *метрикой Леви* называется числовая функция

$$\begin{aligned} d_L(F, G) &= \inf\{\epsilon > 0 : G(x - \epsilon) - \epsilon \leq F(x) \leq \\ &\leq G(x + \epsilon) + \epsilon, \quad x \in \mathbb{R}\}. \end{aligned}$$

Показать, что для $d_L(\cdot)$ выполняются свойства а)–в) и, следовательно, $d_L(\cdot)$ является метрикой в пространстве $S = \{F\}$ всевозможных функций распределения.

6. (Продолжение предыдущего примера.) Пусть $X, Y \in \mathbb{Z}$ – целочисленные случайные величины:

$$d_{TV}(X, Y) = \sum_{k \in \mathbb{Z}} |\mathbf{P}\{X = k\} - \mathbf{P}\{Y = k\}|.$$

Показать, что $d_{TV}(\cdot)$ удовлетворяет свойствам а), в) метрики и что

$$d_{TV}(X, Y) = 0 \iff \mathbf{P}\{X = Y\} = 1.$$

Таким образом, $d_{TV}(\cdot)$ – метрика на пространстве классов эквивалентности случайных величин S с отношением эквивалентности $X \sim Y \iff \mathbf{P}\{X = Y\} = 1$.

Метрика $d_{TV}(\cdot)$ называется *расстоянием полной вариации*. Показать, что

$$d_{TV}(X, Y) = 2 \sup_{A \subseteq \mathbb{Z}} |\mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\}|.$$

7. Пусть $X, Y \in \mathbb{N}$ – независимые, одинаково распределенные случайные величины со значениями на множестве натуральных чисел с распределением вероятностей $\mathbf{P}\{X = x\} = 2^{-x}, \quad x \in \mathbb{N}$.

Вычислить:

- а) $P\{\min(X, Y) < z\}$, $z \in \mathbb{N}$;
- б) $P\{Y > X\}$;
- в) $P\{Y = x\}$;
- г) $P\{X > kY\}$, где $k \in \mathbb{N}$ – константа;
- д) $P\{X \text{ – делитель } Y\}$;
- е) $P\{X = rY\}$, где $r \in \mathbb{Q}$ – заданная рациональная константа.

8. Пусть ξ – случайная величина с геометрическим распределением вероятностей $\mathcal{L}\{\xi\} = G(p)$. Показать, что

$$P\{\xi = n + k | \xi > n\} = P\{\xi = k\}, \quad k, n \geq 1.$$

Почему это свойство называют свойством *отсутствия памяти*? Существуют ли другие дискретные распределения вероятностей с этим свойством?

9. Пусть ξ, η – независимые случайные величины с биномиальным распределением вероятностей: $\mathcal{L}\{\xi\} = Bi(M, p)$, $\mathcal{L}\{\eta\} = Bi(N, p)$. Показать, что $\mathcal{L}\{\xi + \eta\} = Bi(M + N, p)$.
10. Пусть $X \geq 0$ – некоторая положительная случайная величина с функцией распределения $F(x)$ и плотностью распределения вероятностей $f(x) = F'(x)$. Функция $H(x) = -\log(1 - F(x))$ называется *функцией риска (hazard function)*, а

$$r(x) = \lim_{\Delta \rightarrow 0+} \frac{P\{x \leq X < x + \Delta | X \geq x\}}{\Delta}, \quad x \geq 0,$$

называется *функцией интенсивности (отказа) (hazard rate)*. Дать содержательную интерпретацию этих функций для случая, когда X – длительность бессбойного функционирования некоторой системы, имеющая экспоненциальное распределение вероятностей $\mathcal{L}\{X\} = E\{\lambda\}$. Показать:

- а) $r(x) = H'(x) = f(x)/(1 - F(x))$;
 - б) если $r(\cdot)$ – монотонно возрастающая, то и $H(x)/x$ – монотонно возрастающая;
 - в) $H(x)/x$ – монотонно возрастающая $\Leftrightarrow (1 - F(x))^\alpha \leq 1 - F(\alpha x) \forall x \geq 0, \quad 0 \leq \alpha \leq 1$;
 - г) если $H(x)$ – монотонно возрастающая, то $H(x + y) \geq H(x) + H(y) \forall x, y \geq 0$.
11. Пусть ξ_1, ξ_2 – независимые, одинаково распределенные случайные величины с функцией распределения $F(x)$ и плотностью распределения $f(x) = F'(x)$. Показать, что $\xi_+ = \max(\xi_1, \xi_2)$ имеет функцию распределения $F_+(x) = F^2(x)$ и плотность распределения вероятностей $f_+(x) = 2F(x)f(x)$. Найти плотность распределения вероятностей для $\xi_- = \min(\xi_1, \xi_2)$.
12. Пусть $\mathcal{L}\{\xi\} = R[0, 1]$. Для какой функции $g(\cdot)$ случайная величина $\eta = g(\xi)$ имеет стандартное экспоненциальное распределение: $\mathcal{L}\{\eta\} = E(1)$?
13. Пусть $\mathcal{L}\{\xi\} = N(0, 1)$, $\eta = e^\xi$. Найти плотность распределения вероятностей для η . Этот закон распределения вероятностей называется *логнормальным*.
14. Показать, что стандартная нормальная плотность распределения вероятностей $\phi(x)$ удовлетворяет дифференциальному уравнению: $\phi'(x) + x\phi(x) = 0$. Отсюда показать, что для *отношения Миллса* $\psi(x) = (1 - \Phi(x))/\phi(x)$ справедливы неравенства:

$$x^{-1} - x^{-3} < \psi(x) < x^{-1} - x^{-3} + 3x^{-5}, \quad x > 0,$$

порождающие аппроксимации при $x \rightarrow \infty$:

$$\psi(x) = \frac{1}{x} - \frac{1}{x^3} + O\left(\frac{1}{x^5}\right), \quad \Phi(-x) = \Phi(x) \left(\frac{1}{x} - \frac{1}{x^3} + O\left(\frac{1}{x^5}\right) \right).$$

15. Пусть ξ_1, ξ_2 – независимые случайные величины, а случайная величина $\eta \in \{-1, +1\}$ – некоторая СВ Бернулли. Показать, что случайные величины $\zeta_1 = \eta\xi_1$, $\zeta_2 = \eta\xi_2$, вообще говоря, зависимы, а ζ_1^2, ζ_2^2 независимы.
16. Пусть ξ_1, ξ_2 – независимые, одинаково распределенные случайные величины, $\mathcal{L}\{\xi_1\} = E(1)$. Найти плотность распределения вероятностей для $\eta_1 = \xi_1 + \xi_2$ и для $\eta_2 = \xi_1/\eta_1$. Показать, что $\mathcal{L}\{\eta_2\} = R[0, 1]$.
17. Пусть ξ имеет распределение Коши. Показать, что $\eta = 1/\xi$ также подчиняется закону распределения Коши. Найти другое нетривиальное распределение с этим свойством инвариантности.
18. Пусть ξ_1, \dots, ξ_n – независимые в совокупности случайные величины со стандартным равномерным распределением: $\mathcal{L}\{\xi_i\} = R[0, 1]$, $i = 1, \dots, n$. Вычислить плотность распределения вероятностей суммы $\eta = \xi_1 + \dots + \xi_n$, $n = 2, 3, 4, 5, \dots$.
19. Пусть ξ_1, \dots, ξ_n – независимые, одинаково распределенные случайные величины, $\mathcal{L}\{\xi_1\} = R[0, 1]$. Показать, что плотность распределения вероятностей для случайной величины $\eta = \xi_1 + \dots + \xi_n$ в любой точке $x \in (0, n)$ является полиномом степени $n - 1$ относительно x .
20. Шесть координат (X_i, Y_i) , $i = 1, 2, 3$ трех точек A_1, A_2, A_3 на плоскости независимы и одинаково распределены по закону $N(0, 1)$. Показать, что вероятность того, что точка A_3 лежит внутри круга с диаметром A_1A_2 , равна $1/4$.
21. Пусть ξ_1, ξ_2 – независимые, одинаково распределенные случайные величины Бернулли, $\mathcal{L}\{\xi_1\} = Bi(1, 1/2)$. Доказать, что $\xi_1 + \xi_2$ и $|\xi_2 - \xi_1|$ зависимы, хотя и некоррелированы.
22. Показать, что если ξ имеет геометрическое распределение, т. е. $P(\xi = k) = Pq^k$, $k = 0, 1, 2, \dots$, то для любых k и n $P\{\xi - k = n | \xi \geq k\} = P\{\xi = n\}$.
23. Случайные величины ξ_1 и ξ_2 независимы и имеют геометрическое распределение. Доказать, что

$$P\{\xi_1 = k | \xi_1 + \xi_2 = n\} = \frac{1}{n+1}, \quad k = 0, 1, \dots, n.$$

24. Случайные величины ξ_1 и ξ_2 независимы и имеют распределение Пуассона соответственно с параметрами λ_1, λ_2 . Показать, что $P\{\xi_1 = k | \xi_1 + \xi_2 = n\} = C_n^k p^k q^{n-k}$, $k = 0, \dots, n$, где $p = \lambda_1/(\lambda_1 + \lambda_2) = 1 - q$.
25. Пусть v – случайная величина, распределенная по геометрическому закону. Найти производящие функции случайных величин

$$V_+ = \max(N, v), \quad V_- = \min(N, v),$$

где N – целое неотрицательное число.

26. Случайные величины ξ_1 и ξ_2 независимы и имеют экспоненциальное распределение с параметром λ , т. е. $P\{\xi_i < x\} = 1 - e^{-\lambda x}$ для $x > 0$. Показать, что:
- 1) случайная величина $\xi_1/(\xi_1 + \xi_2)$ равномерно распределена на $[0, 1]$;
 - 2) случайные величины $\xi_1/(\xi_1 + \xi_2)$, $\xi_1 + \xi_2$ независимы.
27. Случайные величины ξ_1 и ξ_2 независимы и каждая имеет нормальное распределение $N(0, 1)$. Показать, что случайная величина $\eta = 1/2(\xi_1^2 + \xi_2^2)$ имеет экспоненциальное распределение с параметром $\lambda = 1$, т. е. $P\{\eta < x\} = 1 - e^{-x}$.
28. Случайные величины ξ_1 и ξ_2 независимы и имеют нормальное распределение. Показать, что случайные величины $\xi + \eta$ и $\xi - \eta$ независимы.
29. Случайные величины ξ_1 и ξ_2 независимы и равномерно распределены на $[0, 1]$. Показать, что случайные величины $\eta_1 = \rho \cos \phi$ и $\eta_2 = \rho \sin \phi$, где $\rho = \sqrt{-2 \ln \xi_1}$, $\phi = 2\pi \xi_2$, независимы и имеют нормальное распределение $N(0, 1)$.
30. Случайные величины ξ_1 и ξ_2 независимы и имеют соответственно нормальные распределения $N(0, \sigma_1^2)$ и $N(0, \sigma_2^2)$. Показать, что случайная величина $\xi = \xi_1 \xi_2 / \sqrt{\xi_1^2 + \xi_2^2}$ имеет нормальное распределение $N(0, \sigma^2)$, где $1/\sigma^2 = 1/\sigma_1^2 + 1/\sigma_2^2$.
31. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и имеют показательное распределение с параметром λ . Доказать, что случайные величины $\max(\xi_1, \xi_2, \dots, \xi_n)$ и $\xi_1 + \xi_2/2 + \xi_3/3 + \dots + \xi_n/n$ одинаково распределены.
32. Случайные величины ξ и η независимы. Доказать, что если функция распределения случайной величины ξ непрерывна, то функция распределения $\xi + \eta$ тоже непрерывна.
33. (Контрпример Й. М. Стоянова.) Пусть ξ_1, ξ_2 – независимые, одинаково распределенные случайные величины со стандартным гауссовским распределением вероятностей $N_1(0, 1)$. Определим случайный вектор в \mathbb{R}^2 :

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ |\xi_2| \operatorname{sign}(\xi_1) \end{pmatrix}.$$

Показать, что одномерные маргинальные распределения x_1, x_2 есть $N_1(0, 1)$, но двумерное распределение X не гауссовское.

Глава 3

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

3.1. СХЕМА ПОСТРОЕНИЯ ИНТЕГРАЛА ЛЕБЕГА ПО ВЕРОЯТНОСТНОЙ МЕРЕ. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ПРОСТОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И ЕГО СВОЙСТВА

Пусть на произвольном вероятностном пространстве (Ω, \mathcal{F}, P) определена произвольная случайная величина $\xi = \xi(\omega), \omega \in \Omega, \xi \in \mathbb{R}$. Часто приходится усреднять случайные величины, математически это сводится к интегрированию по ω . Интеграл Римана в общем случае неприменим по следующим двум причинам:

- 1) Ω – пространство элементарных событий – это некоторое абстрактное множество, где может быть не задана метрика;
- 2) функция $\xi(\omega)$ может иметь счетное множество разрывов, и, по Риману, трудно строить интеграл от такой функции.

В 1902 г. французский математик А. Лебег предложил новую конструкцию интеграла:

$$E\{\xi\} = \int_{\Omega} \xi(\omega) P(d\omega) -$$

интеграл Лебега от случайной величины $\xi(\omega)$ по вероятностной мере P ; при этом значение $E\{\xi\} \in \mathbb{R}$ этого интеграла называется *математическим ожиданием (средним значением)* СВ ξ .

Построим интеграл Лебега в соответствии со следующей схемой: «от простого – к сложному» (рис. 3.1). Сначала построим интеграл от любой дискретной случайной величины с конечным множеством значений, затем с помощью предельного перехода построим интеграл от любой неотрицательной случайной величины. Рассматривая любую случайную величину как разность двух неотрицательных величин, построим интеграл Лебега от любой случайной величины.

Определение 3.1. Дискретная случайная величина $\xi \in \{a_1, \dots, a_N\}$ с конечным множеством значений ($N < \infty$) называется *простой случайной величиной*.

Определение 3.2. Индикатором случайного события $A \in \mathcal{F}$ называется *двоичная случайная величина*

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases} \quad (3.1)$$

Укажем свойства индикатора, легко доказываемые на основе формулы (3.1).

C1. Если обозначить $p = P(A)$, то \mathbb{I}_A – изученная ранее случайная величина Бернулли с распределением вероятностей $Bi(1, p)$.

C2. $\mathbb{I}_{\Omega}(\omega) \equiv 1$.

C3. Если события $A_1, \dots, A_N \in \mathcal{F}$ попарно несовместны, то $\mathbb{I}_{\bigcup_{i=1}^N A_i}(\omega) = \sum_{i=1}^N \mathbb{I}_{A_i}(\omega), \omega \in \Omega$.

C4. $\mathbb{I}_A \cdot \mathbb{I}_B = \mathbb{I}_{A \cap B}$.

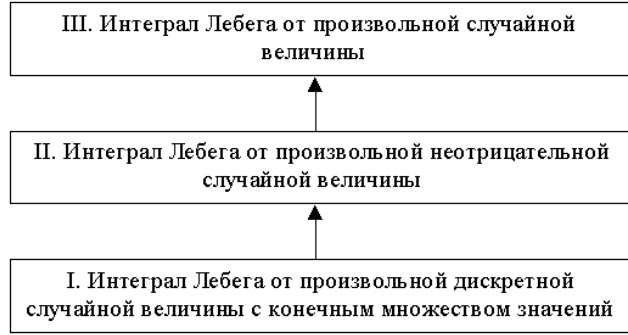


Рис. 3.1. Схема построения интеграла Лебега

Теорема 3.1. Случайная величина $\xi = \xi(\omega)$ является простой случайной величиной тогда и только тогда, когда она представима в виде конечной линейной комбинации $M < \infty$ индикаторов:

$$\xi(\omega) = \sum_{i=1}^M b_i \mathbb{I}_{B_i}(\omega), \quad (3.2)$$

где $B_1, \dots, B_M \in \mathcal{F}$ – попарно несовместные события; b_1, \dots, b_M – действительные числа, не обязательно различные.

Доказательство. *Необходимость.* Пусть $\xi \in \{a_1, \dots, a_N\}$ – простая случайная величина, $A_j = \{\xi = a_j\} \in \mathcal{F}$, $j = 1, \dots, N$. Тогда, сравнивая значения левой и правой частей, устанавливаем справедливость равенства:

$$\xi(\omega) = \sum_{i=1}^M a_i \mathbb{I}_{A_i}(\omega), \quad \omega \in \Omega. \quad (3.3)$$

Это представление и имеет вид (3.2).

Достаточность. Пусть случайная величина ξ имеет представление (3.2). В силу (3.1) это означает, что ξ может принимать лишь значения из конечного множества $\{0, b_1, \dots, b_M\}$. По определению 3.1 это означает, что ξ – простая случайная величина. \square

Определение 3.3. Математическим ожиданием простой случайной величины $\xi \in \{a_1, \dots, a_N\}$, $N < \infty$, имеющей дискретное распределение вероятностей $p_i = P\{\xi = a_i\}$, $i = 1, \dots, N$, называется число

$$E\{\xi\} = E\xi = \sum_{i=1}^N a_i p_i = \sum_{i=1}^N a_i P\{\xi = a_i\}. \quad (3.4)$$

Иначе говоря, интегралом Лебега по вероятностной мере $P(\cdot)$ от дискретной случайной величины, допускающей представление (3.3), называется величина (3.4).

Покажем, что данное определение корректно, т. е. не зависит от способа представления простой случайной величины в виде (3.2).

Теорема 3.2. Если наряду с (3.3) случайная величина ξ представима в виде (3.2), то

$$E\{\xi\} = \sum_{j=1}^M b_j P(B_j). \quad (3.5)$$

Доказательство. В силу «основного» представления (3.3) a_1, \dots, a_N – различные числа из набора чисел $\{b_1, \dots, b_M\}$, $M \geq N$. Поэтому имеем цепочку равенств случайной величины, начиная с (3.2):

$$\xi = \sum_{j=1}^M b_j \mathbb{I}_{B_j} = \sum_{i=1}^N a_i \sum_{j:b_j=a_i} \mathbb{I}_{B_j} = \sum_{i=1}^N a_i \mathbb{I}_{\bigcup_{j:b_j=a_i} B_j} = \sum_{i=1}^N a_i \mathbb{I}_{A_i}.$$

Здесь учтено, что события $\{A_1, \dots, A_N\}$ попарно несовместны, события $\{B_1, \dots, B_M\}$ попарно несовместны, а также учтены свойство СЗ индикаторов, соотношение (3.3) и равенство событий

$$A_i = \{\xi = a_i\} = \bigcup_{j:b_j=a_i} B_j, \quad i = 1, \dots, N. \quad (3.6)$$

Используя (3.4) и (3.6), по свойству вероятности имеем

$$\begin{aligned} E\{\xi\} &= \sum_{i=1}^N a_i P(A_i) = \sum_{i=1}^N a_i P\left(\bigcup_{j:b_j=a_i} B_j\right) = \\ &= \sum_{i=1}^N a_i \sum_{j:b_j=a_i} P(B_j) = \sum_{i=1}^N \sum_{j:b_j=a_i} a_i P(B_j) \equiv \\ &\equiv \sum_{i=1}^N \sum_{j:b_j=a_i} b_j P(B_j) \equiv \sum_{j=1}^M b_j P(B_j), \end{aligned}$$

что совпадает с соотношением (3.5). \square

Свойства математического ожидания простых случайных величин

С1. Математическое ожидание константы c есть сама эта константа: $E\{c\} = c$.

Доказательство. Константа c есть СВ, которая принимает одно значение $a_1=c$ ($N=1$) с вероятностью $p_1=1$. \square

С2. Если $\xi \geq 0$, то $E\{\xi\} \geq 0$.

Доказательство. Если $\xi \geq 0$, то $a_1, \dots, a_N \geq 0$. Поэтому сумма (3.4) этих значений также неотрицательна: $E\{\xi\} = \sum_{i=1}^N a_i p_i \geq 0$. \square

С3. Если $\xi = \mathbb{I}_A$, где $A \in \mathcal{F}$, то $E\{\xi\} = P(A)$.

Доказательство. Применим (3.4) при $N=2$, $a_1=0$, $a_2=1$. \square

С4. Постоянный множитель c можно выносить из-под знака математического ожидания: $E\{c\xi\} = cE\{\xi\}$.

Доказательство. В силу (3.3) $c\xi = \sum_{i=1}^N (ca_i) \mathbb{I}_{A_i}$. Отсюда по теореме 3.2 с учетом (3.4) имеем

$$E\{c\xi\} = \sum_{i=1}^N (ca_i) P(A_i) = c \sum_{i=1}^N a_i P(A_i) = cE\{\xi\}.$$

\square

С5. (Свойство аддитивности.) Для любых простых случайных величин ξ и η случайная величина $\zeta = \xi + \eta$ также является простой случайной величиной и

$$E\{\xi + \eta\} = E\{\xi\} + E\{\eta\}. \quad (3.7)$$

Доказательство. Пусть $\{a_1, \dots, a_N\}, \{b_1, \dots, b_M\}$ – множества значений соответственно случайных величин ξ и η ; $A_i = \{\xi = a_i\}, B_j = \{\eta = b_j\}$ – случайные события, которые (без потери общности) образуют две полные системы случайных событий $\{A_1, \dots, A_N\} \subset \mathcal{F}, \{B_1, \dots, B_N\} \subset \mathcal{F}$:

$$\bigcup_{i=1}^N A_i = \bigcup_{j=1}^M B_j = \Omega; \quad A_i \cap A_{i'} = \emptyset, \quad i \neq i'; \quad B_j \cap B_{j'} = \emptyset, \quad j \neq j'.$$

В силу теоремы 3.1 справедливы представления:

$$\xi = \sum_{i=1}^N a_i \mathbb{I}_{A_i}, \quad \eta = \sum_{j=1}^M b_j \mathbb{I}_{B_j}. \quad (3.8)$$

В силу свойств индикатора, попарной независимости и попарной несовместности $\{B_j\}$ имеем

$$\mathbb{I}_{A_i} \equiv \mathbb{I}_{A_i \cap \Omega} \equiv \mathbb{I}_{\bigcup_{j=1}^M (A_i \cap B_j)} = \sum_{j=1}^M \mathbb{I}_{A_i \cap B_j}. \quad (3.9)$$

Аналогично

$$\mathbb{I}_{B_j} = \sum_{i=1}^N \mathbb{I}_{A_i \cap B_j}. \quad (3.10)$$

Подставляя (3.9), (3.10) в (3.8), получаем

$$\xi + \eta = \sum_{i=1}^N a_i \sum_{j=1}^M \mathbb{I}_{A_i \cap B_j} + \sum_{j=1}^M b_j \sum_{i=1}^N \mathbb{I}_{A_i \cap B_j} = \sum_{i=1}^N \sum_{j=1}^M (a_i + b_j) \mathbb{I}_{A_i \cap B_j}.$$

Отсюда, согласно теореме 3.1, следует, что $\xi + \eta$ – простая случайная величина. Применяя теорему 3.2 к (3.11), приходим к (3.7):

$$\begin{aligned} E\{\xi + \eta\} &= \sum_{i,j} (a_i + b_j) P(A_i \cap B_j) = \sum_{i,j} a_i P(A_i \cap B_j) + \sum_{i,j} b_j P(A_i \cap B_j) = \\ &= \sum_i a_i \sum_j P(A_i \cap B_j) + \sum_j b_j \sum_i P(A_i \cap B_j) = \sum_i a_i P(A_i) + \sum_j b_j P(B_j) = \\ &= E\{\xi\} + E\{\eta\}. \end{aligned}$$

□

С6. Если $\xi \leq \eta$, то $E\{\xi\} \leq E\{\eta\}$.

Доказательство. Рассмотрим СВ $\zeta = \eta - \xi$. По условию $\zeta \geq 0$, поэтому в силу С2, С4, С5 имеем

$$0 \leq E\{\zeta\} = E\{\eta + (-\xi)\} = E\{\eta\} + E\{-\xi\} = E\{\eta\} - E\{\xi\}.$$

□

С7. $|E\{\xi\}| \leq E\{|\xi|\}$.

Доказательство. Согласно (3.3)

$$|\xi| = \sum_{i=1}^N |a_i| \mathbb{I}_{A_i},$$

поэтому в силу (3.4) и теоремы 3.2

$$|E\{\xi\}| = \left| \sum_{i=1}^N a_i P(A_i) \right| \leq \sum_{i=1}^N |a_i| P(A_i) = E\{|\xi|\}.$$

□

С8. Если простые случайные величины ξ , η независимы, то случайная величина $\xi \cdot \eta$ – также простая случайная величина, причем $E\{\xi\eta\} = E\{\xi\}E\{\eta\}$.

Доказательство. Как и при доказательстве С5, воспользуемся представлением (3.19) для простых случайных величин ξ , η . Тогда по свойству С4 индикатора имеем

$$\xi\eta = \sum_{i=1}^N \sum_{j=1}^M a_i b_j \mathbb{I}_{A_i} \mathbb{I}_{B_j} = \sum_{i,j} a_i b_j \mathbb{I}_{A_i \cap B_j}.$$

Отсюда по теореме 3.1 заключаем, что $\xi\eta$ – простая СВ. Тогда по теореме 3.2 и определениям независимости имеем

$$\begin{aligned} E\{\xi\eta\} &= \sum_{i,j} a_i b_j P(A_i \cap B_j) = \sum_{i,j} a_i b_j P\{\xi = a_i, \eta = b_j\} = \\ &= \sum_{i,j} a_i b_j P\{\xi = a_i\} P\{\eta = b_j\} = \\ &= \sum_{i=1}^N a_i P\{\xi = a_i\} \sum_{j=1}^M b_j P\{\eta = b_j\} = E\{\xi\}E\{\eta\}. \end{aligned}$$

□

В заключение выясним содержательный смысл математического ожидания (3.4). Рассмотрим случайный эксперимент (Ω, \mathcal{F}, P) , на множестве результатов которого определена простая случайная величина ξ с множеством N различных значений $\{a_1, \dots, a_N\}$ и распределением вероятностей $p_i = P\{\xi = a_i\}$, $i = 1, \dots, N$. Проводится серия n независимых экспериментов такого вида. Пусть $\omega_j \in \Omega$ – исход j -го эксперимента, а $x_j = \xi(\omega_j)$ – зарегистрированное в j -м эксперименте значение случайной величины ξ .

Вычислим среднее (арифметическое) значение случайной величины ξ в n экспериментах: $\bar{x} = \sum_{j=1}^n x_j / n$. Поскольку $x_j \in \{a_1, \dots, a_N\}$, то

$$\frac{\sum_{j=1}^n x_j}{n} = \frac{\sum_{i=1}^N a_i n_i}{n} = \sum_{i=1}^N a_i \frac{n_i}{n},$$

где n_i – количество значений, равных a_i , среди x_1, \dots, x_n $\left(\sum_{i=1}^N n_i = n \right)$.

В силу статистической устойчивости частот относительная частота наступления события $\{\xi = a_i\}$ сходится к его вероятности:

$$\frac{n_i}{n} \rightarrow P\{\xi = a_i\} = p_i.$$

Поэтому в силу формулы (3.4)

$$\frac{\sum_{j=1}^n x_j}{n} \rightarrow \sum_{i=1}^N a_i p_i = E\{\xi\}.$$

Таким образом, математическое ожидание $E\{\xi\}$ характеризует среднее значение случайной величины ξ в бесконечной серии независимых экспериментов. Поэтому иногда математическое ожидание называют средним значением. Кроме того, \bar{x} можно рассматривать как ожидаемое (прогнозируемое) значение ξ в $(n+1)$ -м эксперименте. Отсюда и происходит термин «математическое ожидание» (ожидаемое значение, вычисленное математически).

Замечание 3.1. Символ E в выражении $E\{\xi\}$ происходит от англ. «expectation», что значит «оператор усреднения».

Замечание 3.2. В некоторых учебных пособиях пользуются обозначением: $M\{\xi\} = E\{\xi\}$, где символ M происходит от англ. «mean» – «среднее».

3.2. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ (ИНТЕГРАЛ ЛЕБЕГА) ДЛЯ ПРОИЗВОЛЬНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Лемма 3.1. Для любой неотрицательной случайной величины $\xi = \xi(\omega)$, определенной на вероятностном пространстве (Ω, \mathcal{F}, P) , существует аппроксимирующая неубывающая последовательность $\xi_k = \xi_k(\omega)$ простых случайных величин на (Ω, \mathcal{F}, P) , сходящаяся к $\xi = \xi(\omega)$:

$$\xi_n(\omega) \uparrow \xi(\omega), \quad n \rightarrow \infty, \quad \omega \in \Omega. \quad (3.11)$$

Доказательство. (Конструктивное.) Укажем способ построения аппроксимирующей последовательности ξ_n . Для $n = 1, 2, \dots$ определим случайные события:

$$A_j^{(n)} = \left\{ \omega : \frac{j-1}{2^n} \leq \xi(\omega) < \frac{j}{2^n} \right\}, \quad j = 1, 2, \dots, n2^n;$$

$$B^{(n)} = \{ \omega : \xi(\omega) \geq n \}$$

и случайную величину

$$\xi_n = \xi_n(\omega) = \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \mathbb{I}_{A_j^{(n)}}(\omega) + n \mathbb{I}_{B^{(n)}}(\omega). \quad (3.12)$$

Из формулы (3.12) вытекают следующие свойства случайной величины $\{\xi_n\}$:

1) $\forall n \in \mathbb{N}$ ξ_n – простая случайная величина:

$$\xi_n \in \left\{ 0, \frac{1}{2^n}, \dots, \frac{n2^n - 1}{2^n}, n \right\};$$

2) $\forall \omega \in \Omega, n \in \mathbb{N}$ $\xi_n(\omega) \leq \xi_{n+1}(\omega) \leq \xi(\omega)$;

3) $\forall \omega \in \Omega$ такого, что $0 \leq \xi(\omega) \leq n$, ошибка аппроксимации $|\xi(\omega) - \xi_n(\omega)| \leq 2^{-n}$.

Из этих свойств следует сходимость (3.11). \square

Заметим, что согласно (3.12) и свойству С6 (п. 3.1) числовая последовательность $E\{\xi_n\}$ неубывающая, и, следовательно, при $n \rightarrow \infty$ имеет предел, конечный или бесконечный.

Определение 3.4. Математическим ожиданием, или интегралом Лебега, по вероятностной мере $P(\cdot)$ называется конечный или бесконечный предел:

$$E\{\xi\} = \int_{\Omega} \xi(\omega) P(d\omega) = \lim_{n \rightarrow \infty} E\{\xi_n\} = \lim_{n \rightarrow \infty} \int_{\Omega} \xi_n(\omega) P(d\omega), \quad (3.13)$$

где аппроксимирующая последовательность случайных величин ξ_n определяется (3.12).

Замечание 3.3. Если подставить (3.12) в (3.13), то получим

$$E\{\xi\} = \int_{\Omega} \xi(\omega) P(d\omega) = \lim_{n \rightarrow \infty} \left(\sum_{j=1}^{n2^n} \frac{j-1}{2^n} P\left\{ \frac{j-1}{2^n} \leq \xi < \frac{j}{2^n} \right\} + nP\{\xi \geq n\} \right).$$

Таким образом, интеграл Лебега, как и интеграл Римана, есть последовательность интегральных сумм. Однако принципы построения интегральных сумм совершенно различны: в интегральных суммах Римана «измельчается» область Ω определения функции $\xi = \xi(\omega)$, а в интегральных суммах Лебега «измельчается» область значений $[0, \infty)$ этой функции.

Покажем, что определение 3.4 корректно, т. е. предел (3.13) не зависит от способа построения аппроксимирующей последовательности ξ_n . Покажем, что если существует некоторая другая последовательность $\eta_n(\omega) \uparrow \xi(\omega)$, $\omega \in \Omega$, $n \rightarrow \infty$, то $\lim_{n \rightarrow \infty} E\{\eta_n\} = E\{\xi\}$.

Лемма 3.2. Пусть $\{\xi_n\}$ и η – простые неотрицательные случайные величины, причем

$$\xi_n \uparrow \xi \geq \eta, \quad (3.14)$$

тогда

$$\lim_{n \rightarrow \infty} E\{\xi_n\} \geq E\{\eta\}. \quad (3.15)$$

Доказательство. Для произвольного $\epsilon > 0$ построим вспомогательную последовательность случайных событий

$$A_n = \{\omega : \xi_n \geq \eta - \epsilon\} \in \mathcal{F}, \quad n = 1, 2, \dots$$

По построению согласно (3.14) имеем

$$A_n \uparrow \Omega, \quad \xi_n = \xi_n \mathbb{1}_{A_n} + \xi_n \mathbb{1}_{\overline{A_n}} \geq \xi_n \mathbb{1}_{A_n} \geq (\eta - \epsilon) \mathbb{1}_{A_n}.$$

Поэтому, используя свойства математического ожидания от простых случайных величин (п. 3.1), находим

$$\begin{aligned} E\{\xi_n\} &\geq E\{(\eta - \epsilon) \mathbb{1}_{A_n}\} = E\{\eta \mathbb{1}_{A_n}\} - \epsilon P(A_n) \equiv E\{\eta(1 - \mathbb{1}_{\overline{A_n}})\} - \\ &- \epsilon P(A_n) = E\{\eta\} - E\{\eta \mathbb{1}_{\overline{A_n}}\} - \epsilon P(A_n) \geq E\{\eta\} - c \cdot P(\overline{A_n}) - \epsilon \cdot 1, \end{aligned} \quad (3.16)$$

где $c = \max_{\omega \in \Omega} \eta(\omega) < \infty$.

Устремляя $n \rightarrow \infty$ и учитывая $P(A_n) \rightarrow 1$, $P(\overline{A_n}) \rightarrow 0$, получим

$$\lim_{n \rightarrow \infty} E\{\xi_n\} \geq E\{\eta\} - \epsilon.$$

Отсюда в силу произвольности $\epsilon > 0$ заключаем справедливость (3.15). \square

Теорема 3.3. Пусть определены две различные аппроксимирующие последовательности простых неотрицательных случайных величин $\xi_n = \xi_n(\omega)$, $n \in \mathbb{N}$, и $\eta_m = \eta_m(\omega)$, $m \in \mathbb{N}$, монотонно сходящиеся к $\xi = \xi(\omega)$ снизу:

$$\xi_n \uparrow \xi, \quad \eta_m \uparrow \xi.$$

Тогда совпадают пределы математических ожиданий для этих последовательностей:

$$\lim_{n \rightarrow \infty} E\{\xi_n\} = \lim_{m \rightarrow \infty} E\{\eta_m\}. \quad (3.17)$$

Доказательство. Поскольку $\forall m \in \mathbb{N} \xi \geq \eta_m$, то $\xi_n \uparrow \xi \geq \eta_m$, и в силу леммы 3.2 имеем

$$\lim_{n \rightarrow \infty} E\{\xi_n\} \geq E\{\eta_m\} \Rightarrow \lim_{n \rightarrow \infty} E\{\xi_n\} \geq \lim_{m \rightarrow \infty} E\{\eta_m\}.$$

Аналогично, используя, что $\forall n \in \mathbb{N} \eta_m \uparrow \xi \geq \xi_n$, имеем

$$\lim_{m \rightarrow \infty} E\{\eta_m\} \geq E\{\xi_n\} \Rightarrow \lim_{m \rightarrow \infty} E\{\eta_m\} \geq \lim_{n \rightarrow \infty} E\{\xi_n\}.$$

Из этих двух предельных соотношений следует (3.17). \square

Следствие 3.1. Определение математического ожидания (3.13) для неотрицательной случайной величины ξ корректно.

Замечание 3.4. Справедливо следующее эквивалентное представление математического ожидания неотрицательной случайной величины $\xi \geq 0$:

$$E\{\xi\} = \sup_{\{s \in S: s \leq \xi\}} E\{s\},$$

где $S = \{s\}$ – множество простых неотрицательных случайных величин.

Итак, для неотрицательных случайных величин математическое ожидание определено. Перейдем теперь к общему случаю произвольной случайной величины.

Определение 3.5. Положительной срезкой случайной величины $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$ называется неотрицательная случайная величина $\xi^+ = \xi^+(\omega) \geq 0$:

$$\xi^+(\omega) = \max(0, \xi(\omega)) = \begin{cases} \xi(\omega), & \xi(\omega) > 0, \\ 0, & \xi(\omega) \leq 0, \end{cases} \quad \omega \in \Omega. \quad (3.18)$$

Отрицательная срезка $\xi^- = \xi^-(\omega) \geq 0$ определяется аналогично:

$$\xi^-(\omega) = -\min(0, \xi(\omega)) = \max(0, -\xi(\omega)) = \begin{cases} -\xi(\omega), & \xi(\omega) < 0, \\ 0, & \xi(\omega) \geq 0, \end{cases} \quad \omega \in \Omega. \quad (3.19)$$

Отметим, что согласно (3.18) и (3.19) $\xi^- \equiv (-\xi)^+$.

Лемма 3.3. Любая случайная величина $\xi = \xi(\omega)$ представима в виде разности двух неотрицательных случайных величин (3.18) и (3.19):

$$\xi(\omega) = \xi^+(\omega) - \xi^-(\omega), \quad \omega \in \Omega, \quad (3.20)$$

при этом

$$|\xi(\omega)| = \xi^+(\omega) + \xi^-(\omega), \quad \omega \in \Omega. \quad (3.21)$$

Доказательство. Соотношения (3.20), (3.21) вытекают из (3.18), (3.19) очевидным образом. \square

Определение 3.6. Математическим ожиданием произвольной случайной величины ξ , допускающей, согласно лемме 3.3, представление в виде разности двух неотрицательных случайных величин (3.20), называется число

$$E\{\xi\} = E\{\xi^+\} - E\{\xi^-\}, \quad (3.22)$$

если $E\{\xi^+\} < +\infty, E\{\xi^-\} < +\infty$. Если же $E\{\xi^+\} = +\infty, E\{\xi^-\} < +\infty$, то полагают $E\{\xi\} = +\infty$; если $E\{\xi^+\} < +\infty, E\{\xi^-\} = +\infty$, то полагают $E\{\xi\} = -\infty$; если же $E\{\xi^+\} = +\infty, E\{\xi^-\} = +\infty$, то говорят, что математическое ожидание случайной величины ξ не существует.

Замечание 3.5. Говорят, что математическое ожидание $E\{\xi\}$ существует или определено, если

$$\min(E\{\xi^+\}, E\{\xi^-\}) < \infty.$$

Замечание 3.6. Говорят, что математическое ожидание $E\{\xi\}$ конечно, если

$$\max(E\{\xi^+\}, E\{\xi^-\}) < \infty : |E\{\xi\}| < \infty \Leftrightarrow E\{|\xi|\} < \infty.$$

Определение 3.7. Определенное согласно (3.22) математическое ожидание случайной величины ξ называется интегралом Лебега от функции $\xi = \xi(\omega)$ по вероятностной мере $P(\cdot)$ и обозначается

$$E\{\xi\} = \int_{\Omega} \xi(\omega) P(d\omega) = \int_{\Omega} \xi dP = \int \xi dP.$$

Определение 3.8. Символом $\int_A \xi dP$, где $A \in \mathcal{F}$, обозначается интеграл Лебега от случайной величины $\xi \mathbb{I}_A$:

$$\int_A \xi dP = \int_{\Omega} \xi \mathbb{I}_A dP.$$

В заключение отметим, что интеграл Лебега $\int_{\Omega} \xi dP$ аналогично строится и в случае, когда мера $P(\cdot)$ не обязательно нормирована, т. е. $P(\Omega) \neq 1, P(\Omega) < \infty$. В том случае, когда мера $P(\cdot)$ является σ -конечной, т. е. когда $P(\Omega) = +\infty$ и существует счетная система подмножеств $\Omega_1, \Omega_2, \dots \subset \Omega$, для которой

$$\bigcup_{i=1}^{\infty} \Omega_i = \Omega, \Omega_i \cap \Omega_j = \emptyset, i \neq j, P(\Omega_i) < \infty,$$

под интегралом Лебега понимается предел

$$\int_{\Omega} \xi dP = \lim_{n \rightarrow \infty} \int_{\bigcup_{i=1}^n \Omega_i} \xi dP = \lim_{n \rightarrow \infty} \sum_{i=1}^n \int_{\Omega_i} \xi dP.$$

3.3. СВОЙСТВА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ (ИНТЕГРАЛА ЛЕБЕГА)

Свойства, которые сейчас будут рассмотрены, относятся к произвольной случайной величине; они обобщают свойства математического ожидания простых случайных величин, рассмотренных ранее.

В нижеприведенных свойствах С1–С6 условимся обозначать: $\xi_n, \eta_n, n \in \mathbb{N}$ – простые неотрицательные СВ; ξ, η – случайные величины, для которых существуют математические ожидания.

С1. Если $\xi \geq 0$, то $E\{\xi\} \geq 0$.

Доказательство. Пусть $\xi_n \uparrow \xi$. Тогда $E\{\xi_n\} \geq 0$ и

$$E\{\xi\} = \lim_{n \rightarrow \infty} E\{\xi_n\} \geq 0.$$

□

Отметим, что условие этого свойства можно ослабить: $\xi \stackrel{\text{п. н.}}{\geq} 0$.

С2. Если $c = \text{const}$, то $E\{c\xi\} = cE\{\xi\}$.

Доказательство. Выделяются три ситуации.

а) Если $c \geq 0$, $\xi \geq 0$ и $\xi_n \uparrow \xi$, то $c\xi_n \uparrow c\xi$ и поэтому имеем

$$E\{c\xi\} = \lim_{n \rightarrow \infty} E\{c\xi_n\} = c \lim_{n \rightarrow \infty} E\{\xi_n\} = cE\{\xi\}.$$

б) Если $c \geq 0$, а ξ – произвольная случайная величина, то по доказанному в пункте а)

$$\begin{aligned} E\{c\xi\} &= E\{(c\xi)^+\} - E\{(c\xi)^-\} = E\{c\xi^+\} - E\{c\xi^-\} = \\ &= cE\{\xi^+\} - cE\{\xi^-\} = c(E\{\xi^+\} - E\{\xi^-\}) = cE\{\xi\}. \end{aligned}$$

в) Если $c < 0$, а ξ – произвольная случайная величина, то по свойствам срезов ξ^\pm и по доказанному в пункте а)

$$\begin{aligned} E\{c\xi\} &= E\{(c\xi)^+\} - E\{(c\xi)^-\} = E\{-c\xi^-\} - E\{-c\xi^+\} = \\ &= -cE\{\xi^-\} + cE\{\xi^+\} = c(E\{\xi^+\} - E\{\xi^-\}) = cE\{\xi\}. \end{aligned}$$

□

С3. (Аддитивность.) Если $\xi \geq 0, \eta \geq 0$, то $E\{\xi + \eta\} = E\{\xi\} + E\{\eta\}$, полагая при этом, что $\infty + \text{const} = \infty, \infty + \infty = \infty$.

Доказательство. Поскольку $\xi \geq 0, \eta \geq 0$ и существуют аппроксимирующие последовательности неотрицательных простых случайных величин $\xi_n \uparrow \xi, \eta_n \uparrow \eta$, то $\xi_n + \eta_n \uparrow \xi + \eta$. Отсюда по определению математического ожидания и свойству аддитивности из п. 3.1 имеем $E\{\xi + \eta\} = \lim_{n \rightarrow \infty} E\{\xi_n + \eta_n\} = \lim_{n \rightarrow \infty} (E\{\xi_n\} + E\{\eta_n\}) = E\{\xi\} + E\{\eta\}$. □

С4. (Аддитивность.) Если ξ, η имеют конечные математические ожидания, то

$$E\{\xi + \eta\} = E\{\xi\} + E\{\eta\}. \quad (3.23)$$

Доказательство. Определим вспомогательные неотрицательные случайные величины:

$$\zeta = \xi^+ + \eta^+ - (\xi + \eta)^+ \geq 0, \quad \zeta' = \xi^- + \eta^- - (\xi + \eta)^- \geq 0. \quad (3.24)$$

Используя (3.24), находим

$$\begin{aligned}\xi + \eta &= (\xi + \eta)^+ - (\xi + \eta)^- = \xi^+ + \eta^+ - \zeta - (\xi^- + \eta^- - \zeta') = \\ &= (\xi^+ - \xi^-) + (\eta^+ - \eta^-) + \zeta' - \zeta = \xi + \eta + \zeta' - \zeta.\end{aligned}$$

Отсюда заключаем: $\zeta' = \zeta \geq 0$, поэтому (3.24) примет вид

$$(\xi + \eta)^+ + \zeta = \xi^+ + \eta^+, \quad (\xi + \eta)^- + \zeta = \xi^- + \eta^-. \quad (3.25)$$

Применив свойство С3 к (3.25), получим

$$E\{(\xi + \eta)^+\} + E\{\zeta\} = E\{\xi^+\} + E\{\eta^+\},$$

$$E\{(\xi + \eta)^-\} + E\{\zeta\} = E\{\xi^-\} + E\{\eta^-\}.$$

Почленно вычтем второе равенство из первого:

$$E\{(\xi + \eta)^+\} - E\{(\xi + \eta)^-\} = (E\{\xi^+\} - E\{\xi^-\}) + (E\{\eta^+\} - E\{\eta^-\}),$$

что с учетом определения 3.6 из п. 3.2 эквивалентно (3.23). \square

С5. Если $\xi \leq \eta$, то $E\{\xi\} \leq E\{\eta\}$, причем если $E\{\xi\} = +\infty$, то $E\{\eta\} = +\infty$, а если $E\{\eta\} = -\infty$, то $E\{\xi\} = -\infty$.

Доказательство. Выделим три ситуации.

а) Если ξ, η имеют конечные математические ожидания, то для вспомогательной случайной величины $\zeta = \eta - \xi \geq 0$ имеем в силу свойств С1, С2, С4

$$0 \leq E\{\zeta\} = E\{\eta - \xi\} = E\{\eta + ((-1)\xi)\} = E\{\eta\} - E\{\xi\}.$$

б) Пусть $\xi \geq 0$ и $E\{\xi\} = +\infty$, т. е. если $\xi_n \uparrow \xi$, то $\lim_{n \rightarrow \infty} E\{\xi_n\} = +\infty$. Пусть $\eta_m \uparrow \eta \geq 0$. Тогда по лемме 3.2 из п. 3.2 $\forall n \in \mathbb{N} \quad \lim_{m \rightarrow \infty} E\{\eta_m\} \geq E\{\xi_n\}$. Переходя к пределу при $n \rightarrow \infty$, получаем $E\{\eta\} = \lim_{m \rightarrow \infty} E\{\eta_m\} = +\infty$.

в) Пусть ξ — произвольная случайная величина такая, что $E\{\xi\} = +\infty$, $E\{\xi^+\} = +\infty$, $E\{\xi^-\} < +\infty$. Из неравенства $\xi \leq \eta$ и свойств срезок заключаем (непосредственной проверкой $\forall \omega \in \Omega$): $\xi^+ \leq \eta^+$, $\xi^- \geq \eta^-$. Отсюда по доказанному в пункте б) получаем $E\{\eta^+\} \geq E\{\xi^+\} = +\infty$, $E\{\eta^+\} = +\infty$, а $E\{\eta^-\} \leq E\{\xi^-\} < +\infty$.

Случай $E\{\eta\} = -\infty$ исследуется аналогично пунктам б), в). \square

С6. (Аддитивность.) Справедливы два утверждения:

а) если $E\{\xi\} = +\infty$, $E\{\eta\} > -\infty$ либо $E\{|\xi|\} < +\infty$, $E\{\eta\} = +\infty$, то $E\{\xi + \eta\} = +\infty$;

б) если $E\{\xi\} = -\infty$, $E\{\eta\} < +\infty$ либо $E\{|\xi|\} < +\infty$, $E\{\eta\} = -\infty$, то $E\{\xi + \eta\} = -\infty$.

Доказательство. Все возможные случаи исследуются аналогично. Для иллюстрации исследуем случай, когда

$$E\{\xi^+\} = +\infty, \quad E\{\xi^-\} < +\infty, \quad E\{\eta^+\} < +\infty.$$

Поскольку $\xi^+ + \eta^+ - \xi^- - \eta^- = (\xi + \eta)^+ - (\xi + \eta)^- \leq (\xi + \eta)^+$, то

$$\xi^+ + \eta^+ \leq (\xi + \eta)^+ + \xi^- + \eta^-.$$

В силу свойства С5 $E\{(\xi + \eta)^+ + \xi^- + \eta^-\} = +\infty$, откуда имеем $E\{(\xi + \eta)^+\} = +\infty$.

Поскольку $(\xi + \eta)^- \leq \xi^- + \eta^-$, то в силу С5 $E\{(\xi + \eta)^-\} \leq E\{\xi^-\} + E\{\eta^-\} < +\infty$.

Поэтому

$$E\{\xi + \eta\} = E\{(\xi + \eta)^+\} - E\{(\xi + \eta)^-\} = +\infty.$$

\square

С7. Следующее свойство сформулируем в виде теоремы.

Теорема 3.4 (о монотонной сходимости). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определены случайные величины $\eta, \xi, \xi_1, \xi_2, \dots$. Тогда справедливы следующие два утверждения:

- а) если $\xi_n \geq \eta$ для всех $n \geq 1$, $E\{\eta\} > -\infty$ и $\xi_n \uparrow \xi$, то $E\{\xi\}$ существует и $E\{\xi_n\} \uparrow E\{\xi\}$;
 б) если $\xi_n \leq \eta$ для всех $n \geq 1$, $E\{\eta\} < +\infty$ и $\xi_n \downarrow \xi$, то $E\{\xi\}$ существует и $E\{\xi_n\} \downarrow E\{\xi\}$.

Доказательство. Докажем сначала первое утверждение. Выделим две ситуации. Пусть $\eta \geq 0$. Построим вспомогательные простые неотрицательные случайные величины $\{\xi_k^{(n)}\}$, обладающие свойством $\xi_k^{(n)} \uparrow \xi_n$ при $k \rightarrow \infty \forall n \in \mathbb{N}$; построим еще простые случайные величины $\eta_k = \max_{1 \leq n \leq k} \xi_k^{(n)}$. Из монотонной сходимости $\xi_k^{(n)} \uparrow$ при $k \rightarrow \infty$ вытекает, что η_k также монотонно сходится: $\eta_k \uparrow$, причем по построению

$$\xi_k^{(n)} \leq \eta_k \leq \xi_k. \quad (3.26)$$

Обозначим $\eta = \lim_{k \rightarrow \infty} \eta_k$. Тогда из предыдущего неравенства имеем

$$\lim_{k \rightarrow \infty} \xi_k^{(n)} \leq \lim_{k \rightarrow \infty} \eta_k \leq \lim_{k \rightarrow \infty} \xi_k,$$

или, что эквивалентно: $\xi_n \leq \eta \leq \xi$, $n = 1, 2, \dots$. При $n \rightarrow \infty$ получаем $\xi \leq \eta \leq \xi$, откуда следует $\eta = \xi$. Таким образом, $\eta_k \uparrow \xi$. Отсюда по определению математического ожидания простых случайных величин получаем

$$\lim_{k \rightarrow \infty} E\{\eta_k\} = E\{\xi\}. \quad (3.27)$$

А так как $\eta_k \leq \xi_k$ в силу (3.26), то согласно (3.27)

$$E\{\xi\} = \lim_{k \rightarrow \infty} E\{\eta_k\} \leq \lim_{k \rightarrow \infty} E\{\xi_k\}. \quad (3.28)$$

С другой стороны, по условию $\xi_k \uparrow \xi$, значит, $E\{\xi_k\} \uparrow$ и $E\{\xi_k\} \leq E\{\xi\} \forall k$. Отсюда

$$\lim_{k \rightarrow \infty} E\{\xi_k\} \leq E\{\xi\}. \quad (3.29)$$

Из сравнения формул (3.28) и (3.29) заключаем:

$$E\{\xi\} = \lim_{k \rightarrow \infty} E\{\xi_k\}.$$

Пусть $\eta \in \mathbb{R}$ – произвольная случайная величина. По условию $E\{\xi\} > -\infty$. Если $E\{\eta\} = +\infty$, то в силу С5 $E\{\xi\} = +\infty$. Если $E\{\eta\} < +\infty$, то имеем неотрицательную монотонно сходящуюся последовательность: $0 \leq \xi_n - \eta \uparrow \xi - \eta$, и по доказанному в предыдущем случае $E\{\xi_n - \eta\} \uparrow E\{\xi - \eta\}$. Это означает, что

$$E\{\xi\} - E\{\eta\} = E\{\xi - \eta\} = \lim_{n \rightarrow \infty} E\{\xi_n - \eta\} = \lim_{n \rightarrow \infty} E\{\xi_n\} - E\{\eta\}.$$

Отсюда следует, что $E\{\xi\} = \lim_{n \rightarrow \infty} E\{\xi_n\}$.

Второе утверждение вытекает из первого, если вместо ξ, η, ξ_n рассмотреть случайные величины $-\xi, -\eta, -\xi_n, n = 1, 2, \dots$. \square

С8. Перед тем как сформулировать это свойство в виде теоремы, докажем вспомогательную лемму.

Лемма 3.4 (лемма Фату). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определены случайные величины η , ξ , ξ_1 , ξ_2 , \dots . Тогда справедливы следующие три утверждения:

а) если $\xi_n \geq \eta$ для всех $n \geq 1$ и $E\{\eta\} > -\infty$, то

$$E\{\lim \xi_n\} \leq \lim E\{\xi_n\};$$

б) если $\xi_n \leq \eta$ для всех $n \geq 1$ и $E\{\eta\} < +\infty$, то

$$\overline{\lim} E\{\xi_n\} \leq E\{\overline{\lim} \xi_n\};$$

в) если $|\xi_n| \geq \eta$ для всех $n \geq 1$ и $E\{\eta\} < +\infty$, то

$$E\{\lim \xi_n\} \leq \lim E\{\xi_n\} \leq \overline{\lim} E\{\xi_n\} \leq E\{\overline{\lim} \xi_n\}. \quad (3.30)$$

Доказательство. Докажем вначале первое утверждение.

Пусть $\zeta_n = \inf_{m \geq n} \xi_m$, тогда

$$\lim \xi_n = \lim_{n \rightarrow \infty} \inf_{m \geq n} \xi_m = \lim_{n \rightarrow \infty} \zeta_n.$$

Это означает, что

$$\zeta_n \uparrow \lim \xi_n \text{ и } \zeta_n \geq \eta \text{ для всех } n \geq 1.$$

Тогда из теоремы о монотонной сходимости имеем (свойство С6)

$$E\{\lim \xi_n\} = E\{\lim \zeta_n\} = \lim E\{\zeta_n\} = \lim E\{\xi_n\} \leq \lim E\{\xi_n\},$$

что и доказывает первое утверждение. Второе утверждение следует из первого (аналогично С6). Третье утверждение – следствие первых двух утверждений. \square

Теорема 3.5 (Лебега о мажорируемой сходимости). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определены случайные величины η , ξ , ξ_1 , ξ_2 , \dots такие, что $|\xi_n| \leq \eta$, $E\{\eta\} < +\infty$ и $\xi_n \rightarrow \xi$. Тогда при $n \rightarrow \infty$

$$E\{|\xi|\} < +\infty; \quad (3.31)$$

$$E\{\xi_n\} \rightarrow E\{\xi\}; \quad (3.32)$$

$$E\{|\xi_n - \xi|\} \rightarrow 0. \quad (3.33)$$

Доказательство. В силу леммы Фату (3-е утверждение) справедлива цепочка соотношений (3.30). По условию данной теоремы $\exists \lim \xi_n = \overline{\lim} \xi_n = \lim \xi_n = \xi$. Поэтому неравенства в (3.30) обращаются в равенства:

$$E\{\lim \xi_n\} = \lim E\{\xi_n\} = \overline{\lim} E\{\xi_n\} = E\{\overline{\lim} \xi_n\} = E\{\xi\},$$

что доказывает (3.32). Поскольку $|\xi_n| \leq \eta$, то $|\xi| \leq \eta$. Поэтому в силу свойства С5 $E\{|\xi|\} \leq E\{\eta\} < +\infty$, что означает (3.31).

Для доказательства утверждения (3.33) определим вспомогательную последовательность случайных величин (уклонений):

$$\zeta_n = |\xi_n - \xi| \geq 0, \quad \zeta_n \rightarrow 0.$$

Поскольку $\zeta_n \leq |\xi_n| + |\xi| \leq 2\eta$, $E\{\eta\} < +\infty$, то по доказанному выше к последовательности ζ_n применимо утверждение вида (3.32): $E\{\zeta_n\} \rightarrow 0$, что эквивалентно (3.33). \square

Следствие 3.2. Пусть η , ξ , ξ_1 , ξ_2 , ... – случайные величины такие, что $|\xi_n| \leq \eta$, $\xi_n \rightarrow \xi$ и $E\{\eta^p\} < \infty$ для некоторого $p > 0$. Тогда

$$E\{|\xi|^p\} < \infty, \quad E\{|\xi_n - \xi|^p\} \xrightarrow{n \rightarrow \infty} 0.$$

Доказательство. Достаточно выбрать $\zeta_n = |\xi_n - \xi|^p \rightarrow 0$ и аналогично доказательству (3.33) учесть, что $\zeta_n \leq 2\eta^p$. \square

Замечание 3.7. Условие $|\xi_n| \leq \eta$, $E\{\eta\} < +\infty$, входящее в лемму Фату и теорему Лебега, можно ослабить, заменяя его условием равномерной интегрируемости [25, 36]

$$\sup_n E\{|\xi_n| \mathbb{1}_{\{|\xi_n| > c\}}\} \rightarrow 0, \quad c \rightarrow \infty.$$

- С9. Пусть ξ , η – независимые случайные величины с конечными математическими ожиданиями: $E\{|\xi|\} < \infty$, $E\{|\eta|\} < \infty$. Тогда $E\{|\xi\eta|\} < \infty$ и математическое ожидание произведения равно произведению математических ожиданий:

$$E\{\xi\eta\} = E\{\xi\}E\{\eta\}. \quad (3.34)$$

Доказательство. Выделим две ситуации. Пусть $\xi \geq 0$, $\eta \geq 0$ и $\xi_n \uparrow \xi$, $\eta_n \uparrow \eta$ – аппроксимирующие последовательности неотрицательных случайных величин, определенные в п. 3.2. По построению (3.14) при каждом $n \in \mathbb{N}$ $\xi_n = g_n(\xi)$, $\eta_n = g_n(\eta)$, где $g_n(\cdot)$ – борелевская функция. Поэтому из свойств независимых случайных величин следует, что ξ_n , η_n – независимые случайные величины. Поскольку по построению $\xi_n \eta_n \uparrow \xi\eta$, то по определению математического ожидания и свойству С5 из п. 3.1 имеем

$$E\{\xi\eta\} = \lim_{n \rightarrow \infty} E\{\xi_n \eta_n\} = \lim_{n \rightarrow \infty} (E\{\xi_n\}E\{\eta_n\}) = E\{\xi\}E\{\eta\}.$$

Пусть ξ , η – произвольные случайные величины с конечными математическими ожиданиями. Тогда по определению математического ожидания

$$\xi\eta = (\xi^+ - \xi^-)(\eta^+ - \eta^-) = \xi^+\eta^+ + \xi^-\eta^- - \xi^+\eta^- - \xi^-\eta^+. \quad (3.35)$$

В формуле (3.35) все слагаемые – неотрицательные случайные величины с конечными математическими ожиданиями, поэтому к каждому слагаемому применим результат, доказанный на предыдущем шаге. Применив свойства С3, С4 к (3.35), получим

$$\begin{aligned} E\{\xi\eta\} &= E\{\xi^+\eta^+\} + E\{\xi^-\eta^-\} - E\{\xi^+\eta^-\} - E\{\xi^-\eta^+\} = \\ &= E\{\xi^+\}E\{\eta^+\} + E\{\xi^-\}E\{\eta^-\} - E\{\xi^+\}E\{\eta^-\} - E\{\xi^-\}E\{\eta^+\} = \\ &= E\{\xi\}E\{\eta\}, \end{aligned}$$

что совпадает с формулой (3.34). \square

- С10. (Математическое ожидание функции от случайной величины.) Пусть $\xi \in \{x_1, x_2, \dots\}$ – дискретная случайная величина, а $y = g(x)$ – борелевская функция. Тогда

$$E\{g(\xi)\} = \sum_{i=1}^{\infty} g(x_i)P\{\xi = x_i\}, \quad (3.36)$$

если ряд в соотношении (3.36) абсолютно сходится. Если:

- а) $\sum_{i=1}^{\infty} g^+(x_i)P\{\xi = x_i\} = \infty$, а $\sum_{i=1}^{\infty} g^-(x_i)P\{\xi = x_i\} < \infty$, то $E\{g(\xi)\} = \infty$;
 б) $\sum_{i=1}^{\infty} g^+(x_i)P\{\xi = x_i\} < \infty$, а $\sum_{i=1}^{\infty} g^-(x_i)P\{\xi = x_i\} = \infty$, то $E\{g(\xi)\} = -\infty$, где
 $g^{\pm}(x) = \max(0, \pm g(x))$.

Доказательство. Воспользуемся эквивалентным представлением дискретной случайной величины $g(\xi)$ (см. п. 3.1):

$$g(\xi) = \sum_{i=1}^{\infty} g(x_i)\mathbb{1}_{\{\xi=x_i\}} = \sum_{i=1}^{\infty} g^+(x_i)\mathbb{1}_{\{\xi=x_i\}} - \sum_{i=1}^{\infty} g^-(x_i)\mathbb{1}_{\{\xi=x_i\}}.$$

Тогда, используя определение и свойства математического ожидания, получаем (3.36). \square

3.4. НЕРАВЕНСТВА ДЛЯ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ

Рассмотрим еще ряд важнейших свойств математических ожиданий (интегралов Лебега), имеющих вид неравенств.

С1. Если для случайной величины ξ существует математическое ожидание $E\{\xi\}$, то

$$|E\{\xi\}| \leq E\{|\xi|\}. \quad (3.37)$$

Доказательство. Поскольку $-|\xi| \leq \xi \leq |\xi|$, то по свойствам С2, С5 из п. 3.3 имеем $-E\{|\xi|\} \leq E\{\xi\} \leq E\{|\xi|\}$, что эквивалентно (3.37). \square

С2. (Неравенство Чебышева.) Если $y = g(x) \geq 0$ – неотрицательная неубывающая функция на $[0, \infty)$, то для любого $\epsilon > 0$ и любой случайной величины ξ , определенной на (Ω, \mathcal{F}, P) , справедливо неравенство

$$P\{|\xi| \geq \epsilon\} \leq \frac{E\{g(|\xi|)\}}{g(\epsilon)}. \quad (3.38)$$

Доказательство. Непосредственная проверка $\forall \omega \in \Omega$ показывает справедливость неравенства

$$g(|\xi|) \geq g(\epsilon)E\{\mathbb{1}_{\{|\xi| \geq \epsilon\}}\} = g(\epsilon)P\{|\xi| \geq \epsilon\},$$

откуда следует (3.38). \square

Следствие 3.3 (неравенство Чебышева). Если $\xi \geq 0$ – произвольная неотрицательная СВ, то для любого $\epsilon > 0$

$$P\{\xi \geq \epsilon\} \leq \frac{E\{\xi\}}{\epsilon}. \quad (3.39)$$

Доказательство. Достаточно положить $g(x) \equiv x$ в соотношении (3.38) и учесть, что $\xi \geq 0$. \square

Следствие 3.4. Для любой случайной величины ξ справедливы неравенства:

$$P\{|\xi| \geq \epsilon\} \leq \frac{E\{|\xi|\}}{\epsilon}, \quad P\{|\xi| \geq \epsilon\} \leq \frac{E\{\xi^2\}}{\epsilon^2}.$$

Доказательство. Достаточно выбрать соответственно $g(x) \equiv x$, $g(x) \equiv x^2$. \square

Определение 3.9. Принято говорить, что некоторое свойство выполнено на вероятностном пространстве (Ω, \mathcal{F}, P) « P – почти наверное», если существует множество $\Omega_0 \in \mathcal{F}$ нулевой меры $P(\Omega_0) = 0$ такое, что это свойство выполнено для каждой точки $\omega \in \Omega \setminus \Omega_0$. Вместо слов « P – почти наверное» часто говорят « P – почти всюду» или просто «почти наверное» (п. н.), «почти всюду» (п. в.).

Следствие 3.5. Если $\xi \geq 0$ и $E\{\xi\} = 0$, то $\xi \stackrel{p. н.}{=} 0$.

Доказательство. В силу неравенства Чебышева (3.39) имеем

$$\forall \epsilon > 0 \quad P\{\xi \geq \epsilon\} = 0.$$

Поэтому $\forall \epsilon_n \downarrow 0 \quad P\{\xi > 0\} \leq \lim_{n \rightarrow \infty} P\{\xi \geq \epsilon_n\} = 0$. Следовательно, $P\{\xi = 0\} = P\{\xi \geq 0\} - P\{\xi > 0\} = 1 - 0 = 1$. □

Следствие 3.6. Если $E\{\xi^2\} = 0$, то $\xi \stackrel{p. н.}{=} 0$.

С3. (Неравенство Йенсена.) Для любой выпуклой функции $y = g(x)$ и любой случайной величины ξ с конечным математическим ожиданием $E\{\xi\}$ справедливо неравенство

$$E\{g(\xi)\} \geq g(E\{\xi\}). \quad (3.40)$$

Доказательство. Воспользуемся следующим свойством выпуклой функции: для любой точки $x_0 \in \mathbb{R}$ существует действительное число $G_0 = G'(x_0)$ такое, что

$$g(x) \geq g(x_0) + G_0 \cdot (x - x_0), \quad x \in \mathbb{R}. \quad (3.41)$$

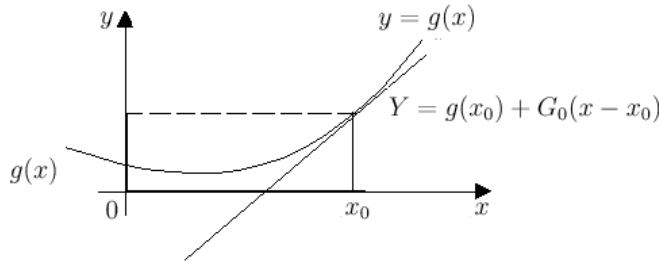


Рис. 3.2. Иллюстрация свойства выпуклости

Неравенство (3.41) графически проиллюстрировано на рис. 3.2. Положим в (3.41) $x_0 = E\{\xi\}$, $x = \xi(\omega)$, применим усреднение к левой и правой частям согласно свойствам (3.24), (3.26) и (3.27) из п. 3.3:

$$\begin{aligned} E\{g(\xi)\} &\geq E\{g(E\{\xi\}) + G_0 \cdot (\xi - E\{\xi\})\} = g(E\{\xi\}) + \\ &+ G_0 \cdot E\{\xi - E\{\xi\}\} = g(E\{\xi\}). \end{aligned}$$

Здесь учтено, что $E\{\xi - E\{\xi\}\} = E\{\xi\} - E\{\xi\} = 0$. □

Следствие 3.7. Для $\forall r \geq 1$ справедливо неравенство

$$|E\{\xi\}|^r \leq E\{|\xi|^r\}.$$

Доказательство. Достаточно выбрать $g(x) = |x|^r$ в неравенстве (3.40). □

Следствие 3.8. *Справедливы неравенства*

$$|E\xi| \leq E\{|\xi|\}, \quad (E\xi)^2 \leq E\{\xi^2\}.$$

Доказательство. Достаточно воспользоваться следствием 3.7 при $r = 1$, $r = 2$. \square

Следствие 3.9. *Если $y = g(x)$ – вогнутая функция, то в неравенстве (3.40) знак меняется на обратный.*

Доказательство. Достаточно рассмотреть выпуклую функцию $\tilde{g}(x) = -g(x)$. \square

С4. (Неравенство Ляпунова.) Для любых положительных чисел $0 < s < t < +\infty$ справедливо неравенство

$$(E\{|\xi|^s\})^{1/s} \leq (E\{|\xi|^t\})^{1/t}. \quad (3.42)$$

Доказательство. Воспользуемся следствием 3.7. Для $\forall \eta \geq 0$, $r \geq 1$ имеем

$$(E\eta)^r \leq E\{\eta^r\}. \quad (3.43)$$

Положим в неравенстве (3.43) $r = t/s > 1$, $\eta = |\xi|^s \geq 0$, тогда из (3.43) имеем $E\{|\xi|^s\}^{t/s} \leq E\{|\xi|^t\}$. Возведем левую и правую части неравенства в положительную степень $1/t > 0$ и получим (3.42). \square

Следствие 3.10. *Справедлива следующая цепочка неравенств Ляпунова:*

$$E\{|\xi|\} \leq \sqrt[2]{E\{\xi^2\}} \leq \sqrt[3]{E\{|\xi|^3\}} \leq \dots$$

С5. (Неравенство Гельдера.) Для любых случайных величин ξ_1 , ξ_2 и положительных констант $p_1, p_2 > 1 : 1/p_1 + 1/p_2 = 1$ справедливо неравенство

$$|E\{\xi_1 \cdot \xi_2\}| \leq E\{|\xi_1 \cdot \xi_2|\} \leq (E\{|\xi_1|^{p_1}\})^{1/p_1} \cdot (E\{|\xi_2|^{p_2}\})^{1/p_2}. \quad (3.44)$$

Доказательство. Из свойства С1 имеем $|E\{\xi_1 \cdot \xi_2\}| \leq E\{|\xi_1 \cdot \xi_2|\}$. Для $\forall x_1, x_2 \geq 0$, $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$ справедливо известное неравенство:

$$x_1^{\lambda_1} \cdot x_2^{\lambda_2} \leq \lambda_1 x_1 + \lambda_2 x_2. \quad (3.45)$$

Выберем в формуле (3.45)

$$\lambda_i = \frac{1}{p_i}, \quad x_i = \frac{|\xi_i|^{p_i}}{E\{|\xi_i|^{p_i}\}} \geq 0, \quad E\{x_i\} = 1, \quad i = 1, 2.$$

Применяя операцию усреднения к левой и правой частям (3.45), получим

$$\begin{aligned} E\left\{\frac{|\xi_1 \xi_2|}{(E\{|\xi_1|^{p_1}\})^{1/p_1} \cdot (E\{|\xi_2|^{p_2}\})^{1/p_2}}\right\} &\leq E\left\{\frac{1}{p_1} x_1 + \frac{1}{p_2} x_2\right\} = \\ &= \frac{1}{p_1} \cdot 1 + \frac{1}{p_2} \cdot 1 = 1. \end{aligned}$$

Домножая на знаменатель, получаем (3.44). \square

Следствие 3.11. *Справедливо неравенство Коши – Буняковского (Шварца):*

$$|E\{\xi_1 \xi_2\}| \leq E\{|\xi_1 \xi_2|\} \leq \sqrt{E\{\xi_1^2\} E\{\xi_2^2\}}.$$

Доказательство. Достаточно положить $p_1 = p_2 = 2$ в неравенстве Гельдера. \square

С6. (Неравенство Минковского.) Если для некоторого $p \geq 1$, $\mathbf{E}\{|\xi|^p\} < \infty$, $\mathbf{E}\{|\eta|^p\} < \infty$, то $\mathbf{E}\{|\xi + \eta|^p\} < \infty$ и

$$(\mathbf{E}\{|\xi + \eta|^p\})^{\frac{1}{p}} \leq (\mathbf{E}\{|\xi|^p\})^{\frac{1}{p}} + (\mathbf{E}\{|\eta|^p\})^{\frac{1}{p}}. \quad (3.46)$$

Доказательство. Воспользуемся известным неравенством:

$$|\xi + \eta|^p \leq 2^{p-1}(|\xi|^p + |\eta|^p).$$

Отсюда в силу свойств математического ожидания имеем

$$\mathbf{E}\{|\xi + \eta|^p\} \leq 2^{p-1}(\mathbf{E}\{|\xi|^p\} + \mathbf{E}\{|\eta|^p\}) < +\infty.$$

Используя неравенство Гельдера (3.44) для $\xi_1 = \xi$, $\xi_2 = \eta$, $p_1 = p$, $p_2 = p/(p-1)$, получаем

$$\begin{aligned} \mathbf{E}\{|\xi + \eta|^p\} &= \mathbf{E}\{|\xi + \eta| |\xi + \eta|^{p-1}\} \leq \mathbf{E}\{|\xi| |\xi + \eta|^{p-1}\} + \\ &+ \mathbf{E}\{|\eta| |\xi + \eta|^{p-1}\} \leq (\mathbf{E}\{|\xi|^p\})^{\frac{1}{p}} (\mathbf{E}\{|\xi + \eta|^{\frac{p}{p-1}(p-1)}\})^{\frac{p-1}{p}} + \\ &+ (\mathbf{E}\{|\eta|^p\})^{\frac{1}{p}} (\mathbf{E}\{|\xi + \eta|^p\})^{\frac{p-1}{p}} = (\mathbf{E}\{|\xi + \eta|^p\})^{\frac{p-1}{p}} \left((\mathbf{E}\{|\xi|^p\})^{\frac{1}{p}} + (\mathbf{E}\{|\eta|^p\})^{\frac{1}{p}} \right). \end{aligned}$$

Деля левую и правую части неравенства на $(\mathbf{E}\{|\xi + \eta|^p\})^{\frac{p-1}{p}}$, получаем (3.46). \square

3.5. ВЫЧИСЛЕНИЕ ИНТЕГРАЛА ЛЕБЕГА. ИНТЕГРАЛЫ ЛЕБЕГА – СТИЛТЬЕСА И РИМАНА – СТИЛТЬЕСА

Пусть $\xi = \xi(\omega)$ – произвольная случайная величина, определенная на некотором вероятностном пространстве (Ω, \mathcal{F}, P) . Рассмотрим другое вероятностное пространство $(\mathbb{R}, \mathcal{B}, P_\xi)$, порожденное случайной величиной ξ . Случайные величины, определенные на этом пространстве, являются борелевскими функциями $g = g(x)$ действительной переменной $x \in \mathbb{R}$. Интеграл Лебега от функции $g(\cdot)$ по мере $P_\xi(\cdot)$ обозначается

$$\mathcal{J} = \int g(x) P_\xi(dx).$$

Известно, что мера $P_\xi = P_\xi(B)$ однозначно выражается через функцию распределения $F_\xi(x)$ и обладает характерным свойством:

$$P_\xi([x, x + \Delta x)) = F_\xi(x + \Delta x) - F_\xi(x) = \Delta F_\xi(x), \quad \Delta x \geq 0. \quad (3.47)$$

Определение 3.10. Поскольку вероятностная мера $P_\xi(\cdot)$ обладает свойством (3.47), то она называется мерой Лебега – Стилтъяса, а соответствующий интеграл \mathcal{J} – интегралом Лебега – Стилтъяса от функции $g(\cdot)$ по функции $F_\xi(\cdot)$:

$$\mathcal{J} = \int g(x) dF_\xi(x).$$

Определение 3.11. Если мера $P_\xi(B) = \text{mes}(B)$ является мерой Лебега (в частности, если $B = [a, b]$ – отрезок, то $\text{mes}(B) = b - a$ – его длина), то интеграл \mathcal{J} записывается в виде

$$\mathcal{J} = \int g(x) dx$$

и называется интегралом Лебега от функции $g(\cdot)$.

Пусть $F(x)$ – произвольная функция распределения.

Определение 3.12. Интегралом Римана – Стильеса от функции $g(\cdot)$ по функции $F(\cdot)$ на отрезке $[a, b]$ называется число

$$\int_a^b g(x) dF(x) = \lim_{\substack{\Delta x \rightarrow 0 \\ N \rightarrow \infty}} \sum_{i=0}^{N-1} g(\tilde{x}_i) (F(x_{i+1}) - F(x_i)), \quad (3.48)$$

если предел последовательности интегральных сумм существует и не зависит от выбора точек $\{x_i\}$, $\{\tilde{x}_i\}$, где $a = x_0 < x_1 < \dots < x_N = b$, $\tilde{x}_i \in [x_i, x_{i+1}]$; $\Delta x = \max_i (x_{i+1} - x_i)$ – диаметр разбиения $[a, b]$.

Определение 3.13. (Несобственным) интегралом Римана – Стильеса $\int_{-\infty}^{+\infty} g(x) dF(x)$

называется число

$$\int_{-\infty}^{+\infty} g(x) dF(x) = \int_{\mathbb{R}} g(x) dF(x) = \lim_{\substack{a \rightarrow -\infty, \\ b \rightarrow +\infty}} \int_a^b g(x) dF(x).$$

Замечание 3.8. Интеграл $\int_{-\infty}^{+\infty} g(x) dF(x)$ существует, если функция $g = g(x)$ непрерывна или кусочно-непрерывна.

Замечание 3.9. Интеграл Римана – Стильеса обладает всеми свойствами, которыми обладает интеграл Римана. Эти свойства доказываются, как и для интеграла Римана. Докажем три специальных свойства.

Теорема 3.6. Если $F(x) = p_1 F_1(x) + p_2 F_2(x)$ – линейная смесь двух функций распределения $F_1(\cdot)$, $F_2(\cdot)$ ($p_1, p_2 \geq 0$, $p_1 + p_2 = 1$), то

$$\int_a^b g(x) dF(x) = p_1 \int_a^b g(x) dF_1(x) + p_2 \int_a^b g(x) dF_2(x).$$

Доказательство. Этот результат вытекает из (3.48) и свойства приращений:

$$F(x_{i+1}) - F(x_i) = p_1 (F_1(x_{i+1}) - F_1(x_i)) + p_2 (F_2(x_{i+1}) - F_2(x_i)).$$

□

Теорема 3.7 (связь с интегралом Римана). Если функция распределения $F(\cdot)$ дифференцируема и существует непрерывная плотность распределения вероятностей $p(x) = F'(x)$, то

$$\int_a^b g(x) dF(x) = \int_a^b g(x) p(x) dx - \text{интеграл Римана}.$$

Доказательство. По теореме о среднем

$$F(x_{i+1}) - F(x_i) = \int_{x_i}^{x_{i+1}} p(x) dx = p(\hat{x}_i) (x_{i+1} - x_i),$$

где $\hat{x}_i \in [x_i, x_{i+1})$ – промежуточная точка. Подставляя это в (3.48), получим

$$\int_a^b g(x) dF(x) = \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{N-1} g(\tilde{x}_i) p(\hat{x}_i) (x_{i+1} - x_i).$$

Полагая $\hat{x}_i = \tilde{x}_i$, получаем интегральную сумму Римана. \square

Теорема 3.8. Если интегрирующая функция $F(x)$ является ступенчатой со скачками в точках $\{y_k\}$, причем величина скачка равна p_k в точке y_k ($k = 1, 2, \dots$), и сходится ряд $\sum_{k=1}^{\infty} |g(y_k)| p_k < \infty$, то

$$\int_{-\infty}^{+\infty} g(x) dF(x) = \sum_{k=1}^{\infty} g(y_k) p_k.$$

Доказательство. Разбиение $\{x_i\}$ в (3.48) осуществим так, чтобы при каждом i промежуток $[x_i, x_{i+1})$ содержал не более одной точки скачка из множества $\{y_k\}$. Тогда

$$F(x_{i+1}) - F(x_i) = \begin{cases} p_k, & \text{если } y_k \in [x_i, x_{i+1}) \text{ для некоторого } i, \\ 0, & \text{в противном случае.} \end{cases}$$

Выберем в (3.48) $\tilde{x}_i = y_k$, если $y_k \in [x_i, x_{i+1})$. Тогда (3.48) примет вид

$$\int_a^b g(x) dF(x) = \sum_{k: y_k \in [a, b]} g(y_k) p_k.$$

Отсюда

$$\int_{-\infty}^{+\infty} g(x) dF(x) = \lim_{\substack{a \rightarrow -\infty, \\ b \rightarrow +\infty}} \int_a^b g(x) dF(x) = \sum_{k=1}^{\infty} g(y_k) p_k,$$

так как по условию ряд справа абсолютно сходится. \square

Теорема 3.9 (формула замены переменных). Пусть $\xi = \xi(\omega)$ – случайная величина на (Ω, \mathcal{F}, P) и $g = g(x)$ – борелевская функция такая, что $E\{|g(\xi)|\} < \infty$. Тогда

$$\int_{\Omega} g(\xi(\omega)) P(d\omega) = \int_{\mathbb{R}} g(x) dF_{\xi}(x). \quad (3.49)$$

Доказательство. Выделим четыре ситуации.

1. Пусть $g(x) = \mathbb{1}_B(x)$, $B \in \mathcal{B}$. Поскольку $\mathbb{1}_B(\xi(\omega)) \equiv \mathbb{1}_{\{\xi \in B\}}(\omega)$, то

$$\begin{aligned} \int_{\Omega} g(\xi(\omega)) P(d\omega) &= \int_{\Omega} \mathbb{1}_B(\xi(\omega)) P(d\omega) = E\{\mathbb{1}_{\{\xi \in B\}}(\omega)\} = \\ &= P\{\xi \in B\} = \int_{\mathbb{R}} \mathbb{1}_B(x) P_{\xi}(dx) = \int_{\mathbb{R}} g(x) P_{\xi}(dx) = \int_{\mathbb{R}} g(x) dF_{\xi}(x), \end{aligned}$$

что совпадает с (3.49).

2. Пусть $g = g(x)$ – простая функция вида

$$g(x) = \sum_{k=1}^n y_k \mathbb{1}_{B_k}(x), \quad x \in \mathbb{R}, \quad n < \infty, \quad B_k = \{x : g(x) = y_k\}.$$

Тогда из свойств интеграла Лебега (см. п. 3.3) и доказанного на предыдущем шаге имеем

$$\begin{aligned} \int_{\Omega} g(\xi(\omega)) P(d\omega) &= \int_{\Omega} \sum_{k=1}^n y_k \mathbb{1}_{B_k}(\xi(\omega)) P(d\omega) = \\ &= \sum_{k=1}^n y_k \int_{\Omega} \mathbb{1}_{B_k}(\xi(\omega)) P(d\omega) = \sum_{k=1}^n y_k \int_{\mathbb{R}} \mathbb{1}_{B_k}(x) dF_{\xi}(x) = \\ &= \int_{\mathbb{R}} \sum_{k=1}^n y_k \mathbb{1}_{B_k}(x) dF_{\xi}(x) = \int_{\mathbb{R}} g(x) dF_{\xi}(x). \end{aligned}$$

3. Пусть $g = g(x) \geq 0$ и $\{g_n\}$ – аппроксимирующая последовательность простых функций таких, что $g_n \uparrow g$. Тогда $g_n(\xi) \uparrow g(\xi)$. Используя предыдущий шаг и определение интеграла Лебега, получаем

$$\begin{aligned} \int g(\xi(\omega)) P(d\omega) &= \lim_{n \rightarrow \infty} \int g_n(\xi(\omega)) P(d\omega) = \\ &= \lim_{n \rightarrow \infty} \int g_n(x) dF_{\xi}(x) = \int g(x) dF_{\xi}(x). \end{aligned}$$

4. Пусть $g = g(x)$ – произвольная борелевская функция, $\mathbf{E}\{|g(\xi)|\} < +\infty$. По определению интеграла Лебега получаем

$$\begin{aligned} \int g(\xi(\omega)) P(d\omega) &= \int g^+(\xi(\omega)) P(d\omega) - \int g^-(\xi(\omega)) P(d\omega) = \\ &= \int g^+(x) dF_{\xi}(x) - \int g^-(x) dF_{\xi}(x) = \int g(x) dF_{\xi}(x), \end{aligned}$$

что совпадает с (3.49). □

Замечание 3.10. Результат доказанной теоремы и схема доказательства остаются в силе, если $\xi = (\xi_i) \in \mathbb{R}^N$ – случайный вектор, а $g = g(x_1, \dots, x_N)$ – действительная функция $N \geq 1$ переменных такая, что $\mathbf{E}\{|g(\xi)|\} < +\infty$.

Чтобы избежать путаницы, через (Л. – С.) $\int g(x) dF_{\xi}(x)$ условимся обозначать интеграл Лебега – Стильеса, а интеграл Римана – Стильеса через (Р. – С.) $\int g(x) dF_{\xi}(x)$. Установим связь между этими интегралами.

Теорема 3.10. Если функция $g = g(x)$ – непрерывна и существует один из интегралов (Л. – С.) $\int g(x) dF_{\xi}(x)$, (Р. – С.) $\int g(x) dF_{\xi}(x)$, то существует и другой, и они равны:

$$(\text{Л. – С.}) \int g(x) dF_{\xi}(x) = (\text{Р. – С.}) \int g(x) dF_{\xi}(x). \quad (3.50)$$

Доказательство. Пусть $-\infty < a < b < +\infty$. Поскольку функция $g = g(x)$ непрерывна на \mathbb{R} , то она будет равномерно непрерывна и ограничена на $[a, b]$. Следовательно, для любого $n = 1, 2, \dots$ найдется набор точек $a = x_0 < x_1 < \dots < x_{N_n} = b$ такой, что $|g(x) - g(x_i)| \leq 1/n$, если $x \in [x_{i-1}, x_i]$, $i = 1, 2, \dots, N_n$. Пусть $c = \text{const}$ такая, что $|g(x)| \leq c$, $x \in [a, b]$. Определим последовательность кусочно-постоянных функций:

$$g_n(x) = g(x_i), \text{ если } x \in (x_{i-1}, x_i), i = 1, 2, \dots, N_n. \quad (3.51)$$

Из (3.51) видно, что $\{g_n(\cdot)\}$ – простые функции, причем

$$|g_n(x) - g(x)| \leq \frac{1}{n}, x \in [a, b]. \quad (3.52)$$

Поскольку согласно (3.52) $|g_n(x)| \leq |g(x)| + 1 \leq c + 1$, то по теореме Лебега о сходимости мажорируемой последовательности получаем

$$\begin{aligned} (\text{Л. - С.}) \int_a^b g(x) dF_\xi(x) &= \int_a^b g(x) P_\xi(dx) = \lim_{n \rightarrow \infty} \int_a^b g_n(x) P_\xi(dx) = \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^{N_n} g_n(x_i) P\{\xi \in [x_{i-1}, x_i]\} = \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^{N_n} g(x_i) (F_\xi(x_i) - F_\xi(x_{i-1})) = \\ &= (\text{Р. - С.}) \int_a^b g(x) dF_\xi(x). \end{aligned}$$

Теперь перенесем этот результат на несобственные интегралы:

$$\begin{aligned} (\text{Л. - С.}) \int_{-\infty}^{+\infty} g(x) dF_\xi(x) &= \lim_{\substack{a \rightarrow -\infty, \\ b \rightarrow +\infty}} (\text{Л. - С.}) \int_a^b g(x) dF_\xi(x) = \\ &= \lim_{\substack{a \rightarrow -\infty, \\ b \rightarrow +\infty}} (\text{Р. - С.}) \int_a^b g(x) dF_\xi(x) = (\text{Р. - С.}) \int_{-\infty}^{+\infty} g(x) dF_\xi(x), \end{aligned}$$

что совпадает с соотношением (3.50). \square

Следствие 3.12. Если ξ – СВ с функцией распределения $F_\xi(x)$, а $g = g(x)$ – непрерывная функция такая, что $E\{|g(\xi)|\} < \infty$, то

$$E\{g(\xi)\} = (\text{Р. - С.}) \int g(x) dF_\xi(x). \quad (3.53)$$

Доказательство. Формула (3.53) вытекает из определения математического ожидания как интеграла Лебега, формулы (3.50) и теоремы 3.10. \square

Следствие 3.13. Если $\xi \in \{x_1, x_2, \dots, x_N\}$, $N \leq \infty$, – ДСВ с дискретным распределением вероятностей $P\{\xi = x_i\} = p_i$, $\sum_{i=1}^N p_i = 1$, то математическое ожидание вычисляется как сумма ряда:

$$E\{g(\xi)\} = \sum_{i=1}^N g(x_i) p_i = \sum_{i=1}^N g(x_i) P\{\xi = x_i\}. \quad (3.54)$$

Следствие 3.14. Если $g = g(x)$ – непрерывная функция, СВ ξ имеет плотность распределения $p_\xi(x) = F'_\xi(x)$ и $\int |g(x)|p_\xi(x)dx < \infty$, то математическое ожидание вычисляется как интеграл Римана:

$$E\{g(\xi)\} = \int g(x)p_\xi(x) dx. \quad (3.55)$$

Замечание 3.11. Теорема 3.10 и формулы (3.53)–(3.55) обобщаются на многомерный случай, когда $\xi = (\xi_i) \in \mathbb{R}^N$ – случайный вектор, а $g = g(x_1, \dots, x_N)$ – действительная функция $N \geq 1$ переменных.

3.6. УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ И ЕГО СВОЙСТВА

На практике часто возникает необходимость усреднения случайной величины $\xi = \xi(\omega)$ при условии наступления события B или равенства некоторой другой случайной величины заданному числу $\eta = y$. Для решения этой задачи используются понятия условного математического ожидания: относительно случайного события, относительно случайной величины и относительно σ -алгебры.

Определение 3.14. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены некоторая случайная величина $\xi = \xi(\omega)$ и некоторое случайное событие $B \in \mathcal{F}$, $P(B) > 0$. Пусть далее $F_\xi(x)$, $x \in \mathbb{R}$, – функция распределения случайной величины ξ , $P_B(A) = P(A|B)$, $A \in \mathcal{F}$, – условная вероятностная мера, а $F_\xi(x|B)$ – условная функция распределения. Тогда условным математическим ожиданием (условным средним) случайной величины ξ при условии события B называется величина интеграла Лебега (Лебега – Стильтеса):

$$E\{\xi|B\} = \int_{\Omega} \xi(\omega) P_B(d\omega) = \int_{-\infty}^{+\infty} x dF_\xi(x|B), \quad (3.56)$$

при этом ранее введенное математическое ожидание

$$E\{\xi\} = \int_{-\infty}^{+\infty} x dF_\xi(x) \quad (3.57)$$

называется безусловным математическим ожиданием.

Из (3.56) и (3.57) видно, что условное математическое ожидание отличается от безусловного лишь интегрирующей функцией. Поэтому формулы вычисления условного математического ожидания аналогичны формулам вычисления безусловного математического ожидания, полученным ранее:

$$E\{\xi|B\} = \sum_{i=1}^N a_i p_i = \sum_{i=1}^N a_i P\{\xi = a_i|B\},$$

если ξ – дискретная случайная величина;

$$E\{\xi|B\} = \int_{-\infty}^{\infty} x p_\xi(x|B) dx,$$

если ξ – абсолютно непрерывная случайная величина с условной плотностью распределения $p_\xi(x|B) = F'_\xi(x|B)$.

Теорема 3.11. Пусть на (Ω, \mathcal{F}, P) определены случайная величина $\xi = \xi(\omega)$ с математическим ожиданием $E\{\xi\}$ и некоторая полная система случайных событий $\{H_i : i = 1, \dots, N\}$, $N \leq +\infty$. Тогда условные и безусловное математические ожидания связаны соотношением

$$E\{\xi\} = \sum_{i=1}^N P(H_i) E\{\xi|H_i\}. \quad (3.58)$$

Доказательство. Воспользуемся формулой полной вероятности: $F_\xi(x) = P\{\xi < x\} = \sum_{i=1}^N P(H_i) \cdot P\{\xi < x|H_i\} = \sum_{i=1}^N P(H_i) F_\xi(x|H_i)$. Подставляя это выражение в соотношение (3.57), приходим к (3.58). \square

Определение 3.15. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены некоторая случайная величина $\xi = \xi(\omega)$ и некоторый случайный вектор $\eta = \eta(\omega) \in \mathbb{R}^n$. Пусть далее $F_{\xi|\eta}(x|y)$, $x \in \mathbb{R}^1$, $y \in \mathbb{R}^n$, есть условная функция распределения случайной величины ξ при условии $\eta = y$. Тогда условным математическим ожиданием ξ при условии $\eta = y$ называется величина интеграла Лебега – Стильтеса:

$$E\{\xi|\eta = y\} = \int_{-\infty}^{+\infty} x dF_{\xi|\eta}(x|y). \quad (3.59)$$

Определение 3.16. Пусть с помощью определения 3.15 построена борелевская функция $z = g(y) : \mathbb{R}^n \rightarrow \mathbb{R}^1$, заданная следующим образом:

$$g(y) = E\{\xi|\eta = y\}, \quad y \in \mathbb{R}^n. \quad (3.60)$$

Тогда условным математическим ожиданием случайной величины ξ относительно η называется случайная величина $\alpha = \alpha(\omega)$, $\omega \in \Omega$, являющаяся суперпозицией функций

$$E\{\xi|\eta\} :: \alpha(\omega) = g(\eta(\omega)), \quad \omega \in \Omega.$$

В отличие от предыдущих определений условного математического ожидания, здесь $E\{\xi|\eta\}$ является случайной величиной.

Свойства условного математического ожидания

С1. Условные математические ожидания (3.56), (3.59) обладают всеми свойствами безусловного математического ожидания.

Доказательство. Условное математическое ожидание представляет собой такой же функционал, как и безусловное математическое ожидание (3.57), отличается лишь видом интегрирующей функции. \square

С2. Если ξ и η независимы, то условное математическое ожидание совпадает с безусловным:

$$E\{\xi|\eta = y\} = E\{\xi\}, \quad y \in \mathbb{R}^n.$$

Доказательство. По свойствам независимости случайных величин: $F_{\xi|\eta}(x|y) = F_\xi(x)$, поэтому из (3.59) и (3.57) следует доказываемое. \square

С3. (Формула полного математического ожидания.) Справедливо соотношение

$$E\{\xi\} = E\{E\{\xi|\eta\}\}. \quad (3.61)$$

Доказательство. Рассмотрим частный случай: $\eta \in Y = \{y_1, \dots, y_N\}$, $N \leq +\infty$, – дискретная случайная величина. По определению 3.16 и с учетом свойств математического ожидания из (3.60) имеем

$$E\{E\{\xi|\eta\}\} = E\{g(\eta)\} = \sum_{i=1}^N g(y_i) \cdot P\{\eta = y_i\}.$$

Построим полную систему случайных событий:

$$H_i = \{\eta = y_i\}, \quad i = 1, \dots, N;$$

используя (3.60) и теорему 3.11, получаем

$$\sum_{i=1}^N g(y_i) \cdot P\{\eta = y_i\} = \sum_{i=1}^N E\{\xi|\eta = y_i\} \cdot P(H_i) = E\{\xi\}.$$

□

Замечание 3.12. Формула (3.61) представляет собой двухэтапное усреднение:

$$E\{\xi\} = E\{E\{\xi|\eta\}\},$$

т. е. сначала осуществляется усреднение ξ при фиксированном значении η , а затем усреднение по всевозможным значениям η .

С4. (Обобщение свойства СЗ.) Для любой ограниченной борелевской функции $h(\cdot)$ справедливо равенство

$$E\{h(\eta)E\{\xi|\eta\}\} = E\{h(\eta)\xi\}. \quad (3.62)$$

Доказательство.

1) Рассмотрим вначале случай дискретного распределения вероятностей η . Воспользуемся схемой доказательства свойства СЗ:

$$\begin{aligned} E\{h(\eta)E\{\xi|\eta\}\} &= \sum_{i=1}^N P\{\eta = y_i\} E\{h(\eta)E\{\xi|\eta\}|\eta = y_i\} = \\ &= \sum_{i=1}^N P\{\eta = y_i\} h(y_i) E\{E\{\xi|\eta\}|\eta = y_i\} = \sum_{i=1}^N P\{\eta = y_i\} h(y_i) E\{\xi|\eta = y_i\}. \end{aligned}$$

С другой стороны,

$$\begin{aligned} E\{h(\eta)\xi\} &= \sum_{i=1}^N P\{\eta = y_i\} E\{h(\eta)\xi|\eta = y_i\} = \\ &= \sum_{i=1}^N P\{\eta = y_i\} h(y_i) E\{\xi|\eta = y_i\}. \end{aligned}$$

Из совпадения этих выражений получаем (3.62).

2) Рассмотрим теперь случай абсолютно непрерывного распределения вероятностей (ξ, η) с совместной плотностью $p_{\xi,\eta}(x, y)$ и условной плотностью $p_{\xi|\eta}(x|y)$. Аналогично построению условной функции распределения в п. 2.4 получаем

$$E\{\xi|\eta = y\} = \int x p_{\xi|\eta}(x|y) dx.$$

Поэтому имеем цепочку равенств, доказывающую (3.62):

$$\begin{aligned} E\{h(\eta)\xi\} &= \int \int h(y)xp_{\xi|\eta}(x, y)dxdy = \\ &= \int \int h(y)xp_{\xi|\eta}(x|y)p_{\eta}(y)dxdy = \int E\{\xi|\eta = y\}h(y)p_{\eta}(y)dy = \\ &= E\{h(\eta)E\{\xi|\eta\}\}. \end{aligned}$$

□

Замечание 3.13. Часто [14, 36] соотношение (3.62) принимается в качестве определения условного математического ожидания $E\{\xi|\eta\}$, являющегося \mathcal{F}_{η} -измеримой случайной величиной.

- С5. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены случайная величина $\xi = \xi(\omega)$ и случайный вектор $\eta = \eta(\omega) \in \mathbb{R}^N$. Пусть далее $z = \psi(y) : \mathbb{R}^N \rightarrow \mathbb{R}^1$ — некоторая борелевская функция. Тогда справедливо

$$E\{\psi(\eta) \cdot \xi|\eta\} \stackrel{\text{п.н.}}{=} \psi(\eta) \cdot E\{\xi|\eta\}.$$

Доказательство. Сравнение значений левой и правой частей этого соотношения при всевозможных фиксированных $\eta = y$ и даст требуемый результат. □

- С6. (Соотношение Вальда.) Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены:
- а) последовательность одинаково распределенных случайных величин ξ_1, ξ_2, \dots с конечным математическим ожиданием $E\{\xi_1\}$;
 - б) не зависящая от всех $\{\xi_k\}$ неотрицательная целочисленная случайная величина $\eta = \eta(\omega) \geq 0$ с конечным математическим ожиданием $E\{\eta\}$;
 - в) сумма случайного числа случайных величин $S_{\eta} ::= \xi_1 + \xi_2 + \dots + \xi_{\eta}$ (η — случайный номер, $S_0 ::= 0$).

Тогда справедливо следующее соотношение Вальда:

$$E\{S_{\eta}\} = E\{\eta\} \cdot E\{\xi_1\}.$$

Доказательство. Воспользуемся формулой полного математического ожидания:

$$\begin{aligned} E\{S_{\eta}\} &= E\{E\{S_{\eta}|\eta\}\} = \sum_{i=0}^{\infty} P\{\eta = i\} \cdot E\{\xi_1 + \dots + \xi_{\eta}|\eta = i\} = \\ &= \sum_{i=0}^{\infty} P\{\eta = i\} \cdot E\{\xi_1 + \dots + \xi_i\} = \sum_{i=0}^{\infty} i P\{\eta = i\} \cdot E\{\xi_1\} = E\{\eta\} \cdot E\{\xi_1\}. \end{aligned}$$

□

3.7. УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ОТНОСИТЕЛЬНО σ -АЛГЕБРЫ И ЕГО СВОЙСТВА

Для усреднения случайных величин в некоторых приложениях оказывается недостаточно введенных в п. 3.6 условного математического ожидания относительно случайного события и относительно случайной величины. В этом пункте исследуем условное математическое ожидание относительно σ -алгебры, а также другие вероятностные характеристики, связанные с этим понятием.

Прежде всего докажем одно вспомогательное свойство условного математического ожидания случайной величины $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$, определенной на вероятностном пространстве (Ω, \mathcal{F}, P) , относительно некоторого случайного события $B \in \mathcal{F}$, имеющего положительную вероятность $P(B) > 0$.

Теорема 3.12. *Если $P(B) > 0$, то справедлива формула*

$$E\{\xi|B\} = \frac{1}{P(B)} \int_B \xi(\omega) P(d\omega) = \frac{1}{P(B)} \int_{\Omega} \mathbb{1}_B(\omega) \xi(\omega) P(d\omega). \quad (3.63)$$

Доказательство. Разобьем доказательство на четыре этапа аналогично схеме конструирования интеграла Лебега.

На первом этапе покажем справедливость формулы (3.63) для индикаторной функции события $A \in \mathcal{F}$ – простейшей случайной величины Бернулли $\xi = \mathbb{1}_A(\omega)$:

$$\begin{aligned} \frac{1}{P(B)} \int_B \xi(\omega) P(d\omega) &= \frac{1}{P(B)} \int_{\Omega} \mathbb{1}_{A \cap B}(\omega) P(d\omega) = \frac{P(A \cap B)}{P(B)} = P(A|B) = \\ &= P_B(A) = \int_{\Omega} \mathbb{1}_A(\omega) P_B(d\omega) = \int_{\Omega} \xi(\omega) P_B(d\omega) = \int_{-\infty}^{\infty} x dF_{\xi|B}(x|B) = E\{\xi|B\}. \end{aligned}$$

Последнее равенство вытекает из определения (3.56).

На втором этапе для простой случайной величины $\xi = \xi(\omega) = \sum_{k=1}^N a_k \mathbb{1}_{A_k}(\omega)$ по ранее установленным свойствам математического ожидания в силу первого этапа имеем

$$\frac{1}{P(B)} \int_B \xi(\omega) P(d\omega) = \frac{1}{P(B)} \sum_{k=1}^N a_k \int_B \mathbb{1}_{A_k}(\omega) P(d\omega) = \sum_{k=1}^N a_k E\{\mathbb{1}_{A_k}(\omega)|B\} = E\{\xi|B\}.$$

На третьем этапе для неотрицательной случайной величины $\xi(\omega) \geq 0$ найдется монотонно возрастающая последовательность простых случайных величин $\xi_n \uparrow \xi$. Тогда $\mathbb{1}_B(\omega) \xi_n(\omega) \uparrow \mathbb{1}_B(\omega) \xi(\omega)$, поэтому в силу предыдущего этапа существует предел

$$\begin{aligned} \frac{1}{P(B)} \int_B \xi(\omega) P(d\omega) &= \frac{1}{P(B)} \int_{\Omega} \mathbb{1}_B(\omega) \xi(\omega) P(d\omega) = \\ &= \lim_{n \rightarrow \infty} \frac{1}{P(B)} \int_{\Omega} \mathbb{1}_B(\omega) \xi_n(\omega) P(d\omega) = \lim_{n \rightarrow \infty} E\{\xi_n|B\} = E\{\xi|B\}. \end{aligned}$$

На четвертом этапе для произвольной случайной величины ξ воспользуемся известным представлением $\xi(\omega) = \xi^+(\omega) - \xi^-(\omega)$ и результатом третьего этапа, а также свойствами математического ожидания:

$$\begin{aligned} \frac{1}{P(B)} \int_B \xi(\omega) P(d\omega) &= \frac{1}{P(B)} \int_B \xi^+(\omega) P(d\omega) - \frac{1}{P(B)} \int_B \xi^-(\omega) P(d\omega) = \\ &= E\{\xi^+|B\} - E\{\xi^-|B\} = E\{\xi|B\}, \end{aligned}$$

что совпадает с формулой (3.63). \square

При построении условного математического ожидания относительно σ -алгебры, как и при построении интеграла Лебега, воспользуемся принципом «от простого к сложному».

Пусть (Ω, \mathcal{F}, P) – произвольное вероятностное пространство, на котором определена произвольная случайная величина $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$, имеющая математическое ожидание $E\{\xi\}$. Пусть далее $H_1, \dots, H_N \in \mathcal{F}$ – произвольная полная система случайных событий ($N \leq \infty$). Всевозможные объединения $\bigcup_k H_{i_k}$ и пустое множество \emptyset , очевидно, образуют наименьшую σ -алгебру, содержащую $\{H_1, \dots, H_N\}$. Обозначим эту σ -алгебру $\mathcal{F}_0 \subset \mathcal{F}$.

Определение 3.17. Условным математическим ожиданием случайной величины ξ относительно счетной σ -алгебры $\mathcal{F}_0 \subset \mathcal{F}$ называется дискретная случайная величина $E\{\xi|\mathcal{F}_0\} : \Omega \rightarrow \mathbb{R}$, определенная соотношением

$$E\{\xi|\mathcal{F}_0\} = \beta(\omega) ::= \begin{cases} E\{\xi|H_1\}, & \omega \in H_1, \\ \vdots \\ E\{\xi|H_N\}, & \omega \in H_N. \end{cases} \quad (3.64)$$

Иногда, чтобы подчеркнуть зависимость от ω , используется обозначение $E\{\xi|\mathcal{F}_0\}(\omega)$.

Свойства условного математического ожидания

C1. Функция $\beta(\omega)$ измерима относительно «своей» σ -алгебры \mathcal{F}_0 .

Доказательство. Для любого борелевского множества $B \in \mathcal{B}$ согласно (3.64) найдется набор индексов $1 \leq i_1 < \dots < i_M \leq N$ ($M \leq N$) такой, что

$$\{\omega : \beta(\omega) \in B\} = \bigcup_{k=1}^M H_{i_k} \in \mathcal{F}_0.$$

Это и означает \mathcal{F}_0 -измеримость функции $\beta(\omega)$. □

C2. Для любого случайного события $A \in \mathcal{F}_0$ выполняется равенство

$$\int_A \xi(\omega) P(d\omega) = \int_A E\{\xi|\mathcal{F}_0\} P(d\omega). \quad (3.65)$$

Доказательство. Применим к правой части свойство аддитивности интеграла Лебега $\left(A \equiv \bigcup_{k=1}^N (A \cap H_k)\right)$, теорему 3.11 и формулу (3.64):

$$\begin{aligned} \int_A E\{\xi|\mathcal{F}_0\} P(d\omega) &= \sum_{k=1}^N \int_{A \cap H_k} \xi(\omega) P(d\omega) = \\ &= \sum_{k=1}^N P(A \cap H_k) E\{\xi|H_k\} = P(A) \cdot \sum_{k=1}^N E\{\xi|H_k\} \times \\ &\times P\{H_k|A\} = P(A) \cdot E\{\xi|A\} = \int_A \xi(\omega) P(d\omega). \end{aligned}$$

□

Перейдем теперь к рассмотрению общего случая для σ -подалгебры $\mathcal{F}_0 \subset \mathcal{F}$. Как известно, случайная величина $\zeta = \zeta(\omega) : \Omega \rightarrow \mathbb{R}$ является \mathcal{F}_0 -измеримой, если для любого интервала $\Delta \in \mathcal{B}$

$$\{\omega : \zeta(\omega) \in \Delta\} \in \mathcal{F}_0.$$

Определение 3.18. Случайная величина $\zeta = E\{\xi|\mathcal{F}_0\}$ называется условным математическим ожиданием случайной величины ξ относительно σ -алгебры $\mathcal{F}_0 \subset \mathcal{F}$, если:

- 1) случайная величина ζ является \mathcal{F}_0 -измеримой;
- 2) для всех $A \in \mathcal{F}_0$ выполняется равенство, аналогичное (3.65):

$$\int_A \xi(\omega) P(d\omega) = \int_A \zeta(\omega) P(d\omega). \quad (3.66)$$

Замечание 3.14. Эти два условия однозначно с вероятностью 1 определяют случайную величину $\zeta(\omega)$. Действительно, если $\zeta_1 = \zeta_1(\omega)$ также удовлетворяет условиям 1) и 2), то, поскольку $\{\zeta_1 - \zeta \geq 0\}, \{\zeta - \zeta_1 > 0\} \in \mathcal{F}_0$, имеем цепочку равенств:

$$\begin{aligned} \int_{\Omega} |\zeta(\omega) - \zeta_1(\omega)| P(d\omega) &= \int_{\{\zeta_1 - \zeta \geq 0\}} (\zeta_1(\omega) - \zeta(\omega)) P(d\omega) + \\ &+ \int_{\{\zeta - \zeta_1 > 0\}} (\zeta(\omega) - \zeta_1(\omega)) P(d\omega) = \int_{\{\zeta_1 - \zeta \geq 0\}} (\xi(\omega) - \zeta_1(\omega)) P(d\omega) + \\ &+ \int_{\{\zeta - \zeta_1 > 0\}} (\xi(\omega) - \zeta_1(\omega)) P(d\omega) = 0. \end{aligned}$$

Отсюда следует: $\zeta_1(\omega) \stackrel{\text{п. н.}}{=} \zeta(\omega)$.

Отметим еще, что $E\{\xi|\mathcal{F}_0\}$ существует, если $E\{|\xi|\} < \infty$.

Теорема 3.13. Если $\mathcal{F}_0 = \mathcal{F}_\eta$ — σ -алгебра, порожденная случайной величиной $\eta = \eta(\omega)$, то (3.66) и (3.62) эквивалентны.

Доказательство. Формула (3.66) — частный случай (3.62) при $h(\eta) = \mathbb{1}_A(\eta)$.

Пусть теперь выполняется (3.66). Тогда (3.62) верна при $h(\eta) = \mathbb{1}_A(\eta)$. В силу свойства линейности математического ожидания равенство (3.62) будет верно и для простых функций $h(\cdot)$. Далее, действуя как и при построении интеграла Лебега, доказываем, что равенство (3.62) верно и для любых ограниченных борелевских функций. \square

Определение 3.19. Выражение

$$P\{A|\mathcal{F}_0\} = E\{\mathbb{1}_A(\omega)|\mathcal{F}_0\}$$

называется условной вероятностью случайного события $A \in \mathcal{F}$ относительно σ -алгебры \mathcal{F}_0 .

Определение 3.20. Действительная функция

$$F_\xi(x|\mathcal{F}_0) = P\{\xi < x|\mathcal{F}_0\}, \quad x \in \mathbb{R},$$

называется условной функцией распределения случайной величины ξ относительно σ -алгебры \mathcal{F}_0 . Если эта условная функция распределения абсолютно непрерывна:

$$F_\xi(x|\mathcal{F}_0) = \int_{-\infty}^x p_\xi(t|\mathcal{F}_0) dt, \quad x \in \mathbb{R},$$

то $p_\xi(x|\mathcal{F}_0) \geq 0$ называется условной плотностью распределения величины ξ относительно σ -алгебры \mathcal{F}_0 .

**Свойства условного математического ожидания
и условной вероятности относительно σ -алгебры $\mathcal{F}_0 \subset \mathcal{F}$**

- C1. Если $c = \text{const}$ и $\xi \stackrel{\text{п. н.}}{=} c$, то $E\{\xi|\mathcal{F}_0\} \stackrel{\text{п. н.}}{=} c$.
 C2. Если $\xi \stackrel{\text{п. н.}}{\leq} \eta$, то $E\{\xi|\mathcal{F}_0\} \stackrel{\text{п. н.}}{\leq} E\{\eta|\mathcal{F}_0\}$.
 C3. $|E\{\xi|\mathcal{F}_0\}| \stackrel{\text{п. н.}}{\leq} E\{|\xi|\mathcal{F}_0\}$.
 C4. Если a, b – постоянные и $a \cdot E\{\xi\} + b \cdot E\{\eta\}$ определено, то

$$E\{a\xi + b\eta|\mathcal{F}_0\} \stackrel{\text{п. н.}}{=} aE\{\xi|\mathcal{F}_0\} + bE\{\eta|\mathcal{F}_0\}.$$

Доказательство. Вышеприведенные свойства очевидным образом вытекают из определения 3.18 и ранее установленных свойств интеграла Лебега. \square

- C5. Если $\mathcal{F}_0 = \{\emptyset, \Omega\}$ – тривиальная σ -подалгебра \mathcal{F} , то условное математическое ожидание почти наверное совпадает с безусловным математическим ожиданием:

$$E\{\xi|\mathcal{F}_0\} \stackrel{\text{п. н.}}{=} E\{\xi\}. \quad (3.67)$$

Доказательство. Достаточно проверить свойства 1) и 2) в определении 3.18. Константа $E\{\xi\}$ является \mathcal{F}_0 -измеримой величиной, поэтому свойство 1) выполнено. Если $A = \Omega$ или $A = \emptyset$, то, очевидно, выполняется (3.66), так что свойство 2) также выполнено в случае (3.67). \square

- C6. Если $\mathcal{F}_0 = \mathcal{F}$, то справедливо равенство

$$E\{\xi|\mathcal{F}\} \stackrel{\text{п. н.}}{=} \xi. \quad (3.68)$$

Доказательство. Проверим свойства 1), 2) определения 3.18. Случайная величина ξ \mathcal{F} -измерима, и в случае (3.68) соотношение (3.66) выполняется тождественно:

$$\int_A \xi(\omega) P(d\omega) = \int_A \xi(\omega) P(d\omega), \quad A \in \mathcal{F}.$$

\square

- C7. Для двух вложенных σ -подалгебр $\mathcal{F}_1 \subseteq \mathcal{F}_2$ справедливо соотношение

$$E\{E\{\xi|\mathcal{F}_2\}|\mathcal{F}_1\} \stackrel{\text{п. н.}}{=} E\{\xi|\mathcal{F}_1\}. \quad (3.69)$$

Доказательство. Пусть $A \in \mathcal{F}_1$. Тогда

$$\int_A E\{\xi|\mathcal{F}_1\} P(d\omega) = \int_A \xi P(d\omega).$$

В силу вложенности σ -подалгебры \mathcal{F}_1 событие $A \in \mathcal{F}_2$, поэтому

$$\int_A E\{E\{\xi|\mathcal{F}_2\}|\mathcal{F}_1\} P(d\omega) = \int_A E\{\xi|\mathcal{F}_2\} P(d\omega) = \int_A \xi P(d\omega).$$

Следовательно, $\forall A \in \mathcal{F}_1$:

$$\int_A E\{\xi|\mathcal{F}_1\} P(d\omega) = \int_A E\{E\{\xi|\mathcal{F}_2\}|\mathcal{F}_1\} P(d\omega).$$

Из этого интегрального равенства в силу произвола $A \in \mathcal{F}_1$ и свойства интеграла Лебега (см. замечание 3.14) заключаем справедливость (3.69). \square

С8. Справедлива обобщенная формула полного математического ожидания

$$E\{E\{\xi|\mathcal{F}_0\}\} = E\{\xi\}. \quad (3.70)$$

Доказательство. Соотношение (3.70) вытекает из свойств С6, С7, если положить $\mathcal{F}_1 = \{\emptyset, \Omega\}$ и $\mathcal{F}_2 = \mathcal{F}$. \square

Следствие 3.15. Справедлива обобщенная формула полной вероятности:

$$E\{P\{A|\mathcal{F}_0\}\} = P(A), \quad A \in \mathcal{F}.$$

Доказательство. Достаточно положить в (3.70) $\xi(\omega) = \mathbb{1}_A(\omega)$, $\omega \in \Omega$. \square

С9. Если $\mathcal{F}_2 \subseteq \mathcal{F}_1$, то

$$E\{E\{\xi|\mathcal{F}_2\}|\mathcal{F}_1\} \stackrel{\text{п. н.}}{=} E\{\xi|\mathcal{F}_2\}. \quad (3.71)$$

Доказательство. $\forall A \in \mathcal{F}_1$ по определению 3.18 условного математического ожидания $E\{E\{\xi|\mathcal{F}_2\}|\mathcal{F}_1\}$ имеем

$$\int_A E\{E\{\xi|\mathcal{F}_2\}|\mathcal{F}_1\} P(d\omega) = \int_A E\{\xi|\mathcal{F}_2\} P(d\omega).$$

Функция $E\{\xi|\mathcal{F}_2\}$ — \mathcal{F}_2 -измерима, а так как $\mathcal{F}_2 \subseteq \mathcal{F}_1$, то и \mathcal{F}_1 -измерима. Отсюда в силу произвола $A \in \mathcal{F}_1$ следует (3.71). \square

С10. Если случайная величина ξ с конечным математическим ожиданием не зависит от σ -подалгебры $\mathcal{F}_0 \subset \mathcal{F}$ (т. е. ξ не зависит от случайных величин Бернулли $\mathbb{1}_B, B \in \mathcal{F}_0$), то

$$E\{\xi|\mathcal{F}_0\} \stackrel{\text{п. н.}}{=} E\{\xi\}. \quad (3.72)$$

Доказательство. Проверим свойства 1), 2) определения 3.18. Поскольку $E\{\xi\}$ — константа, то она \mathcal{F}_0 -измерима, т. е. свойство 1) выполнено. Для выполнения свойства 2) надо проверить равенство $\forall B \in \mathcal{F}_0$:

$$\int_B \xi P(d\omega) = \int_A E\{\xi\} P(d\omega). \quad (3.73)$$

Правая часть (3.73), очевидно, равна

$$\int_B E\{\xi\} P(d\omega) = P(B) \cdot E\{\xi\}.$$

Левая часть в силу свойства мультипликативности математического ожидания равна

$$\int_B \xi P(d\omega) = E\{\xi \cdot \mathbb{1}_B\} = E\{\xi\} \cdot E\{\mathbb{1}_B\} = P(B) \cdot E\{\xi\},$$

так что (3.73) выполняется. Следовательно, имеет место соотношение (3.72). \square

С11. Если $E\{|\eta|\} < \infty$, $E\{|\xi\eta|\} < \infty$, то \mathcal{F}_0 -измеримый множитель $\eta = \eta(\omega)$, $\omega \in \Omega$, можно выносить из-под знака условного математического ожидания:

$$E\{\xi\eta|\mathcal{F}_0\} \stackrel{\text{п. н.}}{=} \eta \cdot E\{\xi|\mathcal{F}_0\}.$$

С12. Пусть $\xi = \xi(\omega)$ – произвольная случайная величина, определенная на вероятностном пространстве (Ω, \mathcal{F}, P) , $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \dots \subset \mathcal{F}$ – поток σ -подалгебр, а \mathcal{F}_∞ – наименьшая σ -алгебра, содержащая $\bigcup_{k=1}^{\infty} \mathcal{F}_k$. Тогда имеет место предельный переход по условию в условном математическом ожидании:

$$\lim_{k \rightarrow \infty} E\{\xi | \mathcal{F}_k\} \stackrel{\text{п. н.}}{=} E\{\xi | \mathcal{F}_\infty\}.$$

В заключение докажем существование \mathcal{F}_0 -измеримой случайной величины $\zeta(\omega)$, удовлетворяющей (3.66) для любого $A \in \mathcal{F}_0$, с помощью теории меры. Обозначим числовую функцию множеств $Q = Q(A)$, определяемую левой частью (3.66):

$$Q = Q(A) = \int_A \xi(\omega) P(d\omega), \quad A \in \mathcal{F}_0. \quad (3.74)$$

Воспользуемся введенным в п. 3.2 представлением СВ $\xi = \xi(\omega)$ через неотрицательные случайные величины ξ^+ , $\xi^- \geq 0$: $\xi(\omega) = \xi^+(\omega) - \xi^-(\omega)$. Тогда

$$Q = Q(A) = Q^+(A) - Q^-(A), \quad Q^\pm = Q^\pm(A) = \int_A \xi^\pm(\omega) P(d\omega), \quad A \in \mathcal{F}_0.$$

Из свойств интеграла Лебега следует:

- 1) функции $Q^+ = Q^+(A)$, $Q^- = Q^-(A)$, $A \in \mathcal{F}_0$, удовлетворяют всем аксиомам меры и поэтому являются некоторыми мерами, причем $\min(Q^+(\Omega), Q^-(\Omega)) < \infty$;
- 2) мера $Q = Q(A)$ абсолютно непрерывна относительно меры $P = P(A)$, т. е.

если $P(A) = 0$, то $Q(A) = 0$.

Это свойство называется еще **доминируемостью мер** и обозначается так: $Q \ll P$. При этом справедлив и обратный результат, выражаемый теоремой Радона – Никодима.

Теорема 3.14 (теорема Радона – Никодима). Пусть на измеримом пространстве (Ω, \mathcal{F}_0) заданы две меры: μ – σ -конечная мера и λ – мера со знаком (т. е. $\lambda = \lambda_1 - \lambda_2$, где одна из мер λ_1 или λ_2 конечна), являющаяся абсолютно непрерывной относительно μ . Тогда существует \mathcal{F}_0 -измеримая функция $f = f(\omega)$ со значениями из расширенной прямой $\bar{\mathbb{R}} = [-\infty, +\infty]$ такая, что

$$\lambda(A) = \int_A f(\omega) \mu(d\omega), \quad A \in \mathcal{F}_0. \quad (3.75)$$

С точностью до множеств μ -меры нуль-функция $f(\omega)$ единственна: если $h = h(\omega)$ – другая \mathcal{F}_0 -измеримая функция такая, что

$$\lambda(A) = \int_A h(\omega) \mu(d\omega), \quad A \in \mathcal{F}_0,$$

то $\mu\{\omega : f(\omega) \neq h(\omega)\} = 0$. При этом если $\lambda = \lambda(A) \geq 0$ – мера, то $f = f(\omega)$ принимает значения в $\bar{\mathbb{R}}_+ = [0, +\infty]$.

Определение 3.21. Подынтегральная функция $f = f(\omega)$ в интеграле Лебега (3.75) называется **производной Радона – Никодима (или плотностью) меры $\lambda(\cdot)$ относительно меры $\mu(\cdot)$** и обозначается $\frac{d\lambda}{d\mu}$ или $\frac{d\lambda}{d\mu}(\omega)$.

Положим в (3.75) $\mu(\cdot) = P(\cdot)$, а $\lambda(\cdot) = Q(\cdot)$, где мера $Q(\cdot)$ определяется (3.74). Тогда, сравнивая (3.74) и (3.75), согласно теореме Радона – Никодима получаем, что

$$\zeta = f(\omega) = E\{\xi|\mathcal{F}_0\} \equiv \frac{dQ}{dP}(\omega),$$

т. е. условное математическое ожидание СВ ξ относительно σ -алгебры \mathcal{F}_0 есть не что иное как производная Радона – Никодима меры Q относительно меры P (рассматриваемых на (Ω, \mathcal{F}_0)). Отметим еще, что в соответствии с теоремой Радона – Никодима условное математическое ожидание $E\{\xi|\mathcal{F}_0\}$ определяется однозначно лишь с точностью до множеств P -меры нуль.

3.8. МОМЕНТЫ СКАЛЯРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И ИХ СВОЙСТВА

Определение 3.22. Пусть на (Ω, \mathcal{F}, P) определена случайная величина $\xi = \xi(\omega)$ и $k \in \mathbb{N}$. Начальным моментом k -го порядка случайной величины ξ называется число

$$\nu_k = E\{\xi^k\} = \int_{-\infty}^{+\infty} x^k dF_\xi(x).$$

Число $E\{|\xi|^k\}$ называется абсолютным начальным моментом k -го порядка.

Определение 3.23. Случайная величина $\overset{\circ}{\xi} = \overset{\circ}{\xi}(\omega) ::= \xi(\omega) - E\{\xi\}$ называется центрированной случайной величиной, для которой выполняется следующее тождество:

$$E\left\{\overset{\circ}{\xi}\right\} \equiv 0.$$

Определение 3.24. Центральным моментом k -го порядка случайной величины ξ называется число $\mu_k = E\left\{\overset{\circ}{\xi}^k\right\} = E\{(\xi - E\{\xi\})^k\} = \int_{-\infty}^{+\infty} (x - E\{\xi\})^k dF_\xi(x)$, при этом $E\left\{|\overset{\circ}{\xi}|^k\right\}$ – абсолютный центральный момент.

Свойства моментов скалярной случайной величины

- C1. Для любой случайной величины ξ справедливы равенства: $\nu_1 = E\{\xi\}$, $\mu_1 \equiv 0$.
 C2. Если для некоторого $t > 0$ ограничен момент порядка t , то для $\forall s, 0 < s < t$, ограничены моменты низших порядков s .
Доказательство. Достаточно воспользоваться неравенством Ляпунова. \square
 C3. Для гауссовской случайной величины ξ , $\mathcal{L}\{\xi\} = N_1(a, B)$, справедливо равенство

$$\mu_k = \begin{cases} 0, & k - \text{нечетное,} \\ B^{k/2}(k-1)!!, & k - \text{четное.} \end{cases}$$

- C4. (Связь начальных и центральных моментов.) Справедливы следующие соотношения:

$$\nu_k = \sum_{l=0}^k C_k^l \mu_l (\nu_1)^{k-l}; \quad \mu_k = \sum_{l=0}^k C_k^l \nu_l (-\nu_1)^{k-l}, \quad k \in \mathbb{N}.$$

Доказательство. Используя свойства математического ожидания, имеем

$$\begin{aligned} \nu_k &\equiv E \left\{ \left(\overset{\circ}{\xi} + \nu_1 \right)^k \right\} = E \left\{ \sum_{l=0}^k C_k^l \left(\overset{\circ}{\xi} \right)^l (\nu_1)^{k-l} \right\} = \\ &= \sum_{l=0}^k C_k^l (\nu_1)^{k-l} E \left\{ \left(\overset{\circ}{\xi} \right)^l \right\} = \sum_{l=0}^k C_k^l (\nu_1)^{k-l} \mu_l. \end{aligned}$$

Второе соотношение доказывается аналогично. \square

Следствие 3.16. Справедливы следующие формулы:

$$\mu_2 = \nu_2 - \nu_1^2 = E \{ \xi^2 \} - E^2 \{ \xi \} \geq 0, \quad \nu_2 = \mu_2 + \nu_1^2 = E \left\{ \left(\overset{\circ}{\xi} \right)^2 \right\} + E^2 \{ \xi \}.$$

С5. (Проблема моментов.) Если выполняется критерий Карлемана, т. е. если расходится ряд

$$\sum_{k=1}^{+\infty} \frac{1}{2k \sqrt{\mu_{2k}}} = \infty,$$

то набор всевозможных начальных моментов $\{\nu_k\}$ однозначно определяет функцию распределения $F_\xi(\cdot)$.

3.9. ДИСПЕРСИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И ЕЕ СВОЙСТВА

Определение 3.25. Пусть $\xi = \xi(\omega)$ – произвольная случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) , имеющая функцию распределения $F_\xi(x)$. Дисперсией этой случайной величины называется ее центральный момент второго порядка:

$$D\{\xi\} = \sigma_\xi^2 ::= \mu_2 = E\{(\xi - E\{\xi\})^2\} = \int_{-\infty}^{+\infty} (x - E\{\xi\})^2 dF_\xi(x) \geq 0, \quad (3.76)$$

при этом $\sigma_\xi = \sqrt{D\{\xi\}} \geq 0$ называется среднеквадратическим (стандартным) отклонением случайной величины.

Справедливы формулы

$$D\{\xi\} = \sum_{i=1}^N (a_i - E\{\xi\})^2 P\{\xi = a_i\},$$

если $\xi \in \{a_1, \dots, a_N\}$ – дискретная случайная величина, $N \leq +\infty$;

$$D\{\xi\} = \int_{-\infty}^{+\infty} (x - E\{\xi\})^2 p_\xi(x) dx,$$

если ξ – абсолютно непрерывная случайная величина с плотностью распределения $p_\xi(x) = F'_\xi(x)$.

Дисперсия характеризует среднеквадратичное рассеяние (разброс) случайной величины относительно точки $E\{\xi\}$.

Свойства дисперсии

С1. Для основных законов распределения вероятностей дисперсия вычисляется по формулам, приведенным в табл. 3.1.

Задание 3.1. Заполнить табл. 3.1 для всех ранее изученных основных законов распределения вероятностей.

С2. $D\{\xi\} \geq 0$.

С3. $D\{\xi\} = 0 \Leftrightarrow \xi \stackrel{\text{п. н.}}{=} \text{const} = E\{\xi\}$.

Доказательство. Используя свойства математического ожидания, получаем цепочку равенств

$$D\{\xi\} = E\left\{\left(\overset{\circ}{\xi}\right)^2\right\} = 0 \Leftrightarrow \overset{\circ}{\xi} \stackrel{\text{п. н.}}{=} 0 \Leftrightarrow \xi \stackrel{\text{п. н.}}{=} E\{\xi\}.$$

□

Таблица 3.1

$\mathcal{L}\{\xi\}$	$Bi(1, p)$	$Bi(K, p)$	$R[a, b]$	$N_1(a, B)$	$\Pi(\lambda)$...
$E\{\xi\}$	p	Kp	$\frac{a+b}{2}$	a	λ	...
$D\{\xi\}$	$p(1-p)$	$Kp(1-p)$	$\frac{(b-a)^2}{12}$	B	λ	...

С4. При линейном преобразовании случайной величины дисперсия изменяется следующим образом:

$$D\{a\xi + b\} = a^2 D\{\xi\}, \forall a, b \in \mathbb{R}.$$

Доказательство. По формуле (3.76) имеем

$$\begin{aligned} D\{a\xi + b\} &= E\{(a\xi + b - E\{a\xi + b\})^2\} = E\left\{\left(a \overset{\circ}{\xi}\right)^2\right\} = \\ &= a^2 E\left\{\left(\overset{\circ}{\xi}\right)^2\right\} = a^2 D\{\xi\}. \end{aligned}$$

□

Следствие 3.17. $D\{\xi + b\} = D\{\xi\}$.

Следствие 3.18. $D\{-\xi\} = D\{\xi\}$.

С5. Дисперсия суммы двух случайных величин удовлетворяет следующему равенству:

$$D\{\xi + \eta\} = D\{\xi\} + D\{\eta\} + 2E\{(\xi - E\{\xi\})(\eta - E\{\eta\})\}.$$

Доказательство. Имеем цепочку равенств:

$$D\{\xi + \eta\} = E\left\{\left(\overset{\circ}{\xi} + \overset{\circ}{\eta}\right)^2\right\} = D\{\xi\} + D\{\eta\} + 2E\left\{\overset{\circ}{\xi}\overset{\circ}{\eta}\right\}.$$

□

С6. (Неравенство Чебышева относительно дисперсии.) Для любой случайной величины $\xi = \xi(\omega)$, имеющей конечное математическое ожидание $E\{\xi\}$ и дисперсию $D\{\xi\}$, и любого $\epsilon > 0$ справедливо неравенство

$$P\{(\xi - E\{\xi\}) \geq \epsilon\} \leq \frac{D\{\xi\}}{\epsilon^2}. \quad (3.77)$$

Доказательство. Из неравенства Чебышева относительно математического ожидания имеем

$$P\left\{t \left| \overset{\circ}{\xi} \right| \geq \epsilon\right\} \leq \frac{E\left\{\overset{\circ}{\xi}^2\right\}}{\epsilon^2} = \frac{D\{\xi\}}{\epsilon^2}.$$

□

Следствие 3.19 (правило «3σ»). С вероятностью, не меньшей $\frac{8}{9}$, случайная величина ξ концентрируется в 3σ -окрестности своего математического ожидания, т. е. в окрестности $U = (E\{\xi\} - 3\sigma, E\{\xi\} + 3\sigma)$, где $\sigma = \sqrt{D\{\xi\}}$.

Доказательство. Используя (3.77), оценим

$$P\{\xi \notin U\} \equiv P\{|\xi - E\{\xi\}| \geq 3\sigma\} \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9}.$$

□

Задание 3.2. Обобщить правило «3σ» в правило «kσ», $k > 0$.

Замечание 3.15. Правило «3σ» верно для любой случайной величины и поэтому широко используется на практике.

3.10. МОМЕНТЫ МНОГОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ, КОВАРИАЦИЯ, КОВАРИАЦИОННАЯ МАТРИЦА, КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ И ИХ СВОЙСТВА

Определение 3.26. Пусть на пространстве (Ω, \mathcal{F}, P) определены случайный вектор $\xi = (\xi_1, \dots, \xi_n)^T \in \mathbb{R}^n$ и математические ожидания $m_i = E\{\xi_i\}$, $m = (m_1, \dots, m_n)^T = E\{\xi\} \in \mathbb{R}^n$ – вектор математического ожидания. Пусть далее $k \in \mathbb{N}$, k_1, \dots, k_n – неотрицательные целые числа, задающие разбиение k , т. е. $k = k_1 + k_2 + \dots + k_n$. Тогда центральным смешанным моментом порядка k называется число

$$\mu_{k_1, \dots, k_n} ::= E\{(\xi_1 - m_1)^{k_1} \dots (\xi_n - m_n)^{k_n}\}.$$

Определение 3.27. Ковариацией (от англ. covariance – взаимное изменение) случайных величин ξ_i, ξ_j называется центральный смешанный момент второго порядка:

$$\sigma_{ij} = \text{cov}\{\xi_i, \xi_j\} = E\{(\xi_i - m_i)(\xi_j - m_j)\}, \quad i, j = 1, \dots, n. \quad (3.78)$$

При этом $(n \times n)$ -матрица $\Sigma = (\sigma_{ij})$ называется ковариационной матрицей.

Приведем формулы для практических вычислений ковариации. Если $\xi \in \{a^{(1)}, \dots, a^{(N)}\}$ – дискретный случайный вектор, то

$$\sigma_{ij} = \sum_{l=1}^N \left(a_i^{(l)} - m_i\right) \left(a_j^{(l)} - m_j\right) P\left\{\xi_i = a_i^{(l)}, \xi_j = a_j^{(l)}\right\}.$$

Если ξ имеет абсолютно непрерывную плотность, то

$$\sigma_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - m_i)(x_j - m_j) p_{\xi_i, \xi_j}(x_i, x_j) dx_i dx_j, \quad i \neq j, \quad (3.79)$$

где $p_{\xi_i, \xi_j}(x_i, x_j)$ – совместная плотность распределения случайных величин ξ_i, ξ_j .

Свойства ковариации и ковариационной матрицы

C1. Для практических вычислений ковариации удобна формула

$$\sigma_{ij} = E\{\xi_i \xi_j\} - E\{\xi_i\} \cdot E\{\xi_j\}, \quad i, j = 1, \dots, n. \quad (3.80)$$

Доказательство. Используя (3.78) и свойства математического ожидания, получаем требуемое. \square

C2. Справедливы матричные формулы:

$$\Sigma = E\{(\xi - E\xi) \cdot (\xi - E\xi)^T\},$$

$$\Sigma = E\{\xi \cdot \xi^T\} - E\{\xi\} \cdot (E\{\xi\})^T.$$

Доказательство. Из (3.78), (3.80) и матричных правил получим матрицу под знаком E , а вследствие покомпонентного действия E получаем доказываемое. \square

C3. Ковариация и ковариационная матрица – объекты симметричные, т. е. $\sigma_{ij} = \sigma_{ji}$, $\Sigma = \Sigma^T$.

C4. Диагональные элементы ковариационной матрицы являются дисперсиями: $\sigma_{ii} = E\{(\xi_i - E\{\xi_i\})^2\} = D\{\xi_i\}$, $i = 1, \dots, n$.

C5. Внедиагональные элементы ковариационной матрицы ограничены: $|\sigma_{ij}| \leq \sqrt{\sigma_{ii} \cdot \sigma_{jj}}$, $i, j = 1, \dots, n$.

Доказательство. Из неравенства Коши – Буняковского

$$|\sigma_{ij}| = \left| E\left\{ \overset{\circ}{\xi}_i \cdot \overset{\circ}{\xi}_j \right\} \right| \leq \sqrt{E\left\{ \left(\overset{\circ}{\xi}_i \right)^2 \right\} \cdot E\left\{ \left(\overset{\circ}{\xi}_j \right)^2 \right\}} = \sqrt{\sigma_{ii} \cdot \sigma_{jj}}.$$

\square

C6. Ковариационная матрица – матрица неотрицательно определенная, т. е. $\forall u \in \mathbb{R}^n$ квадратичная форма $Q(u) = u^T \Sigma u \geq 0$.

Доказательство. В силу C2 имеем

$$Q(u) = u^T E\left\{ \overset{\circ}{\xi} \cdot \left(\overset{\circ}{\xi} \right)^T \right\} u = E\left\{ \left(u^T \overset{\circ}{\xi} \right)^2 \right\} \geq 0.$$

\square

Задание 3.3. Разработать алгоритм генерации (случайным образом) ковариационной матрицы.

C7. Если ξ_i и ξ_j независимы, то их ковариация равна нулю: $\sigma_{ij} = \text{cov}\{\xi_i, \xi_j\} = 0$.

Доказательство. Воспользуемся мультипликативным свойством математического ожидания:

$$\sigma_{ij} = E\left\{ \overset{\circ}{\xi}_i \cdot \overset{\circ}{\xi}_j \right\} = E\left\{ \overset{\circ}{\xi}_i \right\} \cdot E\left\{ \overset{\circ}{\xi}_j \right\} = 0 \cdot 0 = 0.$$

\square

C8. Если ξ – гауссовский случайный вектор с нормальным распределением: $\mathcal{L}\{\xi\} = N_n(a, B)$, то его ковариационная матрица $\Sigma = B$.

Доказательство. Провести самостоятельно, пользуясь формулой (3.79). \square

Определение 3.28. Коэффициентом корреляции случайных величин ξ_i , ξ_j называется их нормированная ковариация:

$$\rho_{ij} = \text{corr}\{\xi_i, \xi_j\} ::= \frac{\text{cov}(\xi_i, \xi_j)}{\sqrt{D\{\xi_i\}D\{\xi_j\}}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \quad i, j = 1, \dots, n. \quad (3.81)$$

При этом иногда $(n \times n)$ -матрица $\rho = (\rho_{ij})$ называется корреляционной матрицей.

Свойства коэффициента корреляции

С1. Коэффициент корреляции – величина симметричная: $\rho_{ij} = \rho_{ji}$.

С2. Коэффициент корреляции ограничен: $-1 \leq \rho_{ij} \leq +1$.

Доказательство. По свойству ковариации $|\sigma_{ij}| \leq \sqrt{\sigma_{ii}\sigma_{jj}}$, и из (3.81) следует доказываемое. \square

С3. Диагональные элементы корреляционной матрицы равны единице: $\rho_{ii} = 1$, $i = 1, \dots, n$.

С4. Если случайные величины ξ_i и ξ_j независимы, то их коэффициент корреляции равен нулю: $\rho_{ij} = 0$.

Доказательство. Результат следует из соответствующего свойства ковариации. \square

Определение 3.29. Случайные величины ξ_i , ξ_j , $i \neq j$, для которых $\rho_{ij} = 0$, называются некоррелированными.

Следствие 3.20. Если случайные величины ξ_i , ξ_j – независимые, то они некоррелированы. Обратное, вообще говоря, неверно.

Доказательство. Первое утверждение следует из С4. Второе утверждение доказывается контрпримером: $n = 2$, $\mathcal{L}\{\xi_1\} = N(0, 1)$, $\xi_2 = \xi_1^2$. Случайные величины ξ_1 , ξ_2 зависимы, но некоррелированы: $\sigma_{12} = \text{cov}\{\xi_1, \xi_2\} = E\{\xi_1^3\} = 0$. \square

С5. Если $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \in \mathbb{R}^2$ – гауссовский случайный вектор с нормальным распределением $N_2(a, B)$, то понятия независимости и некоррелированности случайных величин ξ_1 , ξ_2 совпадают.

Доказательство. Из свойств ковариационной матрицы имеем

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = B.$$

С другой стороны, из критерия независимости гауссовских случайных величин имеем $B = \text{diag}\{b_{11}, b_{22}\}$, или

$$B = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}.$$

Следовательно, $\sigma_{12} = 0$, что равносильно независимости ξ_1 и ξ_2 . \square

С6. Коэффициент корреляции принимает свои экстремальные значения $\rho_{ij} = \pm 1$ тогда и только тогда, когда случайные величины ξ_i и ξ_j связаны почти наверное линейной зависимостью: $\xi_j \stackrel{\text{п. н.}}{=} a + b\xi_i$, a , b – постоянные.

Доказательство. Введем в рассмотрение так называемые нормированные случайные величины: $\xi_k^* = \xi_k / \sqrt{\sigma_{kk}}$, $k \in \{i, j\}$. Тогда по построению $E\{\xi_k^*\} = 0$, $D\{\xi_k^*\} = 1$. С помощью этих нормированных случайных величин построим еще одну величину: $\zeta = \xi_j^* - \rho_{ij} \xi_i^*$. По построению $E\{\zeta\} = 0$. Тогда вычислим

$$E\{\zeta^2\} = E\{(\xi_j^* - \rho_{ij} \xi_i^*)^2\} = 1 - 2\rho_{ij} \cdot E\{\xi_i^* \cdot \xi_j^*\} + \rho_{ij}^2 = 1 - \rho_{ij}^2.$$

Имея это в виду, получаем

$$|\rho_{ij}| = 1 \Leftrightarrow E\{\zeta^2\} = 0 \Leftrightarrow \zeta \stackrel{\text{п. н.}}{=} 0 \Leftrightarrow \xi_j^* \stackrel{\text{п. н.}}{=} \rho_{ij} \xi_i^*,$$

но $\rho_{ij} = \pm 1$, т. е.

$$\xi_j^* \stackrel{\text{п. н.}}{=} \pm \xi_i^* \Leftrightarrow \frac{\xi_j - m_j}{\sqrt{\sigma_{jj}}} \stackrel{\text{п. н.}}{=} \pm \frac{\xi_i - m_i}{\sqrt{\sigma_{ii}}} \Leftrightarrow \xi_j \stackrel{\text{п. н.}}{=} a + b\xi_i,$$

где

$$a = m_j \mp \sqrt{\frac{\sigma_{jj}}{\sigma_{ii}}} m_i, \quad b = \pm \sqrt{\frac{\sigma_{jj}}{\sigma_{ii}}}.$$

□

Следствие 3.21. Модуль коэффициента корреляции ρ_{ij} есть мера линейной зависимости случайных величин: чем больше $|\rho_{ij}|$, тем сильнее эта зависимость.

С7. Справедлива следующая формула для дисперсии суммы случайных величин:

$$D\{\xi_1 + \xi_2\} = D\{\xi_1\} + D\{\xi_2\} + 2\rho_{12}\sqrt{D\{\xi_1\}D\{\xi_2\}}.$$

Доказательство. По формуле (3.81) имеем

$$\begin{aligned} D\{\xi_1 + \xi_2\} &= E\left\{\left(\overset{\circ}{\xi}_1 + \overset{\circ}{\xi}_2\right)^2\right\} = E\left\{\left(\overset{\circ}{\xi}_1\right)^2\right\} + \\ &+ E\left\{\left(\overset{\circ}{\xi}_2\right)^2\right\} + 2E\left\{\overset{\circ}{\xi}_1 \cdot \overset{\circ}{\xi}_2\right\}. \end{aligned}$$

□

Следствие 3.22. Если СВ ξ_1, ξ_2 – некоррелированные (или независимые), то дисперсия суммы равна сумме дисперсий величин.

Доказательство. В свойстве С7 положим $\rho_{12} = 0$ и получим требуемое. □

3.11. ЭНТРОПИЯ, КОЛИЧЕСТВО ИНФОРМАЦИИ ПО ШЕННОНУ И ДРУГИЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

Определение 3.30. Пусть $\xi = (\xi_i) \in \mathbb{R}^n$ – случайный n -вектор, заданный на вероятностном пространстве (Ω, \mathcal{F}, P) , имеющий плотность распределения $p_\xi(x)$, $x \in \mathbb{R}^n$. Тогда энтропией случайного вектора ξ называется величина

$$H\{\xi\} ::= E\{-\ln p_\xi(\xi)\} = - \int_{\mathbb{R}^n} p_\xi(x) \ln p_\xi(x) dx. \quad (3.82)$$

Если $\xi \in \{a_1, \dots, a_N\}$ – дискретный случайный вектор, тогда

$$H\{\xi\} = - \sum_{k=1}^N P\{\xi = a_k\} \cdot \ln P\{\xi = a_k\}.$$

Свойства энтропии

С1. Энтропия составного случайного вектора: $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbb{R}^{n+m}$ не превосходит суммы энтропий каждого из подвекторов:

$$H\{\xi, \eta\} \leq H\{\xi\} + H\{\eta\}. \quad (3.83)$$

Доказательство. Воспользуемся неравенством Йенсена и вычислим разность между правой и левой частями (3.83) с учетом определения (3.82):

$$H\{\xi\} + H\{\eta\} - H\{\xi, \eta\} = E\left\{-\ln \frac{p_\xi(\xi)p_\eta(\eta)}{p_{\xi,\eta}(\xi, \eta)}\right\}.$$

Функция $z = -\ln(y)$ выпуклая, и по неравенству Йенсена

$$\begin{aligned} E \left\{ -\ln \frac{p_{\xi}(\xi)p_{\eta}(\eta)}{p_{\xi,\eta}(\xi,\eta)} \right\} &\geq -\ln E \left\{ \frac{p_{\xi}(\xi)p_{\eta}(\eta)}{p_{\xi,\eta}(\xi,\eta)} \right\} = \\ &= -\ln \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} p_{\xi,\eta}(x,y) \frac{p_{\xi}(x)p_{\eta}(y)}{p_{\xi,\eta}(x,y)} dx dy = -\ln 1 = 0. \end{aligned}$$

□

С2. Если ξ и η – независимые случайные векторы, то $H\{\xi, \eta\} = H\{\xi\} + H\{\eta\}$.

Доказательство. В доказательстве С1 по критерию независимости $p_{\xi,\eta}(\xi, \eta) = p_{\xi}(\xi)p_{\eta}(\eta)$, и получим равенство. □

Определение 3.31. Количеством информации по Шеннону, содержащейся в случайном векторе ξ о случайном векторе η , называется величина

$$I\{\xi, \eta\} = H\{\xi\} + H\{\eta\} - H\{\xi, \eta\}.$$

Свойства количества информации по Шеннону

С1. $I\{\xi, \eta\} \geq 0$.

Доказательство. Достаточно воспользоваться свойством С1 энтропии. □

Задание 3.4. Если случайный вектор $\xi = (\xi_1, \xi_2)'$ имеет нормальное распределение $N_2(\mu, \Sigma)$, то энтропия равна $H\{\xi_i\} = \ln \sqrt{2\pi e \sigma_{ii}}$, $i = 1, 2$, а количество информации по Шеннону

$$I\{\xi_1, \xi_2\} = \ln \frac{1}{\sqrt{1 - \rho_{12}^2}},$$

где ρ_{12} – коэффициент корреляции ξ_1 и ξ_2 .

С2. $I\{\xi, \eta\}$ – симметричный функционал, т. е.

$$I\{\xi, \eta\} = I\{\eta, \xi\}.$$

С3. Если ξ, η – независимы, то каждая из них несет нулевую информацию о другой случайной величине, т. е. $I\{\xi, \eta\} = 0$.

Доказательство. Достаточно применить свойство С2 энтропии. □

Определение 3.32. Пусть ξ – некоторая случайная величина с плотностью распределения $p_{\xi}(x)$. Тогда точка $M = \arg \max_x p_{\xi}(x)$ называется модой распределения вероятностей.

Определение 3.33. Пусть ξ – СВ с функцией распределения $F(x)$, $0 \leq p \leq 1$. Тогда наименьший корень уравнения $F(z) = p$ обозначается $z_p = F^{-1}(p)$ и называется квантилем уровня p .

Определение 3.34. Квантиль уровня $p = 1/2$ $z_{0,5} = m$ называется медианой распределения вероятностей. Для нее характерно свойство $P\{\xi < m\} = P\{\xi \geq m\} = 1/2$.

Задание 3.5. Показать, что если случайная величина ξ имеет распределение вероятностей $N(\mu, \sigma^2)$, то $M = m = \mu$.

3.12. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ И ЕЕ СВОЙСТВА

Определение 3.35. Пусть $\xi = \xi(\omega) : \Omega \rightarrow \mathbb{R}$ – произвольная случайная величина, заданная на вероятностном пространстве (Ω, \mathcal{F}, P) и имеющая функцию распределения $F_{\xi}(x) = P\{\xi < x\}$, $x \in \mathbb{R}$. Тогда характеристической функцией случайной величины ξ называется комплекснозначная функция действительной переменной t , задаваемая соотношением

$$f_{\xi}(t) = \mathbf{E}\{e^{it\xi}\} = \int_{-\infty}^{+\infty} e^{itx} dF_{\xi}(x), \quad t \in \mathbb{R}, \quad f_{\xi}(t) \in \mathbb{C}. \quad (3.84)$$

В соотношении (3.84) i – мнимая единица, интеграл в правой части – интеграл Лебега – Стильеса.

Замечание 3.16. Математическое ожидание в (3.84) вычисляется с помощью формулы Эйлера

$$\mathbf{E}\{e^{it\xi}\} = \mathbf{E}\{\cos(t\xi)\} + i\mathbf{E}\{\sin(t\xi)\}.$$

Замечание 3.17. Для практических вычислений удобны формулы:

$$f_{\xi}(t) = \sum_{k=1}^N e^{ita_k} \cdot P\{\xi = a_k\}, \quad (3.85)$$

если $\xi \in \{a_1, \dots, a_N\}$ – дискретная случайная величина, $N \leq +\infty$;

$$f_{\xi}(t) = \int_{-\infty}^{+\infty} e^{itx} p_{\xi}(x) dx, \quad (3.86)$$

если ξ – абсолютно непрерывная случайная величина с плотностью распределения вероятностей $p_{\xi}(x) = F'_{\xi}(x)$.

Замечание 3.18. Из формулы (3.86) видно, что $f_{\xi}(t)$ – преобразование Фурье от плотности распределения вероятностей.

Свойства характеристической функции

C1. Для основных законов распределения вероятностей характеристическая функция имеет следующий вид:

а) для случайной величины Бернулли: $\mathcal{L}\{\xi\} = Bi(1, p)$, согласно (3.85), имеем $f_{\xi}(t) = e^{it \cdot 0}(1-p) + e^{it \cdot 1} \cdot p = 1 + p(e^{it} - 1)$, $t \in \mathbb{R}$;

б) $\mathcal{L}(\xi) = Bi(K, p) \Rightarrow f_{\xi}(t) = (1 + p(e^{it} - 1))^K$;

в) $\mathcal{L}(\xi) = \Pi(\lambda) \Rightarrow f_{\xi}(t) = e^{\lambda(e^{it} - 1)}$;

г) $\mathcal{L}(\xi) = N(0, 1) \Rightarrow f_{\xi}(t) = e^{-\frac{1}{2}t^2}$, $p_{\xi}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$;

д) $\mathcal{L}(\xi) = \mathbf{E}(\lambda) \Rightarrow f_{\xi}(t) = \frac{1}{1 - \frac{1}{\lambda}it}$, $t \in \mathbb{R}$.

Задание 3.6. Вычислить характеристические функции для основных законов распределения вероятностей, пользуясь формулами (3.85), (3.86), и внести в таблицу основных законов.

C2. $f_{\xi}(0) = 1$, и характеристическая функция ограничена следующим образом:

$$|f_{\xi}(t)| \leq 1 = f_{\xi}(0), \quad \forall t \in \mathbb{R}. \quad (3.87)$$

Доказательство. Положим $t = 0$ в (3.84): $f_{\xi}(0) = \mathbf{E}\{e^{i \cdot 0}\} = 1$. Оценим левую часть (3.87). По свойству математического ожидания имеем $|\mathbf{E}\{e^{it\xi}\}| \leq \mathbf{E}\{|e^{it\xi}|\} = 1$. \square

Следствие 3.23. Характеристическая функция существует для любой случайной величины.

- С3. При комплексном сопряжении характеристическая функция изменяется следующим образом:

$$\overline{f_{\xi}(t)} = f_{\xi}(-t) = f_{-\xi}(t), \quad \forall t \in \mathbb{R}. \quad (3.88)$$

Доказательство. Согласно формуле (3.84) и замечанию 3.16 имеем $\overline{f_{\xi}(t)} = \mathbf{E}\{e^{-it\xi}\} \equiv \mathbf{E}\{e^{i(-t)\xi}\} \equiv \mathbf{E}\{e^{it(-\xi)}\}$. \square

Следствие 3.24. Если характеристическая функция действительна, то она четна.

Доказательство. Мнимая часть равна нулю, следовательно, согласно (3.88) $\overline{f_{\xi}(t)} = f_{\xi}(t) = f_{\xi}(-t)$. \square

- С4. При линейном преобразовании $\eta = a + b\xi$, где a, b – любые постоянные, характеристическая функция изменяется следующим образом:

$$f_{a+b\xi}(t) = e^{ita} f_{\xi}(bt). \quad (3.89)$$

Доказательство. Из (3.84) и свойств математического ожидания имеем:

$$\begin{aligned} \text{левая часть (3.89)} &= \mathbf{E}\{e^{it(a+b\xi)}\} = \mathbf{E}\{e^{ita} \cdot e^{i(bt)\xi}\} = \\ &= \text{правая часть (3.89)}. \end{aligned}$$

\square

Замечание 3.19. $\mathcal{L}\{\xi\} = N(a, \sigma^2) \Rightarrow \xi = a + \sigma \cdot \overset{\circ}{\xi}$, где $\overset{\circ}{\xi} \sim N(0, 1)$.

Используя С4, имеем $f_{\xi}(t) = e^{ita} \cdot e^{-\frac{1}{2}(\sigma t)^2} = e^{ita - \frac{1}{2}t^2\sigma^2}$.

- С5. Пусть ξ_1, \dots, ξ_n – независимые в совокупности случайные величины, определенные на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) . Характеристическая функция суммы $\eta = \xi_1 + \dots + \xi_n$ равна произведению их характеристических функций:

$$f_{\xi_1 + \dots + \xi_n}(t) = \prod_{k=1}^n f_{\xi_k}(t), \quad t \in \mathbb{R}. \quad (3.90)$$

Доказательство. Воспользуемся формулой (3.84) и свойством мультипликативности математического ожидания, тогда

$$\text{левая часть (3.90)} = \mathbf{E}\{e^{it(\xi_1 + \dots + \xi_n)}\} \equiv \mathbf{E}\left\{\prod_{k=1}^n e^{it\xi_k}\right\}.$$

Поскольку все сомножители независимы в совокупности, то далее имеем

$$\text{левая часть (3.90)} = \prod_{k=1}^n \mathbf{E}\{e^{it\xi_k}\} = \text{правая часть (3.90)}.$$

\square

Следствие 3.25. Если вдобавок ξ_1, \dots, ξ_n одинаково распределены с одной и той же характеристической функцией $f_{\xi}(t)$, то

$$f_{\xi_1 + \dots + \xi_n}(t) = (f_{\xi}(t))^n, \quad t \in \mathbb{R}. \quad (3.91)$$

Замечание 3.20. С помощью (3.91) можно вывести характеристическую функцию для $Bi(K, p)$, используя характеристическую функцию для распределения Бернулли.

С6. Характеристическая функция равномерно непрерывна на всей числовой прямой:

$$\forall \epsilon > 0, \exists \delta = \delta(\epsilon) : |f_{\xi}(t + \delta) - f_{\xi}(t)| \leq \epsilon, \quad \forall t \in \mathbb{R}. \quad (3.92)$$

Доказательство. Выберем $\forall \epsilon > 0$, $t \in \mathbb{R}$, и пока произвольные $\delta > 0$, $x^* > 0$ имеем цепочку неравенств

$$\begin{aligned} |f_{\xi}(t + \delta) - f_{\xi}(t)| &= \left| \int_{-\infty}^{+\infty} e^{itx} (e^{i\delta x} - 1) dF_{\xi}(x) \right| \leq \int_{-\infty}^{+\infty} |e^{itx}| \times \\ &\times |e^{i\delta x} - 1| dF_{\xi}(x) = \int_{|x| < x^*} |e^{i\delta x} - 1| dF_{\xi}(x) + \\ &+ \int_{|x| \geq x^*} |e^{i\delta x} - 1| dF_{\xi}(x) \leq I(\delta, x^*) + 2P\{|\xi| \geq x^*\}. \end{aligned}$$

Поскольку $2P\{|\xi| \geq x^*\} = 2(1 - F_{|\xi|}(x^*))$, $F_{|\xi|}(+\infty) = 1$, то найдется $x^* = x^*(\epsilon) : 2(1 - F_{|\xi|}(x^*)) \leq \epsilon/2$.

При фиксированном x^* и $\delta \rightarrow 0 : I(\delta, x^*) \rightarrow 0$. Следовательно, найдется $\delta = \delta_1(\epsilon, x^*(\epsilon)) =: \delta(\epsilon) : I(\delta, x^*) \leq \epsilon/2$. В результате получаем неравенство (3.92). \square

С7. (Связь характеристической функции с моментами.) Пусть для некоторого натурального $k \in \mathbb{N}$ существует ограниченный начальный момент k -го порядка: $\nu_k = \mathbf{E}\{\xi^k\}$, т. е.

$$\mathbf{E}\{|\xi|^k\} = \int_{-\infty}^{+\infty} |x|^k dF_{\xi}(x) < +\infty. \quad (3.93)$$

Тогда характеристическая функция $f_{\xi}(t)$ непрерывно дифференцируема k раз, причем выполняется следующее соотношение:

$$f_{\xi}^{(k)}(0) = i^k \cdot \nu_k. \quad (3.94)$$

Доказательство. По условию, $\mathbf{E}|\xi|^k < \infty$, поэтому из неравенства Ляпунова следует, что $\mathbf{E}|\xi|^2 < \infty$ при $r < k$. Поскольку

$$\frac{f_{\xi}(t+h) - f_{\xi}(t)}{h} = \mathbf{E} \left\{ e^{it\xi} \left(\frac{e^{ih\xi} - 1}{h} \right) \right\}$$

и при $h \rightarrow 0 \frac{e^{ih\xi} - 1}{h} \rightarrow i\xi$, то из неравенства $\left| e^{it\xi} \frac{e^{ih\xi} - 1}{h} \right| \leq |\xi|$ и теоремы Лебега о сходимости мажорируемой последовательности имеем

$$\mathbf{E} \left\{ e^{it\xi} \left(\frac{e^{ih\xi} - 1}{h} \right) \right\} \xrightarrow{h \rightarrow 0} \mathbf{E} \{ i\xi e^{it\xi} \}.$$

Отсюда получаем

$$\frac{f_{\xi}(t+h) - f_{\xi}(t)}{h} \xrightarrow{h \rightarrow 0} f'_{\xi}(t) = i \mathbf{E} \{ \xi e^{it\xi} \},$$

что совпадает с (3.93) при $k = 1$.

Существование производных $f^{(r)}(t)$, $1 < r \leq k$, и справедливость равенства (3.91) устанавливаются по индукции.

Данное доказательство обосновывает формальное дифференцирование соотношения (3.84) k раз по параметру t :

$$f_{\xi}^{(k)}(t) = \int_{-\infty}^{+\infty} (ix)^k e^{itx} dF_{\xi}(x), \quad t \in \mathbb{R}. \quad (3.95)$$

Непрерывность функции $f_{\xi}^{(k)}(t)$ доказывается, как в свойстве С6. Положим в (3.95) $t = 0$, в результате получим (3.94). \square

Следствие 3.26. *Справедлива следующая удобная формула для вычисления начального момента k -го порядка:*

$$\nu_k = \frac{1}{i^k} f_{\xi}^{(k)}(0), \quad k \in \mathbb{N}. \quad (3.96)$$

Задание 3.7. С помощью (3.96) найти моменты произвольного порядка для гауссовской случайной величины: $N(a, b)$.

Следствие 3.27. *Если конечен момент ν_k , то в окрестности точки $t = 0$ характеристическая функция допускает следующее разложение в виде степенного ряда:*

$$f_{\xi}(t) = 1 + \sum_{j=1}^k \frac{(it)^j}{j!} \cdot \nu_j + o(|t|^k).$$

Доказательство. Достаточно воспользоваться формулой Тейлора и формулой (3.94). \square

С8. (Теорема обращения для характеристической функции.)

Теорема 3.15. *Пусть $\xi = \xi(\omega)$ – произвольная случайная величина на (Ω, \mathcal{F}, P) , имеющая некоторую функцию распределения $F_{\xi}(x)$ и характеристическую функцию $f_{\xi}(t)$. Пусть $C(F_{\xi}) \subseteq \mathbb{R}$ – множество точек непрерывности функции распределения $F_{\xi}(\cdot)$ и $x_1, x_2 \in C(F_{\xi})$, $x_1 < x_2$. Тогда справедлива следующая формула обращения:*

$$F_{\xi}(x_2) - F_{\xi}(x_1) = \lim_{\tau \rightarrow +\infty} \frac{1}{2\pi} \int_{-\tau}^{+\tau} \frac{e^{-itx_1} - e^{-itx_2}}{it} f_{\xi}(t) dt. \quad (3.97)$$

Доказательство. Из определения характеристической функции $f_{\xi}(t)$ имеем цепочку равенств:

$$\begin{aligned} Q(T) &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} f_{\xi}(t) dt = \frac{1}{2\pi} \int_{-T}^T \left(\frac{e^{-itx_1} - e^{-itx_2}}{it} \int_{-\infty}^{+\infty} e^{itx} dF_{\xi}(x) \right) dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-T}^T \frac{e^{i(x-x_1)t} - e^{i(x-x_2)t}}{it} dt \right) dF_{\xi}(x) = \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-T}^T \frac{\sin t(x-x_1) - \sin t(x-x_2)}{t} dt \right) dF_{\xi}(x) = \end{aligned}$$

$$= \frac{1}{\pi} \int_{-\infty}^{+\infty} \left(\int_0^T \frac{\sin t(x-x_1) - \sin t(x-x_2)}{t} dt \right) dF_{\xi}(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \left(\int_{T(x-x_2)}^{T(x-x_1)} \frac{\sin u}{u} du \right) dF_{\xi}(x).$$

Поскольку

$$\lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin u}{u} du = \pi,$$

то

$$\psi(x) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{T(x-x_2)}^{T(x-x_1)} \frac{\sin u}{u} du = \begin{cases} 0, & x < x_1 \vee x > x_2, \\ \frac{1}{2}, & x = x_1 \vee x = x_2, \\ 1, & x_1 < x < x_2. \end{cases}$$

Отсюда следует

$$\begin{aligned} \lim_{T \rightarrow \infty} Q(T) &= \int_{-\infty}^{+\infty} \psi(x) dF_{\xi}(x) = E\{\psi(\xi)\} = \frac{1}{2} P\{\xi = x_1\} + \\ &+ \frac{1}{2} P\{\xi = x_2\} + P\{x_1 < \xi < x_2\}. \end{aligned}$$

Поскольку x_1, x_2 – точки непрерывности функции $F_{\xi}(x)$, то

$$P\{\xi = x_1\} = 0, \quad P\{\xi = x_2\} = 0 \text{ и } \lim_{T \rightarrow \infty} Q(T) = P\{x_1 < \xi < x_2\} = F_{\xi}(x_2) - F_{\xi}(x_1).$$

□

Замечание 3.21. Формула (3.97) обратна к (3.84). Действительно, в (3.84) приращению $dF_{\xi}(\cdot)$ ставится в соответствие характеристическая функция $f_{\xi}(\cdot)$, а в (3.97) функции $f_{\xi}(\cdot)$ – приращение $\Delta F_{\xi}(\cdot)$.

С9. (Теорема единственности для характеристической функции.)

Теорема 3.16. Функция распределения $F_{\xi}(\cdot)$ и характеристическая функция $f_{\xi}(\cdot)$ находятся во взаимно-однозначном соответствии: $F_{\xi}(\cdot) \Leftrightarrow f_{\xi}(\cdot)$.

Доказательство.

- а) Пусть задана $F_{\xi}(\cdot)$. Тогда характеристическая функция однозначно определяется формулой (3.84).
 б) Пусть задана $f_{\xi}(\cdot)$. Тогда воспользуемся формулой обращения (3.97) и разобьем область определения функции распределения $F_{\xi}(\cdot)$ на два непересекающихся подмножества:

$$D(F_{\xi}) = C(F_{\xi}) \cup C_0(F_{\xi}),$$

где $C(F_{\xi})$ – множество точек непрерывности функции распределения; $C_0(F_{\xi})$ – множество точек разрыва.

Рассмотрим два случая:

- 1) $y \in C(F_{\xi})$ – точка непрерывности. В (3.97) устремим $x \rightarrow -\infty$, тогда $F_{\xi}(-\infty) = 0$. По (3.97) имеем

$$F_{\xi}(y) = \lim_{x \rightarrow -\infty} \lim_{\tau \rightarrow +\infty} \frac{1}{2\pi} \int_{-\tau}^{+\tau} \frac{e^{-itx} - e^{-ity}}{it} f_{\xi}(t) dt. \quad (3.98)$$

Таким образом, функция распределения восстанавливается на $C(F_{\xi})$.

- 2) $z \in C_0(F_\xi)$, тогда воспользуемся свойством непрерывности слева: $F_\xi(z) ::= F_\xi(z - 0)$. Тем самым восстановим функцию распределения всюду. \square

Замечание 3.22. Справедливо взаимно-однозначное соответствие: $P_\xi(\cdot) \Leftrightarrow F_\xi(\cdot) \Leftrightarrow f_\xi(\cdot)$.

Следствие 3.28. Если случайная величина $\xi = \xi(\omega)$ имеет абсолютно непрерывное распределение вероятностей с некоторой плотностью распределения $p_\xi(x)$, $x \in \mathbb{R}$, то плотность распределения вероятностей $p_\xi(\cdot)$ и характеристическая функция $f_\xi(\cdot)$ находятся во взаимно-однозначном соответствии, и это взаимно-однозначное соответствие задается парой преобразования Фурье:

$$f_\xi(t) = \int_{-\infty}^{+\infty} e^{itx} p_\xi(x) dx =: \Phi\{p_\xi(\cdot)\}, \quad t \in \mathbb{R}; \quad (3.99)$$

$$p_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} f_\xi(t) dt =: \Phi^{-1}\{f_\xi(\cdot)\}, \quad x \in \mathbb{R}. \quad (3.100)$$

Доказательство. Соотношение (3.99) – это фактически (3.86), а (3.100) получается дифференцированием (3.98). \square

C10. (Теорема Бохнера – Хинчина.)

Теорема 3.17. Пусть $f(t)$, $t \in \mathbb{R}$, – произвольная комплекснозначная непрерывная ограниченная функция такая, что $f(0) = 1$. Для того чтобы функция $f(t)$ являлась характеристической функцией некоторой случайной величины, необходимо и достаточно, чтобы эта функция была неотрицательно определенной, т. е. для $\forall n \in \mathbb{N}$, $\forall t_1, \dots, t_n \in \mathbb{R}$, $\forall \lambda_1, \dots, \lambda_n \in \mathbb{C}$ была неотрицательна следующая квадратичная форма:

$$Q = \sum_{j,k=1}^n f(t_j - t_k) \lambda_j \cdot \overline{\lambda_k} \geq 0. \quad (3.101)$$

Доказательство. Необходимость. Пусть $f(t) = E\{e^{it\xi}\}$ – характеристическая функция случайной величины ξ . Покажем, что выполняется (3.101). Подставим в левую часть (3.101) это выражение и проведем эквивалентные преобразования, используя свойства математического ожидания:

$$\begin{aligned} Q &= \sum_{j,k=1}^n E\{e^{i(t_j - t_k)\xi}\} \lambda_j \cdot \overline{\lambda_k} = \\ &= E\left\{ \sum_{j=1}^n e^{it_j \xi} \lambda_j \cdot \overline{\sum_{k=1}^n e^{it_k \xi} \lambda_k} \right\} = E\left\{ \left| \sum_{j=1}^n e^{it_j \xi} \lambda_j \right|^2 \right\} \geq 0. \end{aligned}$$

Достаточность. Провести самостоятельно. \square

C11. (Свойство Марцинкевича.) Пусть для некоторой случайной величины ξ характеристическая функция имеет вид $f_\xi(t) = e^{P_m(t)}$, $t \in \mathbb{R}$, где $P_m(t)$ – некоторый многочлен степени m от переменной t . Тогда необходимо, чтобы $m \leq 2$.

С характеристической функцией тесно связаны еще два понятия: кумулянтная функция и производящая функция.

Определение 3.36. Кумулянтной функцией случайной величины ξ называется функция $\Phi_\xi(t) = \ln f_\xi(t)$. (В некоторых книгах можно встретить название «семинвариантная функция».)

Задание 3.8. Рассмотреть свойства кумулянтной функции, вытекающие из свойств характеристической функции.

Определение 3.37. Пусть ξ – некоторая неотрицательная целочисленная случайная величина $\xi \in \{0, 1, 2, \dots\}$ с некоторым дискретным распределением вероятностей: $p_k = P\{\xi = k\}$, $k \in \{0, 1, \dots\}$. Производящей функцией случайной величины ξ называется комплекснозначная функция комплексной переменной $z \in \mathbb{C}$, определенная в круге $|z| \leq 1$ в виде функционального ряда следующим образом:

$$g_\xi(z) = E\{z^\xi\} = \sum_{k=0}^{+\infty} z^k \cdot p_k.$$

Следствие 3.29. Производящая и характеристическая функции связаны соотношением $f_\xi(t) = g_\xi(e^{it})$, $t \in \mathbb{R}$.

3.13. УПРАЖНЕНИЯ

1. Для каких случайных величин (дискретных, непрерывных) $E\{1/X\} = 1/E\{X\}$?
2. Доказать, что $D\{\xi\} = E\{D\{\xi|\eta\}\} + D\{E\{\xi|\eta\}\}$.
3. Пусть ξ_1, \dots, ξ_n – независимые, одинаково распределенные случайные величины, причем существует $E\{1/\xi_1\}$; $S_m = \xi_1 + \dots + \xi_m$. Доказать, что если $m \leq n$, то

$$E\left\{\frac{S_m}{S_n}\right\} = \frac{m}{n}.$$

4. Пусть $\xi \geq 0$ – неотрицательная случайная величина с плотностью распределения вероятностей $f(x)$. Доказать, что $E\{\xi^r\} = \int_0^\infty r x^{r-1} P\{\xi > x\} dx$ для любого $r \geq 1$ такого, что указанное математическое ожидание существует.
5. Доказать, что математическое ожидание μ , медиана m , дисперсия σ^2 непрерывной случайной величины ξ удовлетворяют неравенству $(\mu - m)^2 \leq \sigma^2$.
6. Пусть $\mathcal{L}\{\xi\} = N(\mu, \sigma^2)$. Доказать, что $E\{(\xi - \mu)g(\xi)\} = \sigma^2 E\{g'(\xi)\}$, когда обе части существуют.
7. Пусть ξ_1, ξ_2 – независимые одинаково распределенные случайные величины, $\mathcal{L}\{\xi_1\} = R[0, 1]$; $\eta_+ = \max(\xi_1, \xi_2)$, $\eta_- = \min(\xi_1, \xi_2)$. Вычислить $E\{\eta_\pm\}$, $D\{\eta_\pm\}$, $\text{cov}\{\eta_+, \eta_-\}$.
8. Пусть ξ_1, ξ_2 – независимые случайные величины с конечными дисперсиями; $\eta_1 = \xi_1 + \xi_2$, $\eta_2 = \xi_1 \xi_2$. При каком условии η_1 и η_2 некоррелированы?
9. Пусть $\text{corr}\{\xi_1, \xi_2\} = \rho$. Доказать, что $E\{D\{\xi_2|\xi_1\}\} \leq (1 - \rho^2) D\{\xi_2\}$.
10. Пусть ξ_1, ξ_2 имеют двумерное нормальное распределение с нулевыми средними, дисперсиями σ_1^2, σ_2^2 и коэффициентом корреляции ρ . Доказать:
 - а) $E\{\xi_1|\xi_2 = y\} = \frac{\rho\sigma_1}{\sigma_2} y$; $D\{\xi_1|\xi_2 = y\} = \sigma_1^2(1 - \rho^2)$;
 - б) $E\{\xi_1|\xi_1 + \xi_2 = z\} = \frac{(\sigma_1^2 + \rho\sigma_1\sigma_2)z}{(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)}$; $D\{\xi_1|\xi_1 + \xi_2 = z\} = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)}$.

11. Пусть ξ_1, ξ_2 – независимые одинаково распределенные случайные величины, $\mathcal{L}\{\xi_1\} = N(0, 1)$; $\eta = \xi_1 + \xi_2$. Найти $\mathcal{L}\{\eta | \xi_1 > 0, \xi_2 > 0\}$ и показать, что

$$E\{\eta | \xi_1 > 0, \xi_2 > 0\} = 2\sqrt{2/\pi}.$$

12. Пусть ξ_1, ξ_2, ξ_3 имеют трехмерное нормальное распределение с нулевыми средними, единичными дисперсиями и коэффициентами корреляции ρ_1, ρ_2, ρ_3 . Показать, что

$$P\{\xi_1 > 0, \xi_2 > 0, \xi_3 > 0\} = \frac{1}{8} + \frac{1}{4\pi}(\arcsin \rho_1 + \arcsin \rho_2 + \arcsin \rho_3).$$

13. Найти зависимые случайные величины ξ_1, ξ_2 , для которых выполняется следующее соотношение между характеристическими функциями:

$$f_{\xi_1 + \xi_2}(t) = f_{\xi_1}(t)f_{\xi_2}(t), \quad t \in \mathbb{R}.$$

14. Если $f(t)$ – некоторая характеристическая функция, то показать справедливость неравенств:

$$\operatorname{Re}(1 - f(t)) \geq \frac{1}{4}\operatorname{Re}(1 - f(2t));$$

$$1 - |f(2t)| \leq 8(1 - |f(t)|).$$

15. Производящей функцией кумулянтов $K_\xi(\theta)$ случайной величины ξ называется логарифмом производящей функции моментов: $K_\xi(\theta) = \log E\{e^{\theta\xi}\}$. Если она конечна в окрестности точки $\theta = 0$, то $K_\xi(\cdot)$ допускает представление в виде сходящегося ряда Тейлора:

$$K_\xi(\theta) = \sum_{n=1}^{\infty} \frac{1}{n!} k_n(\xi) \theta^n.$$

При этом $k_n(\xi)$ называется n -м кумулянтom или семинвариантом случайной величины ξ . Выразить кумулянты $k_1(\xi), k_2(\xi), k_3(\xi)$ через моменты $E\{\xi\}, E\{\xi^2\}, E\{\xi^3\}$. Доказать, что для независимых случайных величин ξ_1, ξ_2 кумулянты складываются: $k_n(\xi_1 + \xi_2) = k_n(\xi_1) + k_n(\xi_2)$.

16. Используя определения из предыдущего упражнения, доказать, что для $\mathcal{L}\{\xi\} = N(a, b)$ $k_1(\xi) = a, k_2(\xi) = b, k_n(\xi) = 0, n \geq 3$.
17. Доказать, что если $f(\cdot)$ – некоторая характеристическая функция, то $\bar{f}, f^2, |f|^2, \operatorname{Re} f$ – также характеристические функции, но $|f|$ в общем случае не является характеристической функцией.
18. Доказать, что нормальное, Пуассона и гамма-распределения – безгранично делимые.
19. Доказать, что если $f(t)$ – характеристическая функция безгранично делимого закона, то $f(t) \neq 0 \forall t \in \mathbb{R}$.
20. Доказать неравенство для характеристической функции: $|1 - f_\xi(t)| \leq E\{|t\xi|\}, t \in \mathbb{R}$.
21. Пусть случайная величина ξ принимает конечное число неотрицательных значений x_1, \dots, x_n . Доказать, что

$$\lim_{n \rightarrow \infty} \frac{E\{\xi^{n+1}\}}{E\{\xi^n\}} = \max(x_1, \dots, x_n), \quad \lim_{n \rightarrow \infty} \sqrt[n]{E\{\xi^n\}} = \max(x_1, \dots, x_n).$$

22. Пусть случайная величина ξ принимает целые неотрицательные значения и $E\{\xi\} < \infty$. Показать, что

$$E\{\xi\} = \sum_{i \geq 1} P\{\xi \geq i\}.$$

23. Пусть случайная величина ξ принимает значения x_1, \dots, x_m , а случайная величина η принимает значения y_1, \dots, y_n . Кроме того, случайные величины ξ^i и η^j при $i = 1, \dots, m-1$ и $j = 1, \dots, n-1$ некоррелированы, т. е. $E\{\xi^i \eta^j\} = E\{\xi^i\}E\{\eta^j\}$. Показать, что случайные величины ξ и η независимы.

24. Пусть ξ и η – независимые случайные величины, принимающие неотрицательные целые значения, $E\{\xi\} < \infty$. Доказать, что

$$E\{\min(\xi, \eta)\} = \sum_{i \geq 1} P\{\xi \geq i\}P\{\eta \geq i\}.$$

25. Показать, что для любой константы c

$$E\{\xi - c\}^2 \geq E\{\xi - E\{\xi\}\}^2 = D\{\xi\}.$$

26. Показать, что для любых чисел a и b

$$E\{\eta - a\xi - b\}^2 \geq (1 - \rho^2)D\eta, \quad \rho = \text{corr}\{\xi, \eta\}.$$

27. Пусть случайные величины ξ_1, ξ_2 независимы, одинаково распределены и имеют конечные вторые моменты. Показать, что случайные величины $\eta_1 = \xi_1 + \xi_2$ и $\eta_2 = \xi_1 - \xi_2$ некоррелированы.

28. Пусть ξ – случайная величина, $0 < D\{\xi\} < \infty$. Показать, что

$$P\left\{-3, 2 < \frac{\xi - E\{\xi\}}{\sqrt{D\{\xi\}}} < 3, 2\right\} > 0, 9.$$

29. Случайные величины ξ_1, ξ_2 независимы и распределены по нормальному закону $N(\mu, \sigma^2)$. Показать, что

$$E\{\max(\xi_1, \xi_2)\} = \mu + \frac{\sigma}{\sqrt{n}}, \quad E\{\min(\xi_1, \xi_2)\} = \mu - \frac{\sigma}{\sqrt{n}}.$$

30. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и равномерно распределены на $[0, 1]$, v – случайная величина, равная тому k , при котором впервые сумма

$$S_k = \xi_1 + \xi_2 + \dots + \xi_k$$

превосходит 1. Доказать, что $Ev = e$.

31. Доказать, что при сложении независимых случайных величин третьи центральные моменты суммируются, а четвертые – нет.

32. Пусть $\xi_1, \xi_2, \dots, \xi_n$ – случайные величины, имеющие конечные математические ожидания. Доказать, что

$$E\{\max(\xi_1, \dots, \xi_n)\} \geq \max(E\{\xi_1\}, \dots, E\{\xi_n\}),$$

$$E\{\min(\xi_1, \dots, \xi_n)\} \leq \min(E\{\xi_1\}, \dots, E\{\xi_n\}).$$

33. Пусть ξ и η – случайные величины, $P\{\xi > 0\} = P\{\eta > 0\} = 3/4$, $P\{\xi + \eta > 0\} = 1/2$. Доказать, что случайные величины ξ и η зависимы.

34. Доказать, что если случайные величины ξ и η принимают по два значения, то из равенства $E\{\xi\eta\} = E\{\xi\}E\{\eta\}$ следует независимость ξ и η .

35. Доказать, что для любых случайных величин ξ и η , имеющих конечные дисперсии, справедливы неравенства:

$$(\sqrt{D\{\xi\}} - \sqrt{D\{\eta\}})^2 \leq D\{\xi + \eta\} \leq (\sqrt{D\{\xi\}} + \sqrt{D\{\eta\}})^2.$$

36. Случайные величины ξ и η независимы и одинаково распределены. Доказать, что если указанные ниже математические ожидания существуют, то

$$E\left\{\frac{\xi}{\xi + \eta}\right\} = E\left\{\frac{\eta}{\xi + \eta}\right\} = \frac{1}{2}.$$

37. Пусть $P\{|\xi| \leq C\} = 1$. Доказать, что $D\{\xi\} \leq CE\{|\xi|\}$.

38. Случайные величины ξ и η независимы и имеют конечные дисперсии. Доказать, что $D\{\xi\eta\} \geq D\{\xi\}D\{\eta\}$.

39. Пусть $\xi \geq 0$ и $E\frac{1}{\xi^\alpha} < \infty$ при некотором $\alpha > 0$. Доказать, что $\frac{F_\xi(x)}{x^\alpha} \rightarrow 0$ при $x \rightarrow 0$.

40. Случайная величина ξ имеет конечное математическое ожидание. Доказать, что для любого x

$$\max(x, E\{\xi\}) \leq E\{\max(x, \xi)\}.$$

41. Пусть ν, ξ_1, ξ_2, \dots – независимые случайные величины, а случайные величины ξ_1, ξ_2, \dots имеют функцию распределения $F(x)$, случайная величина ν распределена по закону Пуассона с параметром λ . Доказать, что характеристическая функция случайной величины $\eta = \xi_1 + \dots + \xi_\nu$ равна

$$\exp\left\{\lambda \int_{-\infty}^{+\infty} (e^{itu} - 1) dF(u)\right\}.$$

42. Пусть ξ_1 и ξ_2 – независимые случайные величины, имеющие геометрическое распределение соответственно с параметрами p_1 и p_2 . Показать, что случайная величина $\xi = \min(\xi_1, \xi_2)$ имеет геометрическое распределение, и найти параметр этого распределения.

43. Производящие функции. Пусть ξ_1 и ξ_2 – независимые случайные величины, принимающие неотрицательные значения, и $\xi_1 + \xi_2$ имеет биномиальное распределение. Показать, что каждая из величин ξ_1 и ξ_2 имеет биномиальное распределение.

44. Пусть $\mathcal{L}\{\xi\} = N(\mu, \sigma^2)$. Показать, что

$$\text{corr}\{\xi, F_\xi(\xi)\} = \sqrt{3/\pi}.$$

45. *Контрпример.* Пусть ξ_1, ξ_2 – н. о. р. гауссовские СВ, $\mathcal{L}\{\xi_1\} = \mathcal{L}\{\xi_2\} = N_1(0, 1)$. Показать, что СВ $\eta_1 = \xi_1, \eta_2 = \xi_1 \cdot \text{sign}(\xi_2)$ также имеют гауссовские распределения $N_1(0, 1)$ и некоррелированы, однако их совместное распределение не является гауссовским.

Глава 4

СЛУЧАЙНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ И ИХ СХОДИМОСТЬ

4.1. ОПРЕДЕЛЕНИЕ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ И ВИДОВ ЕЕ СХОДИМОСТИ

Ранее мы рассматривали случайные эксперименты, в которых каждому исходу $\omega \in \Omega$ ставилось в соответствие конечное множество числовых величин $\xi_1, \dots, \xi_N \in \mathbb{R}$, $1 \leq N < +\infty$. При этом математической моделью таких результатов экспериментов было понятие случайного вектора:

$$\xi = \xi(\omega) = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_N \end{pmatrix} \in \mathbb{R}^N.$$

Однако на практике N часто не удается ограничить. Например, исследуем динамику курса обмена валют. В таком случае наблюдается последовательность $\xi_1, \dots, \xi_k, \dots \in \mathbb{R}$, где ξ_k – обменный курс в k -й торговый день. Удобной математической моделью для таких случаев является понятие случайной последовательности.

Определение 4.1. Случайной последовательностью $\xi_k = \xi_k(\omega) \in \mathbb{R}$, $\omega \in \Omega$, $k = 1, 2, \dots$, определенной на вероятностном пространстве (Ω, \mathcal{F}, P) , называется счетное параметрическое семейство случайных величин $\xi_1 = \xi_1(\omega)$, $\xi_2 = \xi_2(\omega)$, \dots , заданных на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) ; при этом параметр $k \in \mathbb{N}$ этого семейства пробегает множество натуральных чисел.

Основной вопрос, связанный со случайными последовательностями, – это вопрос их сходимости. В теории вероятностей существуют четыре основных вида сходимости.

Определение 4.2. Пусть $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, – некоторая случайная последовательность, определенная на (Ω, \mathcal{F}, P) , $\xi = \xi(\omega)$ – некоторая случайная величина, которая может быть определена и на другом пространстве. Пусть далее $F_{\xi_k}(x)$, $F_\xi(x)$, $x \in \mathbb{R}$, – соответствующие функции распределения, а $C(F_\xi) \subseteq \mathbb{R}$ – множество точек непрерывности функции $F_\xi(\cdot)$. Принято говорить, что случайная последовательность ξ_k при $k \rightarrow \infty$ сходится к случайной величине ξ по распределению (Distribution), и принято кратко обозначать это следующим образом:

$$\xi_k \xrightarrow[k \rightarrow \infty]{D} \xi,$$

если имеет место сходимость последовательности функций распределения, т. е. если выполняется соотношение

$$\lim_{k \rightarrow \infty} F_{\xi_k}(x) = F_\xi(x), \quad x \in C(F_\xi).$$

Задание 4.1. Пусть $\mathcal{L}\{\xi_k\} = N(a + 1/k, b)$, $\mathcal{L}\{\xi\} = N(a, b)$. Показать, что имеет место сходимость по распределению.

Определение 4.3. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены случайная последовательность $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, и случайная величина $\xi = \xi(\omega)$. Принято говорить, что случайная последовательность ξ_k

при $k \rightarrow \infty$ сходится к случайной величине ξ по вероятности (Probability), и этот факт принято кратко обозначать

$$\xi_k \xrightarrow[k \rightarrow \infty]{P} \xi,$$

если

$$\forall \epsilon > 0 \quad \lim_{k \rightarrow \infty} P\{|\xi_k - \xi| > \epsilon\} = 0.$$

Задание 4.2. Пусть $\xi_k = \xi + k_{-1}\eta$, $|\eta| \leq c < +\infty$. Показать, что имеет место сходимость по вероятности.

Определение 4.4. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определена случайная последовательность $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, и случайная величина $\xi = \xi(\omega)$. Пусть $r \in \mathbb{N}$ – произвольное фиксированное натуральное число. Принято говорить, что случайная последовательность ξ_k при $k \rightarrow \infty$ сходится к случайной величине ξ в среднем порядка r , и принято кратко обозначать

$$\xi_k \xrightarrow[k \rightarrow \infty]{L_r} \xi,$$

если ограничены абсолютные моменты порядка r :

$$E\{|\xi_k|^r\} < +\infty, \quad E\{|\xi|^r\} < +\infty$$

и выполняется предельное соотношение

$$\lim_{k \rightarrow \infty} E\{|\xi_k - \xi|^r\} = \lim_{k \rightarrow \infty} \int_{\Omega} |\xi_k(\omega) - \xi(\omega)|^r P(d\omega) = 0.$$

Определение 4.5. Если $r = 2$, то L_2 -сходимость называется сходимостью в среднем квадратическом и обозначается

$$\xi_k \xrightarrow[k \rightarrow \infty]{\text{ср. кв.}} \xi \quad \text{либо} \quad \text{l.i.m.}_{k \rightarrow \infty} \xi_k = \xi \quad (\text{limit in the mean}).$$

Определение 4.6. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены случайная последовательность $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, и случайная величина $\xi = \xi(\omega)$. Принято говорить, что случайная последовательность ξ_k при $k \rightarrow \infty$ сходится к случайной величине ξ почти наверное (почти всюду), и этот факт принято кратко обозначать

$$\xi_k \xrightarrow[k \rightarrow \infty]{\text{п. н.}} \xi, \quad \text{либо} \quad \xi_k \xrightarrow[k \rightarrow \infty]{P=1} \xi, \quad \text{либо} \quad \xi_k \xrightarrow[k \rightarrow \infty]{} \xi (P - \text{п. в.}),$$

если выполняется предельное соотношение

$$P\left\{\omega : \lim_{k \rightarrow \infty} \xi_k(\omega) = \xi(\omega)\right\} = 1.$$

4.2. СХОДИМОСТЬ ПОЧТИ НАВЕРНОЕ. ЗАКОН $0 \vee 1$ БОРЕЛЯ

Имеется ряд эквивалентных определений для сходимости почти наверное, введенной в п. 4.1.

Теорема 4.1. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определены случайная последовательность $\xi_k = \xi_k(\omega)$, случайная величина $\xi = \xi(\omega)$ и для $\forall \epsilon > 0$ последовательность случайных событий $A_k^\epsilon = \{\omega : |\xi_k(\omega) - \xi(\omega)| > \epsilon\}$.

Тогда справедлива следующая цепочка эквивалентных определений сходимости почти наверное:

$$\begin{aligned} \xi_k \xrightarrow[k \rightarrow \infty]{\text{п. н.}} \xi &\Leftrightarrow P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n^{\epsilon}\right) = 0, \quad \epsilon > 0 \Leftrightarrow P\left(\overline{\lim_{k \rightarrow \infty} A_k^{\epsilon}}\right) = 0, \quad \epsilon > 0 \Leftrightarrow \\ &\Leftrightarrow \lim_{k \rightarrow \infty} P\left\{\sup_{n \geq k} |\xi_n - \xi| > \epsilon\right\} = 0, \quad \epsilon > 0. \end{aligned}$$

Доказательство. Последовательно докажем рассматриваемую цепочку определений.

1) Докажем, что $\xi_k \xrightarrow[k \rightarrow \infty]{\text{п. н.}} \xi \Leftrightarrow P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n^{\epsilon}\right) = 0$. Введем в рассмотрение случайное событие

$$\Omega_* = \left\{ \omega : \lim_{k \rightarrow \infty} \xi_k(\omega) = \xi(\omega) \right\} \in \mathcal{F}.$$

По определению сходимости почти наверное

$$\xi_k \xrightarrow[k \rightarrow \infty]{\text{п. н.}} \xi \Leftrightarrow P(\Omega_*) = 1, \quad \text{или} \quad P(\overline{\Omega_*}) = 0.$$

Исследуем подробнее событие Ω_* . Для любого $\epsilon > 0$ имеем

$$\begin{aligned} \Omega_* &\subseteq \{ \omega : \exists K = K(\epsilon, \omega), \forall n \geq K \quad |\xi_n - \xi| \leq \epsilon \} = \\ &= \left\{ \omega : \exists K = K(\epsilon, \omega), \omega \in \bigcap_{n=K}^{\infty} \overline{A_n^{\epsilon}} \right\}. \end{aligned} \quad (4.1)$$

Перебирая всевозможные $K = 1, 2, \dots$ и учитывая (4.1), имеем

$$\Omega_* \subseteq \bigcup_{K=1}^{\infty} \bigcap_{n=K}^{\infty} \overline{A_n^{\epsilon}}. \quad (4.2)$$

Применим к соотношению (4.2) между событиями правило де Моргана:

$$\overline{\Omega_*} \supseteq \bigcap_{K=1}^{\infty} \bigcup_{n=K}^{\infty} A_n^{\epsilon}.$$

Поэтому по свойствам вероятности следует

$$P(\overline{\Omega_*}) \geq P\left(\bigcap_{K=1}^{\infty} \bigcup_{n=K}^{\infty} A_n^{\epsilon}\right) \geq 0.$$

Отсюда $P(\overline{\Omega_*}) = 0$ тогда и только тогда, когда

$$P\left(\bigcap_{K=1}^{\infty} \bigcup_{n=K}^{\infty} A_n^{\epsilon}\right) = 0, \quad \epsilon > 0.$$

2) Равносильность равенств $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n^{\epsilon}\right) = 0$ и $P\left(\overline{\lim_{k \rightarrow \infty} A_k}\right) = 0$ вытекает из п. 1 по определению верхнего предела последовательности случайных событий (см. гл. 1).

3) Для доказательства равносильности равенств $P\left(\overline{\lim}_{k \rightarrow \infty} A_k^\epsilon\right) = 0, \epsilon > 0 \Leftrightarrow \Leftrightarrow \lim_{k \rightarrow \infty} P\left\{\sup_{n \geq k} |\xi_n - \xi| > \epsilon\right\} = 0$ введем вспомогательную последовательность случайных событий: $B_k^\epsilon = \bigcup_{n=k}^{\infty} A_n^\epsilon$. По построению очевидно, что B_k^ϵ — монотонно убывающая последовательность, значит, $\exists \lim_{k \rightarrow \infty} B_k^\epsilon = \bigcap_{k=1}^{\infty} B_k^\epsilon$. Воспользуемся аксиомой непрерывности вероятностной меры:

$$0 = P\left(\lim_{k \rightarrow \infty} B_k^\epsilon\right) = \lim_{k \rightarrow \infty} P(B_k^\epsilon).$$

Наконец, заметим, что

$$B_k^\epsilon \equiv \left\{\omega : \sup_{n \geq k} |\xi_n - \xi| > \epsilon\right\}.$$

□

Следствие 4.1. Если для любого $\epsilon > 0$ сходится ряд

$$\sum_{n=1}^{\infty} P\{|\xi_n - \xi| > \epsilon\} < +\infty,$$

то $\xi_n \xrightarrow{\text{п. н.}} \xi$.

Доказательство. В силу последнего утверждения теоремы 4.1 и свойства вероятности имеем

$$\begin{aligned} \lim_{k \rightarrow \infty} P\left\{\sup_{n \geq k} |\xi_n - \xi| > \epsilon\right\} &= \lim_{k \rightarrow \infty} P\left(\bigcup_{n=k}^{\infty} \{|\xi_n - \xi| > \epsilon\}\right) \leq \\ &\leq \lim_{k \rightarrow \infty} \sum_{n=k}^{\infty} P\{|\xi_n - \xi| > \epsilon\} = 0. \end{aligned}$$

Здесь учтено, что остаток сходящегося ряда стремится к нулю. □

Теорема 4.2 (критерий Коши сходимости п. н.). Для того чтобы случайная последовательность $\xi_k = \xi_k(\omega)$ на вероятностном пространстве (Ω, \mathcal{F}, P) сходилась почти наверное, необходимо и достаточно, чтобы эта последовательность была фундаментальной (последовательностью Коши) в смысле сходимости почти наверное:

$$\forall \epsilon > 0 \quad \lim_{k \rightarrow \infty} P\left\{\sup_{m, n \geq k} |\xi_m - \xi_n| > \epsilon\right\} = 0.$$

Доказательство. Результат этой теоремы вытекает из критерия Коши сходимости числовой последовательности и схемы доказательства теоремы 4.1. □

Сформулируем еще два достаточных признака сходимости почти наверное.

Теорема 4.3 (I достаточный признак сходимости п. н.). Если при произвольном $\epsilon > 0$ с вероятностью 1 наступает лишь конечное число случайных событий среди $A_n = \{\omega : |\xi_n(\omega) - \xi(\omega)| > \epsilon\}, n \in \mathbb{N}$, то имеет место сходимость почти наверное, т. е. $\xi_n \xrightarrow{\text{п. н.}} \xi$.

Доказательство. Воспользуемся теоремой об эквивалентных определениях сходимости почти наверное. Рассмотрим предельное случайное событие

$$\overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n =$$

$$= \{ \text{наступает бесконечно много случайных событий среди } \{A_n\} \}.$$

Тогда по условию доказываемой теоремы $P\{\overline{\lim}_{n \rightarrow \infty} A_n\} = 0$. Из теоремы следует, что $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} \xi$. \square

При исследовании сходимости приходится иметь дело с событиями

$$\{A_n : n \in \mathbb{N}\} \text{ и } A = \overline{\lim}_{n \rightarrow \infty} A_n.$$

Возникает проблема: как, зная $\{P(A_n)\}$, оценить $P(A)$?

Лемма 4.1 (Бореля – Кантелли). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена произвольная последовательность случайных событий $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, и предельное случайное событие $A = \overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \in \mathcal{F}$. Тогда справедливы следующие два утверждения:

- 1) если сходится ряд $\sum_{n=1}^{\infty} P(A_n)$, то $P(A) = 0$;
- 2) если этот ряд расходится и вдобавок события $\{A_n\}$ независимы в совокупности, то $P(A) = 1$.

Доказательство. Докажем два утверждения по отдельности.

- 1) Оценим сверху вероятность предельного события, используя свойства вероятности.

По построению предельного события A очевидно, что $A \subseteq \bigcup_{n=k}^{\infty} A_n, \forall k \in \mathbb{N}$. Отсюда в силу свойств вероятности имеем

$$P(A) \leq \sum_{n=k}^{\infty} P(A_n). \quad (4.3)$$

Правая часть (4.3) – остаток сходящегося ряда, следовательно, $\sum_{n=k}^{\infty} P(A_n) \xrightarrow{k \rightarrow \infty} 0$. Значит, выбором k эта сумма может быть сделана сколь угодно малой, поэтому $P(A) = 0$.

- 2) Оценим сверху вероятность противоположного события, используя независимость событий и свойства вероятности:

$$P(\overline{A}) = P\left(\bigcup_{k=1}^{\infty} \left(\bigcap_{n=k}^{\infty} \overline{A_n}\right)\right) \leq \sum_{k=1}^{\infty} P\left(\bigcap_{n=k}^{\infty} \overline{A_n}\right) =$$

$$= \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} (1 - P(A_n)) \leq \sum_{k=1}^{\infty} e^{-\sum_{n=k}^{\infty} P(A_n)} = 0.$$

Здесь использовано известное неравенство: $1 - p \leq e^{-p}, p \geq 0$. Следовательно, $P(\overline{A}) = 0$, значит, $P(A) = 1$. \square

Следствие 4.2 (закон $0 \vee 1$ Бореля). Пусть на (Ω, \mathcal{F}, P) определена произвольная последовательность независимых в совокупности случайных событий $\{A_n\}$ и предельное случайное событие

$$A = \overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \\ = \{\text{наступает бесконечно много случайных событий среди } \{A_n\}\}.$$

Тогда $P(A) = 0 \vee 1$ в зависимости от того, сходится или расходится ряд, составленный из вероятностей этих событий.

Теорема 4.4 (II достаточный признак сходимости п. н.). Для того чтобы случайная последовательность $\xi_k = \xi_k(\omega)$ на вероятностном пространстве (Ω, \mathcal{F}, P) имела предел в смысле почти наверное, достаточно, чтобы существовала такая неотрицательная последовательность $\epsilon_k \geq 0$, что имеет место сходимость двух рядов:

$$S_1 = \sum_{k=1}^{\infty} \epsilon_k < +\infty, \\ S_2 = \sum_{k=1}^{\infty} P\{|\xi_{k+1} - \xi_k| > \epsilon_k\} < +\infty.$$

Доказательство. Введем в рассмотрение вспомогательную последовательность случайных событий:

$$A_k = \{\omega : |\xi_{k+1} - \xi_k| > \epsilon_k\}, \quad k = 1, 2, \dots,$$

и введем предельное событие:

$$A = \overline{\lim}_{k \rightarrow \infty} A_k.$$

Воспользуемся первым утверждением леммы Бореля – Кантелли. В силу условий теоремы ряд S_2 сходится, следовательно, $P(A) = 0 \Rightarrow P(\overline{A}) = 1$.

Для $\forall \omega \in \overline{A}$: $P(\overline{A}) = 1$ запишем тождество

$$\xi_k \equiv \xi_1 + \sum_{n=1}^{k-1} (\xi_{n+1} - \xi_n); \quad (4.4)$$

$\forall \omega \in \overline{A} \Rightarrow \exists N : \forall n \geq N$ наступает \overline{A}_n , т. е. $|\xi_{n+1} - \xi_n| \leq \epsilon_n$. По условию доказываемой теоремы ряд S_1 сходится, т. е. в правой части (4.4) члены ряда мажорируются элементами сходящегося ряда, следовательно, ряд справа сходится, т. е. существует предел при $k \rightarrow \infty$. Значит, существует предел левой части (4.4).

Таким образом, для $\forall \omega \in \overline{A}$, $P(\overline{A}) = 1$, $\exists \lim_{k \rightarrow \infty} \xi_k(\omega)$. Иначе говоря, предел существует на множестве меры 1, следовательно, по определению последовательность ξ_k сходится почти наверное. \square

4.3. СХОДИМОСТЬ ПО ВЕРОЯТНОСТИ

Вначале установим соотношения между сходимостью по вероятности и сходимостью почти наверное.

Теорема 4.5. Из сходимости почти наверное вытекает сходимость по вероятности.

Доказательство. Пусть $\xi_m \xrightarrow{п. н.} \xi$. Тогда по теореме об эквивалентных определениях сходимости почти наверное имеем

$$\forall \epsilon > 0 \quad P\left\{\sup_{m \geq n} |\xi_m - \xi| > \epsilon\right\} = P\left(\bigcup_{m \geq n} \{|\xi_m - \xi| > \epsilon\}\right) \xrightarrow{n \rightarrow \infty} 0.$$

Отсюда по свойствам вероятности

$$\forall \epsilon > 0 \quad P\{|\xi_n - \xi| > \epsilon\} \leq P\left(\bigcup_{m \geq n} \{|\xi_m - \xi| > \epsilon\}\right) \xrightarrow{n \rightarrow \infty} 0,$$

что и означает сходимость по вероятности: $\xi_n \xrightarrow{P} \xi$. \square

Теперь сформулируем и докажем критерий Коши сходимости по вероятности: $\xi_n \xrightarrow{P} \xi$. Для этого нам понадобится вспомогательное утверждение.

Лемма 4.2. Для любых случайных величин ξ , η , ζ на (Ω, \mathcal{F}, P) и любого $\epsilon > 0$ справедливо неравенство

$$P\{|\xi - \eta| > \epsilon\} \leq P\left\{|\xi - \zeta| > \frac{\epsilon}{2}\right\} + P\left\{|\zeta - \eta| > \frac{\epsilon}{2}\right\}. \quad (4.5)$$

Доказательство. Поскольку $|\xi - \zeta| \equiv |(\xi - \zeta) + (\zeta - \eta)| \leq |\xi - \zeta| + |\zeta - \eta|$, то имеем следующее соотношение между случайными событиями:

$$\left\{|\xi - \zeta| \leq \frac{\epsilon}{2}\right\} \cap \left\{|\zeta - \eta| \leq \frac{\epsilon}{2}\right\} \subseteq \{|\xi - \eta| \leq \epsilon\}.$$

По правилу де Моргана имеем

$$\{|\xi - \eta| > \epsilon\} \subseteq \left\{|\xi - \zeta| > \frac{\epsilon}{2}\right\} \cup \left\{|\zeta - \eta| > \frac{\epsilon}{2}\right\}.$$

Отсюда по свойствам вероятности получаем (4.5). \square

Теорема 4.6 (критерий Коши P -сходимости). Для того чтобы случайная последовательность $\xi_k = \xi_k(\omega)$, $\omega \in \Omega$, $k \in \mathbb{N}$, имела бы предел по вероятности на вероятностном пространстве (Ω, \mathcal{F}, P) , необходимо и достаточно, чтобы эта последовательность была фундаментальной (последовательностью Коши) по вероятности:

$$\forall \epsilon > 0 \quad \lim_{m, n \rightarrow \infty} P\{|\xi_m - \xi_n| > \epsilon\} = 0. \quad (4.6)$$

Доказательство. Соотношение (4.6) означает, что для любых ϵ_1 , $\epsilon_2 > 0$ найдется число $N = N(\epsilon_2)$ такое, что

$$\forall m, n \geq N \quad P\{|\xi_m - \xi_n| > \epsilon_1\} \leq \epsilon_2. \quad (4.7)$$

Необходимость. Пусть $\xi_n \xrightarrow{P} \xi$. По определению это означает

$$\forall \epsilon > 0 \quad P\{|\xi_n - \xi| > \epsilon\} \xrightarrow{n \rightarrow \infty} 0. \quad (4.8)$$

Тогда из (4.8) и предыдущей леммы получаем

$$P\{|\xi_m - \xi_n| > \epsilon\} \leq P\left\{|\xi_m - \xi| > \frac{\epsilon}{2}\right\} + P\left\{|\xi_n - \xi| > \frac{\epsilon}{2}\right\} \xrightarrow{m, n \rightarrow \infty} 0,$$

что и влечет соотношение (4.6).

Достаточность. Пусть выполняется формула (4.6) или, что эквивалентно, соотношение (4.7). Построим монотонно возрастающую последовательность натуральных чисел $n_1, n_2, \dots \in \mathbb{N}$ следующим рекуррентным образом ($k = 2, 3, \dots$):

$$n_1 = 1, n_k = \min \left(n : n > n_{k-1}, \forall m, l \geq n P \left\{ |\xi_m - \xi_l| > \frac{1}{k^2} \right\} \leq \frac{1}{k^2} \right). \quad (4.9)$$

Такая последовательность $\{n_k\}$ существует, так как выполняется (4.7).

С помощью $\{n_k\}$ построим вспомогательную случайную подпоследовательность $\{\xi_{n_k}\}$. В силу (4.9) для нее выполняется свойство

$$\forall k \in \mathbb{N} \quad P \left\{ |\xi_{n_{k+1}} - \xi_{n_k}| > \frac{1}{k^2} \right\} \leq \frac{1}{k^2}. \quad (4.10)$$

Выбирая в (4.10) $\epsilon_k = 1/k^2$ и применяя II достаточный признак сходимости почти наверное (теорему 4.4), получаем $\xi_{n_k} \xrightarrow[k \rightarrow \infty]{\text{п. н.}} \xi$.

Тогда по теореме 4.5 имеем

$$\xi_{n_k} \xrightarrow[k \rightarrow \infty]{P} \xi. \quad (4.11)$$

Теперь из (4.11), вышедшей из леммы 4.2 и (4.7) находим

$$\forall \epsilon > 0 \quad P\{|\xi_n - \xi| > \epsilon\} \leq P\left\{|\xi_n - \xi_{n_k}| > \frac{\epsilon}{2}\right\} + P\left\{|\xi_{n_k} - \xi| > \frac{\epsilon}{2}\right\} \xrightarrow[n, n_k \rightarrow \infty]{} 0,$$

что согласно (4.8) и означает $\xi_n \xrightarrow{P} \xi$. \square

Теорема 4.7 (о единственности предела). Если $\xi_n \xrightarrow{P} \xi$ и $\xi_n \xrightarrow{P} \eta$, то $\xi \stackrel{\text{п. н.}}{=} \eta$, так что $P\{\xi \neq \eta\} = 0$.

Доказательство. В силу леммы 4.2 и условий данной теоремы для любого $\epsilon > 0$ при $n \rightarrow \infty$ имеем

$$P\{|\xi - \eta| > \epsilon\} \leq P\left\{|\xi - \xi_n| > \frac{\epsilon}{2}\right\} + P\left\{|\eta - \xi_n| > \frac{\epsilon}{2}\right\} \rightarrow 0.$$

Таким образом, $\forall \epsilon > 0 \quad P\{|\xi - \eta| > \epsilon\} = 0$. Тогда в силу следствия аксиомы непрерывности

$$\begin{aligned} P\{\xi \neq \eta\} &= P\left(\bigcup_{k=1}^{\infty} \left\{|\xi - \eta| > \frac{1}{k}\right\}\right) = P\left(\overline{\lim_{k \rightarrow \infty}} \left\{\omega : |\xi - \eta| > \frac{1}{k}\right\}\right) = \\ &= \lim_{k \rightarrow \infty} P\left\{|\xi - \eta| > \frac{1}{k}\right\} = 0. \end{aligned}$$

\square

Теорема 4.8. Из всякой случайной последовательности $\xi_n = \xi_n(\omega)$, $n \in \mathbb{N}$, сходящейся на вероятностном пространстве (Ω, \mathcal{F}, P) по вероятности, можно выделить подпоследовательность $\xi_{n_k} = \xi_{n_k}(\omega)$, $k \in \mathbb{N}$, сходящуюся почти наверное.

Доказательство. Из критерия Коши сходимости по вероятности имеем

$$\forall \epsilon > 0 \quad \lim_{r, s \rightarrow \infty} P\{|\xi_r - \xi_s| > \epsilon\} = 0.$$

Воспользуемся этим свойством и построим последовательность отбираемых номеров n_k следующим рекуррентным способом:

$$n_1 = 1, n_k = \min \left\{ N : N > n_{k-1}, \forall r, s > N, P \left\{ |\xi_r - \xi_s| > \frac{1}{2^k} \right\} < \frac{1}{2^k} \right\}, \\ k = 2, 3, \dots$$

Для исследования сходимости $\{\xi_{n_k}\}$ воспользуемся II достаточным признаком сходимости почти наверное:

$$\begin{aligned} \epsilon_k = \frac{1}{2^k}, \quad S_1 = \sum_{k=1}^{\infty} \frac{1}{2^k} < +\infty, \quad S_2 = \sum_{k=1}^{\infty} P \left\{ \left| \xi_{n_{k+1}} - \xi_{n_k} \right| > \frac{1}{2^k} \right\} \leq \\ \leq \sum_{k=1}^{\infty} \frac{1}{2^k} < +\infty. \end{aligned}$$

Следовательно, II достаточный признак выполняется, а из этого вытекает сходимость почти наверное. \square

Теорема 4.9. Пусть на (Ω, \mathcal{F}, P) определены $K \geq 1$ случайных последовательностей, сходящихся по вероятности к константам ($n \rightarrow \infty$):

$$\xi_n^{(1)} \xrightarrow{P} a_1, \quad \xi_n^{(2)} \xrightarrow{P} a_2, \dots, \xi_n^{(K)} \xrightarrow{P} a_K. \quad (4.12)$$

Если функция $g(x_1, \dots, x_K) : \mathbb{R}^K \rightarrow \mathbb{R}^1$ непрерывна в точке $(a_1, \dots, a_K) \in \mathbb{R}^K$, то при $n \rightarrow \infty$

$$\eta_n = g(\xi_n^{(1)}, \dots, \xi_n^{(K)}) \xrightarrow{P} g(a_1, \dots, a_K). \quad (4.13)$$

Доказательство. Свойство непрерывности функции $g(\cdot)$ в точке (a_1, \dots, a_K) означает, что для любого $\epsilon > 0$ найдется такое $\delta = \delta(\epsilon, a_1, \dots, a_K) > 0$, что как только $|x_i - a_i| \leq \delta$, $i = 1, \dots, K$, так $|g(x_1, \dots, x_K) - g(a_1, \dots, a_K)| \leq \epsilon$. Отсюда следует:

$$\bigcap_{i=1}^K \{ \xi_n^{(i)} - a_i \leq \delta \} \subset \{ |g(\xi_n^{(1)}, \dots, \xi_n^{(K)}) - g(a_1, \dots, a_K)| \leq \epsilon \}.$$

Переходя к противоположным событиям, по правилу де Моргана имеем

$$\{ |g(\xi_n^{(1)}, \dots, \xi_n^{(K)}) - g(a_1, \dots, a_K)| > \epsilon \} \subset \bigcup_{i=1}^K \{ |\xi_n^{(i)} - a_i| > \delta \}.$$

Отсюда по свойствам вероятности из (4.12) заключаем:

$$P \{ |g(\xi_n^{(1)}, \dots, \xi_n^{(K)}) - g(a_1, \dots, a_K)| > \epsilon \} \leq \sum_{i=1}^K P \{ |\xi_n^{(i)} - a_i| > \delta \} \rightarrow 0,$$

что означает (4.13). \square

Теорема 4.10. Пусть на (Ω, \mathcal{F}, P) определены случайные величины $\xi^{(1)}, \dots, \xi^{(K)}$ и $K \geq 1$ сходящихся по вероятности случайных последовательностей ($n \rightarrow \infty$): $\xi_n^{(1)} \xrightarrow{P} \xi^{(1)}, \xi_n^{(2)} \xrightarrow{P} \xi^{(2)}, \dots, \xi_n^{(K)} \xrightarrow{P} \xi^{(K)}$. Если функция $g(x_1, \dots, x_K) : \mathbb{R}^K \rightarrow \mathbb{R}^1$ равномерно-непрерывна в \mathbb{R}^K , то

$$\eta_n = g(\xi_n^{(1)}, \dots, \xi_n^{(K)}) \xrightarrow{P} g(\xi^{(1)}, \dots, \xi^{(K)}).$$

Доказательство. По определению равномерной непрерывности $\forall \epsilon > 0 \exists \delta = \delta(\epsilon), |x_i - y_i| \leq \delta, i = 1, \dots, K \Rightarrow |g(x_1, \dots, x_K) - g(y_1, \dots, y_K)| \leq \epsilon$. Далее доказательство осуществляется, как в теореме 4.9, полагая $x_i = \xi_n^{(i)}, y_i = \xi^{(i)}, i = 1, \dots, K$. \square

Замечание 4.1. Справедливы аналоги теорем 4.9, 4.10, в которых вместо сходимости по вероятности рассматривается сходимость почти наверное.

4.4. СХОДИМОСТЬ В СРЕДНЕМ

Теорема 4.11 (критерий Коши L_r -сходимости). Для того чтобы случайная последовательность $\xi_k = \xi_k(\omega)$ на вероятностном пространстве (Ω, \mathcal{F}, P) сходилась в среднем порядка r , необходимо и достаточно, чтобы она была фундаментальной (последовательностью Коши) в смысле L_r -сходимости, т. е. чтобы

$$E\{|\xi_n|^r\} < \infty, \quad n = 1, 2, \dots; \quad \lim_{m, n \rightarrow \infty} E\{|\xi_m - \xi_n|^r\} = 0. \quad (4.14)$$

Доказательство. Необходимость. Пусть $\xi_n \xrightarrow{L_r} \xi$. По определению L_r -сходимости это означает, что

$$E\{|\xi_n|^r\} < \infty, \quad n = 1, 2, \dots; \quad E\{|\xi|^r\} < \infty; \quad E\{|\xi_n - \xi|^r\} \xrightarrow{n \rightarrow \infty} 0. \quad (4.15)$$

Тогда по неравенству Минковского

$$E\{|\eta + \zeta|^r\} \leq \left((E\{|\eta|^r\})^{\frac{1}{r}} + (E\{|\zeta|^r\})^{\frac{1}{r}} \right)^r$$

находим ($\eta = \xi_m - \xi$, $\zeta = \xi - \xi_n$):

$$E\{|\xi_m - \xi_n|^r\} \leq \left((E\{|\xi_m - \xi|^r\})^{\frac{1}{r}} + (E\{|\xi - \xi_n|^r\})^{\frac{1}{r}} \right)^r \xrightarrow{m, n \rightarrow \infty} 0,$$

так что выполняется соотношение (4.14).

Достаточность доказывается сложнее [36], при помощи схемы доказательства теоремы 4.8 из п. 4.3. \square

Теорема 4.12. Если $q < r$, то из сходимости в среднем порядка r вытекает сходимость низшего порядка q .

Доказательство. По условию теоремы выполняется (4.15). В силу неравенства Ляпунова

$$(E\{|\xi_n|^q\})^{\frac{1}{q}} \leq (E\{|\xi_n|^r\})^{\frac{1}{r}}, \quad (E\{|\xi_n - \xi|^q\})^{\frac{1}{q}} \leq (E\{|\xi_n - \xi|^r\})^{\frac{1}{r}},$$

поэтому имеем

$$E\{|\xi_n|^q\} < \infty, \quad n = 1, 2, \dots; \quad E\{|\xi_n - \xi|^q\} \xrightarrow{n \rightarrow \infty} 0.$$

Это означает, что $\xi_n \xrightarrow{L_q} \xi$. \square

Теорема 4.13. Из сходимости в среднем порядка r вытекает сходимость по вероятности.

Доказательство. По условию теоремы выполняется (4.15). В силу неравенства Чебышева

$$\forall \epsilon > 0 \quad P\{|\xi_n - \xi| > \epsilon\} \leq \frac{1}{\epsilon^r} E\{|\xi_n - \xi|^r\} \rightarrow 0.$$

По определению P -сходимости это означает, что $\xi_n \xrightarrow{P} \xi$. \square

Теорема 4.14. Если $\xi_n \xrightarrow{P} \xi$ и существует постоянная $C > 0$ такая, что $|\xi_n| \leq C$ почти наверное ($P\{|\xi_n| \leq C\} = 1$) для $n = 1, 2, \dots$, то для любого $r \geq 1$

$$\xi_n \xrightarrow[n \rightarrow \infty]{L_r} \xi.$$

Доказательство. Покажем ограниченность почти наверное предельной случайной величины ξ : $P\{|\xi| \leq C\} = 1$. Выберем произвольное $\epsilon > 0$. Тогда в силу условий теоремы имеем

$$\begin{aligned} P\{|\xi| > C + \epsilon\} &= P(\{|\xi| > C + \epsilon\} \cap \{|\xi_n - \xi| > \epsilon\}) + \\ &+ P(\{|\xi| > C + \epsilon\} \cap \{|\xi_n - \xi| \leq \epsilon\}) \leq P\{|\xi_n - \xi| > \epsilon\} + P\{|\xi_n| > C\} = \\ &= P\{|\xi_n - \xi| > \epsilon\} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Отсюда следует, что

$$\forall \epsilon > 0 \quad P\{|\xi| > C + \epsilon\} = 0.$$

Тогда из аксиомы непрерывности вероятностной меры имеем

$$\begin{aligned} P\{|\xi| > C\} &= P\left(\bigcap_{k=1}^{\infty} \left\{|\xi| > C + \frac{1}{k}\right\}\right) = P\left(\lim_{k \rightarrow \infty} \left\{|\xi| > C + \frac{1}{k}\right\}\right) = \\ &= \lim_{k \rightarrow \infty} P\left\{|\xi| > C + \frac{1}{k}\right\} = 0. \end{aligned}$$

Таким образом, $|\xi| \leq C$ почти наверное.

Определим вспомогательную последовательность случайных событий $A_n = \{|\xi_n - \xi| > \delta\}$ для некоторого произвольного $\delta > 0$. Тогда по свойствам математического ожидания с учетом доказанного выше имеем

$$\begin{aligned} E\{|\xi_n - \xi|^r\} &\equiv E\{|\xi_n - \xi|^r \mathbb{1}_{A_n} + |\xi_n - \xi|^r \mathbb{1}_{A_n^c}\} = E\{|\xi_n - \xi|^r \mathbb{1}_{A_n}\} + \\ &+ E\{|\xi_n - \xi|^r \mathbb{1}_{A_n^c}\} \leq \delta^r + (2C)^r \cdot P\{|\xi_n - \xi| > \delta\} \xrightarrow{n \rightarrow \infty} \delta^r. \end{aligned}$$

Отсюда в силу произвола $\delta > 0$ заключаем: $E\{|\xi_n - \xi|^r\} \rightarrow 0$, что влечет $\xi_n \xrightarrow{L_r} \xi$. \square

Теорема 4.15. Если $\xi_n \xrightarrow{L_r} \xi$, то для любого целого q , $1 \leq q \leq r$, имеет место сходимость последовательности начальных моментов q -го порядка:

$$E\{\xi_n^q\} \xrightarrow{n \rightarrow \infty} E\{\xi^q\}. \quad (4.16)$$

Доказательство. По теореме 4.12 имеем

$$\xi_n \xrightarrow{L_q} \xi, \quad 1 \leq q \leq r.$$

Схема доказательства (4.16) одна и та же для любого q . Поэтому докажем (4.16) при $q = r$.

Обозначим

$$\eta_n = \xi_n - \xi, \quad E\{|\eta_n|^r\} \rightarrow 0.$$

Тогда по свойствам математического ожидания имеем

$$\begin{aligned} |E\{\xi_n^r\} - E\{\xi^r\}| &= |E\{\xi_n^r - \xi^r\}| = |E\{(\xi + \eta_n)^r - \xi^r\}| \equiv \\ &\equiv \left| E\left\{ \sum_{i=1}^r C_r^i \xi^{r-i} \eta_n^i \right\} \right| \leq \sum_{i=1}^r C_r^i E\{|\xi|^{r-i} |\eta_n|^i\}. \end{aligned} \quad (4.17)$$

Выбирая $p_1 = r/(r-i) > 1$, $p_2 = r/i > 1$ в последней сумме, воспользуемся неравенством Гельдера:

$$E\{|\xi|^{r-i} |\eta_n|^i\} \leq (E\{|\xi|^r\})^{\frac{r-i}{r}} (E\{|\eta_n|^r\})^{\frac{i}{r}}.$$

Подставив эту оценку в (4.17), получим

$$|E\{\xi_m^r\} - E\{\xi^r\}| \leq \sum_{i=1}^r C_r^i (E\{|\xi|^r\})^{\frac{r-i}{r}} (E\{\eta_n^r\})^{\frac{i}{r}} \rightarrow 0.$$

□

Лемма 4.3 (о сходимости в среднем квадратическом). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определены две случайные последовательности, сходящиеся в ср. кв.:

$$\xi_m \xrightarrow[m \rightarrow \infty]{\text{ср. кв.}} \xi, \quad (4.18)$$

$$\eta_n \xrightarrow[n \rightarrow \infty]{\text{ср. кв.}} \eta. \quad (4.19)$$

Тогда существует

$$\lim_{m, n \rightarrow \infty} E\{\xi_m \cdot \eta_n\} = E\{\xi \cdot \eta\}. \quad (4.20)$$

Доказательство. По определению среднеквадратической сходимости из (4.18) и (4.19) имеем

$$E\{\xi_m^2\} < +\infty, \quad E\{\xi^2\} < +\infty, \quad E\{|\xi_m - \xi|^2\} \xrightarrow{m \rightarrow \infty} 0,$$

$$E\{\eta_n^2\} < +\infty, \quad E\{\eta^2\} < +\infty, \quad E\{|\eta_n - \eta|^2\} \xrightarrow{n \rightarrow \infty} 0.$$

Покажем, что математическое ожидание в (4.20) существует. Воспользуемся неравенством Коши – Буняковского:

$$E\{|\xi_m \eta_n|\} \leq \sqrt{E\{\xi_m^2\} \cdot E\{\eta_n^2\}} < +\infty.$$

Аналогично показывается существование математического ожидания в правой части соотношения (4.20).

Оценим разность между левой и правой частями (4.20). Имеем

$$\begin{aligned} |E\{\xi_m \eta_n - \xi \eta\}| &\equiv |E\{(\xi_m - \xi)(\eta_n - \eta) + (\xi_m - \xi)\eta + (\eta_n - \eta)\xi\}| \leq \\ &\leq E\{|(\xi_m - \xi)(\eta_n - \eta)|\} + E\{|(\xi_m - \xi)\eta|\} + E\{|(\eta_n - \eta)\xi|\} \leq \\ &\leq \sqrt{E\{(\xi_m - \xi)^2\} E\{(\eta_n - \eta)^2\}} + \sqrt{E\{(\xi_m - \xi)^2\} E\{\eta^2\}} + \\ &+ \sqrt{E\{(\eta_n - \eta)^2\} E\{\xi^2\}} \xrightarrow{m, n \rightarrow \infty} 0 + 0 + 0 = 0. \end{aligned}$$

□

Теорема 4.16 (критерий L_2 -сходимости). Для того чтобы случайная последовательность ξ_k на (Ω, \mathcal{F}, P) имела предел в среднем квадратическом, необходимо и достаточно, чтобы существовал конечный предел:

$$\lim_{m, n \rightarrow \infty} E\{\xi_m \xi_n\} = A < +\infty. \quad (4.21)$$

Доказательство. Необходимость. Пусть

$$\xi_n \xrightarrow[n \rightarrow \infty]{\text{ср. кв.}} \xi.$$

Тогда воспользуемся леммой 4.3, в которой положим $\eta_n \equiv \xi_n$. Тогда

$$\exists \lim_{m,n \rightarrow \infty} E\{\xi_m \xi_n\} = E\{\xi \cdot \xi\} = A < +\infty.$$

Достаточность. Пусть выполняется (4.21), тогда проверим выполнение критерия Коши:

$$\begin{aligned} \lim_{m,n \rightarrow \infty} E\{(\xi_m - \xi_n)^2\} &= \lim_{m,n \rightarrow \infty} (E\{\xi_m \xi_m\} - 2E\{\xi_m \xi_n\} + \\ &+ E\{\xi_n \xi_n\}) \rightarrow A - 2A + A = 0. \end{aligned}$$

Отсюда следует, что критерий Коши выполняется и, значит, последовательность сходится в среднем квадратическом. \square

Теорема 4.17. Пусть $\xi_k \xrightarrow[k \rightarrow \infty]{\text{ср. кв.}} \xi$, причем настолько быстро, что сходится ряд

$$\sum_{k=1}^{\infty} E\{(\xi_k - \xi)^2\} < +\infty, \quad (4.22)$$

тогда

$$\xi_k \xrightarrow{\text{п. н.}} \xi.$$

Доказательство. Построим вспомогательную последовательность случайных событий $A_k = \{(|\xi_k - \xi| > \epsilon), \forall \epsilon > 0, \epsilon \text{ зафиксировано.}$ Воспользуемся неравенством Чебышева относительно математического ожидания:

$$P(A_k) \leq \frac{1}{\epsilon^2} E\{|\xi_k - \xi|^2\},$$

и исследуем сходимость ряда с учетом (4.22):

$$\sum_{k=1}^{\infty} P(A_k) \leq \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} E\{(\xi_k - \xi)^2\} < +\infty.$$

Тогда по лемме Бореля – Кантелли $P\{\overline{\lim_{n \rightarrow \infty} A_k}\} = 0$, и, следовательно, имеет место сходимость почти на верное. \square

4.5. РАВНОМЕРНАЯ ИНТЕГРИРУЕМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН

Определение 4.7. Последовательность СВ $\{\xi_n\}$ называется равномерно интегрируемой (РИ), если

$$\sup_n \int_{(|\xi_n| > a)} |\xi_n| dP \xrightarrow{a \rightarrow \infty} 0. \quad (4.23)$$

Теорема 4.18. Если для всех n выполняется неравенство $|\xi_n| \leq \eta$, где $E\{\eta\} < \infty$, то последовательность $\{\xi_n\}$ равномерно интегрируема.

Доказательство. Поскольку для всех n $(|\xi_n| > a) \subset (\eta > a)$, то

$$\int_{(|\xi_n| > a)} |\xi_n| dP \leq \int_{(\eta > a)} |\xi_n| dP \leq \int_{(\eta > a)} \eta dP \xrightarrow{a \rightarrow \infty} 0.$$

Отсюда

$$\sup_n \int_{(|\xi_n|>a)} |\xi_n| dP \xrightarrow{a \rightarrow \infty} 0.$$

□

Теорема 4.19. Если для всех $n \leq N$ выполняется неравенство $E\{\xi_n\} < \infty$, то совокупность СВ $\{\xi_n\}_{n \leq N}$ равномерно интегрируема.

Доказательство. Доказательство следует из теоремы 4.18, если взять в качестве η случайную величину:

$$\eta = \sum_{n=1}^N |\xi_n|.$$

□

Теорема 4.20 (критерий равномерной интегрируемости). Для того чтобы последовательность $\{\xi_n\}$ была равномерно интегрируема, необходимо и достаточно, чтобы:

1) последовательность $\{\xi_n\}$ была абсолютно равномерно интегрируема, т. е. для любого $\epsilon > 0$ существовало $\delta = \delta_\epsilon > 0$ такое, что при

$$P(A) \leq \delta \sup_n \int_A |\xi_n| dP \leq \epsilon; \quad (4.24)$$

2) выполнялось условие

$$\sup_n E\{|\xi_n|\} < \infty. \quad (4.25)$$

Доказательство. Необходимость. Пусть выполняется (4.23). Тогда

$$\int_A |\xi_n| dP = \int_{A(|\xi_n| \leq a)} |\xi_n| dP + \int_{A(|\xi_n| > a)} |\xi_n| dP \leq aP(A) + \sup_n \int_{(|\xi_n| > a)} |\xi_n| dP. \quad (4.26)$$

Из (4.23) следует, что для произвольного $\epsilon > 0$ существует a_0 такое, что

$$\sup_n \int_{(|\xi_n| > a_0)} |\xi_n| dP \leq \frac{\epsilon}{2}.$$

Тогда, взяв событие A таким, чтобы $a_0 P(A) \leq \epsilon/2$, получим, что правая часть (4.26) не превосходит ϵ . Поскольку правая часть (4.26) не зависит от n , то, выбрав $\delta = \epsilon/2a_0$, получим

$$\sup_n \int_A |\xi_n| dP \leq \epsilon, \text{ если } P(A) \leq \delta.$$

Соотношение (4.24) доказано.

Для доказательства (4.25) возьмем $A = \Omega$. Тогда из (4.23) для достаточно больших a

$$E\{|\xi_n|\} \leq a + \sup_n \int_{(|\xi_n| > a)} |\xi_n| dP \leq a + 1 < \infty.$$

Отсюда

$$\sup_n E\{|\xi_n|\} < \infty.$$

Необходимость доказана.

Достаточность. Пусть выполняются (4.24) и (4.25). Поскольку

$$P\{|\xi_n| > a\} \leq \frac{1}{a} E\{|\xi_n|\},$$

то из (4.25) имеем

$$\sup_n P\{|\xi_n| > a\} \leq \frac{1}{a} \sup_n E\{|\xi_n|\} \xrightarrow{a \rightarrow \infty} 0.$$

Пусть $\epsilon > 0$, $\delta = \delta(\epsilon) > 0$ такое, что (4.24) верно. Выберем a^* таким, чтобы

$$\sup_n P\{|\xi_n| > a^*\} \leq \delta.$$

Тогда для всех n и $a \geq a^*$

$$\int_{(|\xi_n| > a)} |\xi_n| dP \leq \epsilon.$$

Следовательно, $\sup_n \int_{(|\xi_n| > a)} |\xi_n| dP \leq \epsilon$, т. е. $\sup_n \int_{(|\xi_n| > a)} |\xi_n| dP \xrightarrow{a \rightarrow \infty} 0$. \square

Теорема 4.21. Соотношение $\xi_n \xrightarrow{L_1} \xi$ эквивалентно двум соотношениям: $\xi_n \xrightarrow{P} \xi$, и последовательность $\{\xi_n\}$ равномерно интегрируема.

Доказательство. 1. Пусть $\xi_n \xrightarrow{L_1} \xi$. Это означает, что $E\{|\xi|\} < \infty$, $E\{|\xi_n|\} < \infty$, $E\{|\xi_n - \xi|\} \xrightarrow{n \rightarrow \infty} 0$. Последнее соотношение равносильно условию для любых событий A_n

$$\left| \int_{A_n} \xi_n dP - \int_{A_n} \xi dP \right| \xrightarrow{n \rightarrow \infty} 0. \quad (4.27)$$

Действительно:

$$\left| \int_{A_n} \xi_n dP - \int_{A_n} \xi dP \right| \leq \left| \int_{A_n} |\xi_n - \xi| dP \right| \leq \int_{A_n} |\xi_n - \xi| dP \leq \int |\xi_n - \xi| dP = E\{|\xi_n - \xi|\} \xrightarrow{n \rightarrow \infty} 0.$$

С другой стороны, из (4.27) имеем

$$E\{|\xi_n - \xi|\} = \int_{\Omega} |\xi_n - \xi| dP = \int_{(\xi_n - \xi > 0)} |\xi_n - \xi| dP - \int_{(\xi_n - \xi \leq 0)} (\xi_n - \xi) dP \xrightarrow{n \rightarrow \infty} 0.$$

Поскольку $\xi_n \xrightarrow{L_1} \xi$, то $\xi_n \xrightarrow{P} \xi$. Проверим выполнимость условий (4.24) и (4.25). Из критерия Коши и (4.27) имеем $\int_A |\xi_n - \xi_m| dP \xrightarrow{n, m \rightarrow \infty} 0$ для любого события A . Это означает, что для любого $\epsilon > 0$ существует такое число N , что при всех $n, m \geq N$ и любого события $A \in \mathcal{F}$

$$\int_A |\xi_n - \xi_m| dP \leq \frac{\epsilon}{2}.$$

Тогда для выбранного $\epsilon > 0$, $n \geq N$ и любого события A

$$\int_A |\xi_n| dP = \int_A |\xi_n - \xi_N + \xi_N| dP \leq \int_A |\xi_n - \xi_N| dP + \int_A |\xi_N| dP \leq \frac{\epsilon}{2} + \int_A |\xi_N| dP. \quad (4.28)$$

Поскольку $E\{|\xi_N|\} < \infty$, поэтому правую часть (4.28) можно сделать меньше ϵ , если $P(A) \leq \delta_1$, где $\delta_1 = \delta_1(\epsilon) > 0$.

Поскольку правая часть (4.28) не зависит от n , то

$$\sup_{n \geq N} \int_A |\xi_n| dP \leq \epsilon, \quad \text{если} \quad P(A) \leq \delta_1.$$

По теореме 4.19 совокупность СВ $\{\xi_n\}_{n \leq N}$ равномерно интегрируема, то для некоторого $\delta_2 > 0$

$$\sup_{n < N} \int_A |\xi_n| dP \leq \epsilon, \quad \text{если} \quad P(A) \leq \delta_2.$$

Взяв в качестве $\delta = \min(\delta_1, \delta_2)$, получим

$$\sup_n \int_A |\xi_n| dP \leq \epsilon, \quad \text{если} \quad P(A) \leq \delta.$$

Положив в неравенстве (4.28) $A = \Omega$, получаем

$$\sup_n \int |\xi_n| dP = \sup_n E\{|\xi_n|\} < \infty.$$

2. Пусть $\xi_n \xrightarrow{P} \xi$ и последовательность $\{\xi_n\}$ равномерно интегрируема. Покажем, что $E\{|\xi_n|\} < \infty$. Поскольку $\xi_n \xrightarrow{P} \xi$, то существует последовательность n_k такая, что $\xi_{n_k} \xrightarrow{п. н.} \xi$. Тогда по лемме Фату и (4.25) имеем

$$E\{|\xi|\} = E\left\{\left|\lim_{n_k \rightarrow \infty} \xi_{n_k}\right|\right\} = E\left\{\liminf_{n_k \rightarrow \infty} |\xi_{n_k}|\right\} \leq \liminf_{n_k \rightarrow \infty} E\{|\xi_{n_k}|\} \leq \sup_n E\{|\xi_n|\} < \infty.$$

Используя (4.24), получаем

$$\begin{aligned} E\{|\xi_n - \xi|\} &= \int |\xi_n - \xi| dP = \int_{(|\xi_n - \xi| \leq \epsilon)} |\xi_n - \xi| dP + \int_{(|\xi_n - \xi| > \epsilon)} |\xi_n - \xi| dP \leq \\ &\leq \epsilon + \int_{(|\xi_n - \xi| > \epsilon)} |\xi_n| dP + \int_{(|\xi_n - \xi| > \epsilon)} |\xi| dP \xrightarrow{n \rightarrow \infty} \epsilon. \end{aligned}$$

Отсюда заключаем

$$E\{|\xi_n - \xi|\} \xrightarrow{n \rightarrow \infty} 0.$$

□

4.6. СООТНОШЕНИЯ МЕЖДУ ВИДАМИ СХОДИМОСТИ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Установим соотношения между видами сходимости случайных последовательностей, используя результаты пп. 4.1–4.5.

Лемма 4.4. Для любых случайных событий $A, B \in \mathcal{F}$ справедливо неравенство

$$P(A \cap B) \geq P(A) - P(\overline{B}).$$

Доказательство. Воспользуемся формулой сложения вероятностей:

$$P(A \cup B) = P(A) + 1 - P(\bar{B}) - P(A \cap B).$$

Отсюда имеем

$$P(A \cap B) = P(A) - P(\bar{B}) + (1 - P(A \cup B)) \geq P(A) - P(\bar{B}).$$

□

Теорема 4.22 (соотношения между видами сходимости). *Справедлива схема соотношений между видами сходимости случайных последовательностей, приведенная на рис. 4.1 ($r > q$).*

Доказательство. Проведем доказательство в три этапа в соответствии со схемой на рис. 4.1.

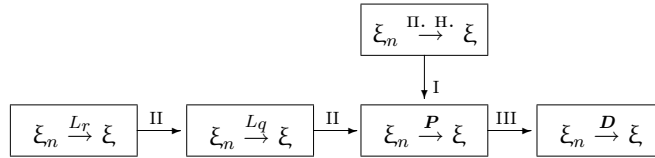


Рис. 4.1. Схема соотношений между видами сходимости

I. Этот результат вытекает из теоремы 4.5.

II. Эти результаты доказаны в теоремах 4.12, 4.13.

III. Пусть $\xi_n \xrightarrow{P} \xi$. Тогда введем для любого фиксированного $\epsilon > 0$ последовательность случайных событий $A_n = \{|\xi_n - \xi| > \epsilon\} \in \mathcal{F}$.

По определению имеем

$$P(A_n) \rightarrow 0, \forall \epsilon > 0. \quad (4.29)$$

По определению D -сходимости надо доказать следующий результат: для произвольной точки непрерывности $x \in C(F_\xi)$ имеет место

$$F_{\xi_n}(x) \xrightarrow{n \rightarrow \infty} F_\xi(x). \quad (4.30)$$

Воспользуемся двумя очевидными соотношениями между событиями:

$$A_n \supseteq \{\xi_n < x\} \cap \{\xi \geq x + \epsilon\}, A_n \supseteq \{\xi_n \geq x\} \cap \{\xi < x - \epsilon\}.$$

Воспользуемся свойством вероятности и леммой 4.4:

$$\begin{aligned} P(A_n) &\geq P(A \cap B) \geq P(A) - P(\bar{B}) = F_{\xi_n}(x) - F_\xi(x + \epsilon), \\ P(A_n) &\geq P(B \cap A) \geq F_\xi(x - \epsilon) - F_{\xi_n}(x). \end{aligned}$$

Разрешим оба неравенства относительно $F_{\xi_n}(x)$:

$$F_\xi(x - \epsilon) - P(A_n) \leq F_{\xi_n}(x) \leq F_\xi(x + \epsilon) + P(A_n).$$

Выберем здесь $x \in C(F_\xi)$, $n \rightarrow \infty$, $\epsilon \rightarrow 0$ и воспользуемся (4.29): $P(A_n) \rightarrow 0$, $\epsilon \rightarrow 0$, x — точка непрерывности, следовательно, существует предел: $\lim_{n \rightarrow \infty} F_{\xi_n}(x)$, причем выполняется соотношение (4.30). □

Приведенная на рис. 4.1 схема соотношений является полной, и без дополнительных условий отсутствуют какие-то другие соотношения. Покажем это примерами.

Пример 4.1. Рассмотрим вероятностное пространство (Ω, \mathcal{F}, P) , где $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}[0, 1]$ – борелевская σ -алгебра подмножеств из Ω , $P(\cdot) = \text{mes}(\cdot)$ – мера Лебега:

$$\xi_n(\omega) = \begin{cases} n^\alpha, & 0 \leq \omega \leq \frac{1}{n}, \\ 0, & \frac{1}{n} < \omega \leq 1, \end{cases}$$

$\alpha > 0$ – некоторый параметр. Очевидно, что $P\{\omega : \xi_n(\omega) \rightarrow 0\} = 1$, поэтому $\xi_n \xrightarrow{p. n.} 0$. Вычислим

$$E\{\xi_n^p\} = n^{\alpha p} \frac{1}{n} + 0 \cdot \left(1 - \frac{1}{n}\right) = n^{\alpha p - 1}.$$

Если $\alpha < 1/p$, то $\xi_n \xrightarrow{L_p} 0$. Если $\alpha \geq 1/p$, то L_p – сходимости отсутствует, т. е. из сходимости почти наверное не следует L_p -сходимость.

Пример 4.2. Пусть ξ, ξ_1, ξ_2, \dots – независимые, одинаково распределенные случайные величины, имеющие стандартную нормальную функцию распределения $\Phi(x)$, $x \in \mathbb{R}$. Поскольку $F_{\xi_n}(x) = F_\xi(x) = \Phi(x)$, то очевидно, что имеет место сходимость по распределению:

$$\xi_n \xrightarrow[n \rightarrow \infty]{D} \xi.$$

В то же время, так как

$$\mathcal{L}\{\xi_1 - \xi\} = N(0, 2),$$

то $\forall \epsilon > 0$

$$P\{|\xi_n - \xi| > \epsilon\} = P\{|\xi_1 - \xi| > \epsilon\} = 2\Phi\left(-\frac{\epsilon}{\sqrt{2}}\right) \not\rightarrow 0 \text{ при } n \rightarrow \infty,$$

так что отсутствует сходимость по вероятности.

Таким образом, в данном примере из сходимости по распределению не следует сходимости по вероятности.

Замечание 4.2. Если случайные величины ξ, ξ_1, ξ_2, \dots с вероятностью 1 являются константами, то для последовательности ξ_n все виды сходимости эквивалентны.

Теорема 4.23. Если $\xi_n \xrightarrow[n \rightarrow \infty]{D} C$, где C – некоторая константа, то $\xi_n \xrightarrow[n \rightarrow \infty]{P} C$, т. е. в случае сходимости к константе виды сходимости по распределению и по вероятности эквивалентны.

Доказательство. По условию теоремы

$$F_{\xi_n}(x) = P\{\xi_n < x\} \xrightarrow[n \rightarrow \infty]{} F(x) = \begin{cases} 0, & x \leq C, \\ 1, & x > C. \end{cases} \quad (4.31)$$

Выберем произвольное $\epsilon > 0$ и с учетом (4.31) исследуем

$$\begin{aligned} P\{|\xi_n - C| > \epsilon\} &= 1 - P\{C - \epsilon \leq \xi_n \leq C + \epsilon\} = 1 - (F_{\xi_n}(C + \epsilon) - \\ &\quad - F_{\xi_n}(C - \epsilon)) \xrightarrow[n \rightarrow \infty]{} 1 - (F(C + 0) - F(C - 0)) = 1 - 1 = 0. \end{aligned}$$

Таким образом, $\xi_n \xrightarrow{P} C$. □

Теорема 4.24. Если случайная последовательность $\{\xi_n\}$ сходится в смысле каких-нибудь двух из трех видов сходимости: по вероятности, почти наверное, в среднем порядка p , то предельные случайные величины с вероятностью 1 совпадают.

Доказательство. Пусть

$$\xi_n \xrightarrow{\text{п. н.}} \xi, \quad \xi_n \xrightarrow{L_p} \eta.$$

Тогда из теоремы 4.22 следует:

$$\xi_n \xrightarrow{P} \xi, \quad \xi_n \xrightarrow{P} \eta.$$

Из теоремы единственности предела последовательности, сходящейся по вероятности (теорема 4.7), заключаем: $P\{\xi = \eta\} = 1$. \square

Замечание 4.3. Проведенное в этой главе изучение видов сходимости позволяет сделать следующие общие выводы. Сходимость почти наверное является характеристикой всей последовательности $\{\xi_n\}$ и основным видом сходимости. Сходимость по вероятности является характеристикой одной случайной величины ξ_n с бесконечно большим номером $n \rightarrow \infty$; ее значения с вероятностью, стремящейся к 1, как угодно мало отличаются от значений предельной случайной величины ξ . В ситуации, когда заранее известно, что $\xi_n \xrightarrow{\text{п. н.}} \xi$, и надо оценить $E\{\xi^p\}$, пользуются сходимостью в среднем порядка p . Если требуется найти лишь вероятностное распределение (функцию распределения) предельной случайной величины ξ , то принято использовать сходимость по распределению.

4.7. УПРАЖНЕНИЯ

1. Доказать, что для всякой непрерывной на отрезке $[0, 1]$ функции $f(x)$ полиномы Бернштейна $B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}$ стремятся при $n \rightarrow \infty$ к $f(x)$.
2. Доказать, что если $f(x)$ – непрерывная и ограниченная функция на $[0, \infty)$, то для $h > 0$

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} f\left(x + \frac{k}{n}\right) \frac{(nh)^k}{k!} e^{-kn} = f(x+h).$$

3. Пусть ξ_1, ξ_2, \dots – последовательность случайных величин, $\delta > 0$. Доказать, что если

$$\sum_{n=1}^{\infty} E\{|\xi_n - \xi|^\delta\} < \infty, \quad \text{то } \xi_n \xrightarrow{\text{п. н.}} \xi.$$

4. Доказать, что если $\xi_n - a_n \xrightarrow{P} 0$, то $m_{\xi_n} - a_n \xrightarrow{n \rightarrow \infty} 0$, где m_{ξ_n} – медиана случайной величины ξ_n , a_n – последовательность действительных чисел.

5. Доказать, что если $\xi_n \xrightarrow{L_p} \xi$, $p \geq 1$, то для любого события $A \in \mathcal{F}$ $\int_A \xi_n dP \rightarrow \int_A \xi dP$.

6. Доказать, что если $\xi_n \xrightarrow{L_p} \xi$, $p \geq 1$, то $|\xi_n|^p \xrightarrow{L_1} |\xi|^p$.

7. Пусть ξ_1, ξ_2, \dots – последовательность независимых случайных величин и $\xi_n \xrightarrow{P} \xi$. Доказать, что ξ имеет вырожденное распределение, т. е. $\xi \stackrel{\text{п. н.}}{=} \text{const}$.

8. Пусть $\xi_n^{(1)} \xrightarrow{\text{п. н.}} \xi_1, \dots, \xi_n^{(k)} \xrightarrow{\text{п. н.}} \xi_k$ и функция $g(x_1, \dots, x_k)$ непрерывна в \mathbb{R}^k . Доказать, что $\eta_n = g(\xi_n^{(1)}, \dots, \xi_n^{(k)}) \xrightarrow{\text{п. н.}} \eta = g(\xi_1, \dots, \xi_k)$.

9. Пусть $r \geq 1$, $c \in \mathbb{R}$ – константа, X, Y – случайные величины, определенные на (Ω, \mathcal{F}, P) , $\|X_r\| = (E\{|X|^r\})^{1/r}$. Показать, что:
- $\|cX\|_r = c\|X\|_r$;
 - $\|X + Y\|_r \leq \|X\|_r + \|Y\|_r$;
 - $\|X\|_r = 0 \Leftrightarrow P\{X = 0\} = 1$.
- Это позволяет сказать, что $\|\cdot\|_r$ – некоторая норма на множестве классов эквивалентности случайных величин с конечным r -м моментом; при этом отношение эквивалентности задается следующим образом: $X \sim Y \Leftrightarrow P\{X = Y\} = 1$.
10. Построить случайные величины X, X_1, X_2, \dots на (Ω, \mathcal{F}, P) , так что $E\{|X_n - X|^2\} \xrightarrow{n \rightarrow \infty} 0$, но $E\{|X_n|\} = \infty$ для любого $n \in \mathbb{N}$.
11. Пусть $X_n \xrightarrow{L_1} X$. Показать, что отсюда следует сходимость средних: $E\{X_n\} \rightarrow E\{X\}$. Справедливо ли обратное?
12. Пусть $X_n \xrightarrow{P} X$, $|X_n| \leq Z$ для любого $n \in \mathbb{N}$, причем $E\{Z\} < \infty$. Показать, что $X_n \xrightarrow{L_1} X$.
13. Пусть $X_n \xrightarrow{D} X$ и $Y_n \xrightarrow{P} c$, где c – некоторая константа. Показать, что $X_n Y_n \xrightarrow{D} cX$ и $X_n/Y_n \xrightarrow{D} X/c$, если $c \neq 0$.
14. Доказать следующую теорему.
Теорема (Слуцкого). Пусть $X_n \xrightarrow{D} 0$, $Y_n \xrightarrow{P} Y$, а $g(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^1$ – непрерывная функция от y для всех x , причем $g(0, y)$ непрерывна по y . Показать, что $g(X_n, Y_n) \xrightarrow{P} g(0, Y)$.
15. Пусть $X_1, X_2, \dots \geq 0$ – независимые случайные величины, распределенные по закону Пуассона: $\mathcal{L}\{X_r\} = \Pi(\lambda_r)$, $r \in \mathbb{N}$. Показать, что $\sum_{r=1}^{\infty} X_r$ сходится или расходится почти наверное в зависимости от того, сходится или расходится числовой ряд $\sum_{r=1}^{\infty} \lambda_r$.
16. Случайная последовательность X_1, X_2, \dots на (Ω, \mathcal{F}, P) называется вполне сходящейся к случайной величине X , если $\forall \epsilon > 0 \quad \sum_{k=1}^{\infty} P\{|X_n - X| > \epsilon\} < \infty$.
 Показать, что для последовательностей независимых случайных величин вполне сходимость эквивалентна почти наверное сходимости. Построить пример последовательности зависимых случайных величин, которая сходится почти наверное, но не является вполне сходящейся. (Использовать результаты упражнения 9.)
17. Пусть X_1, X_2, \dots – независимые, одинаково распределенные случайные величины с конечным математическим ожиданием μ и дисперсией. Показать, что $\frac{1}{C_n^2} \sum_{1 \leq i < j \leq n} X_i X_j \xrightarrow{P} \mu^2$.
18. Показать, что если $X_n \xrightarrow{D} X$, то $aX_n + b \xrightarrow{D} aX + b$ для любых констант $a, b \in \mathbb{R}$.

Глава 5

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

5.1. ЗАКОН БОЛЬШИХ ЧИСЕЛ И УСЛОВИЯ ЕГО ВЫПОЛНЕНИЯ

Пусть задана последовательность независимых случайных величин ξ_1, \dots, ξ_n , имеющих конечное математическое ожидание $E\{\xi_k\} = a$ ($k = 1, \dots, n$). Пусть, например, ξ_k – результат k -го взвешивания некоторого физического тела на аналитических весах, a – истинное значение веса. Тогда $\xi_k - a$ – случайная погрешность измерения в k -м эксперименте. При увеличении числа экспериментов, оказывается, имеет место сходимость последовательности средних арифметических:

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{n \rightarrow \infty} a. \quad (5.1)$$

Если сходимость в (5.1) понимается по вероятности, то соотношение вида (5.1) называется законом больших чисел (ЗБЧ). Если же в (5.1) имеем в виду сходимость почти наверное, то соотношение вида (5.1) называют усиленным законом больших чисел (УЗБЧ).

Теорема 5.1 (критерий ЗБЧ). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность как угодно зависимых случайных величин $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$. Тогда условие

$$E \left\{ \frac{\left(\sum_{k=1}^n (\xi_k - a_k) \right)^2}{n^2 + \left(\sum_{k=1}^n (\xi_k - a_k) \right)^2} \right\} \xrightarrow{n \rightarrow \infty} 0$$

является необходимым и достаточным условием выполнения закона больших чисел:

$$\frac{1}{n} \sum_{k=1}^n (\xi_k - a_k) \xrightarrow[n \rightarrow \infty]{P} 0. \quad (5.2)$$

Замечание 5.1. Если $\exists a = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n a_k$, то закон больших чисел (5.2) можно записать в виде (5.1).

Замечание 5.2. Введем в рассмотрение вспомогательную последовательность средних арифметических:

$$\eta_n = \eta_n(\omega) = \frac{1}{n} \sum_{k=1}^n (\xi_k - a_k).$$

В этом случае формулировка теоремы примет следующий упрощенный вид:

$$\eta_n \xrightarrow{P} 0 \quad (5.3)$$

тогда и только тогда, когда

$$E \left\{ \frac{\eta_n^2}{1 + \eta_n^2} \right\} \rightarrow 0. \quad (5.4)$$

Доказательство. По определению сходимости по вероятности имеем

$$(5.3) \Leftrightarrow \forall \epsilon > 0 \quad P_n(\epsilon) ::= P\{|\eta_n| > \epsilon\} \xrightarrow{n \rightarrow \infty} 0. \quad (5.5)$$

Необходимо показать эквивалентность: (5.5) \Leftrightarrow (5.4). Введем в рассмотрение вспомогательную функцию:

$$g(y) = \frac{y^2}{1+y^2}, \quad 0 \leq g(y) \leq 1, \quad g(y) \leq y^2, \quad g(\cdot) -$$

монотонно возрастающая функция переменной $|y|$.

Получим двухстороннюю оценку для $P_n(\epsilon)$. В силу монотонности функции $g(\cdot)$ имеем оценку сверху:

$$\begin{aligned} P_n(\epsilon) &= \int_{|y|>\epsilon} 1 dF_{\eta_n}(y) \leq \int_{|y|>\epsilon} \frac{g(y)}{g(\epsilon)} dF_{\eta_n}(y) \leq \frac{1+\epsilon^2}{\epsilon^2} \times \\ &\times \int_{-\infty}^{+\infty} \frac{y^2}{1+y^2} dF_{\eta_n}(y) = \frac{1+\epsilon^2}{\epsilon^2} E\left\{ \frac{\eta_n^2}{1+\eta_n^2} \right\}. \end{aligned} \quad (5.6)$$

Оценим $P_n(\epsilon)$ снизу:

$$\begin{aligned} P_n(\epsilon) &= \int_{|y|>\epsilon} 1 dF_{\eta_n}(y) \geq \int_{|y|>\epsilon} g(y) dF_{\eta_n}(y) = \\ &= \int_{-\infty}^{+\infty} g(y) dF_{\eta_n}(y) - \int_{|y|\leq\epsilon} g(y) dF_{\eta_n}(y) \geq \\ &\geq E\left\{ \frac{\eta_n^2}{1+\eta_n^2} \right\} - \int_{|y|\leq\epsilon} y^2 dF_{\eta_n}(y) \geq E\left\{ \frac{\eta_n^2}{1+\eta_n^2} \right\} - \epsilon^2. \end{aligned} \quad (5.7)$$

Объединяя (5.6) и (5.7), получаем

$$E\left\{ \frac{\eta_n^2}{1+\eta_n^2} \right\} - \epsilon^2 \leq P_n(\epsilon) \leq \frac{1+\epsilon^2}{\epsilon^2} E\left\{ \frac{\eta_n^2}{1+\eta_n^2} \right\}, \quad \forall n \in \mathbb{N}, \forall \epsilon > 0.$$

Отсюда заключаем, что при $n \rightarrow \infty$ (5.4) \Leftrightarrow (5.5). \square

Замечание 5.3. Доказанный критерий закона больших чисел на практике часто труднопроверяем в вычислительном отношении.

Сформулируем ряд более простых достаточных условий закона больших чисел.

Теорема 5.2 (достаточное условие Маркова ЗБЧ). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность как угодно зависимых случайных величин $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, имеющих конечные математические ожидания $E\{\xi_k\} = a_k$ и дисперсии. Если выполняется условие

$$\frac{1}{n^2} D\left\{ \sum_{k=1}^n \xi_k \right\} \xrightarrow{n \rightarrow \infty} 0, \quad (5.8)$$

то к последовательности $\{\xi_k\}$ применим закон больших чисел в виде (5.2).

Доказательство. Покажем, что (5.8) влечет выполнение условия (5.4). По свойствам дисперсии имеем

$$E\left\{\frac{\eta_n^2}{1+\eta_n^2}\right\} \leq E\{\eta_n^2\} \equiv D\{\eta_n\} = D\left\{\frac{1}{n}\sum_{k=1}^n \xi_k - \frac{1}{n}\sum_{k=1}^n a_k\right\} = \frac{1}{n^2}D\left\{\sum_{k=1}^n \xi_k\right\} \rightarrow 0,$$

и, следовательно, выполняется (5.4), а значит, и закон больших чисел. \square

Теорема 5.3 (достаточное условие Чебышева ЗБЧ). Если $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, – последовательность попарно независимых случайных величин, имеющих конечные математические ожидания и ограниченные дисперсии:

$$D\{\xi_k\} \leq c < +\infty, \quad (5.9)$$

то к этой последовательности применим закон больших чисел в виде (5.2).

Доказательство. Проверим достаточное условие Маркова, используя свойства дисперсии и формулу (5.9):

$$\frac{1}{n^2}D\left\{\sum_{k=1}^n \xi_k\right\} = \frac{1}{n^2}\sum_{k=1}^n D\{\xi_k\} \leq \frac{c}{n} \rightarrow 0.$$

Следовательно, выполнено достаточное условие Маркова, а значит, и закон больших чисел. \square

Теорема 5.4 (теорема Бернулли). Если $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, – последовательность одинаково распределенных попарно независимых случайных величин Бернулли:

$$\xi_k \in \{0, 1\}, \quad P\{\xi_k = 1\} = p, \quad P\{\xi_k = 0\} = 1 - p, \quad 0 \leq p \leq 1,$$

то выполняется закон больших чисел в виде

$$\frac{1}{n}\sum_{k=1}^n \xi_k \xrightarrow[n \rightarrow \infty]{P} p.$$

Доказательство. Проверим достаточное условие Чебышева:

$$E\{\xi_k\} = a_k = p, \quad D\{\xi_k\} = p(1-p) \leq c = \frac{1}{4} < +\infty.$$

Следовательно, имеет место закон больших чисел. \square

Следствие 5.1 (эквивалентная формулировка). Пусть m – число наступлений случайного события A в схеме n независимых испытаний Бернулли, в каждом из которых событие A наступает с одной и той же вероятностью p . Тогда при увеличении числа испытаний ($n \rightarrow \infty$) относительная частота $\nu_n(A) = m/n$ сходится к вероятности p :

$$\nu_n(A) \xrightarrow{P} p.$$

Доказательство. Достаточно учесть, что

$$m = \sum_{k=1}^n \xi_k.$$

\square

Теорема 5.5 (Пуассона). Пусть ξ_k – последовательность попарно независимых случайных величин Бернулли: $\xi_k \in \{0, 1\}$, $P\{\xi_k = 1\} = p_k$, $P\{\xi_k = 0\} = 1 - p_k$, $0 \leq p_k \leq 1$. Тогда выполняется закон больших чисел в следующем виде:

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow[n \rightarrow \infty]{P} p^*,$$

$$\text{где } p^* = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_k.$$

Доказательство. Схема доказательства та же, что и в теореме 5.4. \square

Следствие 5.2. Теоремы Чебышева, Бернулли и Пуассона остаются в силе, если вместо условия попарной независимости используется условие попарной некоррелированности.

5.2. УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ И УСЛОВИЯ ЕГО ВЫПОЛНЕНИЯ

Определение 5.1. Принято говорить, что случайная последовательность $\xi_k = \xi_k(\omega)$, $k \in \mathbb{N}$, на вероятностном пространстве (Ω, \mathcal{F}, P) удовлетворяет усиленному закону больших чисел, если существует такая числовая последовательность $a_k \in \mathbb{R}$, что при $n \rightarrow \infty$ последовательность средних арифметических сходится почти наверное:

$$\frac{1}{n} \sum_{k=1}^n (\xi_k - a_k) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0. \quad (5.10)$$

Замечание 5.4. Из выполнения усиленного закона больших чисел следует выполнение закона больших чисел.

Замечание 5.5. Выполнение усиленного закона больших чисел требует более жестких условий, чем закон больших чисел. В частности, мы будем требовать независимость в совокупности случайных величин.

Лемма 5.1 (неравенство Гаека – Реньи). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\xi_k = \xi_k(\omega)$, $k \in \mathbb{N}$, имеющих математические ожидания $E\{\xi_k\} = a_k$ и дисперсии $D\{\xi_k\} = \sigma_k^2$. Тогда для $\forall \epsilon > 0$, $\forall m, n \in \mathbb{N}, m < n$, любой невозрастающей последовательности $c_k \geq 0$ справедливо неравенство

$$P \left\{ \max_{m \leq k \leq n} \left| c_k \sum_{i=1}^k (\xi_i - a_i) \right| > \epsilon \right\} \leq \frac{1}{\epsilon^2} \left(c_m^2 \sum_{i=1}^m \sigma_i^2 + \sum_{i=m+1}^n c_i^2 \sigma_i^2 \right). \quad (5.11)$$

Доказательство. Обозначим

$$S_k = \sum_{i=1}^k (\xi_i - a_i), \quad \eta = \sum_{k=m}^{n-1} S_k^2 (c_k^2 - c_{k+1}^2) + S_n^2 c_n^2.$$

Тогда

$$\begin{aligned}
 E\{\eta\} &= \sum_{k=m}^{n-1} (c_k^2 - c_{k+1}^2) E\{S_k^2\} + c_n^2 E\{S_n^2\} = \sum_{k=m}^{n-1} \sum_{i=1}^k \sigma_i^2 (c_k^2 - c_{k+1}^2) + c_n^2 \sum_{i=1}^n \sigma_i^2 = \\
 &= \sum_{i=1}^m \sum_{k=m}^{n-1} \sigma_i^2 (c_k^2 - c_{k+1}^2) + \sum_{i=m+1}^{n-1} \sum_{k=1}^{n-1} \sigma_i^2 (c_k^2 - c_{k+1}^2) + c_n^2 \sum_{i=1}^n \sigma_i^2 = \\
 &= \sum_{i=1}^m \sigma_i^2 (c_m^2 - c_n^2) + \sum_{i=m+1}^{n-1} \sigma_i^2 (c_i^2 - c_n^2) + c_n^2 \sum_{i=1}^n \sigma_i^2 = c_m^2 \sum_{i=1}^m \sigma_i^2 + \sum_{i=m+1}^n \sigma_i^2 c_i^2.
 \end{aligned}$$

Рассмотрим случайные события:

$$A_i = \{\omega : c_k |S_k(\omega)| \leq \epsilon, m \leq k < i-1, c_i |S_i(\omega)| > \epsilon\},$$

$$i = m, m+1, \dots, n.$$

Поскольку события $\{A_i\}$ несовместны, то

$$P\left\{\max_{m \leq k \leq n} c_k \left|\sum_{i=1}^k (\xi_i - a_i)\right| > \epsilon\right\} = P\left\{\bigcup_{i=m}^n A_i\right\} = \sum_{i=m}^n P\{A_i\}.$$

Отсюда для доказательства теоремы достаточно доказать, что

$$E\{\eta\} \geq \epsilon^2 \sum_{i=m}^n P\{A_i\}.$$

Это неравенство следует из неравенств

$$E\{\eta\} \geq E\{\eta \mathbb{1}_{\bigcup_{i=m}^n A_i}\} = \sum_{i=m}^n E\{\eta \mathbb{1}_{A_i}\};$$

$$\begin{aligned}
 E\{\eta \mathbb{1}_{A_i}\} &= \sum_{k=m}^{n-1} (c_k^2 - c_{k+1}^2) E\{S_k^2 \mathbb{1}_{A_i}\} + c_n^2 E\{S_n^2 \mathbb{1}_{A_i}\} \geq \\
 &\geq \sum_{k=i}^{n-1} (c_k^2 - c_{k+1}^2) E\{S_k^2 \mathbb{1}_{A_i}\} + c_n^2 E\{S_n^2 \mathbb{1}_{A_i}\};
 \end{aligned}$$

$$E\{S_k^2 \mathbb{1}_{A_i}\} = E\{(S_k - S_i + S_i)^2 \mathbb{1}_{A_i}\} \geq E\{S_i^2 \mathbb{1}_{A_i}\} + 2E\{(S_k - S_i)S_i \mathbb{1}_{A_i}\} =$$

$$= E\{S_i^2 \mathbb{1}_{A_i}\} + 2E\{S_k - S_i\}E\{S_i \mathbb{1}_{A_i}\} = E\{S_i^2 \mathbb{1}_{A_i}\} \geq$$

$$\geq E\left\{\frac{\epsilon^2}{c_i^2} \mathbb{1}_{A_i}\right\} = \frac{\epsilon^2}{c_i^2} P\{A_i\}.$$

□

Следствие 5.3 (неравенство Колмогорова). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\xi_k = \xi_k(\omega)$, $k \in \mathbb{N}$, имеющих математические ожидания $E\{\xi_k\} = a_k$

и дисперсии $D\{\xi_k\} = \sigma_k^2$. Тогда для $\forall \epsilon > 0$ справедливо следующее неравенство:

$$P\left\{\max_{1 \leq k \leq n} \left| \sum_{j=1}^k (\xi_j - a_j) \right| > \epsilon\right\} \leq \frac{\sum_{j=1}^n \sigma_j^2}{\epsilon^2} \equiv \frac{D\left\{\sum_{j=1}^n \xi_j\right\}}{\epsilon^2}. \quad (5.12)$$

Доказательство. Неравенство (5.12) следует из неравенства Гаека – Реньи, если положить $m = 1$, $c_k \equiv 1$. \square

Следствие 5.4. При $n = 1$ формула (5.12) превращается в неравенство Чебышева.

Замечание 5.6. Таким образом, неравенство Гаека – Реньи является обобщением неравенства Чебышева.

Теорема 5.6 (достаточное условие Колмогорова УЗБЧ). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\xi_k = \xi_k(\omega)$, $k \in \mathbb{N}$, имеющих конечные математические ожидания $E\{\xi_k\} = a_k$ и дисперсии $D\{\xi_k\} = \sigma_k^2$. Если сходится ряд Колмогорова

$$K = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < +\infty, \quad (5.13)$$

то к последовательности ξ_k применим усиленный закон больших чисел в виде (5.10).

Доказательство. Воспользуемся неравенством (5.11), в котором положим $c_k = 1/k$, $n \rightarrow \infty$, $\overset{\circ}{\xi}_j = \xi_j - a_j$. Получим

$$\forall \epsilon > 0 \quad P\left\{\sup_{k \geq m} \left| \frac{1}{k} \sum_{j=1}^k \overset{\circ}{\xi}_j \right| > \epsilon\right\} < \frac{1}{\epsilon^2} \left(\frac{1}{m^2} \sum_{j=1}^m \sigma_j^2 + \sum_{j=m+1}^{\infty} \frac{\sigma_j^2}{j^2} \right) =: \frac{1}{\epsilon^2} (A_m + B_m).$$

Исследуем асимптотику этой числовой последовательности при $m \rightarrow \infty$. Из (5.13) следует: $B_m \rightarrow 0$ и $A_m \rightarrow 0$. Отсюда вытекает, что

$$\forall \epsilon > 0 \quad \lim_{m \rightarrow \infty} P\left\{\sup_{k \geq m} \left| \frac{1}{k} \sum_{j=1}^k \overset{\circ}{\xi}_j \right| > \epsilon\right\} = 0.$$

По одному из эквивалентных определений сходимости почти наверное это и означает (5.10). \square

Лемма 5.2. Пусть $\xi = \xi(\omega)$ – произвольная случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) , и для нее определена последовательность вероятностей событий:

$$p_k = P\{|\xi| \geq k\}, \quad k = 0, 1, 2, \dots; \quad p_0 = 1.$$

Тогда справедливо следующее двухстороннее неравенство:

$$\sum_{k=1}^{\infty} p_k \leq E\{|\xi|\} \leq 1 + \sum_{k=1}^{\infty} p_k.$$

Доказательство. Доказать данную лемму можно, воспользовавшись следующим выражением:

$$E\{|\xi|\} = \int_0^{+\infty} x dF_{|\xi|}(x) \equiv \sum_{k=0}^{\infty} \int_k^{k+1-0} x dF_{|\xi|}(x).$$

\square

Теорема 5.7 (Колмогорова, УЗБЧ для н. о. р. СВ). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности, одинаково распределенных случайных величин $\xi_k = \xi_k(\omega), k \in \mathbb{N}$. Для того чтобы выполнялся усиленный закон больших чисел в следующем виде:

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow[n \rightarrow \infty]{\text{п. н.}} a, \quad (5.14)$$

необходимо и достаточно, чтобы существовало математическое ожидание

$$E\{|\xi_1|\} < \infty, \quad (5.15)$$

причем $E\{\xi_1\} = a$.

Доказательство. *Достаточность.* Разобьем доказательство на четыре части.

1) Определим последовательность вероятностей:

$$p_k = P\{|\xi_1| \geq k\}, k = 0, 1, \dots$$

В силу леммы 5.2 и формулы (5.15) заметим, что сходится ряд, составленный из этих вероятностей:

$$\sum_{k=1}^{\infty} p_k < +\infty. \quad (5.16)$$

2) Построим вспомогательную последовательность ξ_n^* на (Ω, \mathcal{F}, P) и покажем, что она эквивалентна ξ_n в смысле сходимости почти наверное:

$$\xi_n^* - \xi_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0.$$

Вспомогательная последовательность ξ_n^* называется усеченной и строится следующим образом:

$$\xi_n^* = \begin{cases} \xi_n, & |\xi_n| < n, \\ 0, & |\xi_n| \geq n. \end{cases} \quad (5.17)$$

Отсюда имеем $P\{\xi_n^* \neq \xi_n\} = P\{|\xi_n| \geq n\} = p_n$. Выберем $\forall \epsilon > 0, \forall N \in \mathbb{N}$ и оценим вероятность такого события:

$$\begin{aligned} P_N &::= P\left\{\sup_{n \geq N} |\xi_n^* - \xi_n| > \epsilon\right\} \leq P\left\{\sup_{n \geq N} |\xi_n^* - \xi_n| > 0\right\} = \\ &= P\left\{\bigcup_{n=N}^{\infty} \{\xi_n^* \neq \xi_n\}\right\} \leq \sum_{n=N}^{\infty} P\{\xi_n^* \neq \xi_n\} = \sum_{n=N}^{\infty} p_n. \end{aligned}$$

В силу формулы (5.16) P_N оценена остатком сходящегося ряда. Поэтому

$$P_N \xrightarrow[N \rightarrow \infty]{} 0 \Rightarrow \xi_n^* - \xi_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0.$$

Таким образом, рассматриваемые последовательности $\{\xi_n\}, \{\xi_n^*\}$ эквивалентны в смысле сходимости почти наверное.

3) Покажем, что к усеченной последовательности ξ_n^* применим усиленный закон больших чисел. Проверим достаточное условие Колмогорова, т. е. сходимость ряда:

$$K = \sum_{n=1}^{\infty} \frac{D\{\xi_n^*\}}{n^2}. \quad (5.18)$$

Оценим дисперсию в (5.18) с учетом (5.17):

$$\begin{aligned} D\{\xi_n^*\} &\leq E\{(\xi_n^*)^2\} ::= \int_0^{n-0} y^2 dF_{|\xi_1|}(y) \equiv \\ &\equiv \sum_{k=0}^{n-1} \int_k^{k+1-0} y^2 dF_{|\xi_1|}(y) \leq \sum_{k=0}^{n-1} (k+1)^2 P\{k \leq |\xi_1| < k+1\} = \\ &= \sum_{k=0}^{n-1} (k+1)^2 ((1-p_{k+1}) - (1-p_k)) = \sum_{k=0}^{n-1} (k+1)^2 (p_k - p_{k+1}). \end{aligned}$$

Подставляя эту оценку в (5.18), получаем

$$K \leq \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \frac{1}{n^2} (k+1)^2 (p_k - p_{k+1}).$$

Поменяв порядок суммирования, имеем

$$K \leq \sum_{k=0}^{\infty} (k+1)^2 (p_k - p_{k+1}) \sum_{n=k+1}^{\infty} \frac{1}{n^2}.$$

Справедлива оценка суммы внутреннего ряда:

$$\begin{aligned} \sum_{n=k+1}^{\infty} \frac{1}{n^2} &\leq \frac{1}{(k+1)^2} + \sum_{n=k+2}^{\infty} \frac{1}{n(n-1)} \equiv \frac{1}{(k+1)^2} + \sum_{n=k+2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = \\ &= \frac{1}{(k+1)^2} + \frac{1}{k+1} \leq \frac{2}{k+1}. \end{aligned}$$

Используя эту оценку, получаем согласно (5.16)

$$K \leq 2 \sum_{k=0}^{\infty} (k+1) (p_k - p_{k+1}) = 2 \left(\sum_{k=0}^{\infty} k p_k + \sum_{k=0}^{\infty} p_k - \sum_{i=0}^{\infty} i p_i \right) = 2 \left(1 + \sum_{k=1}^{\infty} p_k \right) < +\infty.$$

В силу достаточного условия Колмогорова заключаем, что усиленный закон больших чисел применим к ξ_n^* в следующем виде:

$$\frac{1}{n} \sum_{k=1}^n (\xi_k^* - E\{\xi_k^*\}) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0. \quad (5.19)$$

- 4) Исследуем для (5.19) сходимость последовательности математических ожиданий, используя (5.17):

$$\lim_{k \rightarrow \infty} E\{\xi_k^*\} = \lim_{k \rightarrow \infty} \int_{-k}^k x dF_{\xi_1}(x) = \int_{-\infty}^{+\infty} x dF_{\xi_1}(x) = E\{\xi_1\} = a.$$

Таким образом, (5.19) принимает вид

$$\frac{1}{n} \sum_{k=1}^n \xi_k^* \xrightarrow[n \rightarrow \infty]{\text{п. н.}} a.$$

А так как ξ_k^* эквивалентна ξ_k в смысле сходимости почти наверное, то выполняется (5.14).

Необходимость. Введем вспомогательную последовательность средних арифметических:

$$\eta_n = \frac{1}{n} \sum_{k=1}^n \xi_k, n \in \mathbb{N}.$$

В силу формулы (5.14) имеем $\eta_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} a$.

Из этого следует, что

$$\frac{1}{n} \xi_n \equiv \eta_n - \frac{n-1}{n} \eta_{n-1} \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0.$$

Воспользуемся определением сходимости почти наверное и построим вспомогательную последовательность случайных событий:

$$A_n = \left\{ \left| \frac{1}{n} \xi_n \right| \geq \epsilon \right\} = \{ |\xi_n| \geq n \epsilon \}, n \in \mathbb{N},$$

где выбрано $\epsilon = 1$.

Из определения сходимости почти наверное имеем $P\{\overline{\lim}_{n \rightarrow \infty} A_n\} = 0$. Заметим, что $\{A_n\}$ независимы в совокупности, поэтому из закона $0 \vee 1$ Бореля следует, что ряд $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} p_n$ сходится. И в силу леммы 5.2 $E\{|\xi_1|\} < +\infty$. Далее, пользуясь первой частью доказательства достаточности, получаем

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow[n \rightarrow \infty]{\text{п. н.}} E\{\xi_1\}.$$

Из формулы (5.14) и единственности предела следует, что $E\{\xi_1\} = a$. \square

Теорема 5.8 (теорема Бореля). Пусть m – случайное число наступлений события A в схеме n независимых испытаний, в каждом из которых A наступает с одной и той же вероятностью p . Тогда при $n \rightarrow \infty$ относительная частота сходится к вероятности почти наверное:

$$\nu_n = \frac{m}{n} \xrightarrow[n \rightarrow \infty]{\text{п. н.}} p.$$

Доказательство. Проверим условия теоремы Колмогорова. Так как ξ_k – случайная величина Бернулли, то ее математическое ожидание существует: $E\{\xi_1\} = a = p$. \square

Замечание 5.7. Теорема Бореля усиливает теорему Бернулли.

Теорема 5.9 (теорема Хинчина). Пусть на (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности, одинаково распределенных случайных величин ξ_k , имеющих математическое ожидание $E\{\xi_k\} = a$. Тогда к этой последовательности применим закон больших чисел:

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow[n \rightarrow \infty]{P} a.$$

Доказательство. По теореме Колмогорова выполняется соотношение (5.14), а из сходимости почти наверное следует сходимость по вероятности. \square

Замечание 5.8. Теорема Хинчина есть ослабление теоремы Колмогорова.

5.3. СХОДИМОСТЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ И ХАРАКТЕРИСТИЧЕСКИХ ФУНКЦИЙ

Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена случайная последовательность $\xi_n = \xi_n(\omega)$, $n \in \mathbb{N}$, и определены соответствующие последовательности функций распределения:

$$F_n(x) ::= P\{\xi_n < x\}, \quad x \in \mathbb{R},$$

и характеристических функций:

$$f_n(t) ::= E\{e^{it\xi}\} = \int_{-\infty}^{+\infty} e^{itx} dF_n(x), \quad t \in \mathbb{R}.$$

Пусть $F(x)$, $x \in \mathbb{R}$, – некоторая неубывающая ограниченная функция, $0 \leq F(x) \leq 1$; $C(F) \subseteq \mathbb{R}$ – множество точек непрерывности функции F .

Определение 5.2. Принято говорить, что последовательность функций распределения $F_n(x)$, $x \in \mathbb{R}$, слабо сходится к некоторой неубывающей ограниченной функции $F(x)$, $x \in \mathbb{R}$, при $n \rightarrow \infty$, и этот факт принято кратко обозначать

$$F_n(\cdot) \Rightarrow F(\cdot), \quad (5.20)$$

если имеет место поточечная сходимость последовательности функций распределения на множестве $C(F)$, т. е. если выполняется предельное соотношение

$$\forall x \in C(F) \quad F_n(x) \rightarrow F(x). \quad (5.21)$$

Замечание 5.9. Предельная функция $F(\cdot)$ – не обязательно функция распределения, так как ее колебание $\omega(F) = \sup F(x) - \inf F(x) \in [0, 1]$ может оказаться меньше единицы.

Замечание 5.10. Если $\omega(F) = 1$, то предельная функция $F(\cdot)$ – функция распределения некоторой случайной величины ξ , и тогда соотношение (5.20) эквивалентно ранее изученной сходимости по распределению: $\xi_n \xrightarrow{D} \xi$.

Теорема 5.10 (критерий слабой сходимости ФР). Для того чтобы $F_n(\cdot) \Rightarrow F(\cdot)$, необходимо и достаточно, чтобы имела место поточечная сходимость последовательности функций распределения $F_n(\cdot)$ на некотором всюду плотном в \mathbb{R} множестве $Z \subseteq \mathbb{R}$:

$$\forall z \in Z, \quad F_n(z) \rightarrow F(z). \quad (5.22)$$

Замечание 5.11. Множество $Z \subseteq \mathbb{R}$ всюду плотно в \mathbb{R} тогда и только тогда, когда $\forall x \in \mathbb{R}, \forall \epsilon > 0$, ϵ -окрестность точки x содержит хотя бы один элемент из Z .

Доказательство. *Необходимость.* Пусть $F_n(\cdot) \Rightarrow F(\cdot)$. Найдем такое всюду плотное множество $Z \subset \mathbb{R}$, для которого выполняется (5.22). Выберем $Z ::= C(F)$. Поскольку F – неубывающая ограниченная функция, то множество точек ее разрыва $C_0(F)$ не более чем счетно. Тогда $C(F) = \mathbb{R}/C_0(F)$ всюду плотно в \mathbb{R} .

Достаточность. Пусть выполняется (5.22). В силу свойств всюду плотного множества Z подберем числовые последовательности:

$$z_1 \uparrow x, z_2 \downarrow x, z_1, z_2 \in Z, \quad z_1 \leq x \leq z_2, \quad x \in C(F). \quad (5.23)$$

В силу свойства функции распределения из (5.23) имеем

$$F_n(z_1) \leq F_n(x) \leq F_n(z_2).$$

Устремляя $n \rightarrow \infty$, в силу (5.22) имеем

$$\lim_{n \rightarrow \infty} F_n(z_1) = F(z_1) \leq \varliminf_{n \rightarrow \infty} F_n(x) \leq \overline{\lim}_{n \rightarrow \infty} F_n(x) \leq \lim_{n \rightarrow \infty} F_n(z_2) = F(z_2).$$

Устремим теперь: $z_1 \uparrow x$, $z_2 \downarrow x$. Поскольку x – точка непрерывности функции F , то $\varliminf_{n \rightarrow \infty} F_n(x) = \overline{\lim}_{n \rightarrow \infty} F_n(x) = F(x)$, что означает выполнение (5.21). \square

Лемма 5.3 (первая теорема Хелли). Из всякой последовательности функций распределения $F_n(x)$, $x \in \mathbb{R}$, $n \in \mathbb{N}$, можно выделить подпоследовательность $F_{n_k}(x)$, $x \in \mathbb{R}$, $k \in \mathbb{N}$, которая слабо сходится к некоторой неубывающей ограниченной функции $F: F_{n_k}(\cdot) \Rightarrow F(\cdot)$.

Доказательство. Для доказательства следует воспользоваться леммой Больца – Вейерштрасса. \square

Лемма 5.4 (вторая теорема Хелли). Пусть некоторая последовательность функций распределения слабо сходится к некоторой неубывающей ограниченной функции: $F_n(\cdot) \Rightarrow F(\cdot)$, причем предельная функция $F(\cdot)$ является некоторой функцией распределения (т. е. $\omega(F) = 1$). Пусть далее $y = g(x)$ – некоторая непрерывная ограниченная функция: $|g(x)| \leq M < +\infty$. Тогда имеет место сходимость последовательности интегралов

$$\int_{-\infty}^{+\infty} g(x) dF_n(x) \rightarrow \int_{-\infty}^{+\infty} g(x) dF(x).$$

Доказательство. На первом этапе доказывается сходимость последовательности интегралов на конечном промежутке $[a, b]$. На этом промежутке для подынтегральной функции $g(x)$ строится аппроксимирующая функция $g_\epsilon(x)$:

$$\sup_{a \leq x \leq b} |g_\epsilon(x) - g(x)| \leq \epsilon,$$

для которой показывается, что

$$\int_a^b g_\epsilon(x) dF_n(x) \rightarrow \int_a^b g_\epsilon(x) dF(x).$$

На втором этапе используются свойства функции распределения: $F(a) \rightarrow 0$ при $a \rightarrow -\infty$, $1 - F(b) \rightarrow 0$ при $b \rightarrow +\infty$. \square

Теорема 5.11 (теорема непрерывности). Пусть на (Ω, \mathcal{F}, P) определены последовательность случайных величин $\xi_n = \xi_n(\omega)$, $n = 1, 2, \dots$, и соответствующие последовательности функций распределения $F_n(x)$, $x \in \mathbb{R}$, и характеристических функций $f_n(t)$, $t \in \mathbb{R}$. Тогда при $n \rightarrow \infty$ справедливы следующие два утверждения:

1) если последовательность функций распределения слабо сходится к некоторой функции распределения $F(\cdot): F_n(\cdot) \Rightarrow F(\cdot)$, то имеет место поточечная сходимость последовательности характеристических функций:

$$f_n(t) \xrightarrow{n \rightarrow \infty} f(t), t \in \mathbb{R}, \text{ причем } f(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x) -$$

характеристическая функция, соответствующая предельной функции распределения $F(\cdot)$;

- 2) если имеет место поточечная сходимость последовательности характеристических функций: $f_n(t) \rightarrow f(t)$, $t \in \mathbb{R}$, причем предельная функция $f(t)$ непрерывна в точке $t = 0$, то имеет место слабая сходимость последовательности функций распределения: $F_n(\cdot) \Rightarrow F(\cdot)$; при этом $F(\cdot)$ – функция распределения, которой соответствует характеристическая функция $f(t)$.

Доказательство.

- 1) Выберем во второй теореме Хелли: $g(x) = e^{itx}$, t – фиксированное; g – ограничена и непрерывна: $|g(x)| = 1$. Итак, из второй теоремы Хелли имеем

$$\int_{-\infty}^{+\infty} e^{itx} dF_n(x) \rightarrow \int_{-\infty}^{+\infty} e^{itx} dF(x), \quad \text{следовательно,}$$

$$f_n(t) \xrightarrow{n \rightarrow \infty} f(t), \quad t \in \mathbb{R}.$$

- 2) Из доказанной выше первой теоремы Хелли следует, что существует последовательность n_k такая, что $\{F_{n_k}\}$ будет слабо сходиться к некоторой функции F , которая непрерывна слева и неубывающая. Покажем, что функция F – функция распределения. Для этого требуется доказать, что $F(-\infty) = 0$, $F(+\infty) = 1$ или $\omega(F) = F(+\infty) - F(-\infty) = 1$.

Предположим противное:

$$F(+\infty) - F(-\infty) = \delta < 1.$$

Поскольку для всех n $f_n(0) = 1$, то и $f(0) = 1$.

Из непрерывности $f(t)$ в точке $t = 0$ следует, что для любого $\epsilon > 0$ найдется $\tau > 0$ такое, что

$$\left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f(t) dt \right| \geq 1 - \frac{\epsilon}{2}.$$

Поскольку $f_{n_k} \rightarrow f(t)$, $t \in \mathbb{R}$, то существует n_0 такое, что $|f_{n_k}(t) - f(t)| \leq \epsilon/2$ при $n_k \geq n_0$. Тогда

$$1 - \frac{\epsilon}{2} \leq \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f(t) dt \right| \leq \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f_{n_k}(t) dt \right| +$$

$$+ \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} (f(t) - f_{n_k}(t)) dt \right| \leq \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f_{n_k}(t) dt \right| + \frac{\epsilon}{2}.$$

Отсюда при $n_k \geq n_0$

$$\left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f_{n_k}(t) dt \right| \geq 1 - \epsilon. \quad (5.24)$$

Используя неравенства

$$\left| \frac{\sin \tau x}{\tau x} \right| \leq 1, \quad |\sin \tau x| \leq 1,$$

получаем

$$\begin{aligned} \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f_{n_k}(t) dt \right| &= \left| \frac{1}{2\tau} \int_{-\tau}^{\tau} \int e^{itx} dF_{n_k}(x) dt \right| = \left| \int \frac{1}{2\tau} \int_{-\tau}^{\tau} e^{itx} dF_{n_k}(x) dt \right| = \\ &= \left| \int \frac{\sin \tau x}{\tau x} dF_{n_k}(x) \right| = \left| \int_{-X \leq x < X} \frac{\sin \tau x}{\tau x} dF_{n_k}(x) \right| + \left| \int_{x \geq X, x < -X} \frac{\sin \tau x}{\tau x} dF_{n_k}(x) \right| \leq \\ &\leq F_{n_k}(X) - F_{n_k}(-X) + \frac{1}{\tau X}. \end{aligned}$$

Выберем X и $-X$ точками непрерывности функции F . Тогда

$$F_{n_k}(X) \xrightarrow{n_k \rightarrow \infty} F(X) \text{ и } F_{n_k}(-X) \xrightarrow{n_k \rightarrow \infty} F(-X).$$

Тогда по предположению

$$F_{n_k}(X) - F_{n_k}(-X) \rightarrow F(X) - F(-X) \leq \delta.$$

Отсюда получаем, что для любого $\epsilon_1 > 0$ существует номер $n_1 = n_1(\epsilon, X)$ такой, что при $n_k \geq n_1$

$$F_{n_k}(X) - F_{n_k}(-X) \leq \delta + \frac{\epsilon_1}{2}.$$

Выберем X из условия

$$\frac{1}{\tau X} \leq \frac{\epsilon_1}{2}.$$

Тогда при $n \geq \max(n_0, n_1)$

$$\left| \frac{1}{2\tau} \int_{-\tau}^{\tau} f_{n_k}(t) dt \right| \leq \delta + \epsilon_1. \quad (5.25)$$

Из соотношений (5.24) и (5.25) получаем неравенство $1 - \epsilon \leq \delta + \epsilon_1$, верное при всех $\epsilon > 0$ и $\epsilon_1 > 0$, что невозможно. Получили противоречие. Таким образом, функция F — функция распределения.

Поскольку $f_{n_k}(t) \rightarrow \int_{-\infty}^{+\infty} e^{itx} dF(x)$ по второй теореме Хелли и $f_{n_k}(t) \rightarrow f(t)$, следовательно,

$$f(t) = \int e^{itx} dF(x).$$

Это означает, что f — характеристическая функция, соответствующая функции распределения F .

Покажем, что $F_n \Rightarrow F$. Допустим противное: $F_n \not\Rightarrow F$. Тогда будет существовать последовательность n_k^* такая, что $F_{n_k^*} \Rightarrow F^*$, где $F^* \neq F$ и F^* — функция распределения. По второй теореме Хелли получаем

$$f_{n_k^*}(t) \rightarrow \int e^{itx} dF^*(x) \neq \int e^{itx} dF(x) = f(t).$$

Получили противоречие, так как

$$f_{n_k^*}(t) \xrightarrow{n_k^* \rightarrow \infty} f(t).$$

□

Замечание 5.12. Теорема 5.11 показывает, что соответствие между функцией распределения и характеристической функцией является не только взаимно-однозначным, но и взаимно непрерывным:

$$F_n(\cdot) \Rightarrow F(\cdot) \Rightarrow f_n(\cdot) \rightarrow f(\cdot);$$

$$f_n(\cdot) \rightarrow f(\cdot) \Rightarrow F_n(\cdot) \Rightarrow F(\cdot).$$

5.4. УСЛОВИЕ ЛИНДЕБЕРГА И ЕГО ВЕРОЯТНОСТНЫЙ СМЫСЛ

Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\xi_k = \xi_k(\omega)$, $k = 1, 2, \dots$, имеющих функции распределения:

$$F_{\xi_k}(y) = P\{\xi_k < y\} = \int_{-\infty}^y dF_{\xi_k}(z), \quad y \in \mathbb{R},$$

конечные математические ожидания:

$$a_k = E\{\xi_k\} = \int_{-\infty}^{+\infty} y dF_{\xi_k}(y)$$

и конечные дисперсии:

$$\sigma_k^2 = D\{\xi_k\} = E\{(\xi_k - a_k)^2\} = \int_{-\infty}^{+\infty} (y - a_k)^2 dF_{\xi_k}(y). \quad (5.26)$$

Определение 5.3. Усеченным математическим ожиданием случайной величины ξ_k называется величина

$$E\{\xi_k; |\xi_k| \geq c\} ::= \int_{|y| \geq c} y dF_{\xi_k}(y),$$

где $c \geq 0$ – параметр усечения. Очевидно: $E\{\xi_k; |\xi_k| \geq 0\} \equiv E\{\xi_k\}$.

Рассмотрим последовательность частичных сумм случайных величин:

$$S_n = \sum_{k=1}^n \xi_k, \quad n = 1, 2, \dots$$

Определим соответствующее математическое ожидание $A_n = E\{S_n\} = \sum_{k=1}^n a_k$ и дисперсию

$$B_n^2 = D\{S_n\} = \sum_{k=1}^n \sigma_k^2. \quad (5.27)$$

Определение 5.4. Функцией Линдеберга для суммы S_n вышеопределенных случайных величин называется функция двух переменных:

$$L = L(n, \tau) ::= \frac{1}{B_n^2} \sum_{k=1}^n \int_{|y - a_k| > \tau B_n} (y - a_k)^2 dF_{\xi_k}(y), \quad n \in \mathbb{N}, \quad \tau \geq 0. \quad (5.28)$$

Согласно формулам (5.26) и (5.27)

$$B_n^2 = \sum_{k=1}^n \int_{-\infty}^{+\infty} (y - a_k)^2 dF_{\xi_k}(y).$$

Поэтому функция Линдеберга обладает следующими свойствами:

C1. $0 \leq L(n, \tau) \leq 1$.

C2. $L(n, 0) = 1$, $L(n, +\infty) = 0$.

C3. $L(n, \tau)$ – невозрастающая по переменной τ функция.

Определение 5.5. Говорят, что случайная последовательность ξ_k удовлетворяет условию Линдеберга, если

$$\forall \tau > 0, L(n, \tau) \xrightarrow{n \rightarrow \infty} 0. \quad (5.29)$$

Упростим обозначения. Для этого введем нормированные случайные величины:

$$\xi_{kn} = \frac{\xi_k - a_k}{B_n}, \quad E\{\xi_{kn}\} = 0, \quad D\{\xi_{kn}\} = \frac{\sigma_k^2}{B_n^2}, \quad k = 1, \dots, n.$$

Рассмотрим нормированные частичные суммы:

$$S_n^0 = \sum_{k=1}^n \xi_{kn} = \frac{S_n - A_n}{B_n}, \quad E\{S_n^0\} = 0, \quad D\{S_n^0\} = 1.$$

В функции Линдеберга (5.28) сделаем замену переменных: $(y - a_k)/B_n = x$ и получим

$$L = L(n, \tau) = \sum_{k=1}^n \int_{|x| > \tau} x^2 dF_{\xi_{kn}}(x) = \sum_{k=1}^n E\{\xi_{kn}^2; |\xi_{kn}| > \tau\}, \quad n \in \mathbb{N}, \tau \geq 0. \quad (5.30)$$

Оказывается, что при увеличении n распределение $\mathcal{L}\{S_n^0\}$ приближается к гауссовскому распределению, хотя слагаемые имеют произвольное распределение вероятностей. Это явление называют «чудом Лапласа». Основным условием выполнения «чуда Лапласа» является условие Линдеберга (5.29). Выясним его вероятностный смысл.

Лемма 5.5 (вероятностный смысл условия Линдеберга). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\{\xi_k\}$, имеющих конечные математические ожидания $\{a_k\}$ и конечные дисперсии $\{\sigma_k^2\}$. Тогда если выполняется условие Линдеберга (5.29), то справедливы следующие два свойства:

1) имеет место равномерная малость дисперсий слагаемых в сумме S_n :

$$\max_{1 \leq k \leq n} \frac{\sigma_k^2}{B_n^2} = \max_{1 \leq k \leq n} D\{\xi_{kn}\} \xrightarrow{n \rightarrow \infty} 0; \quad (5.31)$$

2) нормированные случайные величины $\{\xi_{kn}\}$ группируются в окрестности точки $x = 0$ так, что выполняется соотношение

$$\forall \tau > 0 \quad P\left\{\max_{1 \leq k \leq n} |\xi_{kn}| > \tau\right\} \xrightarrow{n \rightarrow \infty} 0. \quad (5.32)$$

Доказательство. Докажем (5.31), (5.32) по отдельности.

1) По построению $\{\xi_{kn}\}$ и с учетом (5.30) имеем

$$\begin{aligned} \max_{1 \leq k \leq n} D\{\xi_{kn}\} &= \max_{1 \leq k \leq n} E\{\xi_{kn}^2\} \equiv \\ &\equiv \max_{1 \leq k \leq n} (E\{\xi_{kn}^2; |\xi_{kn}| \leq \tau\} + E\{\xi_{kn}^2; |\xi_{kn}| > \tau\}) \leq \\ &\leq \tau^2 + \max_{1 \leq k \leq n} E\{\xi_{kn}^2; |\xi_{kn}| > \tau\} \leq \tau^2 + \mathcal{L}(n, \tau) \xrightarrow{n \rightarrow \infty} \tau^2. \end{aligned}$$

Отсюда в силу произвола τ получаем (5.31).

2) По свойствам вероятности и с учетом (5.30) имеем

$$\begin{aligned} P\left\{\max_{1 \leq k \leq n} |\xi_{kn}| > \tau\right\} &\equiv P\left(\bigcup_{k=1}^n \{|\xi_{kn}| > \tau\}\right) \leq \\ &\leq \sum_{k=1}^n P\{|\xi_{kn}| > \tau\} = \sum_{k=1}^n \int_{|x| > \tau} 1 dF_{\xi_{kn}}(x) \leq \\ &\leq \frac{1}{\tau^2} \sum_{k=1}^n E\{\xi_{kn}^2; |\xi_{kn}| > \tau\} = \frac{1}{\tau^2} L(n, \tau) \xrightarrow{n \rightarrow \infty} 0, \forall \tau. \end{aligned}$$

Отсюда следует выполнение (5.32). □

5.5. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ЛИНДЕБЕРГА – ФЕЛЛЕРА

Лемма 5.6. Для любых $\alpha \in \mathbb{R}$, $\beta \in \mathbb{C}$, $|\beta| \leq 1/2$ справедливы неравенства:

$$\begin{aligned} |e^{i\alpha} - 1| &\leq |\alpha|; \\ |e^{i\alpha} - 1 - i\alpha| &\leq \frac{\alpha^2}{2}; \\ |e^{i\alpha} - 1 - i\alpha + \frac{\alpha^2}{2}| &\leq \frac{|\alpha|^3}{6}; \\ |\ln(1 + \beta) - \beta| &\leq |\beta|^2. \end{aligned}$$

Доказательство. Проводится с использованием формулы Тейлора и представлением остаточного члена в форме Лагранжа. □

Теорема 5.12 (ЦПТ Линдеберга). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин ξ_k , $k = 1, 2, \dots$, имеющих конечные математические ожидания $E\{\xi_k\} = a_k$ и дисперсии $D\{\xi_k\} = \sigma_k^2$. Тогда если выполняется условие Линдеберга (5.29), то выполняется условие равномерной малости дисперсий (5.31) и при $n \rightarrow \infty$ последовательность функций распределения нормированных сумм $S_n^0 = (S_n - A_n)/B_n$ случайных величин сходится к функции распределения стандартного нормального закона:

$$\forall x \in \mathbb{R} \quad F_{S_n^0}(x) = P\{S_n^0 < x\} \rightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz. \quad (5.33)$$

Замечание 5.13. Стандартный нормальный закон $N(0, 1)$ – закон распределения вероятностей некоторой случайной величины $\xi \in \mathbb{R}$ с функцией распределения $\Phi(x)$, плотностью распределения $\phi(x) = \Phi'(x)$ и характеристической функцией $f(t) = e^{-\frac{t^2}{2}}$.

Замечание 5.14. $C(\Phi) = \mathbb{R}$, следовательно, формула (5.33) эквивалентна слабой сходимости $F_{S_n^0}(\cdot) \Rightarrow \Phi(\cdot)$, эквивалентной, в свою очередь, сходимости по распределению $S_n^0 \xrightarrow{D} \xi$.

Доказательство. Состоит из четырех этапов.

Этап 1. Из (5.29) в силу леммы 5.5 следует (5.31), т. е. имеет место равномерная малость дисперсий.

Этап 2. Вместо (5.33) будем доказывать эквивалентное утверждение о сходимости последовательности характеристических функций:

$$f_{S_n^0}(t) \xrightarrow{n \rightarrow \infty} f(t) = e^{-\frac{t^2}{2}}, \quad t \in \mathbb{R}. \quad (5.34)$$

Предельная функция $f(t)$ непрерывна в точке $t = 0$; и отсюда по теореме непрерывности для характеристической функции будет следовать эквивалентность (5.33) и (5.34). В свою очередь, (5.34) эквивалентно сходимости последовательности кумулянтных функций:

$$\ln f_{S_n^0}(t) \xrightarrow{n \rightarrow \infty} -\frac{t^2}{2}, \quad t \in \mathbb{R}. \quad (5.35)$$

Итак, вместо (5.33) будем доказывать (5.35).

Этап 3. Построим асимптотическое разложение для кумулянтной функции в произвольной точке $t \in \mathbb{R}$, используя свойства характеристической функции и условие Линдеберга:

$$\ln f_{S_n^0}(t) = \ln \prod_{k=1}^n f_{\xi_{kn}}(t) = \sum_{k=1}^n \ln(1 + \Delta_{kn}(t)), \quad (5.36)$$

где

$$\Delta_{kn}(t) ::= f_{\xi_{kn}}(t) - 1 = \mathbf{E}\{e^{it\xi_{kn}} - 1\}.$$

Исследуем асимптотику $\Delta_{kn}(t)$ при $n \rightarrow \infty$. Построим оценку, используя лемму 5.6:

$$|\Delta_{kn}(t)| \equiv |\mathbf{E}\{e^{it\xi_{kn}} - 1 - it\xi_{kn}\}| \leq \mathbf{E}\left\{\frac{t^2}{2}\xi_{kn}^2\right\} = \frac{t^2}{2} \cdot \frac{\sigma_k^2}{B_n^2}. \quad (5.37)$$

Из формулы (5.37) и условия (5.31) имеем

$$\max_{1 \leq k \leq n} |\Delta_{kn}(t)| \leq \frac{t^2}{2} \max_{1 \leq k \leq n} \frac{\sigma_k^2}{B_n^2} \rightarrow 0.$$

Отсюда следует, что $\exists \bar{n} < \infty : \forall n > \bar{n}, \forall k \in \{1, \dots, n\} \quad |\Delta_{kn}(t)| < 1/2$.

В (5.36) воспользуемся линейной формулой Тейлора

$$\ln f_{S_n^0}(t) = \sum_{k=1}^n \Delta_{kn}(t) + r_n, \quad t \in \mathbb{R},$$

где с учетом (5.37) для остаточного члена справедлива следующая оценка:

$$\begin{aligned} |r_n| &= \left| \sum_{k=1}^n (\ln(1 + \Delta_{kn}(t)) - \Delta_{kn}(t)) \right| \leq \sum_{k=1}^n |\Delta_{kn}(t)| \cdot |\Delta_{kn}(t)| \leq \\ &\leq \max_{1 \leq k \leq n} |\Delta_{kn}(t)| \cdot \frac{t^2}{2} \cdot \sum_{k=1}^n \frac{\sigma_k^2}{B_n^2} = \frac{t^2}{2} \max_{1 \leq k \leq n} |\Delta_{kn}(t)| \rightarrow 0. \end{aligned}$$

Таким образом, $r_n \xrightarrow{n \rightarrow \infty} 0, \forall t \in \mathbb{R}$.

Этап 4. Оценим по модулю разность между левой и правой частями (5.35):

$$\begin{aligned} \Delta_n &= \left| \ln f_{S_n^0}(t) - \left(-\frac{t^2}{2} \right) \right| \equiv \left| r_n + \sum_{k=1}^n \mathbf{E} \{ e^{it\xi_{kn}} - 1 - it\xi_{kn} \} + \frac{t^2}{2} \times \sum_{k=1}^n \mathbf{E} \{ \xi_{kn}^2 \} \right| \leq \\ &\leq |r_n| + \sum_{k=1}^n \mathbf{E} \left\{ \left| e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2} \right| \right\}. \end{aligned}$$

Выберем $\forall \tau > 0$ и воспользуемся леммой 5.6:

$$\begin{aligned} \Delta_n &\leq |r_n| + \sum_{k=1}^n \mathbf{E} \left\{ \left| e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2} \right|; |\xi_{kn}| \leq \tau \right\} + \\ &+ \sum_{k=1}^n \mathbf{E} \left\{ \left| e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2} \right|; |\xi_{kn}| > \tau \right\} \leq \\ &\leq |r_n| + \sum_{k=1}^n \mathbf{E} \left\{ \frac{|t|^3 \tau \xi_{kn}^2}{6}; |\xi_{kn}| \leq \tau \right\} + \\ &+ \sum_{k=1}^n \mathbf{E} \left\{ \left| e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2} \right|; |\xi_{kn}| > \tau \right\} \leq \\ &\leq |r_n| + \frac{|t|^3}{6} \tau \cdot 1 + 2 \cdot \frac{t^2}{2} \cdot L(n, \tau) + \frac{|t|^3 \tau}{6}, \quad \forall n > \bar{n}, \forall t \in \mathbb{R}, \forall \tau > 0. \end{aligned} \quad (5.38)$$

Устремим в (5.38) $n \rightarrow \infty$ и учтем условие Линдеберга (5.29). В результате из (5.38) получаем

$$\lim_{n \rightarrow \infty} \Delta_n \leq \frac{|t|^3 \tau}{6}, \quad \forall t \in \mathbb{R}, \tau > 0.$$

В силу произвола τ заключаем: $\Delta_n \rightarrow 0$. Следовательно, выполняется соотношение (5.35), а значит, и (5.33). \square

Теорема 5.13 (ЦПТ Феллера). Пусть на (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\{\xi_k\}$, имеющих конечные математические ожидания $\{a_k\}$ и дисперсии $\{\sigma_k^2\}$. Тогда при $n \rightarrow \infty$, если выполнено условие равномерной малости дисперсий (5.31) и имеет место сходимость последовательности функций распределения нормированных сумм к стандартному нормальному закону (5.33), то выполняется условие Линдеберга (5.29).

Замечание 5.15. Теоремы Линдеберга и Феллера и лемму 5.6 обычно объединяют в одну теорему, которая называется центральной предельной теоремой Линдеберга – Феллера.

Замечание 5.16. Центральная предельная теорема допускает обобщение в двух направлениях:

- 1) для слабовзависимых случайных величин $\{\xi_k\}$;
- 2) на случай суммы случайных векторов (многомерных случайных величин).

Замечание 5.17. ЦПТ интенсивно используется в математической статистике.

5.6. ЧАСТНЫЕ СЛУЧАИ ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЫ

Определение 5.6. Если выполняется сходимость (5.33) к нормальному закону, то принято говорить, что сумма $S_n = \sum_{k=1}^n \xi_k$ при $n \rightarrow \infty$ распределена асимптотически

нормально с математическим ожиданием $A_n = \sum_{k=1}^n a_k$ и дисперсией $B_n^2 = \sum_{k=1}^n \sigma_k^2$.

Теорема 5.14 (теорема Ляпунова). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности случайных величин $\xi_k, k \in \mathbb{N}$, имеющих конечные математические ожидания $E\{\xi_k\} = a_k$ и дисперсии $D\{\xi_k\} = \sigma_k^2$. Пусть далее для некоторого $\delta > 0$ существуют конечные абсолютные моменты порядка $2 + \delta$: $c_k^{2+\delta} := E\{|\xi_k - a_k|^{2+\delta}\}$ и определены суммы моментов: $B_n^2 = \sum_{k=1}^n \sigma_k^2$, $C_n^{2+\delta} = \sum_{k=1}^n c_k^{2+\delta}$. Тогда при $n \rightarrow \infty$, если выполнено условие Ляпунова, т. е. если стремится к нулю дробь Ляпунова:

$$\frac{C_n}{B_n} \xrightarrow{n \rightarrow \infty} 0,$$

то последовательность сумм $S_n = \sum_{k=1}^n \xi_k$ распределена асимптотически нормально с математическим ожиданием $A_n = \sum_{k=1}^n a_k$ и дисперсией B_n^2 :

$$F_{S_n^0}(x) = P\left\{\frac{S_n - A_n}{B_n} < x\right\} \rightarrow \Phi(x), \quad x \in \mathbb{R}. \quad (5.39)$$

Доказательство. Проверим условие Линдеберга и воспользуемся центральной предельной теоремой Линдеберга:

$$\begin{aligned} L(n, \tau) &::= \frac{1}{B_n^2} \sum_{k=1}^n E\{|\xi_k - a_k|^2 \cdot 1; |\xi_k - a_k| > \tau \cdot B_n\} \leq \\ &\leq \frac{1}{B_n^2} \sum_{k=1}^n E\left\{|\xi_k - a_k|^2 \cdot \left(\frac{|\xi_k - a_k|}{\tau B_n}\right)^\delta; |\xi_k - a_k| > \tau B_n\right\} \leq \frac{1}{\tau^\delta} \cdot \frac{C_n^{2+\delta}}{B_n^{2+\delta}} \rightarrow 0, \quad \tau > 0. \end{aligned}$$

Следовательно, по центральной предельной теореме Линдеберга заключаем справедливость (5.39). \square

Следствие 5.5 (неравенство Берри – Эссеена). Справедливо следующее неравенство:

$$\sup_{x \in \mathbb{R}} |F_{S_n^0}(x) - \Phi(x)| \leq \lambda \cdot \left(\frac{C_n}{B_n}\right)^{2+\delta}, \quad \frac{1}{\sqrt{2\pi}} < \lambda < 1.$$

Теорема 5.15 (ЦПТ Леви – Линдеберга). Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности, одинаково распределенных случайных величин $\xi_k, k \in \mathbb{N}$, имеющих конечное математическое ожидание $E\{\xi_1\} = a$ и конечную дисперсию $D\{\xi_1\} = \sigma^2$. Тогда при $n \rightarrow \infty$ имеет место асимптотическая нормальность сумм этих случайных величин:

$$P\left\{\frac{S_n - na}{\sqrt{n}\sigma} < x\right\} \rightarrow \Phi(x), \quad x \in \mathbb{R}. \quad (5.40)$$

Доказательство. Проверим условие Линдеберга, учитывая одинаковую распределенность случайных величин:

$$\begin{aligned} A_n &= n \cdot a, \quad B_n^2 = n \cdot \sigma^2, \quad L(n, \tau) = \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{|y-a| > \tau\sigma\sqrt{n}} (y-a)^2 dF_{\xi_1}(y) = \\ &= \frac{1}{\sigma^2} \int_{|y-a| > \tau\sigma\sqrt{n}} (y-a)^2 dF_{\xi_1}(y) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

По центральной предельной теореме Линдеберга заключаем справедливость (5.40). \square

Теорема 5.16 (теорема Муавра – Лапласа). Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) определена последовательность независимых в совокупности одинаково распределенных случайных величин Бернулли: $\xi_k = \xi_k(\omega)$, $k \in \mathbb{N}$, $P\{\xi_k = 1\} = p$, $P\{\xi_k = 0\} = 1 - p$, $0 < p < 1$. Тогда сумма $S_n = \sum_{k=1}^n \xi_k$ этих случайных величин Бернулли при $n \rightarrow \infty$ распределена асимптотически нормально с математическим ожиданием $A_n = np$ и дисперсией $B_n^2 = np(1-p)$:

$$F_{S_n^0}(x) = P\left\{ \frac{S_n - np}{\sqrt{np(1-p)}} < x \right\} \rightarrow \Phi(x), \quad x \in \mathbb{R}. \quad (5.41)$$

Доказательство. Проверим условия теоремы Леви – Линдеберга. По свойствам распределения Бернулли: $\exists a = E\{\xi_k\} = p$, $\exists \sigma^2 = D\{\xi_k\} = p(1-p) < \infty$. Таким образом, соотношение (5.41) вытекает из предыдущей теоремы. \square

Следствие 5.6. Пусть m – случайное число наступлений некоторого случайного события A в схеме n независимых испытаний Бернулли, в каждом из которых событие A наступает с одной и той же вероятностью p . Тогда справедливо следующее предельное соотношение:

$$P\left\{ \frac{m - np}{\sqrt{np(1-p)}} < x \right\} \xrightarrow{n \rightarrow \infty} \Phi(x), \quad x \in \mathbb{R}.$$

Следствие 5.7. Справедлива следующая формула для приближенных вычислений:

$$P\{X \leq m < Y\} \approx \Phi\left(\frac{Y - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{X - np}{\sqrt{np(1-p)}}\right),$$

причем точность приближения зависит от $X, Y \in \mathbb{R}$ ($X < Y$) и возрастает при увеличении числа экспериментов n .

Доказательство. Основано на следствии 5.6 и следующем тождестве:

$$P\{X \leq m < Y\} \equiv P\left\{ \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{m - np}{\sqrt{np(1-p)}} < \frac{Y - np}{\sqrt{np(1-p)}} \right\}.$$

\square

5.7. СХЕМА СЕРИЙ. СХОДИМОСТЬ К ЗАКОНАМ ГАУССА И ПУАССОНА

Определение 5.7. Принято говорить, что множество случайных величин $\Xi = \{\xi_{kn}\}$ задано в схеме серий, если это множество устроено следующим образом:

первая серия: ξ_{11} ;

вторая серия: ξ_{12}, ξ_{22} ;

...

n -я серия: $\xi_{1n}, \xi_{2n}, \dots, \xi_{nn}$. Причем внутри каждой серии случайные величины независимы в совокупности и одинаково распределены.

Примером схемы серий служит множество нормированных величин в теореме Леви – Линдберга: $\xi_{kn} = \frac{\xi_k - a}{\sqrt{n}\sigma}$.

Основной вопрос – это вопрос о предельном распределении $\mathcal{L}\{S_n\}$, где $S_n = \sum_{k=1}^n \xi_{kn}$.

Рассмотрим частный случай, когда $\{\xi_{kn}\}$ – случайные величины Бернулли:

$$\xi_{kn} \in \{0, 1\}, \quad P\{\xi_{kn} = 1\} = p_n, \quad P\{\xi_{kn} = 0\} = 1 - p_n, \quad k = 1, \dots, n. \quad (5.42)$$

Выделим две ситуации:

1. Пусть $p_n = p = \text{const}$, $0 < p < 1$, $np \rightarrow \infty$. Тогда выполнены условия теоремы Муавра – Лапласа, и в этом случае $\mathcal{L}\{S_n\}$ сходится к закону Гаусса:

$$P\left\{\frac{S_n - np}{\sqrt{np(1-p)}} < x\right\} \xrightarrow{n \rightarrow \infty} \Phi(x), \quad x \in \mathbb{R}.$$

2. Пусть $p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \xrightarrow{n \rightarrow \infty} 0$, $0 < \lambda < \infty$, $np_n \rightarrow \lambda$.

Теорема 5.17 (теорема Пуассона). Пусть $\Xi = \{\xi_{kn}\}$ – определенная выше схема серий случайных величин Бернулли (5.42), причем вероятность успеха

$$p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right), \quad 0 < \lambda < \infty, \quad np_n \rightarrow \lambda.$$

Тогда при $n \rightarrow \infty$ распределение вероятностей сумм $S_n = \sum_{k=1}^n \xi_{kn}$ этих случайных величин сходится к распределению Пуассона $\Pi(\lambda)$ с параметром λ :

$$F_{S_n}(x) = P\{S_n < x\} \rightarrow F_{\Pi}(x, \lambda) = \sum_{k=0}^{x-1} \frac{\lambda^k e^{-\lambda}}{k!}, \quad x \in \mathbb{N}. \quad (5.43)$$

Доказательство. Воспользуемся той же идеей, что и при доказательстве центральной предельной теоремы Линдберга – Феллера, а именно воспользуемся теоремой непрерывности для характеристических функций. Имеем

$$f_{\xi_{kn}}(t) = E\{e^{it\xi_{kn}}\} = 1 + p_n(e^{it} - 1); \quad f_{\Pi}(t, \lambda) = e^{\lambda(e^{it} - 1)}. \quad (5.44)$$

Воспользуемся свойством характеристической функции:

$$f_{S_n}(t) = (f_{\xi_{1n}}(t))^n = \left(1 + \frac{\lambda(e^{it} - 1)}{n} + o\left(\frac{1}{n}\right)\right) \xrightarrow{n \rightarrow \infty} e^{\lambda(e^{it} - 1)} = f_{\Pi}(t, \lambda), \quad t \in \mathbb{R}.$$

Характеристическая функция Пуассона (5.44) непрерывна в точке $t = 0$. Поэтому по теореме непрерывности для характеристических функций из установленной сходимости следует формула (5.43), т. е. сходимость последовательности функций распределения. \square

Следствие 5.8. *Справедливо следующее предельное соотношение:*

$$P\{S_n = x\} \xrightarrow{n \rightarrow \infty} \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in 0, 1, \dots$$

Возникает вопрос, какие законы распределения вероятностей могут быть предельными для $\mathcal{L}\{S_n\}$, $n \rightarrow \infty$. Если предельное распределение $\mathcal{L}\{S_n\}$, $n \rightarrow \infty$, существует, то оно относится к классу безграничных делимых законов распределения вероятностей.

Определение 5.8. *Принято говорить, что случайная величина ξ имеет безгранично делимый закон распределения вероятностей, если для $\forall n \in \mathbb{N}$ найдется такое разложение случайной величины ξ в виде $\xi \stackrel{D}{=} \xi_{1n} + \dots + \xi_{nn}$, причем все слагаемые независимы в совокупности и имеют одинаковый закон распределения того же типа, что и ξ , отличающийся лишь значением параметра. Другими словами: $\sqrt[n]{f_\xi(t)}$ – характеристическая функция того же типа, что и $f_\xi(t)$, и отличается лишь значением параметра.*

Например, рассмотрим $\sqrt[n]{f_\Pi(t, \lambda)} = e^{\frac{\lambda}{n}(e^{it} - 1)} = f_\Pi\left(t, \frac{\lambda}{n}\right)$.

Задание 5.1. По таблице законов распределения вероятностей классифицировать все законы по признаку, являются они безгранично делимыми или нет.

5.8. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ СЛУЧАЙНЫХ ВЕКТОРОВ И СЛАБОЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Как уже отмечалось, центральная предельная теорема допускает обобщения на многомерный случай и случай слабозависимых слагаемых. Дадим краткую формулировку этих результатов.

Теорема 5.18 (ЦПТ для случайных векторов). *Пусть $X_1, X_2, \dots \in \mathbb{R}^N$ – определенные на вероятностном пространстве (Ω, \mathcal{F}, P) , независимые в совокупности случайные N -векторы, одинаково распределенные и имеющие математическое ожидание $E\{X_k\} = \mu \in \mathbb{R}^N$ и ковариационную $(N \times N)$ -матрицу Σ . Тогда случайный N -вектор*

$$\frac{1}{\sqrt{n}} \Sigma^{-\frac{1}{2}} \sum_{k=1}^n (X_k - \mu)$$

при $n \rightarrow \infty$ распределен асимптотически нормально по N -мерному стандартному нормальному закону $N_N(0, I_N)$.

Доказательство. Укажем лишь схему доказательства.

На первом этапе, используя теорему непрерывности для характеристических функций, необходимо доказать следующий результат.

Пусть на (Ω, \mathcal{F}, P) определены случайные N -векторы $X, X_1, X_2, \dots \in \mathbb{R}^N$. Сходимость по распределению последовательности случайных векторов $X_n \xrightarrow[n \rightarrow \infty]{D} X$ имеет место тогда и только тогда, когда для любого N -вектора $a \in \mathbb{R}^N$ имеет место сходимость последовательности случайных величин $a^T X_n$:

$$\forall a \in \mathbb{R}^N \quad a^T X_n \xrightarrow[n \rightarrow \infty]{D} a^T X.$$

На втором этапе для случайных величин $\{a^T X_n\}$ используется центральная предельная теорема Линдеберга для одинаково распределенных случайных величин. \square

Приведем теперь вариант центральной предельной теоремы для слабовзависимых случайных величин – случайных величин с сильным перемешиванием.

Пусть X_1, X_2, \dots – последовательность случайных величин с $\mathbf{E}\{X_j\} = 0, j \geq 1$, \mathfrak{S}_a^b – σ -алгебра, порожденная случайными величинами, $X_j, j \in [a, b]$,

$$\alpha(\tau) = \sup_{t, A \in \mathfrak{S}_0^t, B \in \mathfrak{S}_{t+\tau}^\infty} |P(AB) - P(A)P(B)| -$$

коэффициент перемешивания по Розенблатту. Обозначим

$$\sigma_n^2 = \mathbf{E} \left\{ \left(\sum_{j=1}^n X_j \right)^2 \right\}, \quad S_n = \frac{1}{\sigma_n} \sum_{j=1}^n X_j,$$

$$\Delta_n = \sup_z \left| P(S_n < z) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx \right|.$$

Теорема 5.19 (ЦПТ для слабовзависимых СВ [14]). Пусть существуют такие постоянные $C > 0, C_0 > 0, K > 0, 0 < \delta \leq 1, \beta > 1$, что

$$\sup_{k \leq n} \mathbf{E} \left\{ |X_k|^{2+\delta} \right\} \leq C, \quad \sigma_n^2 \geq C_0 n, \quad \alpha(\tau) \leq K \tau^{-\frac{2+\delta}{\delta} \beta}.$$

Тогда $\Delta_n \leq B n^{-\frac{\delta}{2} \frac{\beta-1}{\beta+2\delta}}$, где ограниченная величина B зависит только от C, C_0, K, δ, β .

5.9. УПРАЖНЕНИЯ

1. Пусть X_1, X_2, \dots – независимые случайные величины такие, что $P\{X_n = n\} = P\{X_n = -n\} = \frac{1}{2n \log n}$, $P\{X_n = 0\} = 1 - \frac{1}{n \log n}$. Показать, что для $\{X_n\}$ выполняется закон больших чисел, но не выполняется усиленный закон больших чисел.
2. Построить последовательность $\{X_r : r \geq 1\}$ независимых случайных чисел с нулевым математическим ожиданием такую, что $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{\text{п. н.}} -\infty$.
3. Пусть случайные величины X_1, X_2, \dots таковы, что $D\{X_n\} < c$ для всех $n \in \mathbb{N}$ и некоторой константы $0 < c < \infty$. Показать, что $\frac{1}{n} \sum_{k=1}^n (X_i - \mathbf{E}\{X_i\}) \xrightarrow[n \rightarrow \infty]{P} 0$, если коэффициенты корреляции удовлетворяют хотя бы одному из следующих условий:
а) $\text{corr}\{X_i, X_j\} \leq 0$ для всех $i \neq j$;
б) $\text{corr}\{X_i, X_j\} \rightarrow 0$, если $|j - i| \rightarrow \infty$.
4. Пусть X_1, X_2, \dots – независимые, одинаково распределенные случайные величины с плотностью распределения вероятностей $p(x) = \begin{cases} 0, & |x| \leq 2, \\ \frac{c}{x^2 \log |x|}, & |x| > 2, \end{cases}$ где c – нормировочная константа. Показать, что $\mathbf{E}\{X_i\}$ не существует, но $n^{-1} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} 0$. Показать, что усиленный закон больших чисел не имеет места.

5. Пусть X_1, X_2, \dots – независимые дискретные случайные величины. $P\{X_k = k\} = P\{X_k = -k\} = 1/2k^2$, $P\{X_k = 1\} = P\{X_k = -1\} = 1/2(1 - 1/k^2)$, $S_n = \sum_{k=1}^n X_k$. Показать, что $\mathcal{L}\{S_n/\sqrt{n}\} \xrightarrow{n \rightarrow \infty} N(0, 1)$, но $D\{S_n/\sqrt{n}\} \xrightarrow{n \rightarrow \infty} 2$.
6. Пусть X_1, X_2, \dots – независимые, одинаково распределенные случайные величины, $P\{X_i = 1\} = P\{X_i = -1\} = 1/2$. Сходится ли почти наверное при $n \rightarrow \infty$ ряд $\sum_{i=1}^n X_i/i$?
7. Пусть X_1, X_2, \dots – случайная последовательность, для которой выполняется закон больших чисел. Будет ли выполняться закон больших чисел для последовательности $|X_1|, |X_2|, \dots$?
8. Пусть X_1, X_2, \dots – случайная последовательность, $E\{X_i\} = 0$ ($i = 1, 2, \dots$), причем $X_n \xrightarrow[n \rightarrow \infty]{P} 0$. Выполняется ли закон больших чисел $n^{-1} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} 0$?
9. Пусть X_1, X_2, \dots – независимые, одинаково распределенные случайные величины, $E\{|X_i|^r\} < \infty$, где $1 \leq r < 2$. Показать, что $\frac{1}{\sqrt[n]{n}} \sum_{i=1}^n (X_i - E\{X_i\}) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0$.
10. Пусть X_1, X_2, \dots – независимые, одинаково распределенные случайные величины, $E\{|X_i|^r\} < \infty$, где $0 < r < 1$. Показать, что $\frac{1}{n^{1/r}} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\text{п. н.}} 0$.
11. Пусть X_1, X_2, \dots – независимые, одинаково распределенные случайные величины, $\mathcal{L}\{X_i\} = \Pi(\lambda)$. Найти предел: $\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n (X_{2i-1} - X_{2i})}{\sqrt{n}} < x\right\}$, $x \in \mathbb{R}$.
12. Пусть X_1, X_2, \dots – независимые случайные величины, $\mathcal{L}\{X_i\} = N(0, \sigma_i^2)$, $\sigma_1^2 = 1$, $\sigma_i^2 = 2^{i-2}$. Показать, что условие Линдеберга не выполнено, но тем не менее ЦПТ имеет место.
13. Пусть ξ_1, ξ_2, \dots – независимые, одинаково распределенные случайные величины с нулевыми математическими ожиданиями и единичными дисперсиями. Доказать, что величины $\eta_n = \sqrt{n} \frac{\xi_1 + \dots + \xi_n}{\xi_1^2 + \dots + \xi_n^2}$ и $\varsigma_n = \sqrt{n} \frac{\xi_1 + \dots + \xi_n}{\sqrt{\xi_1^2 + \dots + \xi_n^2}}$ асимптотически нормально распределены при $n \rightarrow \infty$.
14. Случайные величины ξ_1, ξ_2, \dots независимы, причем $\{\xi_{2k-1}, k = 1, 2, \dots\}$ одинаково распределены и $\{\xi_{2k}, k = 1, 2, \dots\}$ одинаково распределены, $D\{\xi_1\} > 0$, $D\{\xi_2\} > 0$, $E\{\xi_1^2\} < \infty$, $E\{\xi_2^2\} < \infty$. Положим $S_n = \xi_1 + \dots + \xi_n$. Доказать, что при $n \rightarrow \infty$ $\eta_n = (S_n - E\{S_n\})/\sqrt{D\{S_n\}}$ имеет нормальное распределение.
15. Пусть Y_1, Y_2, \dots – независимые, одинаково распределенные случайные величины, каждая из которых равномерно распределена на $\{0, 1, \dots, 9\}$. Пусть $X_n = \sum_{i=1}^n Y_i 10^{-i}$ – десятичная дробь, построенная с помощью n случайных цифр. Используя характеристическую функцию, показать сходимость к стандартному равномерному закону: $\mathcal{L}\{X_n\} \xrightarrow[n \rightarrow \infty]{} \mathcal{R}[0, 1]$. Показать, что $X_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} Y$ для некоторой случайной величины Y , $\mathcal{L}\{Y\} = \mathcal{R}[0, 1]$.

ЧАСТЬ 2

ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ

Глава 6

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И ОБЩИЕ СВОЙСТВА СЛУЧАЙНЫХ ПРОЦЕССОВ

6.1. ОПРЕДЕЛЕНИЕ И ЗАДАНИЕ СЛУЧАЙНЫХ ПРОЦЕССОВ

Определение 6.1. *Случайным процессом (СП) $\xi(t) = \xi(t, \omega)$, $\omega \in \Omega$, $t \in T$, называется семейство случайных величин $\{\xi(t)\}$, зависящих от параметра $t \in T$ и определенных на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) .*

Если $T \subset \mathbb{R}^n$, $n > 1$, то СП $\xi(t)$ называют *случайным полем*, если T – конечное или счетное множество, то $\xi(t)$ называют *случайной последовательностью*; если T – числовой интервал (конечный или бесконечный), то $\xi(t)$ называют *случайным процессом*. Параметр t часто интерпретируется как время.

В дальнейшем будем рассматривать случай, когда $T \subset \mathbb{R}$, т. е. когда $\xi(t)$ является случайным процессом или случайной последовательностью.

Таким образом, $\xi(t) = \xi(t, \omega)$, $\omega \in \Omega$, $t \in T$, является функцией двух переменных. Функция времени $\xi(t, \omega)$, соответствующая фиксированному ω , называется *траекторией (реализацией, выборочной функцией)* СП $\xi(t)$.

Существует несколько способов задания случайного процесса.

1. *Аналитический* способ задания случайного процесса:

$$\xi(t) = g(t, \eta_1, \dots, \eta_k),$$

где $\{\eta_i\}$ – набор случайных величин.

Например, $\xi(t) = \eta \cos(\zeta t + \theta)$ – гармоническое колебание, где η, ζ, θ – случайные величины.

2. *Рекуррентный* способ задания случайного процесса:

$$\xi(t) = g(t, \xi(t-1), \dots, \xi(t-k), \eta(t), \dots, \eta(t-l)),$$

т. е. значение случайного процесса в момент t определяется значениями случайного процесса $\xi(t)$ в предыдущие моменты времени и значениями некоторых случайных величин, не зависящих от случайного процесса $\xi(t)$. Такой способ задания случайного процесса используется в основном для случайных последовательностей.

3. Задание случайного процесса *с помощью конечномерных распределений*.

Пусть $\xi(t)$, $t \in T$, – случайный процесс. Тогда для любого n и любых точек $t_1, \dots, t_n \in T$ набор случайных величин $(\xi(t_1), \dots, \xi(t_n))$ является n -мерной случайной величиной с n -мерной функцией распределения

$$\begin{aligned} F_{\xi(t_1), \dots, \xi(t_n)}(x_1, \dots, x_n) &= F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \\ &= P\{\xi(t_1) < x_1, \xi(t_2) < x_2, \dots, \xi(t_n) < x_n\} = P\left\{\bigcap_{i=1}^n (\xi(t_i) < x_i)\right\}. \end{aligned}$$

Эти функции удовлетворяют свойствам 1–5 функций распределения n -мерных случайных величин (гл. 2), а также свойству 6 (симметрии), т. е. при любой перестановке индексов $\{i_1, \dots, i_n\}$ набора $(1, \dots, n)$ справедливо равенство

$$F_{t_{i_1}, \dots, t_{i_n}}(x_{i_1}, \dots, x_{i_n}) = F_{t_1, \dots, t_n}(x_1, \dots, x_n).$$

Определение 6.2. Семейство функций $\{F_{t_1, \dots, t_n}(x_1, \dots, x_n)\}$ называется семейством конечномерных распределений случайного процесса $\xi(t)$.

Определение 6.3. Семейство функций $\{F_{t_1, \dots, t_n}(x_1, \dots, x_n)\}$ называется согласованным, если оно удовлетворяет свойствам 1–6 функций распределения n -мерных случайных величин $\forall n \in \mathbb{N}$.

Теорема 6.1 (теорема Колмогорова о согласованности). Если при любом n и любых $t_1, \dots, t_n \in T$ семейство функций $\{F_{t_1, \dots, t_n}(x_1, \dots, x_n)\}$ является согласованным, то существует вероятностное пространство (Ω, \mathcal{F}, P) и случайный процесс $\xi(t), t \in T$, определенный на этом вероятностном пространстве, для которого данное семейство функций будет семейством конечномерных распределений.

Пример 6.1. Случайный процесс $\xi(t), t \geq 0$, называется гауссовским, если для любых n и любых t_1, \dots, t_n характеристическая функция

$$f_{\xi(t_1), \dots, \xi(t_n)}(\lambda_1, \dots, \lambda_n) = E \exp \left\{ i \sum_{k=1}^n \lambda_k \xi(t_k) \right\}$$

конечномерных распределений имеет вид

$$f_{\xi(t_1), \dots, \xi(t_n)} = \exp \left\{ i \sum_{k=1}^n a(t_k) \lambda_k - \frac{1}{2} \sum_{k, m=1}^n b(t_k, t_m) \lambda_k \lambda_m \right\}, \quad (6.1)$$

где матрица $B = [b(t_k, t_m)]$ симметрическая, неотрицательно определенная. При этом $a(t) = E\{\xi(t)\}$, $b(t, u) = \text{cov}(\xi(t), \xi(u))$. Если матрица B имеет определитель $|B| \neq 0$, то n -мерная случайная величина $(\xi(t_1), \dots, \xi(t_n))$ имеет плотность распределения

$$p(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} |B|^{-\frac{1}{2}} \cdot \exp \left\{ -\frac{1}{2|B|} \sum_{k, m=1}^n b^*(t_k, t_m) (x_k - a(t_k)) (x_m - a(t_m)) \right\},$$

где $b^*(t_k, t_m)$ – алгебраическое дополнение элемента $b(t_k, t_m)$.

Замечание 6.1. Из свойств функции распределения следует, что для любого борелевского множества $B \subset \mathbb{R}^n$ $P\{(\xi(t_1), \dots, \xi(t_n)) \in B\}$ однозначно определяется с помощью функции распределения $F_{t_1, \dots, t_n}(x_1, \dots, x_n)$. Тогда, используя аксиомы вероятности, с помощью конечномерных распределений можно определить поведение случайного процесса в конечном или счетном числе точек.

Часто возникает необходимость знать вероятность того, что выборочная функция случайного процесса обладает тем или иным свойством, которое определяется ее поведением на некотором несчетном множестве точек t . К таким событиям относятся события $(\xi(t) < c, t \in T)$, $(\xi(t)$ непрерывна в точке $t_0 \in T)$, $(\xi(t)$ непрерывна на множестве $T)$ и др. Ниже (пример 6.2) покажем, что вероятность таких событий нельзя найти с помощью конечномерных распределений.

4. Задание случайного процесса моментами первого и второго порядков.

Пусть $\xi(t), t \in T$, – случайный процесс. Обозначим $a(t) = E\{\xi(t)\}$, $r(t, u) = \text{cov}(\xi(t), \xi(u))$. Функция $a(t)$ называется математическим ожиданием случайного процесса, функция $r(t, u)$ – ковариационной функцией случайного процесса.

Теорема 6.2. Ковариационная функция случайного процесса является неотрицательно определенной функцией, т. е. для любого n , любых $t_1, \dots, t_n \in T$ и любых комплексных чисел z_1, \dots, z_n справедливо неравенство

$$\sum_{k, l=1}^n r(t_k, t_l) z_k \bar{z}_l \geq 0.$$

Доказательство. Имеем:

$$\begin{aligned} & \sum_{k,l=1}^n r(t_k, t_l) z_k \bar{z}_l = \sum_{k,l=1}^n E\{(\xi(t_k) - a(t_k))(\xi(t_l) - a(t_l))\} z_k \bar{z}_l = \\ & = E \left\{ \sum_{k=1}^n (\xi(t_k) - a(t_k)) z_k \overline{\sum_{l=1}^n (\xi(t_l) - a(t_l)) z_l} \right\} = E \left\{ \left| \sum_{k=1}^n (\xi(t_k) - a(t_k)) z_k \right|^2 \right\} \geq 0. \end{aligned}$$

□

Теорема 6.3. Если функция $r(t, u)$ является неотрицательно определенной функцией, то существует случайный процесс $\xi(t)$, для которого $r(t, u)$ будет ковариационной функцией.

Доказательство. В качестве такого процесса можно взять гауссовский процесс с функцией $b(t, u) = r(t, u)$. □

6.2. ЭКВИВАЛЕНТНЫЕ, ТОЖДЕСТВЕННЫЕ И СЕПАРАБЕЛЬНЫЕ ПРОЦЕССЫ

Определение 6.4. Случайные процессы $\xi(t), \eta(t) \in T$ называются эквивалентными, если при каждом $t \in T$ $P\{\xi(t) = \eta(t)\} = 1$.

Определение 6.5. Случайные процессы $\xi(t), \eta(t) \in T$ называются тождественными, если

$$P\{\xi(t) = \eta(t), t \in T\} = P\{\omega : \xi(t, \omega) = \eta(t, \omega), t \in T\} = 1,$$

т. е. вероятность совпадений траекторий равна 1.

Теорема 6.4. У эквивалентных процессов конечномерные распределения совпадают.

Доказательство. Покажем, что для любого k и любых $t_1, \dots, t_k \in T$

$$P \left\{ \bigcap_{l=1}^k (\xi(t_l) = \eta(t_l)) \right\} = 1.$$

Поскольку

$$P \left\{ \bigcap_{l=1}^k (\xi(t_l) = \eta(t_l)) \right\} = 1 - P \left\{ \bigcup_{l=1}^k (\xi(t_l) \neq \eta(t_l)) \right\} \geq 1 - \sum_{l=1}^k P\{\xi(t_l) \neq \eta(t_l)\} = 1, \text{ то}$$

$$P \left\{ \bigcap_{l=1}^k (\xi(t_l) = \eta(t_l)) \right\} = 1.$$

Покажем равенство конечномерных распределений. Поскольку из того, что $P(B) = 1$, следует $P(A) = P(AB)$, получаем

$$\begin{aligned} F_{\xi(t_1), \dots, \xi(t_k)}(x_1, \dots, x_k) &= P \left\{ \bigcap_{l=1}^k (\xi(t_l) < x_l) \right\} = P \left\{ \bigcap_{l=1}^k (\xi(t_l) < x_l) \bigcap_{l=1}^k (\xi(t_l) = \eta(t_l)) \right\} = \\ &= P \left\{ \bigcap_{l=1}^k (\eta(t_l) < x_l) \bigcap_{l=1}^k (\xi(t_l) = \eta(t_l)) \right\} = P \left\{ \bigcap_{l=1}^k (\eta(t_l) < x_l) = F_{\eta(t_1), \dots, \eta(t_k)}(x_1, \dots, x_k) \right\}. \end{aligned}$$

□

Следствие 6.1. Для случайных последовательностей понятия эквивалентности и тождественности случайных процессов совпадают.

Пример 6.2. Рассмотрим вероятностное пространство (Ω, \mathcal{F}, P) , где $\Omega = [0, 1]$, \mathcal{F} – сигма-алгебра борелевских подмножеств $[0, 1]$, P – мера Лебега. Определим случайные процессы:

$$\xi(t, \omega) = 0, \quad t \in [0, 1], \quad \omega \in \Omega,$$

$$\eta(t, \omega) = \begin{cases} 1, & t = \omega, \\ 0, & t \neq \omega; \end{cases} \quad \zeta(t, \omega) = \begin{cases} 1, & t = \frac{\omega}{n}, n = 1, 2, \dots, \\ 0, & t \neq \frac{\omega}{n}, n = 1, 2, \dots \end{cases}$$

Поскольку

$$P\{\xi(t) \neq \eta(t)\} = P\{t\} = 0,$$

$$P\{\xi(t) \neq \zeta(t)\} = P\left\{\bigcup_{n: 0 \leq tn \leq 1} \{t_n\}\right\} = 0,$$

то случайные процессы $\xi(t)$, $\eta(t)$, $\zeta(t)$ эквивалентны. С другой стороны,

$$P\{\xi(t) < 1, t \in [0, 1]\} = P\{[0, 1]\} = 1,$$

$$P\{\eta(t) < 1, t \in [0, 1]\} = P\{\emptyset\} = 0.$$

Таким образом, по конечномерным распределениям нельзя определить вероятность $P\{\xi(t) < c\}$.

Аналогично,

$$P\{\xi(t) \text{ непрерывен справа в точке } \{0\}\} = P\{[0, 1]\} = 1;$$

$$P\{\zeta(t) \text{ непрерывен справа в точке } \{0\}\} = P\{0\} = 0;$$

$$P\{\xi(t) \text{ непрерывен на } [0, 1]\} = 1;$$

$$P\{\eta(t) \text{ непрерывен на } [0, 1]\} = 0.$$

Заметим, что случайные процессы $\xi(t)$, $\eta(t)$ и $\zeta(t)$ не являются тождественными.

Определение 6.6. Случайный процесс $\xi(t)$ называется сепарабельным, если существует такое счетное множество S , называемое множеством сепарабельности, что для любого открытого интервала I с вероятностью единица выполняются соотношения:

$$\sup_{t \in I} \xi(t) = \sup_{t \in I \cap S} \xi(t), \quad \inf_{t \in I} \xi(t) = \inf_{t \in I \cap S} \xi(t).$$

Смысл этих соотношений состоит в том, что траектории сепарабельного процесса могут быть восстановлены по их значениям на некотором счетном множестве точек.

Замечание 6.2. Для сепарабельных процессов рассмотренные выше вероятности событий могут быть найдены с помощью конечномерных распределений.

На практике сепарабельными процессами будут процессы, имеющие непрерывные траектории, кусочно-непрерывные траектории, либо процессы с непрерывными справа или слева траекториями.

6.3. НЕПРЕРЫВНОСТЬ И ДИФФЕРЕНЦИРУЕМОСТЬ СЛУЧАЙНОГО ПРОЦЕССА В СРЕДНЕМ КВАДРАТИЧНОМ

В дальнейшем будем считать, что $a(t) = E\{\xi(t)\} = 0$. В противном случае можно рассмотреть процесс $\xi(t) - a(t)$.

Определение 6.7. Случайный процесс $\xi(t)$ называется непрерывным в среднем квадратичном в точке t_0 , если $\xi(t_0) = \text{l.i.m.}_{t \rightarrow t_0} \xi(t)$.

Теорема 6.5. Для того чтобы случайный процесс $\xi(t)$ был непрерывным в среднем квадратичном в точке t_0 , необходимо и достаточно, чтобы ковариационная функция $r(t, u)$ была непрерывна в точке $t = u = t_0$. Если функция $r(t, u)$ непрерывна в точках $t = u$, то она непрерывна во всех точках.

Доказательство. Доказательство теоремы следует из критерия сходимости в среднем квадратичном: последовательность $\{\xi_n\}$ сходится в среднем квадратичном тогда и только тогда, когда

$$E\{\xi_n \xi_m\} \xrightarrow{n, m \rightarrow \infty} C, \text{ причем } C = E\{\xi^2\}, \xi = \text{l.i.m.}_{n \rightarrow \infty} \xi_n.$$

Справедливо также: если $\xi = \text{l.i.m.}_{n \rightarrow \infty} \xi_n$, $\eta = \text{l.i.m.}_{n \rightarrow \infty} \eta_n$, то $E\{\xi_n \eta_n\} \xrightarrow{n \rightarrow \infty} E\{\xi \eta\}$.

Необходимость. Пусть $\xi(t_0) = \text{l.i.m.}_{t \rightarrow t_0} \xi(t)$. Тогда

$$r(t_0 + \Delta t, t_0 + \Delta u) = E\{\xi(t_0 + \Delta t) \xi(t_0 + \Delta u)\} \xrightarrow{\Delta t \rightarrow 0, \Delta u \rightarrow 0} E\{\xi^2(t_0)\} = r(t_0, t_0),$$

т. е. функция $r(t, u)$ непрерывна в точке $t = u = t_0$.

Достаточность. Пусть функция $r(t, u)$ непрерывна в точке $t = u = t_0$. Тогда

$$E\{(\xi(t) - \xi(t_0))^2\} = r(t, t) - r(t_0, t) - r(t, t_0) + r(t_0, t_0) \xrightarrow{t \rightarrow t_0} 0,$$

т. е. случайный процесс $\xi(t)$ непрерывен в среднем квадратичном в точке t_0 . \square

Заметим, что непрерывность в среднем квадратичном не влечет за собой непрерывность выборочных функций. В качестве контрпримера можно взять процесс Пуассона, рассмотренный ниже.

Определение 6.8. Случайный процесс $\xi(t)$ называется дифференцируемым в среднем квадратичном в точке t_0 , если существует случайная величина $\xi'(t_0)$ такая, что

$$\frac{\xi(t_0 + h) - \xi(t_0)}{h} \xrightarrow[h \rightarrow 0]{\text{ср. кв.}} \xi'(t_0).$$

Случайная величина $\xi'(t_0)$ называется производной в среднем квадратичном случайного процесса $\xi(t)$ в точке t_0 .

Теорема 6.6. Для того чтобы случайный процесс $\xi(t)$ был дифференцируем в среднем квадратичном в точке t_0 , необходимо и достаточно, чтобы ковариационная функция $r(t, u)$ имела вторую производную

$$\left. \frac{\partial^2 r(t, u)}{\partial t \partial u} \right|_{t=u=t_0}.$$

Если функция $r(t, u)$ имеет вторую производную $\frac{\partial^2 r(t, u)}{\partial t \partial u}$ в точках $t = u$, то она имеет вторую производную во всех точках, причем эта производная является ковариационной функцией случайного процесса $\xi'(t)$.

Доказательство. Существование производной в среднем квадратичном случайного процесса в точке t_0 эквивалентно тому, что

$$E \left\{ \frac{\xi(t_0 + \Delta t) - \xi(t_0)}{\Delta t} \cdot \frac{\xi(t_0 + \Delta u) - \xi(t_0)}{\Delta u} \right\} = \frac{1}{\Delta t \Delta u} (r(t_0 + \Delta t, t_0 + \Delta u) - r(t_0 + \Delta t, t_0) - r(t_0, t_0 + \Delta u) + r(t_0, t_0)) \xrightarrow[\Delta u \rightarrow 0]{\Delta t \rightarrow 0} E\{\xi'(t_0)\}^2 \quad (6.2)$$

по критерию сходимости. Но левая часть (6.2) стремится к $\frac{\partial^2 r(t_0, t_0)}{\partial t \partial u}$.

Из свойств сходимости в среднем квадратичном получаем: если существует $\frac{\partial^2 r(t, u)}{\partial t \partial u}$ в точках $t = u$, то для любых точек t_0, u_0 существуют производные в среднем квадратичном $\xi'(t_0)$ и $\xi'(u_0)$ и

$$\begin{aligned} & \mathbf{E} \left\{ \frac{\xi(t_0 + \Delta t) - \xi(t_0)}{\Delta t} \cdot \frac{\xi(u_0 + \Delta u) - \xi(u_0)}{\Delta u} \right\} = \\ &= \frac{1}{\Delta t \Delta u} (r(t_0 + \Delta t, u_0 + \Delta u) - r(t_0 + \Delta t, u_0) - r(t_0, u_0 + \Delta u) + \\ & \quad + r(t_0, u_0)) \xrightarrow[\Delta u \rightarrow 0]{\Delta t \rightarrow 0} \mathbf{E} \{ \xi'(t_0) \xi'(u_0) \}. \end{aligned} \quad (6.3)$$

Но левая часть (6.3) стремится к $\frac{\partial^2 r(t_0, u_0)}{\partial t \partial u}$.

Отсюда получаем

$$\frac{\partial^2 r(t_0, u_0)}{\partial t \partial u} = \mathbf{E} \{ \xi'(t_0) \xi'(u_0) \} = r_{\xi'(t)}(t_0, u_0).$$

□

6.4. СТОХАСТИЧЕСКИЕ ИНТЕГРАЛЫ В СРЕДНЕМ КВАДРАТИЧНОМ

Пусть $\xi(t)$ – случайный процесс с ковариационной функцией $r(t, u)$ и $g(t)$ – некоторая неслучайная борелевская функция.

Определим интегралы вида

$$\begin{aligned} I_1 &= \int_a^b \xi(t) g(t) dt, \\ I_2 &= \int_a^b g(t) d\xi(t), \end{aligned}$$

где $[a, b]$ – конечный или бесконечный интервал.

Рассмотрим вначале случай, когда $-\infty < a < b < \infty$. Пусть $a < t_0 < t_1 < \dots < t_n < b$.

$$\lambda_n = \max_k (t_k - t_{k-1});$$

$$S_1 = \sum_{k=1}^n g(\tilde{t}_k) \xi(\tilde{t}_k) (t_{k+1} - t_k);$$

$$S_2 = \sum_{k=1}^n g(\tilde{t}_k) (\xi(t_{k+1}) - \xi(t_k)),$$

где $\tilde{t}_k \in (t_{k-1}, t_k]$.

Определим интегралы I_1 и I_2 как

$$I_1 = l.i.m._{\lambda_n \rightarrow 0} S_1, \quad I_2 = l.i.m._{\lambda_n \rightarrow 0} S_2,$$

где предел не должен зависеть от разбиения отрезка $[a, b]$ и выбора точек \tilde{t}_k .

Теорема 6.7.

1. Интеграл I_1 существует тогда и только тогда, когда существует интеграл Римана

$$Q_1 = \int_a^b \int_a^b g(t)g(u)r(t,u)dtdu.$$

2. Интеграл I_2 существует тогда и только тогда, когда существует интеграл Римана – Стильбеса

$$Q_2 = \int_a^b \int_a^b g(t)g(u)r(t,u)d_{t,u}r(t,u),$$

причем $Q_1 = \mathbf{E}\{I_1^2\}$, $Q_2 = \mathbf{E}\{I_2^2\}$.

Доказательство. Рассмотрим другое разбиение отрезка $[a, b] : a = u_0 < u_1 < \dots < u_m < b$, $\lambda_m = \max_k(u_k - u_{k-1})$, $\tilde{u}_k \in (u_{k-1}, u_k]$, и интегральные суммы:

$$S'_1 = \sum_{l=1}^m g(\tilde{u}_l)\xi(\tilde{u}_l)(u_{l+1} - u_l);$$

$$S'_2 = \sum_{l=1}^m g(\tilde{u}_l)(\xi(u_{l+1}) - \xi(u_l)).$$

Тогда, по критерию сходимости в среднем квадратичном, интеграл I_1 существует тогда и только тогда, когда существует предел

$$\mathbf{E}\{S'_1 S'_1\} = \sum_{k=1}^n \sum_{l=1}^m g(\tilde{t}_k)g(\tilde{u}_l)r(\tilde{t}_k, \tilde{u}_l)(t_{k+1} - t_k)(u_{l+1} - u_l), \text{ когда } \lambda_n, \lambda_m \rightarrow 0.$$

Но такой предел есть интеграл Q_1 и $\mathbf{E}\{I_1^2\} = Q_1$.

Аналогично,

$$\mathbf{E}\{S'_2 S'_2\} = \sum_{k=1}^n \sum_{l=1}^m g(\tilde{t}_k)g(\tilde{u}_l)(r(t_{k+1}, u_{l+1}) - r(t_{k+1}, u_l) - r(t_k, u_{l+1}) + r(t_k, u_l)) \xrightarrow[\lambda_m \rightarrow 0]{\lambda_n \rightarrow 0} Q_2,$$

причем $Q_2 = \mathbf{E}\{I_2^2\}$.

Под интегралами по бесконечному промежутку понимается предел в среднем квадратичном интегралов по конечному промежутку, когда концы интервалов стремятся к бесконечности. \square

6.5. НЕПРЕРЫВНОСТЬ ТРАЕКТОРИЙ СЛУЧАЙНОГО ПРОЦЕССА

Теорема 6.8. Пусть случайный процесс $\xi(t)$, $t \in [0, 1]$, удовлетворяет условию: для любого $h \geq 0$ выполняется

$$P\{|\xi(t+h) - \xi(t)| \geq g(h)\} \leq q(h), \quad (6.4)$$

где $g(t)$ и $q(t)$ – четные неубывающие на интервале $[0, +\infty)$ функции, удовлетворяющие условиям:

$$\sum_{k=0}^{\infty} g(2^{-k}) < +\infty \quad \text{и} \quad \sum_{k=0}^{\infty} 2^k q(2^{-k}) < +\infty. \quad (6.5)$$

Тогда для процесса $\xi(t)$ существует эквивалентный ему случайный процесс $\eta(t)$, траектории которого непрерывны с вероятностью 1.

Если случайный процесс $\xi(t)$ сепарабельный, то траектории самого $\xi(t)$ непрерывны.

Доказательство. Основано на приближении траектории случайного процесса $\xi(t)$ ломаными прямыми $\xi_n(t)$, которые сходятся равномерно к траекториям процесса $\xi(t)$ с вероятностью 1.

Для каждого n положим $t_{nr} = \frac{r}{2^n}$, $r = 0, 1, \dots, 2^n$.

$$\xi_n(t) = \xi(t_{n,r}) + 2^n(t - t_{n,r}) [\xi(t_{n,r+1}) - \xi(t_{n,r})] \quad \text{при} \quad t_{n,r} \leq t \leq t_{n,r+1}.$$

Рассмотрим один из промежутков $[t_{nr}, t_{nr+1}]$ (рис. 6.1).

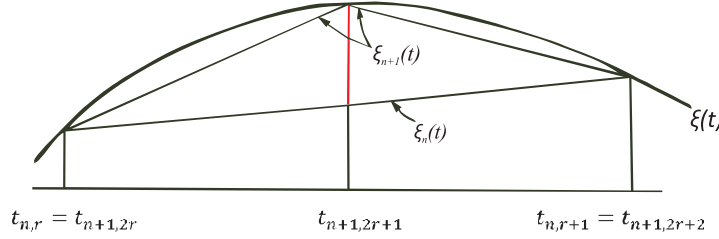


Рис. 6.1. Траектории процессов $\xi(t)$, $\xi_n(t)$ и $\xi_{n+1}(t)$

Здесь $t_{nr} = t_{n+1,2r}$ и $t_{nr+1} = t_{n+1,2r+2}$. Рассмотрим теперь

$$\eta = \max_{t_{nr} \leq t \leq t_{nr+1}} |\xi_{n+1}(t) - \xi_n(t)| = |\xi(t_{n+1,2r+1}) - \frac{1}{2}(\xi(t_{n+1,2r+2}) + \xi(t_{n+1,2r}))| \leq \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2,$$

где

$$\eta_1 = \frac{1}{2}|\xi(t_{n+1,2r+1}) - \xi(t_{n+1,2r})|, \quad \eta_2 = \frac{1}{2}|\xi(t_{n+1,2r+1}) - \xi(t_{n+1,2r+2})|.$$

Поскольку $(\eta_1 \leq a)(\eta_2 \leq a) \subset (\eta \leq a)$, то, переходя к противоположным событиям, получаем

$$(\eta_1 > a) + (\eta_2 > a) \supset (\eta > a).$$

В качестве a возьмем $a = g(t_{n+1,2r+1} - t_{n+1,2r}) = g(2^{-n-1})$. Используя свойство вероятностей, получаем:

$$\begin{aligned} P \left\{ \max_{t_{nr} \leq t \leq t_{nr+1}} |\xi_{n+1}(t) - \xi_n(t)| > g(2^{-n-1}) \right\} &\leq \\ &\leq P \left\{ |\xi_{t_{n+1,2r+1}} - \xi_{t_{n+1,2r}}| > g(2^{-n-1}) \right\} + \\ &+ P \left\{ |\xi_{t_{n+1,2r+1}} - \xi_{t_{n+1,2r+2}}| > g(2^{-n-1}) \right\} \leq 2q(2^{-n-1}). \end{aligned} \quad (6.6)$$

Рассмотрим теперь

$$\begin{aligned} P\{A_n\} &= P \left\{ \max_{0 \leq t \leq 1} |\xi_{n+1}(t) - \xi_n(t)| > g(2^{-n-1}) \right\} = \\ &= P \left\{ \bigcup_{r=0}^{2^n-1} \left(\max_{t_{n,r} \leq t \leq t_{n,r+1}} |\xi_{n+1}(t) - \xi_n(t)| > g(2^{-n-1}) \right) \right\} \leq \\ &\leq \sum_{r=0}^{2^n-1} 2q(2^{-n-1}) = 2^{n+1}q(2^{-n-1}). \end{aligned}$$

Из (6.5) получаем, что $\sum_{n=1}^{\infty} P\{A_n\} < \infty$. Тогда по лемме Бореля – Кантелли следует $P\{\overline{\lim} A_n\} = 0$ или $P\{\underline{\lim} A_n\} = 1$.

Из определения нижнего предела следует, что для всех $\omega \in \underline{\lim} A_n$ существует такой номер $n_0 = n_0(\omega)$, что для всех $n \geq n_0$ справедливо неравенство

$$\max_{0 \leq t \leq 1} |\xi_{n+1}(t, \omega) - \xi_n(t, \omega)| \leq g(2^{-n-1}).$$

Отсюда и из (6.5) следует, что для этих ω ряд $\sum_{n=1}^{\infty} |\xi_{n+1}(t, \omega) - \xi_n(t, \omega)|$ сходится равномерно. Тогда

$$\sum_{n=1}^{\infty} |\xi_{n+1}(t, \omega) - \xi_n(t, \omega)| = \lim_{m \rightarrow \infty} \sum_{n=1}^m |\xi_{n+1}(t, \omega) - \xi_n(t, \omega)| = \lim_{m \rightarrow \infty} \xi_m(t, \omega) - \xi_1(t, \omega).$$

Отсюда следует, что с вероятностью единица существует предел $\lim_{m \rightarrow \infty} \xi_m(t, \omega) = \eta(t)$ при каждом фиксированном t , причем эта сходимость равномерная. Поскольку траектории $\xi_n(t, \omega)$ непрерывны и сходятся равномерно, то траектория $\eta(t, \omega)$ непрерывна. Поскольку $P\{\underline{\lim} A_n\} = 1$, то с вероятностью единица процесс $\eta(t)$ непрерывен. Покажем теперь, что случайные процессы $\xi(t)$ и $\eta(t)$ эквивалентны, т. е. для всех t $P\{\xi(t) = \eta(t)\} = 1$. Заметим, что по построению $\xi(t_{nr}) = \eta(t_{nr})$. Пусть теперь t – произвольная точка из отрезка $[0, 1]$. Тогда существует последовательность r_n такая, что $t_{n,r_n} \xrightarrow{n \rightarrow \infty} t$, причем $|t_{n,r_n} - t| \leq 2^{-n}$. Поскольку по условию теоремы $P\{B_n\} = P\{|\xi(t_{n,r_n}) - \xi(t)| > g(t_{n,r_n} - t)\} \leq g(t_{n,r_n} - t)$. Тогда из (6.5) следует

$$\sum_n g(t_{n,r_n} - t) \leq \sum_n g(2^{-n}) < +\infty.$$

Аналогично получаем, что

$$\sum_n q(t_{n,r_n} - t) \leq \sum_n q(2^{-n}) < +\infty.$$

Теперь применяем те же рассуждения, что и ранее, получаем

$$\xi(t_{n,r_n}) \xrightarrow{\text{П. Н.}} \xi_t.$$

С другой стороны, $\xi(t_{n,r_n}) = \eta(t_{n,r_n}) \xrightarrow{\text{П. Н.}} \eta_t$. Тогда из теоремы единственности следует, что при каждом

$$t \in [0, 1] \quad P\{\xi_t = \eta_t\} = 1.$$

Для доказательства второй части достаточно показать, что траектории ξ и η совпадают с вероятностью 1. Поскольку траектории случайного процесса $\eta(t)$ с вероятностью единица непрерывны, то для почти всех ω и любого $\epsilon > 0$ существует $\delta = \delta(\omega, t_0)$ такое, что как только $|t - t_0| \leq \delta$, то $|\eta(t) - \eta(t_0)| \leq \epsilon$.

Покажем, что в произвольной точке $t_0 > 0$ траектории совпадают. Возьмем интервал $I = (t_0 - \delta, t_0 + \delta)$. Тогда в силу сепарабельности с вероятностью единица будут выполняться неравенства

$$\xi(t_0) \leq \sup_{t \in I} \xi(t) = \sup_{t \in I \cap S} \xi(t) = \sup_{t \in I \cap S} \eta(t) \leq \sup_{t \in I} \eta(t) \leq \eta(t_0) + \epsilon,$$

где S – множество сепарабельности. Итак, $\xi(t_0) \leq \eta(t_0) + \epsilon$. Поскольку ϵ любое, то $\xi(t_0) \leq \eta(t_0)$. Аналогично получим, что $\xi(t_0) \geq \eta(t_0)$. Таким образом, в произвольной точке $t_0 > 0$ с вероятностью единица $\xi(t_0) = \eta(t_0)$. А так как траектории процесса $\eta(t)$ непрерывны, то и траектории процесса $\xi(t)$ тоже непрерывны. \square

6.6. УПРАЖНЕНИЯ

1. Известно, что $E\{\xi(t)\} = 2t + 1$; $r_\xi(t, t') = e^{-(t'-t)^2}$; $\eta(t) = \frac{d\xi(t)}{dt}$. Найти математическое ожидание, ковариационную функцию и дисперсию процесса $\eta(t)$.
2. Доказать, что случайный процесс $\xi(t) = e^{-\alpha t} \sin(\omega t + \phi)$, где α и ω – положительные постоянные, а ϕ – случайная величина, равномерно распределенная на отрезке $[0, 2\pi]$, дифференцируем при всех $t > 0$.
3. Пусть $\xi(t)$ – стационарный и гауссовский процесс с математическим ожиданием m и ковариационной функцией $r(\tau)$. Написать выражения одномерной и двухмерной плотностей этого процесса.
4. Будут ли непрерывны и дифференцируемы стационарные процессы, имеющие ковариационные функции:
 - 1) $De^{-a|\tau|} \cos b\tau$;
 - 2) $De^{-a|\tau|} \left(\cos b\tau + \frac{a}{b} \sin b|\tau| \right)$?
5. Пусть $\xi^{(1)}(t), \dots, \xi^{(n)}(t)$ – независимые случайные процессы такие, что $E\{\xi^{(i)}(t)\} = 0$, $i = 1, \dots, n$, $r_1(t, s), \dots, r_n(t, s)$ – соответствующие ковариационные функции. Найти ковариационную функцию процесса $\xi^{(1)}(t) + \dots + \xi^{(n)}(t)$.
6. Пусть ξ – гауссовская случайная величина, $L(\xi) = N(m, \sigma^2)$, $b \in \mathbb{R}$. Найти ковариационную функцию процесса

$$\xi(t) = \xi t + b, t \geq 0.$$

7. Пусть $\xi^{(1)}(t)$, $\xi^{(2)}(t)$ – два независимых случайных процесса с ковариационными функциями $r_1(t, s)$ и $r_2(t, s)$ соответственно. Найти ковариационную функцию процесса $\eta_t = \xi^{(1)}(t)\xi^{(2)}(t)$.

Глава 7

СТАЦИОНАРНЫЕ В ШИРОКОМ СМЫСЛЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ

7.1. ОСНОВНЫЕ ПОНЯТИЯ И СВОЙСТВА

В этой главе будут рассмотрены комплекснозначные случайные процессы

$$\xi(t) = \xi_1(t) + i\xi_2(t) \in \mathbb{C}, \quad t \in \mathbb{R},$$

где $\xi_1(t)$ и $\xi_2(t)$ – вещественнозначные случайные процессы, $i^2 = -1$.

Изучение комплекснозначных процессов удобно в теоретических исследованиях; полученные результаты переносятся на вещественнозначные случайные процессы.

Ковариационная функция $r(t, u)$ для таких процессов определяется следующим образом:

$$r(t, u) = E\{(\xi(t) - E\{\xi(t)\})(\overline{\xi(u) - E\{\xi(u)\}})\},$$

где $\overline{\xi(u)} \in \mathbb{C}$ обозначает комплексное сопряжение $\xi(u)$.

Определение 7.1. СП $\xi(t)$ называется стационарным в широком смысле, если выполнены следующие три условия:

- 1) $E\{|\xi(t)|^2\} < \infty$;
- 2) $E\{\xi(t)\} = m$;
- 3) $z(t, u) = r(t - u, 0) = r(t - u)$.

Свойства ковариационной функции

C1. $z(t, u) = \overline{r(u, t)}$.

C2. Функция $r(t, u)$ является неотрицательно определенной, т. е. для любого n , любых действительных чисел t_1, \dots, t_n и любых комплекснозначных чисел z_1, \dots, z_n

$$\sum_{k,l=1}^n r(t_k, t_l) z_k \bar{z}_l \geq 0.$$

C3. $|r(t, u)|^2 \leq r(t, t)r(u, u)$.

Доказательства свойств C1 и C2 следуют из линейности математического ожидания и равенства

$$E\{\xi(t)\} = E\{\xi_1(t)\} + iE\{\xi_2(t)\}.$$

Третье свойство следует из неравенства Шварца

$$|E\{\xi(t)\overline{\xi(u)}\}|^2 \leq E\{|\xi(t)|^2\}E\{|\xi(u)|^2\}.$$

В дальнейшем без потери общности будем полагать, что

$$E\{\xi(t)\} = m = 0.$$

7.2. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ КОВАРИАЦИОННОЙ ФУНКЦИИ СЛУЧАЙНОГО ПРОЦЕССА

Если $r(t)$ – ковариационная функция стационарного в широком смысле и непрерывного в среднем квадратичном СП, то существует вещественная, неубывающая, непрерывная справа функция F такая, что

$$r(t) = \int e^{i\lambda t} dF(\lambda), \quad (7.1)$$

$F(-\infty) = 0$, $F(+\infty) = r(0)$. Функция $F(\lambda)$ называется *спектральной функцией* СП $\xi(t)$. Равенство (7.1) называется спектральным представлением ковариационной функции СП.

Если $F(\lambda) = \int_{-\infty}^{\lambda} f(x) dx$, то функция $f(x) \geq 0$ называется *спектральной плотностью*

СП $\xi(t)$.

Доказательство теоремы является следствием следующей теоремы.

Теорема 7.1 (теорема Бохнера). *Если функция $r(t)$ – неотрицательно определена и непрерывна в нуле, $r(0) < \infty$, то справедливо представление (7.1).*

Доказательство. Рассмотрим функцию

$$g(\lambda, A) = \frac{1}{2\pi A} \int_0^A \int_0^A r(t-u) e^{-it\lambda} e^{itu} dt du. \quad (7.2)$$

В силу неотрицательной определенности функции $r(t)$ интегральные суммы интеграла (7.2) неотрицательны. Следовательно, функция $g(\lambda, A) \geq 0$.

Сделав замену переменных в интеграле (7.2), получаем

$$g(\lambda, A) = \frac{1}{2\pi} \int_{-A}^A \mu\left(\frac{x}{A}\right) r(x) e^{-i\lambda x} dx,$$

где $\mu(x) = \begin{cases} 1 - |x|, & \text{если } |x| \leq 1, \\ 0, & \text{если } |x| > 1. \end{cases}$

Покажем, что

$$\int_{-\infty}^{+\infty} g(\lambda, A) d\lambda < \infty.$$

Поскольку для всех $M > 0$

$$\begin{aligned} \int_{-\infty}^{+\infty} \mu\left(\frac{\lambda}{2M}\right) g(\lambda, A) d\lambda &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mu\left(\frac{t}{A}\right) r(t) dt \int_{-\infty}^{+\infty} \mu\left(\frac{\lambda}{2M}\right) e^{-it\lambda} d\lambda = \\ &= \frac{1}{\pi} M \int_{-\infty}^{+\infty} \mu\left(\frac{t}{A}\right) r(t) \left(\frac{\sin Mt}{Mt}\right)^2 dt \leq \frac{1}{\pi} r(0) \int_{-\infty}^{+\infty} \left(\frac{\sin t}{t}\right)^2 dt = r(0), \end{aligned}$$

то, переходя в этом неравенстве к пределу при $M \rightarrow \infty$, получаем

$$\int_{-\infty}^{+\infty} g(\lambda, A) d\lambda \leq r(0). \quad (7.3)$$

Поскольку функция $g(\lambda, A)$ есть преобразование Фурье функции $\mu\left(\frac{t}{A}\right)r(t)$, и выполняется (7.3), то существует обратное преобразование Фурье

$$\mu\left(\frac{t}{A}\right)r(t) = \int_{-\infty}^{+\infty} g(\lambda, A) e^{it\lambda} d\lambda.$$

В частности, при $t = 0$

$$\int_{-\infty}^{+\infty} g(\lambda, A) d\lambda = r(0).$$

Таким образом, функцию $\frac{g(\lambda, A)}{r(0)}$ можно рассматривать как плотность распределения вероятностей некоторой случайной величины. Тогда функция $\mu\left(\frac{t}{A}\right) \frac{r(t)}{r(0)}$ при каждом A будет характеристической функцией этой случайной величины. Поскольку она непрерывна в нуле, то по теореме о непрерывности для характеристических функций получаем, что

$$\lim_{A \rightarrow \infty} \mu\left(\frac{t}{A}\right) \frac{r(t)}{r(0)} = \frac{r(t)}{r(0)}$$

есть характеристическая функция.

Следовательно, по теореме единственности для характеристической функции существует функция распределения $F_0(\lambda)$ такая, что

$$\frac{r(t)}{r(0)} = \int_{-\infty}^{+\infty} e^{it\lambda} dF_0(\lambda).$$

Взяв в качестве $F(\lambda)$ функцию $r(0)F_0(\lambda)$, получим равенство (7.1).

Полученные ранее результаты для функции распределения $F(x) = P(\xi < x)$ верны и для функции распределения $F(x) = P(\xi \leq x)$, которая непрерывна справа. \square

7.3. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ СТАЦИОНАРНОГО В ШИРОКОМ СМЫСЛЕ СЛУЧАЙНОГО ПРОЦЕССА

Определение 7.2. Случайный процесс $\xi(\lambda)$ называется процессом с ортогональными приращениями, если для любых $\lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$

$$E\{(\xi(\lambda_4) - \xi(\lambda_3))(\overline{\xi(\lambda_2) - \xi(\lambda_1)})\} = 0.$$

Теорема 7.2. Если $\xi(t)$ – стационарный в широком смысле, непрерывный в среднем квадратичном случайный процесс, то существует случайный процесс $\zeta(\lambda)$ с ортогональными приращениями такой, что

$$\xi(t) = \int e^{i\lambda t} d\zeta(\lambda), \quad \mathbf{E}\{|\zeta(\lambda)|^2\} = F(\lambda). \quad (7.4)$$

Равенство (7.4) называется спектральным представлением СП $\xi(t)$.

Доказательство. Пусть $F(\lambda)$ – спектральная функция СП $\xi(t)$. Рассмотрим пространство

$$H_F = \left(g(\lambda) : \int |g(\lambda)|^2 dF(\lambda) < \infty \right).$$

В пространстве H_F введем скалярное произведение и норму

$$(g_1, g_2) = \int g_1(\lambda) \overline{g_2(\lambda)} dF(\lambda), \quad \|g\|_{H_F} = \left(\int |g(\lambda)|^2 dF(\lambda) \right)^{\frac{1}{2}}.$$

Таким образом, H_F – гильбертово пространство. Рассмотрим пространство

$$H_\xi = \left(\xi(t), \sum_k a_k \xi_{t_k}, \text{l.i.m.} \sum_k a_k \xi_{t_k} \right) -$$

пространство, состоящее из случайных величин $\{\xi(t) : t \in \mathbb{R}\}$, их линейных комбинаций, а также пределов в среднем квадратичном этих линейных комбинаций. Если $\eta_1, \eta_2 \in H_\xi$, то скалярное произведение и норму в этом пространстве определим следующим образом:

$$(\eta_1, \eta_2) = \mathbf{E}\{\eta_1 \overline{\eta_2}\}, \quad \|\eta\|_{H_\xi} = \mathbf{E}^{\frac{1}{2}}\{|\eta|^2\}.$$

Между пространствами H_F и H_ξ установим взаимнооднозначное соответствие: случайной величине $\xi(t)$ поставим в соответствие функцию $g(\lambda) = e^{i\lambda t}$. Покажем, что скалярные произведения и нормы соответствующих элементов будут совпадать:

$$(\xi(t), \xi(u))_{H_\xi} = \mathbf{E}\{\xi(t) \overline{\xi(u)}\} = r(t-u) = \int e^{i\lambda(t-u)} dF(\lambda) = (e^{it\lambda}, e^{iu\lambda})_{H_F}.$$

Распространим это соответствие на конечные линейные комбинации элементов $\xi(t_k)$ и $e^{i\lambda t_k}$, а именно элементу

$$\eta = \alpha_1 \xi(t_1) + \dots + \alpha_n \xi(t_n) \quad (7.5)$$

сопоставим элемент

$$g(\lambda) = \alpha_1 e^{i\lambda t_1} + \dots + \alpha_n e^{i\lambda t_n}. \quad (7.6)$$

Из линейности математического ожидания и интегралов будет следовать, что для соответствующих пар η_1, η_2 и g_1, g_2

$$(\eta_1, \eta_2)_{H_\xi} = (g_1, g_2)_{H_F} \quad (7.7)$$

и для соответствующих норм

$$\|\eta_1 - \eta_2\|_{H_\xi} = \|g_1 - g_2\|_{H_F}. \quad (7.8)$$

Последнее соотношение показывает, что установленное соответствие взаимнооднозначно, так как η_1 и η_2 не могут совпадать, если не совпадают g_1 и g_2 , и обратно. Из равенства

$$\mathbf{E}\{|\eta_m - \eta_n|^2\} = \int |g_m(\lambda) - g_n(\lambda)|^2 dF(\lambda)$$

следует, что если последовательность $\{\eta_n\}$ сходится в среднем квадратичном к некоторой случайной величине η , то и соответствующие функции $\{g_n(\lambda)\}$ сходятся в среднем квадратичном относительно спектральной функции $F(\lambda)$.

Если $g(\lambda) = \text{l.i.m. } g_n(\lambda)$, то элементу η поставим в соответствие элемент $g(\lambda)$. Аналогично, любой элемент из H_F есть предел в среднем квадратичном некоторой последовательности элементов вида (7.6). Из свойств сходимости в среднем квадратичном следует, что соотношения (7.7) и (7.8) будут выполняться для всех $\eta_1, \eta_2 \in H_\xi$ и $g_1, g_2 \in H_F$.

Для любого λ_0 функция $g(\lambda - \lambda_0)$, где $g(\lambda) = \mathbb{I}_{(-\infty, 0]}(\lambda)$, принадлежит H_F . Пусть $\zeta(\lambda_0)$ – соответствующий элемент из H_ξ . Тогда приращению $\zeta(\lambda_2) - \zeta(\lambda_1)$ будет соответствовать разность $g(\lambda - \lambda_2) - g(\lambda - \lambda_1)$. Если интервалы (λ_1, λ_2) и (λ_3, λ_4) не пересекаются, то в силу инвариантности скалярного произведения получаем

$$\begin{aligned} & \mathbf{E}\{(\zeta(\lambda_4) - \zeta(\lambda_3))(\overline{\zeta(\lambda_2) - \zeta(\lambda_1)})\} = \\ & = \int (g(\lambda - \lambda_4) - g(\lambda - \lambda_3))(\overline{g(\lambda - \lambda_2) - g(\lambda - \lambda_1)}) dF(\lambda) = 0, \end{aligned}$$

т. е. $\zeta(\lambda)$ – процесс с ортогональными приращениями. Полагая $\lambda_3 = \lambda_1, \lambda_4 = \lambda_2$, получаем $\mathbf{E}\{|\zeta(\lambda_2) - \zeta(\lambda_1)|^2\} = F(\lambda_2) - F(\lambda_1)$, $\mathbf{E}\{|\zeta(\lambda)|^2\} = F(\lambda)$.

Пусть для некоторого $A > 0$ задано разбиение $[-A, A]$:

$$-A = \lambda_0 < \lambda_1 < \dots < \lambda_n = A, \mu_n = \max_j (\lambda_j - \lambda_{j-1}).$$

Тогда случайной величине $\eta_n = \sum_{j=1}^n e^{it\lambda_j} (\zeta(\lambda_j) - \zeta(\lambda_{j-1}))$ соответствует функция

$$g_n(\lambda) = \sum_{j=1}^n e^{it\lambda_j} (g(\lambda - \lambda_j) - g(\lambda - \lambda_{j-1})).$$

Поскольку при $\mu_n \xrightarrow{n \rightarrow \infty} 0$

$$\eta_n \rightarrow \int_{-A}^A e^{it\lambda} d\zeta(\lambda), \quad g_n(\lambda) \rightarrow e^{it\lambda} \mathbb{I}_{[-A, A]}(\lambda),$$

то устремляя A к бесконечности, получим, что случайной величине $\int_{-\infty}^{+\infty} e^{it\lambda} d\zeta(\lambda)$ соответствует функция $e^{it\lambda}$. Поскольку функции $e^{it\lambda}$ соответствует случайная величина $\xi(t)$, то, в силу однозначного соответствия, получим

$$\xi(t) = \int e^{it\lambda} d\zeta(\lambda). \quad (7.9)$$

□

Заметим, что если случайной величине η соответствует функция $g(\lambda)$, то

$$\eta(t) = \int g(\lambda) d\zeta(\lambda), \quad (7.10)$$

причем если $\eta_1 \leftrightarrow g_1(\lambda)$, $\eta_2 \leftrightarrow g_2(\lambda)$, то

$$\mathbf{E}\{\eta_1 \overline{\eta_2}\} = (\eta_1, \eta_2)_{H_\xi} = (g_1, g_2)_{H_F} = \int g_1(\lambda) \overline{g_2(\lambda)} dF(\lambda). \quad (7.11)$$

7.4. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ДЕЙСТВИТЕЛЬНЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ

Рассмотрим случай, когда $\xi(t) = \overline{\xi(t)}$. Для процесса $\xi(t)$ ковариационная функция – действительная и

$$r(t) = r(-t) = \int \cos t\lambda dF(\lambda), \quad \int \sin t\lambda dF(\lambda) = 0.$$

Поскольку случайной величине $\eta = \alpha_1 \xi(t_1) + \dots + \alpha_n \xi(t_n)$ соответствует функция $g(\lambda) = \alpha_1 e^{it_1 \lambda} + \dots + \alpha_n e^{it_n \lambda}$, а комплексно-сопряженной случайной величине $\bar{\eta} = \overline{\alpha_1 \xi(t_1)} + \dots + \overline{\alpha_n \xi(t_n)}$ соответствует функция $\overline{\alpha_1 e^{it_1 \lambda}} + \dots + \overline{\alpha_n e^{it_n \lambda}} = \overline{g(-\lambda)}$, то отсюда следует, что если случайной величине $\eta \in H_\xi$ соответствует функция $g(\lambda) \in H_F$, то случайной величине $\bar{\eta}$ будет соответствовать функция $\overline{g(-\lambda)}$.

Обозначим $\Delta = (\lambda_1, \lambda_2]$, $\zeta(\Delta) = \zeta(\lambda_2) - \zeta(\lambda_1)$, $g(\Delta) = g(\lambda_2 - \lambda_1) - g(\lambda_1 - \lambda_1)$.

Поскольку случайной величине $\zeta(\Delta)$ соответствует функция $g(\Delta)$, то из показанного выше случайной величине $\overline{\zeta(\Delta)}$ будет соответствовать функция $g(-\Delta)$, а ей соответствует случайная величина $\zeta(-\Delta)$, где $-\Delta = [-\lambda_2, -\lambda_1]$. Отсюда получаем $\overline{\zeta(\Delta)} = \zeta(-\Delta)$, т. е. $\zeta_1(\Delta) - i\zeta_2(\Delta) = \zeta_1(-\Delta) + i\zeta_2(-\Delta)$. Поэтому

$$\zeta_1(\Delta) = \zeta_1(-\Delta), \quad \zeta_2(\Delta) = -\zeta_2(-\Delta). \quad (7.12)$$

Пусть Δ_1 и Δ_2 – два произвольных интервала. Обозначим $\Delta_3 = \Delta_1 \cap \Delta_2$, $\Delta'_1 = \Delta_1 \setminus \Delta_3$, $\Delta'_2 = \Delta_2 \setminus \Delta_3$. Тогда $\zeta(\Delta_1) = \zeta(\Delta'_1) + \zeta(\Delta_3)$, $\zeta(\Delta_2) = \zeta(\Delta'_2) + \zeta(\Delta_3)$ и $\Delta'_1 \Delta'_2 = \emptyset$.

Поскольку процесс $\zeta(\lambda)$ – процесс с ортогональными приращениями, то

$$E\{\zeta(\Delta_1)\overline{\zeta(\Delta_2)}\} = E\{\zeta(\Delta_3)\overline{\zeta(\Delta_3)}\} \geq 0.$$

Следовательно, для любых интервалов Δ_1 и Δ_2 $\text{Im}\{E\{\zeta(\Delta_1)\overline{\zeta(\Delta_2)}\}\} = E\{\zeta_1(\Delta_1)\zeta_2(\Delta_2)\} - E\{\zeta_2(\Delta_1)\zeta_1(\Delta_2)\} = 0$. Заменяя в этом соотношении Δ_1 на $-\Delta_1$ и используя (7.12), получим $E\zeta_1(\Delta_1)\zeta_2(\Delta_2) = 0$ для любых интервалов Δ_1 и Δ_2 , т. е. процессы $\zeta_1(\lambda)$ и $\zeta_2(\lambda)$ являются взаимно ортогональными.

Пусть интервалы Δ_1 и Δ_2 такие, что $\Delta_1 \cap \Delta_2 = \emptyset$, $-\Delta_1 \cap \Delta_2 = \emptyset$. Поскольку $E\{\zeta(\Delta_1)\overline{\zeta(\Delta_2)}\} = 0$, то $\text{Re}\{E\{\zeta(\Delta_1)\overline{\zeta(\Delta_2)}\}\} = E\{\zeta_1(\Delta_1)\zeta_2(\Delta_2)\} + E\{\zeta_2(\Delta_1)\zeta_1(\Delta_2)\} = 0$. Заменяя в этом соотношении Δ_1 на $-\Delta_1$ и используя (7.12), получим $E\{\zeta_1(\Delta_1)\zeta_1(\Delta_2)\} = 0$, $E\{\zeta_2(\Delta_1)\zeta_2(\Delta_2)\} = 0$, т. е. процессы $\zeta_1(\lambda)$ и $\zeta_2(\lambda)$ являются процессами с ортогональными приращениями.

Используя (7.12) и равенство (7.9), получим спектральное представление случайного процесса $\xi(t)$

$$\xi(t) = \int \cos t\lambda d\zeta_1(\lambda) + \int \sin t\lambda d\zeta_2(\lambda).$$

7.5. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ СЛУЧАЙНОГО ПРОЦЕССА

Пусть $\xi(t)$ – стационарный в широком смысле, непрерывный в среднем квадратичном случайный процесс, имеющий спектральное представление

$$\xi(t) = \int e^{it\lambda} d\zeta(\lambda), \quad E\{\xi(t)\overline{\xi(0)}\} = r(t) = \int e^{it\lambda} dF(\lambda).$$

Рассмотрим простейшее линейное преобразование СП $\xi(t)$ $\eta(t) = \sum_k \lambda_k \xi(t + t_k)$, где $\{t_k\}$ – некоторое множество действительных чисел. Используя спектральное представление СП $\xi(t)$, получим

$$\eta(t) = \sum_k \lambda_k \int e^{it_k \lambda} d\zeta(\lambda) = \int \sum_k \lambda_k e^{it_k \lambda} d\xi(\lambda) = \int g(\lambda) e^{it\lambda} d\zeta(\lambda),$$

где $g(\lambda) = \sum_k \lambda_k e^{it_k \lambda}$.

Определение 7.3. Случайный процесс $\eta(t)$ называется линейным преобразованием СП $\xi(t)$, если существует такая функция $g(\lambda)$, что

$$\int |g(\lambda)|^2 dF(\lambda) < \infty \quad (7.13)$$

и справедливо представление

$$\eta(t) = \int g(\lambda) e^{it\lambda} d\zeta(\lambda). \quad (7.14)$$

Функция $g(\lambda)$ называется частотной характеристикой линейного преобразования.

Теорема 7.3. Если СП $\xi(t)$ – стационарный в широком смысле, то и СП $\eta(t)$ – также стационарный в широком смысле.

Доказательство. Используя (7.10) и (7.11), получаем

$$\mathbf{E}\{\eta(t+u)\overline{\eta(u)}\} = \mathbf{E}\left\{\int g(\lambda) e^{i(t+u)\lambda} d\zeta(\lambda) \overline{\int g(\lambda) e^{iu\lambda} d\zeta(\lambda)}\right\} = \int |g(\lambda)|^2 e^{it\lambda} dF(\lambda) = r_\eta(t).$$

Отсюда получаем $\mathbf{E}\{|\eta(t)|^2\} = r_\eta(0) = \int |g(\lambda)|^2 dF(\lambda) < \infty$. \square

Из этих соотношений следует, что если СП $\xi(t)$ имеет спектральную плотность $f_\xi(\lambda)$, то СП $\eta(t)$ имеет спектральную плотность

$$f_\eta(\lambda) = |g(\lambda)|^2 f_\xi(\lambda). \quad (7.15)$$

Рассмотрим частные случаи линейных преобразований.

Случай 7.1. Дифференцирование СП.

Теорема 7.4. Если для рассмотренного выше СП $\xi(t)$ $\int \lambda^2 dF(\lambda) < \infty$, то СП $\xi(t)$ дифференцируем в среднем квадратичном, и его производная есть линейное преобразование СП $\xi(t)$ с частотной характеристикой $g(\lambda) = i\lambda$.

Доказательство. Имеем цепочку равенств:

$$\begin{aligned} \mathbf{E}\left\{\left|\frac{\xi(t+h) - \xi(t)}{h} - \int i\lambda e^{it\lambda} d\xi(\lambda)\right|^2\right\} &= \mathbf{E}\left\{\left|\int \frac{e^{i(t+h)\lambda} - e^{it\lambda}}{h} d\xi(\lambda) - \int i\lambda e^{it\lambda} d\xi(\lambda)\right|^2\right\} = \\ &= \mathbf{E}\left\{\left|\left(\int \frac{e^{ih\lambda} - 1 - ih\lambda}{h} d\xi(\lambda)\right) e^{it\lambda}\right|^2\right\} = \int \left|\frac{e^{ih\lambda} - 1 - ih\lambda}{h}\right|^2 dF(\lambda). \end{aligned} \quad (7.16)$$

Поскольку $|(e^{ih\lambda} - 1 - ih\lambda)/h|^2 \leq \lambda^2$ и $(e^{ih\lambda} - 1 - ih\lambda)/h \rightarrow 0$ при $h \rightarrow 0$, то правая часть (7.16) будет стремиться к нулю при $h \rightarrow 0$. Отсюда будет следовать результат теоремы. \square

Верна более общая теорема.

Теорема 7.5. Если для СП $\xi(t)$ $\int \lambda^{2n} dF(\lambda) < \infty$, то СП $\xi(t)$ n раз дифференцируем в среднем квадратичном и его n -я производная есть линейное преобразование СП $\xi(t)$ с частотной характеристикой $g(\lambda) = (i\lambda)^n$.

Доказательство теоремы аналогично доказательству теоремы 7.3 и проводится по индукции.

Случай 7.2. Фильтрация СП $\xi(t)$.

Определение 7.4. СП $\eta(t)$ называется фильтрацией СП $\xi(t)$, если справедливо следующее соотношение, называемое фильтром:

$$\eta(t) = \int h(t, s) \xi(s) ds. \quad (7.17)$$

Если $h(t, s) = 0$ при $s > t$, то фильтр называется физически осуществимым. В этом случае значение $\eta(t)$ зависит только от значений СП $\{\xi(s) : s \leq t\}$. Функция $h(t, s)$ называется импульсной характеристикой фильтра.

Рассмотрим случай, когда $h(t, s) = h(t - s)$, $h(u) = 0$ при $u < 0$. Процесс $\eta(t)$ согласно (7.17) является линейным преобразованием СП $\xi(t)$ с частотной характеристикой

$$g(\lambda) = \int h(x) e^{-i\lambda x} dx.$$

Теорема 7.6. Если СП $\eta_1(t)$ является фильтрацией СП $\xi(t)$, а $\eta_2(t)$ является фильтрацией СП $\eta_1(t)$ с импульсными характеристиками $h_1(t)$ и $h_2(t)$ и частотными характеристиками $g_1(\lambda)$ и $g_2(\lambda)$ соответственно, то СП $\eta_2(t)$ является фильтрацией СП $\xi(t)$ с импульсной характеристикой $h(t) = h_1 * h_2(t)$ (свертка функций) и частотной характеристикой $g(\lambda) = g_1(\lambda)g_2(\lambda)$.

Доказательство. По условию теоремы

$$\begin{aligned} \eta_1(t) &= \int h_1(t - s) \xi(s) ds, \quad \eta_2(t) = \int h_2(t - u) \eta_1(u) du = \int h_2(t - u) \times \\ &\times \int h_1(u - s) \xi(s) ds du = \int \left(\int h_2(t - u) h_1(u - s) du \right) \xi(s) ds = \int h(t - s) \xi(s) ds. \end{aligned}$$

Для частотных характеристик имеем

$$\begin{aligned} \eta_1(t) &= \int g_1(\lambda) e^{it\lambda} d\zeta(\lambda), \quad \eta_2(t) = \int h_2(t - s) \eta_1(s) ds = \int h_2(t - s) \int g_1(\lambda) e^{is\lambda} d\zeta(\lambda) ds = \\ &= \int g_1(\lambda) \int (h_2(t - s) e^{-i(t-s)\lambda} ds) e^{it\lambda} d\zeta(\lambda) = \int g_1(\lambda) g_2(\lambda) e^{it\lambda} d\zeta(\lambda). \end{aligned}$$

□

7.6. ПРОГНОЗИРОВАНИЕ СЛУЧАЙНОГО ПРОЦЕССА

Пусть $\xi(t)$ – рассмотренный выше СП. Рассмотрим задачу оценивания случайной величины $\xi(t + h)$, в будущий момент времени $t + h$, $h > 0$, по значениям СП $\xi(s)$ для $s \leq t$.

Обозначим $H(\xi, t)$ – сужение пространства H_ξ на множество случайных величин $\{\xi(s) : s \leq t\}$. Будем искать оценку $\xi^*(t, h)$ случайной величины $\xi(t + h)$ в пространстве $H(\xi, t)$ допустимых оценок $\hat{\xi}(t, h)$, которая минимизирует среднеквадратичную ошибку, т. е.

$$E\{|\xi(t + h) - \xi^*(t, h)|^2\} = \min_{\hat{\xi} \in H(\xi, t)} E\{|\xi(t + h) - \hat{\xi}|^2\}. \quad (7.18)$$

Теорема 7.7. Если случайная величина $\xi^*(t, h)$ удовлетворяет соотношению

$$E\{(\xi(t+h) - \xi^*(t, h))\bar{\eta}\} = 0 \quad (7.19)$$

для любой случайной величины $\eta \in H(\xi, t)$, то она удовлетворяет соотношению (7.18).

Доказательство. Пусть случайная величина $\xi^*(t, h)$ удовлетворяет соотношению (7.19). Покажем, что она минимизирует (7.18).

Пусть $\hat{\xi} \in H(\xi, t)$. Тогда имеем цепочку равенств:

$$\begin{aligned} E\{|\xi(t+h) - \hat{\xi}|^2\} &= E\{|\xi(t+h) - \xi^*(t, h) + \xi^*(t, h) - \hat{\xi}|^2\} = \\ &= E\{|\xi(t+h) - \xi^*(t, h)|^2\} + E\{|\xi^*(t, h) - \hat{\xi}|^2\} + \\ &+ E\{(\xi(t+h) - \xi^*(t, h))(\xi^*(t, h) - \hat{\xi})\} + E\{(\xi(t+h) - \xi^*(t, h))(\xi^*(t, h) - \hat{\xi})\}. \end{aligned} \quad (7.20)$$

Поскольку случайная величина $\xi^*(t, h) - \hat{\xi} \in H(t, \xi)$, то из (7.19) следует, что два последних слагаемых в (7.20) равны нулю. Следовательно, минимум левой части (7.20) достигается, если $\hat{\xi} = \xi^*(t, h)$. \square

Следствие 7.1. Соотношение (7.19) эквивалентно соотношению

$$E\{(\xi(t+h) - \xi^*(t, h))\bar{\xi}(s)\} = 0 \quad (7.21)$$

для любой случайной величины $\xi(s)$, где $s \leq t$.

Доказательство. Доказательство следует из линейности математического ожидания и свойств сходимости в среднем квадратичном. \square

Заметим, что соотношение (7.21) допускает эквивалентный вид:

$$r(t+h-s) = E\{\xi^*(t, h)\bar{\xi}(s)\}. \quad (7.22)$$

Следовательно, для нахождения оценки $\xi^*(t, h)$ необходимо проверить равенство (7.22) для всех $\xi(s)$, $s \leq t$.

Пример 7.1. СП $\xi(t)$ имеет ковариационную функцию $r(t) = e^{-|t|}$. Покажем, что оценкой случайной величины $\xi(t+h)$, $t > 0$, $h > 0$ является случайная величина $\xi^*(t, h) = \xi(t)e^{-h}$. Проверим выполнимость соотношения (7.22):

$$E\{\xi^*(t, h)\bar{\xi}(s)\} = e^{-h} E\{\xi(t)\bar{\xi}(s)\} = e^{-h} r(t-s) = e^{-h-t+s} = r(t+h-s).$$

Пример 7.2. СП $\xi(t)$, $t \geq 0$, имеет ковариационную функцию $r(t) = e^{-t}(\cos t + \sin t)$.

Имеем $r(t+h) = e^{-t-h}(\cos(t+h) + \sin(t+h)) = e^{-h}((\cos h + \sin h)r(t) + (\sin h)r'(t))$. В качестве оценки $\xi^*(t, h)$ случайной величины $\xi(t+h)$ возьмем случайную величину $\xi^*(t, h) = e^{-h}((\cos h + \sin h)\xi(t) + (\sin h)\xi'(t))$. Поскольку $E\{\xi'(t)\bar{\xi}(s)\} = r'(t-s)$, то $E\{\xi^*(t, h)\bar{\xi}(s)\} = e^{-h}((\cos h + \sin h)r(t-s) + (\sin h)r'(t-s)) = r(t+h-s)$.

Следовательно, в силу (7.22) $\xi^*(t, h)$ есть оптимальный прогноз.

Пример 7.3. СП $\xi(t)$ имеет вид $\xi(t) = \int_{-\infty}^t g(t-\lambda)d\zeta(\lambda)$, где $\zeta(\lambda)$ – процесс с ортогональными приращениями, соответствующий процессу $\xi(t)$. Поскольку $\xi(t+h) = \int_{-\infty}^{t+h} g(t+h-\lambda)d\zeta(\lambda)$, то в качестве $\xi^*(t, h)$ возьмем случайную величину

$$\xi^*(t, h) = \int_{-\infty}^t g(t+h-s)d\zeta(\lambda).$$

Проверим выполнимость соотношения (7.22). Поскольку процесс $\zeta(\lambda)$ является процессом с ортогональными приращениями, то

$$\begin{aligned} \mathbf{E}\{\xi^*(t, h)\bar{\xi}(s)\} &= \mathbf{E}\left\{\int_{-\infty}^t g(t+h-\lambda)d\zeta(\lambda) \overline{\int_{-\infty}^s g(s-\lambda)d\zeta(\lambda)}\right\} = \\ &= \mathbf{E}\left\{\int_{-\infty}^s g(t+h-\lambda)\overline{g(s-\lambda)}dF(\lambda)\right\} = \mathbf{E}\left\{\int_{-\infty}^{t+h} g(t+h-\lambda)d\zeta(\lambda) \overline{\int_{-\infty}^s g(s-\lambda)d\zeta(\lambda)}\right\} = \\ &= \mathbf{E}\{\xi(t+h)\bar{\xi}(s)\} = r(t+h-s). \end{aligned}$$

Заметим, что в примере 7.1 для прогноза необходимо знание СП $\xi(t)$ только в одной точке t , в примере 7.2 знание значения СП $\xi(t)$ в окрестности точки t , а в примере 7.3 знание всех значений СП $\xi(t)$ до момента времени t .

7.7. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ КОВАРИАЦИОННОЙ ФУНКЦИИ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ И САМОЙ СТАЦИОНАРНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Будем рассматривать стационарную в широком смысле случайную последовательность (ССП) $\{\xi(n)\}$, $n \in \mathbb{Z}$, $\mathbf{E}\{\xi(n)\} = 0$, $r(n) = \mathbf{E}\{\xi(n+m)\bar{\xi}(m)\} = \mathbf{E}\{\xi(n)\bar{\xi}(0)\}$. Ковариационная функция $r(n)$ обладает следующими свойствами (см. п. 7.1):

1. $r(n) = r(-n)$;
2. $|r(n)| \leq r(0)$;
3. $r(n)$ – неотрицательно определенная функция, т. е. $\sum_{k,l=1}^n r(k-l)z_k\bar{z}_l \geq 0$.

Теорема 7.8 (теорема Герглотца). Если функция $r(n)$ неотрицательно определена, то она представима в виде

$$r(n) = \int_{-\pi}^{\pi} e^{i\lambda n} dF(\lambda),$$

где $F(\lambda)$ – некоторая неубывающая, непрерывная справа функция переменной $\lambda \in [-\pi, \pi]$, $F(-\pi) = 0$, $F(\pi) = r(0)$.

Доказательство. Рассмотрим функцию

$$g(\lambda, N) = \frac{1}{2\pi N} \sum_{k,l=0}^N r(k-l)e^{-i\lambda k}e^{i\lambda l}.$$

В силу неотрицательной определенности $r(n)$ $g(\lambda, N) \geq 0$. Сделав замену суммирования, получим

$$g(\lambda, N) = \frac{1}{2\pi} \sum_{|m|=0}^N \left(1 - \frac{|m|}{N}\right) r(m)e^{-i\lambda m}.$$

Рассмотрим функцию

$$F_N(\lambda) = \int_{-\pi}^{\lambda} g(x, N)d\lambda.$$

Тогда имеем

$$\int_{-\pi}^{\pi} e^{i\lambda n} dF_N(\lambda) = \frac{1}{2\pi} \sum_{|m|=0}^N r(m) \left(1 - \frac{|m|}{N}\right) \int_{-\pi}^{\pi} e^{i\lambda n} e^{-i\lambda m} d\lambda = r(n) \left(1 - \frac{|n|}{N}\right). \quad (7.23)$$

Полагая $n = 0$, получаем $F_N(\pi) - F_N(-\pi) = r(0)$. Поскольку $F_N(-\pi) = 0$, то $F_N(\pi) = r(0)$. Тогда функция $\tilde{F}_N(\lambda) = \frac{F_N(\lambda)}{r(0)}$ является функцией распределения некоторой случайной величины.

Из первой теоремы Хелли следует, что существует последовательность N_k такая, что $\tilde{F}_{N_k}(\lambda)$ будет слабо сходиться к некоторой функции распределения $\tilde{F}(\lambda)$. Тогда из (7.23) и второй теоремы Хелли получим

$$\int_{-\pi}^{\pi} e^{i\lambda n} d\tilde{F}(\lambda) = \lim_{N_k \rightarrow \infty} \int_{-\pi}^{\pi} e^{i\lambda n} \tilde{F}_{N_k}(\lambda) = \frac{r(n)}{r(0)}.$$

Взяв в качестве $F(\lambda) = r(0)\tilde{F}(\lambda)$, получим результат теоремы. □

Функция $F(\lambda)$ называется *спектральной функцией ССП* $\{\xi(n)\}$.

Если $F(\lambda) = \int_{-\pi}^{\lambda} f(x)dx$, то неотрицательная функция $f(\lambda) \geq 0$ называется *спектральной плотностью ССП* $\{\xi(n)\}$.

Теорема 7.9. Если $\{\xi(n)\}$ стационарна в широком смысле, то справедливо представление

$$\xi(n) = \int_{-\pi}^{\pi} e^{i\lambda n} d\zeta(\lambda), \quad (7.24)$$

где $\zeta(\lambda)$, $\lambda \in [-\pi, \pi]$, – случайный процесс с ортогональными приращениями;
 $E\{|\zeta(\lambda)|^2\} = F(\lambda)$.

Доказательство. Доказательство теоремы повторяет доказательство теоремы 7.2, причем $t \in \mathbb{Z}$, $\lambda \in [-\pi, \pi]$.

В случае действительного ССП справедливо представление

$$\xi(n) = \int_{-\pi}^{\pi} \cos \lambda n d\zeta_1(\lambda) + \int_{-\pi}^{\pi} \sin \lambda n d\zeta_2(\lambda),$$

где $\zeta_1(\lambda)$ и $\zeta_2(\lambda)$ – случайные процессы с ортогональными приращениями такие же, как и в п. 7.4. Доказательство повторяет доказательство представления из п. 7.4. □

7.8. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Пусть $\{\xi(n)\}$ – стационарная в широком смысле случайная последовательность со спектральной функцией $F(\lambda)$ и спектральным представлением

$$\xi(n) = \int_{-\pi}^{\pi} e^{i\lambda n} d\zeta(\lambda).$$

Определение 7.5. Последовательность $\{\eta_n\}$ называется линейным преобразованием ССП $\{\xi(n)\}$, если существует функция $g(\lambda)$ такая, что

$$\int_{-\pi}^{\pi} |g(\lambda)|^2 dF(\lambda) < \infty, \quad \eta(n) = \int_{-\pi}^{\pi} g(\lambda) e^{i\lambda n} d\zeta(\lambda). \quad (7.25)$$

Функция $g(\lambda)$ называется частотной характеристикой линейного преобразования. Найдем ковариационную функцию случайной последовательности $\{\eta(n)\}$:

$$r_{\eta}(n, m) = E\{\eta(n)\overline{\eta(m)}\} = E\left\{\int_{-\pi}^{\pi} g(\lambda) e^{i\lambda n} d\zeta(\lambda) \overline{\int_{-\pi}^{\pi} g(\lambda) e^{i\lambda m} d\zeta(\lambda)}\right\} = \int_{-\pi}^{\pi} |g(\lambda)|^2 e^{i\lambda(n-m)} dF(\lambda).$$

Отсюда следует, что $\{\eta(n)\}$ является стационарной в широком смысле случайной последовательностью с ковариационной функцией:

$$r(n) = \int_{-\pi}^{\pi} |g(\lambda)|^2 e^{i\lambda n} dF(\lambda).$$

Таким образом, если ССП $\{\xi(n)\}$ имеет спектральную плотность $f_{\xi}(\lambda)$, то ССП $\{\eta(n)\}$ имеет спектральную плотность

$$f_{\eta}(\lambda) = |g(\lambda)|^2 f_{\xi}(\lambda). \quad (7.26)$$

Рассмотрим частный случай линейного преобразования, называемый *фильтром*:

$$\eta(n) = \sum_{m=-\infty}^{+\infty} h(n-m) \xi(m). \quad (7.27)$$

Функция $h(s)$, $s \in \mathbb{Z}$, — некоторая комплекснозначная функция, называемая импульсной переходной функцией фильтра. Соотношение (7.27) можно рассматривать как значение выходного сигнала $\eta(n)$, если на вход системы, описываемой фильтром (7.27), подается сигнал $\xi(m)$, $-\infty < m < \infty$. Для физически осуществимых систем значение выходного сигнала в момент времени n определяется лишь значениями входного сигнала $\xi(m)$ при $m \leq n$.

В дальнейшем будем рассматривать фильтр с импульсной переходной функцией $h(s) = 0$ при всех $s < 0$, который называется *физически осуществимым фильтром*.

В этом случае $\eta(n) = \sum_{m=0}^{\infty} h(m) \xi(n-m)$. Тогда

$$\eta(n) = \sum_{m=0}^{\infty} h(m) \int_{-\pi}^{\pi} e^{i\lambda(n-m)} d\zeta(\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} \sum_{m=0}^{\infty} h(m) e^{-i\lambda m} d\zeta(\lambda).$$

Отсюда получаем, что последовательность $\{\eta(n)\}$ является линейным преобразованием ССП $\{\xi(n)\}$ с частотной характеристикой

$$g(\lambda) = \sum_{m=0}^{\infty} h(m) e^{-i\lambda m}. \quad (7.28)$$

Условие (7.25) примет вид

$$\int_{-\pi}^{\pi} |g(\lambda)|^2 dF(\lambda) = \sum_{n,m=0}^{\infty} h(n) \overline{h(m)} r(n-m) < \infty. \quad (7.29)$$

7.9. ПОСЛЕДОВАТЕЛЬНОСТИ АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО

Рассмотрим ССП $\{\epsilon(n)\}$, $n \in \mathbb{Z}$, с $E\{\epsilon(n)\} = 0$,

$$r_{\epsilon}(n-m) = E\{\epsilon(n) \overline{\epsilon(m)}\} = \delta_{mn}. \quad (7.30)$$

Последовательность $\{\epsilon(n)\}$ называется «белым шумом». Из (7.30) следует, что $r(n) = r_{\epsilon}(n) = \int_{-\pi}^{\pi} e^{i\lambda n} 1/(2\pi) d\lambda$. Следовательно, ССП $\{\epsilon(n)\}$ имеет спектральную плотность $f(\lambda) = 1/(2\pi)$, $\lambda \in [-\pi, \pi]$.

Рассмотрим теперь случайную последовательность

$$\xi(n) = \sum_{k=0}^{\infty} a_k \epsilon(n-k), \quad (7.31)$$

которая называется *последовательностью скользящего среднего* и является фильтрацией последовательности $\{\epsilon(n)\}$ с частотной характеристикой $g(\lambda) = \sum_{k=0}^{\infty} a_k e^{-i\lambda k}$. Чтобы последовательность $\{\xi(n)\}$ была стационарной в широком смысле, необходимо выполнение условия (7.29), которое в нашем случае имеет вид

$$\sum_{k=0}^{\infty} |a_k|^2 < \infty. \quad (7.32)$$

В случае, когда

$$\xi(n) = \sum_{k=0}^p a_k \epsilon(n-k),$$

последовательность $\{\xi(n)\}$ называется *последовательностью скользящего среднего порядка p* и обозначается $CC(p)$. Последовательность $CC(p)$ имеет спектральную плотность (см. (7.26)):

$$f_{\xi}(\lambda) = \frac{1}{2\pi} |P(e^{-i\lambda})|^2, \quad P(z) = \sum_{k=0}^p a_k z^k. \quad (7.33)$$

Последовательность $\{\xi(n)\}$, задаваемая соотношением

$$\sum_{k=0}^q b_k \xi(n-k) = \epsilon(n),$$

называется *последовательностью авторегрессии* и обозначается $AP(q)$.

Последовательность $\{\xi(n)\}$, задаваемая соотношением

$$\sum_{k=0}^p a_k \epsilon(n-k) = \sum_{k=0}^q b_k \xi(n-k), \quad (7.34)$$

называется *смешанной моделью авторегрессии и скользящего среднего* и обозначается $\text{АРСС}(q, p)$.

Покажем, что последовательность $\text{АРСС}(q, p)$ является последовательностью скользящего среднего, т. е. представима в виде

$$\xi(n) = \sum_{k=0}^{\infty} c_k \epsilon(n-k). \quad (7.35)$$

Для удобства будем считать, что $p = q = s$. Подставляя (7.35) в (7.34) и сравнивая коэффициенты при одинаковых элементах последовательности $\{\epsilon(n)\}$, получаем систему уравнений относительно $\{c_k\}$:

$$\begin{cases} c_0 = a_0, \\ c_1 + b_1 c_0 = a_1, \\ \dots\dots\dots \\ c_s + b_1 c_{s-1} + \dots + b_s c_0, \\ c_p + b_1 c_{p-1} + \dots + b_s c_{p-s} = 0 \text{ при } p > s. \end{cases} \quad (7.36)$$

Для нахождения коэффициентов $\{c_k : k = 0, 1, \dots\}$ введем производящие функции $A(z) = \sum_{k=0}^s a_k z^k$, $B(z) = \sum_{k=0}^s b_k z^k$, $C(z) = \sum_{k=0}^{\infty} c_k z^k$. Тогда, умножая k -е ($k = 0, 1, 2, \dots$) уравнение (7.36) на z^k и суммируя по k , получим

$$A(z) = \sum_{k=0}^{\infty} \sum_{l=0}^s c_{k-l} b_l z^k = B(z)C(z).$$

Поскольку $A(z)$ и $B(z)$ – полиномы степени s , то

$$C(z) = \frac{A(z)}{B(z)} = a_0 + z \frac{A_1(z)}{B(z)}, \quad (7.37)$$

где $A_1(z)$ – полином степени $s-1$. Условие стационарности в широком смысле последовательности $\{\xi(n)\}$:

$$\sum_{n=0}^{\infty} |c_n|^2 < \infty. \quad (7.38)$$

Теорема 7.10. *Если корни полинома $B(z)$ лежат вне круга $|z| = 1$, то выполняется условие (7.38), т. е. последовательность $\{\xi(n)\}$ стационарна в широком смысле.*

Доказательство. Воспользуемся равенством (7.37). Для простоты доказательства будем считать, что корни полинома $B(z)$ – простые. Для кратных корней доказательство проводится аналогично. Имеем

$$C(z) = a_0 + z \left(\frac{A_1}{z_1 - z} + \dots + \frac{A_s}{z_s - z} \right) = a_0 + \sum_{n=1}^{\infty} \left(\frac{A_1}{z_1^n} + \dots + \frac{A_s}{z_s^n} \right) z^n.$$

Отсюда $c_n = A_1 z_1^{-n} + \dots + A_s z_s^{-n}$. Если $|z_n| > 1$, $n = 1, 2, \dots$, то выполняется условие (7.38). \square

Из доказанной теоремы следует, что последовательность $\text{APCC}(q, p)$ имеет спектральную плотность

$$f_\xi(\lambda) = \frac{1}{2\pi} \sum_{k=0}^{\infty} c_k e^{-i\lambda k}.$$

Выразим эту спектральную плотность через коэффициенты a_k и b_k .

Для последовательности $\text{AP}(q)$ имеем $1/(2\pi) = |Q(e^{-i\lambda})|^2 f_\xi(\lambda)$, где $Q(z) = \sum_{k=0}^q b_k z^k$.

Следовательно, для последовательности $\text{AP}(q)$ $f_\xi(\lambda) = (2\pi)^{-1} |Q(e^{-i\lambda})|^{-2}$.

Если левую часть (7.34) обозначить $\eta(n)$, то из (7.26) имеем $f_\eta(\lambda) = |Q(e^{-i\lambda})|^2 f_\xi(\lambda)$. Поскольку $f_\eta(\lambda) = 1/(2\pi) |P(e^{-i\lambda})|^2$, то спектральная плотность $f_\xi(\lambda)$ последовательности $\text{APCC}(q, p)$ будет иметь вид

$$f_\xi(\lambda) = \frac{1}{2\pi} \left| \frac{P(e^{-i\lambda})}{Q(e^{-i\lambda})} \right|^2. \quad (7.39)$$

7.10. ПРОГНОЗИРОВАНИЕ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Пусть $\{\xi(n)\}$ – стационарная в широком смысле случайная последовательность, $n \in \mathbb{Z}$. Пусть известны значения случайных величин $\{\xi(m) : m \leq 0\}$. Требуется найти оценку (прогноз) случайной величины $\xi(n)$, $n > 0$.

Обозначим $H(\xi, 0)$ – пространство случайных величин $\{\xi(m) : m \leq 0\}$, их линейных комбинаций и пределов в среднем квадратичном этих линейных комбинаций.

Будем искать оценку $\xi^*(n)$, принадлежащую пространству $H(\xi, 0)$ и удовлетворяющую соотношению

$$\mathbf{E}\{|\xi(n) - \xi^*(n)|^2\} = \min_{\tilde{\xi} \in H(\xi, 0)} \mathbf{E}\{|\xi(n) - \tilde{\xi}|^2\}. \quad (7.40)$$

Теорема 7.11. Если случайная величина $\xi^*(n) \in H(\xi, 0)$ удовлетворяет соотношению

$$\mathbf{E}\{(\xi(n) - \xi^*(n))\overline{\xi(m)}\} = 0 \quad (7.41)$$

для всех $m \leq 0$, то она удовлетворяет соотношению (7.40).

Доказательство теоремы повторяет доказательство теоремы 7.7 и следствия 7.1 для случая дискретной переменной $t = n \in \mathbb{Z}$.

Рассмотрим случай, когда

$$\xi(n) = \sum_{k=0}^{\infty} a_k \epsilon(n-k), \quad (7.42)$$

где $\{\epsilon(n)\}$ – «белый шум». Поскольку пространство $H(\xi, 0) = H(\epsilon, 0)$, то соотношение (7.41) будет эквивалентно соотношению

$$\mathbf{E}\{(\xi(n) - \xi^*(n))\overline{\epsilon(m)}\} = 0, \quad m \leq 0. \quad (7.43)$$

Этому соотношению удовлетворяет случайная величина

$$\xi^*(n) = \sum_{k=n}^{\infty} a_k \epsilon(n-k). \quad (7.44)$$

$$\begin{aligned} \text{Действительно, } \xi^*(n) \in H(\epsilon, 0) \text{ и } E \left\{ \left(\sum_{k=0}^{\infty} a_k \epsilon(n-k) - \sum_{k=n}^{\infty} a_k \epsilon(n-k) \right) \overline{\epsilon(m)} \right\} = \\ = E \left\{ \left(\sum_{k=0}^{n-1} a_k \epsilon(n-k) \right) \overline{\epsilon(m)} \right\} = \sum_{k=0}^{n-1} a_k E \{ \epsilon(n-k) \overline{\epsilon(m)} \} = 0. \quad \square \end{aligned}$$

Покажем, как получить оценку $\xi^*(n)$ через значения случайных величин $\{\xi(m) : m \leq 0\}$. Случайная последовательность $\{\xi(n)\}$, представляемая в виде (7.42), имеет спектральную плотность $f_{\xi}(\lambda) = 1/(2\pi)|\Phi(e^{-i\lambda})|^2$, где $\Phi(z) = \sum_{k=0}^{\infty} a_k z^k$. Воспользуемся спектральным представлением ССП $\{\xi(n)\}$:

$$\xi(n) = \int_{-\pi}^{\pi} e^{i\lambda n} d\zeta(\lambda).$$

Каждой случайной величине $\tilde{\xi} \in H(\xi, 0)$ будет соответствовать функция $\tilde{g}(\lambda)$ такая, что

$$\tilde{\xi} = \int_{-\pi}^{\pi} \tilde{g}(\lambda) d\zeta(\lambda),$$

причем $\tilde{g}(\lambda) \in H(F, 0)$ — замкнутое линейное пространство, порожденное функциями $e^{i\lambda n}$, $n \leq 0$.

Теорема 7.12. Случайной величине $\xi^*(n)$ соответствует функция

$$g_n^*(\lambda) = e^{i\lambda n} \frac{\Phi_n(e^{-i\lambda})}{\Phi(e^{-i\lambda})}, \quad \text{где } \Phi_n = \sum_{k=n}^{\infty} a_k z^k.$$

Доказательство. Проверим выполнимость (7.43). Имеем цепочку равенств

$$\begin{aligned} E\{(\xi(n) - \xi^*(n)) \overline{\xi(m)}\} &= E \left\{ \int_{-\pi}^{\pi} (e^{i\lambda n} - g^*(\lambda)) d\zeta(\lambda) \int_{-\pi}^{\pi} e^{-i\lambda m} d\zeta(\lambda) \right\} = \\ &= \int_{-\pi}^{\pi} (e^{i\lambda n} - g^*(\lambda)) e^{-i\lambda m} dF(\lambda) = \int_{-\pi}^{\pi} (e^{-i\lambda n} - g^*(\lambda)) e^{-i\lambda m} f_{\xi}(\lambda) d\lambda = \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (e^{-i\lambda(n-m)} \left(1 - \frac{\Phi_n(e^{-i\lambda})}{\Phi(e^{-i\lambda})} \right) |\Phi(e^{-i\lambda})|^2 d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} (e^{i\lambda(n-m)} (\Phi(e^{-i\lambda}) - \\ &- \Phi_n(e^{-i\lambda})) \overline{\Phi(e^{-i\lambda})} d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} (e^{-i\lambda(n-m)} \sum_{k=0}^{n-1} a_k e^{-i\lambda k} \sum_{l=0}^{\infty} \bar{a}_l e^{i\lambda l} d\lambda = \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=0}^{n-1} a_k \sum_{l=0}^{\infty} \bar{a}_l \int_{-\pi}^{\pi} e^{-i\lambda(n-m-k+l)} d\lambda = 0. \end{aligned}$$

□

Разлагая функцию $g_n^*(\lambda)$ в ряд Фурье: $g_n^*(\lambda) = C_0 + C_{-1}e^{-i\lambda} + \dots + C_{-k}e^{-i\lambda k} + \dots$, находим

$$\xi^*(n) = \int_{-\pi}^{\pi} g_n^*(\lambda) d\zeta(\lambda) = C_0 \xi(0) + C_{-1} \xi(-1) + C_2 \xi(-2) + \dots$$

7.11. ИНТЕРПОЛЯЦИЯ И ФИЛЬТРАЦИЯ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Пусть $\{\xi(n)\}$ – стационарная в широком смысле случайная последовательность со спектральной плотностью $f(\lambda)$.

Интерполяция. Пусть известны значения $\xi(n)$ для всех $n \neq 0$. Требуется найти оценку (интерполяцию) пропущенного значения $\xi(0)$.

Обозначим $H^0(\xi)$ – замкнутое линейное пространство, порожденное случайными величинами $\{\xi(n) : n \neq 0\}$, $H^0(F)$ – замкнутое линейное пространство, порожденное функциями $\{e^{i\lambda n} : n \neq 0\}$. Будем искать оценку ξ_0^* случайной величины $\xi(0)$ из пространства $H^0(\xi)$, удовлетворяющую условию

$$E\{|\xi(0) - \xi_0^*|^2\} = \min_{\tilde{\xi} \in H^0(\xi)} E\{|\xi(0) - \tilde{\xi}|^2\}. \quad (7.45)$$

Условие (7.45), по доказанному ранее, эквивалентно условию

$$E\{(\xi(0) - \xi_0^*)\bar{\xi}(n)\} = 0, \quad n \neq 0. \quad (7.46)$$

Используя спектральное представление, имеем

$$\xi_0^* = \int_{-\pi}^{\pi} g_0^*(\lambda) d\zeta(\lambda), \quad \text{где } g_0(\lambda) \in H^0(F).$$

Теорема 7.13 (теорема Колмогорова). Если $\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)} < \infty$, то функция $g_0^*(\lambda)$ имеет вид

$$g_0^*(\lambda) = 1 - \frac{\alpha}{f(\lambda)}, \quad \alpha = \int_{-\pi}^{\pi} \frac{2\pi}{\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}}.$$

Доказательство. Преобразуем равенство (7.46):

$$\begin{aligned} E\{(\xi(0) - \xi_0^*)\bar{\xi}(n)\} &= E\{\xi(0)\bar{\xi}(n)\} - E\{\xi_0^*\bar{\xi}(n)\} = \\ &= \int_{-\pi}^{\pi} e^{-i\lambda n} dF(\lambda) - \int_{-\pi}^{\pi} g_0^*(\lambda) e^{-i\lambda n} dF(\lambda) = \int_{-\pi}^{\pi} e^{-i\lambda n} (1 - g_0^*(\lambda)) f(\lambda) d\lambda = 0. \end{aligned}$$

Поскольку это равенство должно выполняться для всех $n \neq 0$, то $(1 - g_0^*(\lambda))f(\lambda) = \alpha = \text{const}$. Отсюда $g_0^*(\lambda) = 1 - \alpha/f(\lambda)$.

Поскольку функция $g_0^*(\lambda) \in H^0(F)$, то в разложении ее в ряд Фурье будет отсутствовать свободный член. Следовательно, $\int_{-\pi}^{\pi} g_0^*(\lambda) d\lambda = 0$.

Отсюда имеем

$$0 = \int_{-\pi}^{\pi} \left(1 - \frac{\alpha}{f(\lambda)}\right) d\lambda = 2\pi - \alpha \int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}, \quad \alpha = \frac{2\pi}{\int_{-\pi}^{\pi} \frac{d\lambda}{f(\lambda)}}.$$

□

Разлагая функцию $g_0^*(\lambda)$ в ряд Фурье: $g_0^*(\lambda) = \sum_{n \neq 0} c_n e^{i\lambda n}$, получаем оценку

$$\xi_0^* = \sum_{n \neq 0} c_n \xi(n).$$

Фильтрация случайной последовательности. Пусть $(\{\theta(n)\}, \{\xi(n)\})$, $n \in \mathbb{Z}$, — частично наблюдаемая двухкомпонентная последовательность, где $\{\theta(n)\}$ — ненаблюдаемая, а $\{\xi(n)\}$ — наблюдаемая компонента. Пусть $\{\theta(n)\}$ и $\{\xi(n)\}$ — стационарные в широком смысле случайные последовательности с нулевыми средними и спектральными представлениями $\theta(n) = \int_{-\pi}^{\pi} e^{i\lambda n} d\zeta_{\theta}(\lambda)$, $\xi(n) = \int_{-\pi}^{\pi} e^{i\lambda n} d\zeta_{\xi}(\lambda)$ и спектральными плотностями $f_{\theta}(\lambda)$, $f_{\xi}(\lambda)$ соответственно.

Предположим

$$E\{\theta_n \overline{\xi_m}\} = r(n-m) = \int_{-\pi}^{\pi} e^{i\lambda(n-m)} dF_{\theta\xi}(\lambda), \quad F_{\theta\xi}(\lambda) = \int_{-\pi}^{\lambda} f_{\theta\xi}(x) dx.$$

Функция $f_{\theta\xi}(\lambda)$ называется *взаимной спектральной плотностью последовательностей* $\{\theta(n)\}$ и $\{\xi(n)\}$.

Задача фильтрации состоит в построении оптимальной в среднеквадратическом смысле оценки $\hat{\theta}_n$ величины θ_n по наблюдаемым значениям последовательности $\{\xi(n)\}$.

Будем искать оценку $\theta_n^* \in H(\xi)$, удовлетворяющую условию

$$E\{|\theta_n - \theta_n^*|^2\} = \min_{\tilde{\theta}_n \in H(\xi)} E\{|\theta_n - \tilde{\theta}_n|^2\}. \quad (7.47)$$

По доказанному выше это условие будет эквивалентно условию

$$E\{(\theta_n - \theta_n^*) \overline{\xi(m)}\} = 0, \quad m \in \mathbb{Z}. \quad (7.48)$$

Пусть $\theta_n^* = \int_{-\pi}^{\pi} g_n^*(\lambda) d\zeta_{\xi}(\lambda)$. Используя спектральное представление, получаем из (7.48):

$$0 = E\{\theta_n \overline{\xi(m)}\} - E\{\theta_n^* \overline{\xi(m)}\} = \int_{-\pi}^{\pi} e^{i\lambda(n-m)} (f_{\theta\xi}(\lambda) - g_n^*(\lambda) e^{-i\lambda n} f_{\xi}(\lambda)) d\lambda.$$

Поскольку это равенство должно выполняться для всех m , то $f_{\theta\xi}(\lambda) - g_n^*(\lambda) e^{-i\lambda n} f_{\xi}(\lambda) = 0$. Отсюда

$$g_n^*(\lambda) = \frac{e^{i\lambda n} f_{\theta\xi}(\lambda)}{f_{\xi}(\lambda)}. \quad (7.49)$$

Разлагая функцию $g_n^*(\lambda)$ в ряд Фурье

$$g_n^*(\lambda) = \sum_k c_k e^{i\lambda k}, \quad (7.50)$$

получаем

$$\theta_n^* = \sum_k c_k \xi(k). \quad (7.51)$$

Рассмотрим пример. Пусть $\xi(n) = \theta(n) + \eta(n)$, $n \in \mathbb{Z}$, где сигнал $\{\theta(n)\}$ и шум $\{\eta(n)\}$ являются некоррелированными последовательностями со спектральными плотностями $f_\theta(\lambda)$ и $f_\eta(\lambda)$ соответственно. Требуется найти θ_n^* по значениям $\{\xi(m) : m \in \mathbb{Z}\}$.

Согласно (7.49) $g_n^*(\lambda) = \frac{e^{i\lambda n} f_{\theta\xi}(\lambda)}{f_\xi(\lambda)}.$

Поскольку

$$\begin{aligned} E\{\theta(n)\overline{\xi(m)}\} &= E\{\theta(n)(\overline{\theta(m)} + \overline{\eta(m)})\} = E\{\theta(n)\overline{\theta(m)}\} + E\{\theta(n)\overline{\eta(m)}\} = \\ &= E\{\theta(n)\overline{\theta(m)}\} = \int_{-\pi}^{\pi} e^{i\lambda(n-m)} f_\theta(\lambda) d\lambda; \end{aligned}$$

$$\begin{aligned} E\{\xi(n)\overline{\xi(m)}\} &= E\{(\theta(n) + \eta(n))(\overline{\theta(m)} + \overline{\eta(m)})\} = \\ &= E\{\theta(n)\overline{\theta(m)}\} + E\{\eta(n)\overline{\eta(m)}\} = \int_{-\pi}^{\pi} e^{i\lambda(n-m)} (f_\theta(\lambda) + f_\eta(\lambda)) d\lambda, \end{aligned}$$

то $f_\xi(\lambda) = f_\theta(\lambda) + f_\eta(\lambda)$, $f_{\theta\xi}(\lambda) = f_\theta(\lambda)$. Тогда

$$g_n^*(\lambda) = \frac{e^{i\lambda n} f_\theta(\lambda)}{f_\theta(\lambda) + f_\eta(\lambda)}.$$

Используя (7.50) и (7.51), получим оценку θ_n^* .

7.12. УПРАЖНЕНИЯ

1. Случайный процесс $\xi(t)$ представляет собой случайную величину η , $E\eta = a$, $D\eta = \sigma^2$. Найти $E\{\xi(t)\}$, $D\{\xi(t)\}$, ковариационную функцию. Определить, является ли процесс стационарным.
2. Пусть ϕ – случайная величина с плотностью $P_\phi(x) = \cos x$, $0 \leq x \leq \pi/2$; α и w – положительные постоянные. Построим случайный процесс $\xi(t) = a \sin(wt + \phi)$. Будет ли этот процесс стационарным?
3. Пусть ϕ – случайная величина, равномерно распределенная на отрезке $[0, 2\pi]$, α и w – положительные постоянные. Доказать, что случайный процесс $\xi(t) = a \sin(wt + \phi)$ – стационарный (в широком смысле).
4. Стационарный и гауссовский процесс $\xi(t)$ имеет математическое ожидание $m_x = 5$ и ковариационную функцию $r(t, u) = e^{-2|\tau|}(\cos 2\tau + \sin 2|\tau|)$. Найти:
 - а) одномерную плотность процесса $\eta(t) = \frac{d\xi(t)}{dt}$;
 - б) вероятность того, что $|\eta(t)| < \sqrt{3}$.

5. Доказать, что следующие функции являются положительно определенными:

$$r_1(t, s) = \min(t, s) - ts, \quad t, s \in [0, 1];$$

$$r_2(t, s) = e^{-|t-s|}, \quad t, s \in \mathbb{R}.$$

6. Пусть $\phi_1(t), \dots, \phi_n(t)$ – произвольные действительные функции, C_1, \dots, C_n – неотрицательные числа. Доказать, что функция $r(t_1, t_2) = \sum_{i=1}^n C_i \phi_i(t_1) \phi_i(t_2)$ является ковариационной функцией некоторого случайного процесса.
7. Случайный процесс $\xi(t)$ строится следующим образом. В точке $t = 0$ он случайным образом с одинаковой вероятностью принимает одно из значений $+1$ или -1 и остается постоянным до точки $t = 1$, в точке $t = 1$ снова равновероятно и независимо от предыдущего значения принимает одно из значений $+1$ или -1 и сохраняет его до точки $t = 2$, и т. д. Определить математическое ожидание, дисперсию и ковариационную функцию $\xi(t)$. Является ли случайный процесс стационарным в широком смысле?
8. Случайный процесс $\xi(t)$ строится так же, как и в предыдущем упражнении, с той лишь разницей, что точки, в которых происходит «розыгрыш» нового значения, не закреплены на оси времени, а занимают на ней случайное положение, сохраняя между собой постоянное расстояние, равное единице. Распределение начала отсчета относительно первого момента «розыгрыша» – равномерное на $[0, 1]$. Найти $E\{\xi(t)\}$, $D\{\xi(t)\}$, ковариационную функцию. Определить, является ли $\xi(t)$ стационарным в широком смысле.
9. Доказать, что если $\xi(t)$ – стационарный в широком смысле случайный процесс, то его дисперсия $D\{\xi(t)\} = 2 \int_0^\infty f(\lambda) d\lambda$.
10. Найти спектральные плотности, соответствующие ковариационным функциям:
- $r(\tau) = \begin{cases} 1 - \frac{|\tau|}{\tau_0}, & \text{при } |\tau| \leq \tau_0, \\ 0, & |\tau| > \tau_0; \end{cases}$
 - $r(\tau) = e^{-|\tau|} \cos \beta \tau$;
 - $r(\tau) = \sigma^2 e^{-\alpha|\tau|}$;
 - $r(\tau) = \sigma^2 e^{-\alpha|\tau|} (1 + \alpha|\tau|)$;
 - $r(\tau) = \sigma^2 e^{-\alpha|\tau|} (\cos \beta \tau + \frac{\alpha}{\beta} \sin \beta|\tau|)$.
11. Случайный процесс $\xi(t)$ имеет спектральную плотность, равную нулю вне промежутка $[\lambda_1, \lambda_2]$ и постоянную на этом промежутке. Найти коэффициент корреляции между $\xi(t_1)$ и $\xi(t_2)$; t_1, t_2 произвольны.
12. Случайный процесс $\xi(t)$ имеет спектральную плотность $f(\lambda) = \frac{a^2}{a^2 + \lambda^2}$. Найти ковариационную функцию.
13. Будут ли дифференцируемы случайные процессы, имеющие следующие спектральные плотности:
- $f(\lambda) = \frac{a^2}{\lambda^2 + b^2}$;
 - $f(\lambda) = \sum_j \frac{a_j^2}{\lambda^2 + b_j^2}$.

Глава 8

СЛУЧАЙНЫЕ ПРОЦЕССЫ С НЕЗАВИСИМЫМИ ПРИРАЩЕНИЯМИ

8.1. ВИД ХАРАКТЕРИСТИЧЕСКОЙ ФУНКЦИИ СЛУЧАЙНОГО ПРОЦЕССА С НЕЗАВИСИМЫМИ ПРИРАЩЕНИЯМИ

Рассмотрим СП $\xi(t)$, где $t \geq 0$.

Определение 8.1. СП $\xi(t)$ называется стохастически непрерывным в точке t_0 , если

$$\xi(t) \xrightarrow[t \rightarrow t_0]{P} \xi(t_0).$$

Определение 8.2. СП $\xi(t)$ называется процессом с независимыми приращениями, если для любого n и любых $0 \leq t_1 < t_2 < \dots < t_n$ случайные величины $\xi(t_1), \xi(t_2) - \xi(t_1), \dots, \xi(t_n) - \xi(t_{n-1})$ независимы.

Определение 8.3. СП $\xi(t)$ называется однородным, если для любых $t \geq 0, \tau > 0$ распределение случайной величины $\xi(t + \tau) - \xi(t)$ не зависит от t .

Замечание 8.1. Из стохастической непрерывности СП не следует непрерывность траекторий СП. Для таких процессов возможны разрывы траекторий в случайные моменты времени (см. п. 8.3).

Замечание 8.2. Из свойств сходимостей следует, что из стохастической непрерывности СП в точке t_0 следует непрерывность характеристической функции $f_{\xi(t)}(\lambda)$ по t в точке t_0 .

Теорема 8.1. Если СП $\xi(t), \xi(0) = 0$, стохастически непрерывен, однороден и является процессом с независимыми приращениями, то его характеристическая функция $f_{\xi(t)}(\lambda) = E\{e^{i\lambda\xi(t)}\}$ имеет вид

$$f_{\xi(t)}(\lambda) = (f_{\xi(1)}(\lambda))^t.$$

Доказательство. Пусть $0 \leq u < t$. Тогда

$$f_t(\lambda) = f_{\xi(t)}(\lambda) = E\{e^{i\lambda\xi(t)}\} = E\{e^{i\lambda(\xi(t) - \xi(u) + \xi(u))}\} = E\{e^{i\lambda(\xi(t) - \xi(u))} e^{i\lambda\xi(u)}\}.$$

Используя независимость приращений и однородность СП $\xi(t)$, получаем

$$f_t(\lambda) = E\{e^{i\lambda(\xi(t) - \xi(u))}\} E\{e^{i\lambda\xi(u)}\} = E\{e^{i\lambda\xi(t-u)}\} E\{e^{i\lambda\xi(u)}\} = f_{t-u}(\lambda) f_u(\lambda).$$

Положим $g(t) = f_t(\lambda)$. Из предыдущего равенства получаем

$$g(t) = g(t-u) g(u).$$

Полагая $u = t_1, t - u = t_2$, имеем

$$g(t_1 + t_2) = g(t_1) g(t_2).$$

Отсюда по индукции

$$g(t_1 + t_2 + \dots + t_n) = g(t_1) g(t_2) \times \dots \times g(t_n) \quad (8.1)$$

для $t_1 \geq 0, t_2 \geq 0, \dots, t_n \geq 0$.

Полагая в (8.1) $t_j = 1/n$ для всех $j = 1, \dots, n$, получаем $g(1) = (g(1/n))^n$, т. е. $g(1/n) = (g(1))^{1/n}$. Заменяя в (8.1) n на k и полагая $t_j = 1/n$ для всех $j = 1, \dots, k$, получим

$$g\left(\frac{k}{n}\right) = \left(g\left(\frac{1}{n}\right)\right)^k = (g(1))^{\frac{k}{n}},$$

т. е. для всех рациональных чисел r , $g(r) = (g(1))^r$.

Пусть t — произвольное число больше нуля. Тогда найдется последовательность рациональных чисел $\{r_n\}$ такая, что $r_n \xrightarrow{n \rightarrow \infty} t$. В силу непрерывности функции $g(t)$ имеем

$$g(t) = \lim_{n \rightarrow \infty} g(r_n) = \lim_{n \rightarrow \infty} (g(1))^{r_n} = (g(1))^t.$$

□

8.2. ВИНЕРОВСКИЙ ПРОЦЕСС

Определение 8.4. СП $\xi(t)$, $t \geq 0$, называется винеровским процессом (ВП), если он удовлетворяет следующим условиям:

1. $\xi(0) = 0$.
 2. $\xi(t)$ — однородный процесс с независимыми приращениями.
 3. $\mathcal{L}(\xi(1)) = N(a, \sigma^2)$.
 4. Траектории СП $\xi(t)$ с вероятностью единица непрерывны.
- В случае $a = 0$, $\sigma = 1$ СП $\xi(t)$ называется стандартным винеровским процессом и обозначается $w(t)$.

Из теоремы 8.1 следует, что

$$f_{\xi(t)}(\lambda) = (f_{\xi(1)}(\lambda))^t = \left(e^{i\lambda a - \frac{1}{2}\lambda^2 \sigma^2}\right)^t = e^{i\lambda at - \frac{1}{2}\lambda^2 \sigma^2 t},$$

т. е. $\mathcal{L}(\xi(t)) = N(at, \sigma^2 t)$.

Из свойств нормально распределенной случайной величины следует

$$w(t) = \frac{\xi(t) - at}{\sigma\sqrt{t}}.$$

Замечание 8.3. В определении ВП условие непрерывности СП можно заменить на условие сепарабельности.

Доказательство следует из теоремы 6.8, если взять в качестве $g(h) = h^\alpha$, $\alpha < 1/2$. Действительно,

$$\begin{aligned} P\{|w(t+h) - w(t)| > h^\alpha\} &= P\{|w(h)| > h^\alpha\} = \\ &= \frac{1}{\sqrt{2\pi h}} \int_{|x| > h^\alpha} e^{-\frac{x^2}{2h}} dx = \frac{1}{\sqrt{2\pi}} \int_{|y| > h^{\alpha-\frac{1}{2}}} e^{-\frac{y^2}{2}} dy = q(h). \end{aligned}$$

В дальнейшем нам понадобится нижеследующая лемма.

Лемма 8.1. При $x \rightarrow \infty$ $\int_x^\infty e^{-\frac{y^2}{2}} dy = O\left(\frac{1}{x} e^{-\frac{x^2}{2}}\right)$.

Доказательство следует из Правила Лопиталя.

Тогда по лемме $q(h) = O\left(h^{1/2-\alpha} e^{-1/2 h^{2\alpha-1}}\right)$. Следовательно, функции $q(h)$ и $g(h)$ удовлетворяют условиям теоремы 6.8.

Свойства винеровского процесса

С1. ВП $w(t)$ является гауссовским процессом с $E\{w(t)\} = 0$ и ковариационной функцией $r(t, u) = \min(t, u)$.

Доказательство. Пусть $0 \leq t_1 < t_2 < \dots < t_n$. Тогда

$$\begin{aligned} f_{w(t_1), w(t_2), \dots, w(t_n)}(\lambda_1, \dots, \lambda_n) &= E\{\exp\{i(\lambda_1 w(t_1) + \lambda_2 w(t_2) + \dots + \lambda_n w(t_n))\}\} = \\ &= E\{\exp\{i[\lambda_n(w(t_n) - w(t_{n-1}))) + (\lambda_n + \lambda_{n-1})(w(t_{n-1}) - w(t_{n-2})) + \\ &\quad + \dots + (\lambda_n + \dots + \lambda_1)(w(t_1))]\}\} = E\{\exp\{i\lambda_n(w(t_n) - w(t_{n-1})))\}\} \times \\ &\quad \times E\{\exp\{i(\lambda_n + \lambda_{n-1})(w(t_{n-1}) - w(t_{n-2})))\}\} \times \dots \times \\ &\quad \times E\{\exp\{i(\lambda_n + \dots + \lambda_1)(w(t_1))\}\} = \exp\left\{-\frac{1}{2}\lambda_n^2(t_n - t_{n-1})\right\} \times \\ &\times \exp\left\{-\frac{1}{2}(\lambda_n + \lambda_{n-1})^2(t_{n-1} - t_{n-2})\right\} \times \dots \times \exp\left\{-\frac{1}{2}(\lambda_n + \lambda_{n-1} + \dots + \lambda_1)^2 t_1\right\} = \\ &= \exp\left\{-\frac{1}{2}[\lambda_n^2(t_n - t_{n-1}) + (\lambda_n + \lambda_{n-1})^2(t_{n-1} - t_{n-2}) + \dots + \right. \\ &\quad \left. + (\lambda_n + \lambda_{n-1} + \dots + \lambda_1)^2 t_1]\right\}. \end{aligned} \quad (8.2)$$

По определению гауссовского процесса следует, что процесс $w(t)$ гауссовский. По определению винеровского процесса $E\{w(t)\} = 0$. Пусть $t < u$. Тогда

$$\begin{aligned} r(t, u) &= E\{w(t)w(u)\} = E\{(w(u) - w(t) + w(t))w(t)\} = \\ &= E\{(w(u) - w(t))w(t)\} + E\{(w(t))^2\} = E\{w(u) - w(t)\}E\{w(t)\} + E\{(w(t))^2\} = \\ &= E\{(w(t))^2\} = t. \end{aligned}$$

□

С2. ВП $w(t)$ является симметрическим СП, т. е. конечномерные распределения случайных процессов $w(t)$ и $-w(t)$ совпадают.

Доказательство следует из равенства (8.2). Поскольку

$$\begin{aligned} f_{w(t_1), w(t_2), \dots, w(t_n)}(\lambda_1, \dots, \lambda_n) &= f_{w(t_1), w(t_2), \dots, w(t_n)}(-\lambda_1, \dots, -\lambda_n) = \\ &= f_{-w(t_1), -w(t_2), \dots, -w(t_n)}(\lambda_1, \dots, \lambda_n). \end{aligned}$$

□

С3. ВП $w(t)$ обладает свойством строгой марковости.

Пусть $0 \leq t_1 < t_2 < \dots < t_n < t$, B – борелевское множество, $x_1, \dots, x_n \in \mathbb{R}$. Используя свойства ВП $w(t)$, имеем

$$\begin{aligned} P\{w(t) \in B | w(t_n) = x_n, \dots, w(t_1) = x_1\} &= \\ &= P\{w(t) - w(t_n) \in B - \{x_n\} | w(t_n) - w(t_{n-1}) = x_n - x_{n-1}, \dots, \\ &\quad w(t_2) - w(t_1) = x_2 - x_1, w(t_1) = x_1\} = P\{w(t) - w(t_n) \in B - \{x_n\}\} = \\ &= P\{w(t) \in B | w(t_n) = x_n\} = \frac{1}{\sqrt{2\pi(t - t_n)}} \int_B e^{-\frac{(x - x_n)^2}{2(t - t_n)}} dx. \end{aligned} \quad (8.3)$$

Отсюда следует свойство марковости процесса $w(t)$ при $s < t$:

$$P\{w(t) \in B | \mathcal{F}_s\} = P\{w(t) \in B | w(s)\} = \frac{1}{\sqrt{2\pi(t - s)}} \int_B e^{-\frac{(x - w(s))^2}{2(t - s)}} dx,$$

где $\mathcal{F}_s = \sigma(w(t), t \leq s)$. Из равенства (8.3) и независимости приращения СП $w(t)$ следует при $s < t$, что

$$P\{w(t) - w(s) \in B | \mathcal{F}_s\} = P\{w(t) - w(s) \in B\}. \quad (8.4)$$

Пусть τ – марковский момент, т. е. при каждом t событие $(\tau \leq t) \in \mathcal{F}_t$,

$$\mathcal{F}_\tau = (A \in \mathcal{F} : A(\tau = t) \in \mathcal{F}_t).$$

Замечание 8.4. Марковским моментом будет момент достижения СП некоторого множества B :

$$\tau = \min(t : \xi(t) \in B).$$

Свойство строгой марковости означает, что для любого $t > 0$ и марковского момента $\tau \geq 0$

$$P\{w(t + \tau) - w(\tau) \in B | \mathcal{F}_\tau\} = P\{w(t) \in B\}.$$

Последнее соотношение равносильно следующему: для любого события $A \in \mathcal{F}_\tau$

$$P\{A \cap (w(t + \tau) - w(\tau) \in B)\} = P\{A\}P\{w(t) \in B\}. \quad (8.5)$$

Доказательство. Если марковский момент τ принимает конечное число значений $\tau = t_1, t_2, \dots, t_n$, то, используя (8.4), имеем

$$\begin{aligned} P\{A \cap (w(t + \tau) - w(\tau) \in B)\} &= \sum_j P\{A(\tau = t_j) (w(t + \tau) - w(\tau) \in B)\} = \\ &= \sum_j P\{A(\tau = t_j) (w(t + t_j) - w(t_j) \in B)\} = \\ &= \sum_j P\{A(\tau = t_j)\} P\{w(t + t_j) - w(t_j) \in B\} = \\ &= \sum_j P\{A(\tau = t_j)\} P\{w(t) \in B\} = P\{A\}P\{w(t) \in B\}. \end{aligned} \quad (8.6)$$

В общем случае равенство (8.5) доказывается приближением случайной величины τ марковскими моментами, принимающими конечное число значений.

Рассмотрим последовательность марковских моментов

$$\tau^{(n)} = \sum_{k=1}^{n^2} \frac{k}{n} \mathbb{1}_{\frac{k}{n} \leq \tau < \frac{k+1}{n}} + n \mathbb{1}_{(\tau > n)}.$$

Поскольку $\tau^{(n)}$ стремится к τ снизу и

$$A\left(\tau^{(n)} = \frac{k}{n}\right) = A\left(\tau \in \left(\frac{k}{n}, \frac{k+1}{n}\right)\right) \in \mathcal{F}_{\frac{k}{n}}; \quad (8.7)$$

$$w(t + \tau^{(n)}) - w(\tau^{(n)}) \xrightarrow{\text{П. Н.}} w(t + \tau) - w(\tau), \quad (8.8)$$

где $A \in \mathcal{F}_\tau$, то, используя свойства сходимостей последовательностей случайных величин и соотношений (8.6), (8.7), (8.8), имеем

$$\begin{aligned} P\{A \cap (w(t + \tau) - w(\tau) \in B)\} &= \lim_{n \rightarrow \infty} P\{A \cap (w(t + \tau^{(n)}) - w(\tau^{(n)}) \in B)\} = \\ &= P\{A\}P\{w(t) \in B\}. \end{aligned}$$

□

В дальнейшем приведенные свойства выполняются с вероятностью единица.

С4. Траектории ВП $w(t)$ за конечное время достигнут любого уровня.

Доказательство. Пусть $\tau_a = \min(t : w(t) = a)$.

В силу симметрии, однородности и строгой марковости ВП $w(t)$ имеем $P\{w(t) > a | \tau_a < t\} = 1/2$. Отсюда

$$\frac{1}{2} = \frac{P\{w(t) > a, \tau_a < t\}}{P\{\tau_a < t\}} = \frac{P\{w(t) > a\}}{P\{\tau_a < t\}}.$$

Следовательно,

$$P\{\tau_a < t\} = 2P\{w(t) > a\} = \sqrt{\frac{2}{\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}} dx = \sqrt{\frac{2}{\pi}} \int_{\frac{a}{\sqrt{t}}}^\infty e^{-\frac{y^2}{2}} dy.$$

По аксиоме непрерывности

$$P\{\tau_a < \infty\} = \lim_{t \rightarrow \infty} P\{\tau_a < t\} = \lim_{t \rightarrow \infty} \sqrt{\frac{2}{\pi}} \int_{\frac{a}{\sqrt{t}}}^\infty e^{-\frac{y^2}{2}} dy = 1.$$

□

Найдем плотность распределения вероятностей СВ τ_a :

$$P_{\tau_a}(t) = \frac{dP\{\tau_a < t\}}{dt} = \frac{at^{-\frac{3}{2}}}{\sqrt{2\pi}} e^{-\frac{a^2}{2t}}. \quad (8.9)$$

С5. Траектории ВП $w(t)$, исходя из любой точки, за сколь угодно малое время примут значения, лежащие как выше, так и ниже этой точки.

Доказательство. В силу однородности ВП это свойство достаточно доказать для точки $t = 0$.

Пусть $\eta(t) = \max_{0 \leq s \leq t} w(s)$. Поскольку

$$P\{\eta(t) > a\} = P\{\tau_a < t\} = \sqrt{\frac{2}{\pi}} \int_{\frac{a}{\sqrt{t}}}^\infty e^{-\frac{y^2}{2}} dy,$$

то по аксиоме непрерывности получаем

$$P\{\eta(t) > 0\} = \lim_{a \rightarrow 0} P\{\eta(t) > a\} = 1.$$

В силу симметрии ВП $w(t)$ $P\{\eta(t) < 0\} = 1$.

Это равенство верно для любого $t > 0$.

□

Найдем плотность распределения вероятностей СВ $\eta(t)$:

$$P_{\eta(t)}(x) = -\frac{dP\{\eta(t) > x\}}{dx} = \sqrt{\frac{2}{\pi t}} e^{-\frac{x^2}{2t}}. \quad (8.10)$$

С6. Точка максимума траектории ВП $w(t)$ на любом промежутке подчиняется закону арксинуса, т. е. с большой вероятностью примет значение или в начале промежутка, или в конце промежутка.

Доказательство. Не ограничивая общности, будем рассматривать промежуток $[0, t]$. Пусть $\tau = \tau(t) = \arg \max_{0 \leq u \leq t} w(u)$, $s < t$. Используя свойства условной плотности вероятности и (8.10), имеем

$$\begin{aligned} P_{\eta(t)}(x | \tau_a = s) &= P_{\max_{s \leq u \leq t} w(u) + a}(x | \tau_a = s) = P_{\max_{s \leq u \leq t} w(u) + a}(x - a | \tau_a = s) = \\ &= P_{\max_{0 \leq u \leq t-s} w(u)}(x - a) = P_{\eta(t-s)}(x - a) = \sqrt{\frac{2}{\pi(t-s)}} e^{-\frac{(x-a)^2}{2(t-s)}}. \end{aligned}$$

Отсюда и из равенства (8.10) получаем

$$P_{\eta(t), \tau_a}(x, s) = P_{\tau_a}(s) P_{\eta(t)}(x | \tau_a = s) = \frac{as^{-\frac{3}{2}}}{\pi\sqrt{t-s}} e^{-\frac{a^2}{2s} - \frac{(x-a)^2}{2(t-s)}},$$

тогда

$$P_{\eta(t), \tau}(x, s) = P_{\eta(t)}(x) P_{\tau}(s | \eta(t) = x) = P_{\eta(t), \tau_x}(x, s) = \frac{xs^{-\frac{3}{2}}}{\pi\sqrt{t-s}} e^{-\frac{x^2}{2s}}.$$

Следовательно,

$$P_{\tau}(s) = \int_0^\infty P_{\eta(t), \tau}(x, s) dx = \frac{s^{-\frac{3}{2}}}{\pi\sqrt{t-s}} \int_0^\infty x e^{-\frac{x^2}{2s}} dx = \frac{1}{\pi\sqrt{s(t-s)}}.$$

$$\text{Таким образом } \mathcal{F}_{\tau}(s) = \int_{-\infty}^s P_{\tau}(x) dx = \frac{2}{\pi} \arcsin \sqrt{s/t}, \text{ если } 0 \leq s \leq t. \quad \square$$

С7. Траектории ВП $w(t)$ ни в одной точке не дифференцируемы.

Доказательство. В силу однородности ВП $w(t)$ достаточно доказать для точки $t = 0$. Предположим, что существует предел $\lim_{t \rightarrow 0} w(t)/t = w'(0)$. Тогда будет существовать предел

$$\lim_{n \rightarrow \infty} \frac{w(2^{-n})}{2^{-n}} = w'(0).$$

В этом случае

$$\lim_{n \rightarrow \infty} \frac{w(2^{-n+1}) - w(2^{-n})}{2^{-n}} = \lim_{n \rightarrow \infty} 2 \frac{w(2^{-n+1})}{2^{-n+1}} - \lim_{n \rightarrow \infty} \frac{w(2^{-n})}{2^{-n}} = w'(0).$$

Рассмотрим события

$$A_n = (w(2^{-n+1}) - w(2^{-n}) > \sqrt{2^{-n}}).$$

Поскольку интервалы $(2^{-n}, 2^{-n+1}]$, $n > 0$, не пересекаются, то события A_n будут независимы. Найдем вероятности этих событий:

$$P\{A_n\} = P\{w(2^{-n}) > \sqrt{2^{-n}}\} = \frac{1}{\sqrt{2\pi}2^{-n}} \int_{\sqrt{2^{-n}}}^\infty e^{-\frac{x^2}{2 \cdot 2^{-n}}} dx = \frac{1}{\sqrt{2\pi}} \int_1^\infty e^{-\frac{y^2}{2}} dy = C.$$

Отсюда $\sum_{n=1}^\infty P\{A_n\} = \infty$. Тогда по закону «0 или 1» $P\{\overline{\lim} A_n\} = 1$.

Следовательно, для всех событий $\omega \in \overline{\lim} A_n$ неравенство

$$\frac{w(2^{-n+1}, \omega) - w(2^{-n}, \omega)}{\sqrt{2^{-n}}} > 1$$

будет выполняться для бесконечно большого числа значений n . Поэтому для этих ω последовательность

$$\frac{w(2^{-n+1}, \omega) - w(2^{-n}, \omega)}{2^{-n}} = \frac{w(2^{-n+1}, \omega) - w(2^{-n}, \omega)}{\sqrt{2^{-n}}} \cdot \frac{\sqrt{2^{-n}}}{2^{-n}}$$

не будет иметь предела. Получили противоречие. \square

С8. Траектории ВП $w(t)$ на любом конечном промежутке имеют неограниченную вариацию.

Доказательство. Рассмотрим промежуток $[s, t]$ и разбиение этого промежутка точками $s = t_1 < t_2 < \dots < t_n = t$. Пусть $\lambda_n = \max_k (t_k - t_{k-1})$, $V_n = \sum_{k=1}^n |w(t_k) - w(t_{k-1})|$. Покажем, что $V_n \xrightarrow{\text{П. Н.}} \infty$, если λ_n достаточно быстро стремится к нулю. Вычислим

$$\begin{aligned} E\{|w(t)|\} &= \frac{1}{\sqrt{2\pi t}} \int |x| e^{-\frac{x^2}{2t}} dx = \left[\frac{x}{\sqrt{t}} = y \right] = \\ &= \sqrt{t} \frac{1}{\sqrt{2\pi}} \int |y| e^{-\frac{y^2}{2}} dy = \sqrt{\frac{2}{\pi}} \sqrt{t} = C_1 \sqrt{t}. \end{aligned}$$

Тогда

$$\begin{aligned} E\{V_n\} &= \sum_{k=1}^n E\{|w(t_k) - w(t_{k-1})|\} = \sum_{k=1}^n E\{|w(t_k - t_{k-1})|\} = \\ &= C_1 \sum_{k=1}^n \sqrt{t_k - t_{k-1}} = C_1 \sum_{k=1}^n \frac{t_k - t_{k-1}}{\sqrt{t_k - t_{k-1}}} \geq \frac{C_1}{\sqrt{\lambda_n}} (t - s) \xrightarrow{\lambda_n \rightarrow 0} \infty; \\ D\{V_n\} &= \sum_{k=1}^n D\{|w(t_k) - w(t_{k-1})|\} = \sum_{k=1}^n D\{|w(t_k - t_{k-1})|\} = \\ &= \sum_{k=1}^n E\{|w(t_k - t_{k-1})|^2\} - \sum_{k=1}^n E^2\{|w(t_k - t_{k-1})|\} = \\ &= \sum_{k=1}^n (t_k - t_{k-1}) - \sum_{k=1}^n C_1^2 (t_k - t_{k-1}) = C_2 (t - s). \end{aligned}$$

По неравенству Чебышева имеем для достаточно больших n

$$\begin{aligned} P\{V_n \leq N\} &= P\{V_n - E\{V_n\} \geq N - E\{V_n\}\} \leq P\{|V_n - E\{V_n\}| \geq E\{V_n\} - N\} \leq \\ &\leq \frac{D\{V_n\}}{(E\{V_n\} - N)^2} \leq \frac{C_2 (t - s)}{\left(\frac{C_1 (t - s)}{\sqrt{\lambda_n}} - N\right)^2} \xrightarrow{\lambda_n \rightarrow 0} 0. \end{aligned}$$

Отсюда следует $V_n \xrightarrow{P} \infty$. Заметим, что последовательность λ_n можно выбрать такой, что

$$\sum_n P\{V_n \leq N\} < \infty.$$

Следовательно,

$$P\{\overline{\lim}(V_n \leq N)\} = 0.$$

Или $P\{\underline{\lim}(V_n > N)\} = 1$. Отсюда следует, что $V_n \xrightarrow{\text{П. Н.}} \infty$. \square

С9. ВП $w(t)$ подчиняется закону повторного логарифма, т. е.

$$P\left\{\overline{\lim}_{t \rightarrow \infty} \left(\frac{w(t)}{\sqrt{2t \ln \ln t}} \right) = 1\right\} = 1.$$

Доказательство. Из определения верхнего предела следует, что если a является верхним пределом числовой последовательности a_n , то для любого $\epsilon > 0$ должны выполняться соотношения:

1. Существует n_0 такое, что при всех $n \geq n_0$ $a_n \leq a + \epsilon$;
2. $a_n \geq a - \epsilon$ для бесконечно большого числа значений n .

Отсюда следует, что для доказательства закона повторного логарифма нужно доказать выполнимость следующих соотношений с вероятностью единица:

1. Для любого $\epsilon > 0$ существует $t_0 = t_0(\omega, \epsilon)$ такое, что при всех $t \geq t_0$

$$|w(t)| \leq (1 + \epsilon) \sqrt{2t \ln \ln t}. \quad (8.11)$$

2. Для любого $\epsilon > 0$ соотношение

$$|w(t)| \geq (1 - \epsilon) \sqrt{2t \ln \ln t} \quad (8.12)$$

выполняется для бесконечного числа значений t .

Докажем вначале выполнимость соотношения

$$w(t) \leq (1 + \epsilon) \sqrt{2t \ln \ln t} \quad (8.13)$$

для достаточно больших t .

Пусть $a > 1$. Рассмотрим события

$$A_k = \left(\max_{a^{k-1} \leq u \leq a^k} w(u) > (1 + \epsilon) \sqrt{2a^{k-1} \ln \ln a^{k-1}} \right).$$

Поскольку

$$A_k \subset B_k = \left(\max_{u \leq a^k} w(u) > (1 + \epsilon) \sqrt{2a^{k-1} \ln \ln a^{k-1}} \right),$$

то, используя лемму 8.1 и распределение случайной величины $\eta(t)$, получим

$$\begin{aligned} P\{B_k\} &= P\left\{\eta(a_k) > (1 + \epsilon) \sqrt{2a^{k-1} \ln \ln a^{k-1}}\right\} = \sqrt{\frac{2}{\pi}} \int_{(1+\epsilon)\sqrt{\frac{2}{a} \ln \ln a^{k-1}}}^{\infty} e^{-\frac{y^2}{2}} dy = \\ &= O\left(\frac{1}{\sqrt{\ln \ln a^{k-1}}} \exp\left\{-(1 + \epsilon)^2 \frac{1}{a} \ln \ln a^{k-1}\right\}\right) = O\left(\frac{1}{\sqrt{\ln k}} (k \ln a)^{-\frac{(1+\epsilon)^2}{a}}\right). \end{aligned}$$

Положив $a = 1 + \epsilon$, получим $\sum_{n=1}^{\infty} P\{B_n\} < \infty$. Следовательно, $\sum_n P\{A_n\} < \infty$.

По лемме Бореля – Кантелли $P\{\overline{\lim} A_n\} = 0$ или $P\{\underline{\lim} \bar{A}_n\} = 1$. Из определения нижнего предела последовательности событий получаем, что для всех $\omega \in \underline{\lim} \bar{A}_n$ существует номер $k_0 = k_0(\omega)$ такой, что при всех $k \geq k_0$ будут выполняться соотношения

$$\max_{a^{k-1} \leq u \leq a^k} w(u, \omega) \leq (1 + \epsilon) \sqrt{2a^{k-1} \ln \ln a^{k-1}}.$$

Отсюда следует выполнимость соотношения (8.12). Поскольку процесс $w(t)$ симметричный, то из доказанного следует, что процесс $w(t)$ будет удовлетворять соотношению

$$w(t) \geqslant -(1 + \epsilon) \sqrt{2t \ln \ln t}$$

для всех достаточно больших t , т. е. будет выполняться соотношение (8.11).

Аналогично, для доказательства соотношения (8.12) достаточно доказать неравенство

$$w(t) \geqslant (1 - \epsilon) \sqrt{2t \ln \ln t} \quad (8.14)$$

для бесконечного числа значений t .

Пусть $b > 1$. Рассмотрим события

$$C_k = \left(w(b^k) - w(b^{k-1}) > \left(1 - \frac{\epsilon}{2}\right) \sqrt{2b^k \ln \ln b^k} \right).$$

Поскольку интервалы $[b^{k-1}, b^k]$, $k \geqslant 1$, не пересекаются, то события C_k будут независимы. Используя лемму 8.1 и распределение случайной величины $w(t)$, получим

$$P\{C_k\} = P\left\{w(b^{k-1}(b-1)) > \left(1 - \frac{\epsilon}{2}\right) \sqrt{2b^k \ln \ln b^k}\right\} = O\left(\frac{1}{\sqrt{\ln k}} k^{-\frac{(1-\frac{\epsilon}{2})^2}{1-\frac{1}{b}}}\right).$$

Если выбрать $b \geqslant \frac{2}{\epsilon}$, то

$$\sum_{k=1}^{\infty} P\{C_k\} = \infty.$$

Следовательно, по закону «0 или 1» $P\{\overline{\lim} C_k\} = 1$. Это означает, что для любого $\omega \in \overline{\lim} C_k$ существует последовательность $\{k_n\}$, $k_n = k_n(\omega)$ такая, что

$$w(b^{k_n}, \omega) - w(b^{k_n-1}, \omega) > \left(1 - \frac{\epsilon}{2}\right) \sqrt{2b_n^k \ln \ln b_n^k}.$$

С другой стороны, по доказанному имеем: для любого $\delta > 0$ и для почти всех w существует $t_0 = t_0(\omega, \delta)$, что для всех $t \geqslant t_0$ $w(t) > -(1 + \delta) \sqrt{2t \ln \ln t}$.

Тогда неравенство

$$w(b^k) = w(b^k) - w(b^{k-1}) + w(b^{k-1}) > \left(1 - \frac{\epsilon}{2}\right) \sqrt{2b^k \ln \ln b^k} - (1 + \delta) \sqrt{2t \ln \ln t}$$

будет выполняться для бесконечного числа значений k . Выбором числа δ и достаточно большим b правую часть этого неравенства можно сделать меньшей, чем

$$(1 - \epsilon) \sqrt{2b^k \ln \ln b^k}.$$

□

C10. Если $w(t)$ – стандартный винеровский процесс, то и $W(t) = tw\left(\frac{1}{t}\right)$ – стандартный винеровский процесс.

Доказательство. Докажем, что процесс $W(t)$ – гауссовский. Пусть $0 \leq t_1 < t_2 < \dots < t_n$. Тогда случайная величина

$$\begin{aligned} \lambda_1 W(t_1) + \lambda_2 W(t_2) + \dots + \lambda_n W(t_n) &= \lambda_1 t_1 w\left(\frac{1}{t_1}\right) + \lambda_2 t_2 w\left(\frac{1}{t_2}\right) + \dots + \\ &+ \lambda_n t_n w\left(\frac{1}{t_n}\right) = \lambda_1 t_1 \left(w\left(\frac{1}{t_1}\right) - w\left(\frac{1}{t_2}\right) \right) + (\lambda_1 t_1 + \lambda_2 t_2) \left(w\left(\frac{1}{t_2}\right) - w\left(\frac{1}{t_3}\right) \right) + \\ &+ \dots + (\lambda_1 t_1 + \lambda_2 t_2 + \dots + \lambda_n t_n) \left(w\left(\frac{1}{t_n}\right) \right) \end{aligned}$$

имеет нормальный закон распределения. Найдя ее характеристическую функцию, получим выражение вида (6.1), т. е. процесс $W(t)$ является гауссовским процессом. Найдём его характеристики:

$$\mathbf{E}\{W(t)\} = \mathbf{E}\left\{tw\left(\frac{1}{t}\right)\right\} = t\mathbf{E}\left\{w\left(\frac{1}{t}\right)\right\} = 0.$$

$$\begin{aligned} r(t, u) &= \mathbf{E}\{W(t)W(u)\} = \mathbf{E}\left\{tw\left(\frac{1}{t}\right)uw\left(\frac{1}{u}\right)\right\} = \\ &= tu\mathbf{E}\left\{w\left(\frac{1}{t}\right)w\left(\frac{1}{u}\right)\right\} = tu \min\left(\frac{1}{t}, \frac{1}{u}\right) = \min(t, u). \end{aligned}$$

□

C11. ВП $w(t)$ подчиняется локальному закону повторного логарифма, т. е.

$$\mathbf{P}\left\{\overline{\lim}_{t \rightarrow 0} \frac{|w(t)|}{\sqrt{2t \ln \ln \frac{1}{t}}} = 1\right\} = 1.$$

Доказательство. Применим закон повторного логарифма к процессу $W(t)$. Имеем

$$\begin{aligned} 1 &= \mathbf{P}\left\{\overline{\lim}_{t \rightarrow \infty} \frac{|W(t)|}{\sqrt{2t \ln \ln t}} = 1\right\} = \\ &= \mathbf{P}\left\{\overline{\lim}_{t \rightarrow \infty} \frac{t|w(\frac{1}{t})|}{\sqrt{2t \ln \ln t}} = 1\right\} = \mathbf{P}\left\{\overline{\lim}_{u \rightarrow 0} \frac{|w(u)|}{\sqrt{2u \ln \ln \frac{1}{u}}} = 1\right\}. \end{aligned}$$

□

8.3. ПУАССОНОВСКИЙ ПРОЦЕСС

Определение 8.5. Случайный процесс $\xi(t)$, $t > 0$, называется пуассоновским процессом, если он удовлетворяет следующим условиям:

1. $\xi(0) = 0$.
2. $\xi(t)$ – однородный процесс с независимыми приращениями.
3. $\xi(t)$ – стохастически непрерывный процесс.
4. Случайная величина $\xi(1)$ распределена по закону Пуассона с параметром λ .

Из теоремы 8.1 следует, что

$$f_{\xi(t)}(u) = \left(f_{\xi(1)}(u)\right)^t = e^{\lambda t(e^{iu} - 1)},$$

т. е. случайная величина $\xi(t)$ имеет распределение Пуассона с параметром λt .

Замечание 8.5. Из определения пуассоновского процесса следует, что он является неубывающим процессом, траектории которого – кусочно-постоянные.

Покажем, что скачки данного процесса равны единице, т. е. он удовлетворяет условию ординарности:

$$\xi(t) \in N_0 \quad \text{и} \quad P\{\xi(t+h) - \xi(t) > 1\} = o(h). \quad (8.15)$$

Из определения процесса $\xi(t)$ имеем

$$\begin{aligned} P\{\xi(t+h) - \xi(t) > 1\} &= P\{\xi(h) > 1\} = 1 - P\{\xi(h) = 0\} - P\{\xi(h) = 1\} = \\ &= 1 - e^{-\lambda h} - \lambda h e^{-\lambda h} = o(h), \end{aligned}$$

т. е. выполняется соотношение (8.15).

Замечание 8.6. В определении пуассоновского процесса условие 4 можно заменить на условие ординарности.

Доказательство. Обозначим $P_k(t) = P\{\xi(t) = k\}$. Для нахождения $P_k(t)$ применим метод дифференциальных уравнений. Из определения стохастической непрерывности СП $\xi(t)$ имеем при $t > t_0$

$$P\{\xi(t) - \xi(t_0) = 0\} = P\{\xi(t) - \xi(t_0) \leq \epsilon\} \xrightarrow{t \rightarrow t_0} 1, \quad 0 < \epsilon < 1.$$

Отсюда

$$\begin{aligned} 0 \leq P_0(t_0) - P_0(t) &= P\{\xi(t_0) = 0\} - P\{\xi(t) = 0\} = P\{\xi(t_0) = 0, \xi(t) = 0\} + \\ &+ P\{\xi(t_0) = 0, \xi(t) \neq 0\} - P\{\xi(t) = 0, \xi(t_0) = 0\} = \\ &= P\{\xi(t_0) = 0, \xi(t) - \xi(t_0) \neq 0\} \leq P\{\xi(t) - \xi(t_0) \neq 0\} \xrightarrow{t \rightarrow t_0} 0, \end{aligned}$$

т. е. $P_0(t)$ – непрерывная функция. Используя свойства условия 1 и 2, имеем

$$\begin{aligned} P_0(t+u) &= P\{\xi(t+u) = 0\} = P\{\xi(t+u) = 0, \xi(t) = 0\} = \\ &= P\{\xi(t+u) - \xi(t) = 0, \xi(t) - \xi(0) = 0\} = \\ &= P\{\xi(t+u) - \xi(t) = 0\} P\{\xi(t) - \xi(0) = 0\} = P_0(u) P_0(t). \end{aligned}$$

Используя доказательство теоремы 8.1, имеем $P_0(t) = e^{-\lambda t}$, где λ – некоторое число больше 0. Случаи $P_0(t) = 0$ и $P_0(t) = 1$ исключаем, так как они не представляют практического интереса. Используя формулу полной вероятности и условие ординарности, получаем

$$\begin{aligned} P_k(t + \Delta t) &= P\{\xi(t + \Delta t) = k\} = \sum_{l=0}^k P\{\xi(t) = l\} P\{\xi(t + \Delta t) = k | \xi(t) = l\} = \\ &= \sum_{l=0}^k P_l(t) P\{\xi(t + \Delta t) - \xi(t) = k - l | \xi(t) - \xi(0) = l\} = \\ &= \sum_{l=0}^k P_l(t) P\{\xi(t + \Delta t) - \xi(t) = k - l\} = P_k(t) P\{\xi(t + \Delta t) - \xi(t) = 0\} + \\ &+ P_{k-1}(t) P\{\xi(t + \Delta t) - \xi(t) = 1\} + \sum_{l=0}^{k-2} P_l(t) P\{\xi(t + \Delta t) - \xi(t) = k - l\} = \\ &= P_k(t) P(\Delta t) + P_{k-1}(t) P_1(\Delta t) + o(\Delta t). \end{aligned} \quad (8.16)$$

Поскольку $\sum_{k=0}^{\infty} P_k(\Delta t) = 1$, $\sum_{k=2}^{\infty} P_k(\Delta t) = o(\Delta t)$, то

$$P_1(\Delta t) = 1 - P_0(\Delta t) + o(\Delta t) = 1 - e^{-\lambda \Delta t} + o(\Delta t) = \lambda \Delta t + o(\Delta t).$$

Подставляя эти соотношения в (8.16), получаем

$$P_k(t + \Delta t) = P_k(t)(1 - \lambda \Delta t) + P_{k-1}(t)\lambda \Delta t + o(\Delta t).$$

Отсюда

$$P'_k(t) = -\lambda \Delta t P_k(t) + \lambda P_{k-1}(t), \quad k \geq 1.$$

Сделаем замену $Q_k(t) = e^{\lambda t} P_k(t)$. Тогда

$$Q_k(t) = \lambda e^{\lambda t} P_k(t) + e^{\lambda t} P'_k(t) = \lambda Q_k(t) - \lambda Q_k(t) + \lambda Q_{k-1}(t) = \lambda Q_{k-1}(t).$$

Поскольку $Q_k(0) = P_k(0) = 1$ при $k = 0$ и равно 0, если $k \neq 0$, то $Q_k(t) = \frac{(\lambda t)^k}{k!}$, т. е.

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

□

Найдем вид конечномерных распределений процесса $\xi(t)$. Пусть $0 \leq t_1 < t_2 < \dots < t_n$, $0 = k_1 < k_2 < \dots < k_n \in N_0$. Тогда

$$\begin{aligned} & P\{\xi(t_1) = k_1, \xi(t_2) = k_2, \dots, \xi(t_n) = k_n\} = \\ & = P\{\xi(t_1) - \xi(0) = k_1, \xi(t_2) - \xi(t_1) = k_2 - k_1, \dots, \xi(t_n) - \xi(t_{n-1}) = k_n - k_{n-1}\} = \\ & = P\{\xi(t_1) - \xi(0) = k_1\} P\{\xi(t_2) - \xi(t_1) = k_2 - k_1\} \times \dots \times \\ & \times P\{\xi(t_n) - \xi(t_{n-1}) = k_n - k_{n-1}\} = P\{\xi(t_1) = k_1\} P\{\xi(t_2 - t_1) = k_2 - k_1\} \times \dots \times \\ & \times P\{\xi(t_n - t_{n-1}) = k_n - k_{n-1}\} = \frac{(\lambda t_1)}{k_1!} e^{-\lambda t_1} \frac{\lambda(t_2 - t_1)}{(k_2 - k_1)!} e^{-\lambda(t_2 - t_1)} \times \dots \times \\ & \times \frac{\lambda(t_n - t_{n-1})}{(k_n - k_{n-1})!} e^{-\lambda(t_n - t_{n-1})}. \end{aligned}$$

8.4. СТОХАСТИЧЕСКИЙ ИНТЕГРАЛ ИТО

Пусть $(\mathcal{F}_t)_{t \geq 0}$ — неубывающая последовательность σ -алгебр. Винеровский процесс $w(t)$ измерим относительно \mathcal{F}_t , и $w(t + \tau) - w(t)$ не зависит от событий из \mathcal{F}_t . В частности, это будет верно, если $\mathcal{F}_t = \mathcal{F}_t^w = \sigma(w(s), s \leq t)$.

Заметим, что в общем случае $\mathcal{F}_t^w \subset \mathcal{F}_t$.

Определение 8.6. Функция $\phi(t) = \phi(t, \omega)$ называется неупреждающей, если она при каждом t измерима относительно \mathcal{F}_t .

Определение 8.7. Функция $\phi(t)$ называется простой, если существует такое разбиение промежутка $[0, \infty)$ точками $0 \leq t_0 < t_1 < \dots$ на промежутки $[t_k, t_{k+1})$, что при $t \in [t_k, t_{k+1})$ $\phi(t) = \phi(t_k)$.

Пусть $\phi(t)$ – неупреждающая простая функция, $-\infty < a < b < +\infty$, $t_0 \leq a < t_1 < \dots < t_m < b \leq t_{m+1}$. Тогда определим

$$\begin{aligned} \int_a^b \phi(t) dw(t) &= \phi(t_0)(w(t_1) - w(a)) + \\ &+ \sum_{k=1}^{m-1} \phi(t_k)(w(t_{k+1}) - w(t_k)) + \phi(t_m)(w(b) - w(t_m)). \end{aligned}$$

Для удобства записи положим $a = t_0, b = t_{m+1}$. Тогда

$$\int_a^b \phi(t) dw(t) = \sum_{k=0}^m \phi(t_k)(w(t_{k+1}) - w(t_k)). \quad (8.17)$$

Этот интеграл обладает следующими свойствами.

Пусть функции ϕ и ψ – простые и неупреждающие. Можно считать, что они являются постоянными на одних и тех же промежутках. В противном случае можно рассматривать новые промежутки постоянства функций, полученные путем пересечения промежутков постоянства функций ϕ и ψ .

C1. Для любых констант α и β

$$\int_a^b (\alpha\phi(t) + \beta\psi(t)) dw(t) = \alpha \int_a^b \phi(t) dw(t) + \beta \int_a^b \psi(t) dw(t).$$

C2. Если $a < c < b$, то

$$\int_a^b \phi(t) dw(t) = \int_a^c \phi(t) dw(t) + \int_c^b \phi(t) dw(t).$$

Доказательство этих свойств следует из (8.17).

$$C3. \quad E \left\{ \int_a^b \phi(t) dw(t) \int_a^b \psi(t) dw(t) \right\} = \int_a^b E \{ \phi(t) \psi(t) \} dt.$$

Доказательство. Используя представление (8.17), получаем

$$\begin{aligned} E \left\{ \int_a^b \phi(t) dw(t) \int_a^b \psi(t) dw(t) \right\} &= \\ &= \sum_{k,l=0}^m E \{ \phi(t_k) \psi(t_l) (w(t_{k+1}) - w(t_k)) (w(t_{l+1}) - w(t_l)) \}. \end{aligned} \quad (8.18)$$

Пусть $k < l$. Тогда из свойств винеровского процесса $w(t)$ получаем, что случайные величины $\phi(t_k) \psi(t_k) (w(t_{k+1}) - w(t_k))$ и $w(t_{l+1}) - w(t_l)$ будут независимы. А так как $E \{ w(t_{l+1}) - w(t_l) \} = 0$, то в (8.18) слагаемые с $k \neq l$ будут равны 0.

Следовательно,

$$\begin{aligned} \mathbf{E} \left\{ \int_a^b \Phi(t) dw(t) \int_a^b \Psi(t) dw(t) \right\} &= \sum_{k=0}^m \mathbf{E} \{ \Phi(t_k) \Psi(t_k) (w(t_{k+1}) - w(t_k))^2 \} = \\ &= \sum_{k=0}^m \mathbf{E} \{ \Phi(t_k) \Psi(t_k) \} \mathbf{E} \{ (w(t_{k+1}) - w(t_k))^2 \} = \sum_{k=0}^m \mathbf{E} \{ \Phi(t_k) \Psi(t_k) \} (t_{k+1} - t_k) = \\ &= \int_a^b \mathbf{E} \{ \Phi(t) \Psi(t) \} dt. \end{aligned}$$

□

С4. Для любых промежутков $[a, b]$ и $[c, d]$

$$\mathbf{E} \left\{ \int_a^b \Phi(t) dw(t) \int_c^d \Psi(t) dw(t) \right\} = \int_{\max(a,c)}^{\min(b,d)} \mathbf{E} \{ \Phi(t) \Psi(t) \} dt. \quad (8.19)$$

Доказательство следует из того, что если интервалы (a, b) и (c, d) не пересекаются, то (см. доказательство С3) левая часть соотношения (8.19) будет равна 0.

Доопределим $\int_a^b \Phi(t) dw(t)$ для неупреждающих функций $\Phi(t)$, удовлетворяющих условию

$$\int_a^b \mathbf{E} \{ \Phi^2(t) \} dt < \infty. \quad (8.20)$$

Построим последовательность простых неупреждающих функций $\Phi_n(t)$, удовлетворяющих условию

$$\int_a^b \mathbf{E} \{ (\Phi(t) - \Phi_n(t))^2 \} dt \xrightarrow{n \rightarrow \infty} 0. \quad (8.21)$$

Положим

$$\Phi_n(t) = \sum_{k=-n2^n}^{n2^n} \Phi\left(\frac{k}{2^n}\right) \mathbb{1}_{\frac{k}{2^n}, \frac{k+1}{2^n}}(t).$$

Поскольку из условия (8.20) следует

$$\int_a^b \mathbf{E} \{ \Phi^2(t) \mathbb{1}_{(|\Phi(t)| > n)} \} dt \xrightarrow{n \rightarrow \infty} 0$$

и по теореме Лебега для каждого $N > 0$

$$\int_a^b \mathbf{E} \{ (\Phi(t) - \Phi_n(t))^2 \mathbb{1}_{(|\Phi(t)| \leq N)} \} dt \xrightarrow{n \rightarrow \infty} 0,$$

то

$$\int_a^b \mathbf{E} \{ (\Phi(t) - \Phi_n(t))^2 \} dt \xrightarrow{n \rightarrow \infty} 0.$$

Определим $\int_a^b \Phi(t) dw(t)$ от неупреждающих функций, удовлетворяющих (8.20) следующим образом:

$$\int_a^b \Phi(t) dw(t) = \lim_{n \rightarrow \infty} \int_a^b \Phi_n(t) dw(t), \quad (8.22)$$

где $\Phi_n(t)$ удовлетворяют (8.21).

Покажем, что данное определение является корректным, т. е. предел в (8.22) существует и не зависит от последовательности $\{\Phi_n(t)\}$, удовлетворяющей (8.21).

Проверим критерий Коши сходимости в среднем квадратичном. Используя (8.17), имеем

$$\begin{aligned} \mathbf{E} \left\{ \left(\int_a^b \Phi_n(t) dw(t) - \int_a^b \Phi_m(t) dw(t) \right)^2 \right\} &= \mathbf{E} \left\{ \left(\int_a^b (\Phi_n(t) - \Phi_m(t)) dw(t) \right)^2 \right\} = \\ &= \int_a^b \mathbf{E} \{ (\Phi_n(t) - \Phi_m(t))^2 \} dt \leq \\ &\leq 2 \int_a^b \mathbf{E} \{ \Phi_n(t) - \Phi(t) \}^2 dt + 2 \int_a^b \mathbf{E} \{ \Phi_n(t) - \Phi(t) \} dt \xrightarrow{n, m \rightarrow \infty} 0. \end{aligned}$$

Пусть последовательность $\{\Psi_n(t)\}$ неупреждающих функций удовлетворяет (8.21).

Тогда покажем, что $\int_a^b \Phi(t) dw(t) = \lim_{n \rightarrow \infty} \int_a^b \Psi_n(t) dw(t)$.

По определению сходимости в среднем квадратичном имеем

$$\begin{aligned} &\mathbf{E} \left\{ \left(\int_a^b \Psi_m(t) dw(t) - \int_a^b \Phi(t) dw(t) \right)^2 \right\} \leq \\ &\leq 2 \mathbf{E} \left\{ \left(\int_a^b (\Psi_m(t) - \Phi_n(t)) dw(t) \right)^2 \right\} + 2 \mathbf{E} \left\{ \left(\int_a^b (\Phi_n(t) - \Phi(t)) dw(t) \right)^2 \right\}. \quad (8.23) \end{aligned}$$

По определению второе слагаемое в (8.23) стремится к нулю при $n \rightarrow \infty$. Покажем, что

и первое слагаемое в (8.23) имеет пределом 0:

$$\begin{aligned} E \left\{ \left(\int_a^b (\psi_m(t) - \phi_n(t)) dw(t) \right)^2 \right\} &= \int_a^b E \{ (\psi_m(t) - \phi_n(t))^2 \} dt \leq \\ &\leq 2 \int_a^b E \{ (\psi_m(t) - \phi(t))^2 \} dt + 2 \int_a^b E \{ (\phi_n(t) - \phi(t))^2 \} dt \xrightarrow{n, m \rightarrow \infty} 0. \end{aligned}$$

Замечание 8.7. Из свойств сходимости в среднем квадратичном следует, что $\int_a^b \phi(t)dw(t)$ обладает свойствами C1 – C4 как для простых неупреждающих функций.

Построенный интеграл называется стохастическим интегралом Ито. Заметим, что таким образом определенный интеграл нельзя определять так же, как и ранее определенные стохастические интегралы, так как предел интегральных сумм будет зависеть от способа построения интегральных сумм. Покажем это на примере.

Вычислим стохастический интеграл Ито $\int_0^T w(t) dw(t)$.

Пусть $0 \leq t_0 < t_1 < \dots < t_n = T$, $\lambda_n = \max_k (t_{k+1} - t_k)$.

$$Q_1 = \sum_{k=0}^{n-1} w(t_k) (w(t_{k+1}) - w(t_k)).$$

Найдем $l.i.m. Q_1$ при $\lambda_n \rightarrow 0$:

$$Q_1 = \frac{1}{2} \sum_{k=1}^n (w^2(t_{k+1}) - w^2(t_k) - (w(t_{k+1}) - w(t_k))^2) = \frac{1}{2} w^2(T) - \frac{1}{2} \sum_{k=1}^n (w(t_{k+1}) - w(t_k))^2.$$

Обозначим $\eta_n = \sum_{k=1}^n (w(t_{k+1}) - w(t_k))^2$:

$$E\{\eta_n\} = \sum_{k=1}^n E\{(w(t_{k+1}) - w(t_k))^2\} = \sum_{k=1}^n E\{(w(t_{k+1} - t_k))^2\} = \sum_{k=1}^n (t_{k+1} - t_k) = T.$$

$$\begin{aligned} D\{\eta_n\} &= D\left\{ \sum_{k=1}^n (w(t_{k+1}) - w(t_k))^2 \right\} = \sum_{k=1}^n D\{(w(t_{k+1} - t_k))^2\} = \\ &= 2 \sum_{k=1}^n (t_{k+1} - t_k)^2 \leq 2\lambda_n \sum_{k=1}^n (t_{k+1} - t_k) = 2\lambda_n T \xrightarrow{\lambda_n \rightarrow 0} 0. \end{aligned}$$

Итак, $E\{\eta_n\} = T$, $D\{\eta_n\} \xrightarrow{\lambda_n \rightarrow 0} 0$. Покажем $\eta_n \xrightarrow{\lambda_n \rightarrow 0} T$. $E\{(\eta_n - T)^2\} = D\{\eta_n\} \xrightarrow{\lambda_n \rightarrow 0} 0$.

Следовательно, $Q_1 \xrightarrow[\lambda_n \rightarrow 0]{l_2} \frac{1}{2} w_T^2 - \frac{1}{2} T = \int_0^T w(t) dw(t)$.

Рассмотрим

$$Q_2 = \sum_{k=1}^n w(t_{k+1}) (w(t_{k+1}) - w(t_k)) = \\ = \frac{1}{2} \sum_{k=1}^n (w^2(t_{k+1}) - w^2(t_k) + (w(t_{k+1}) - w(t_k))^2) \xrightarrow{\lambda_n \rightarrow 0} \frac{1}{2} w_T^2 + \frac{1}{2} T.$$

Таким образом, предел зависит от выбора промежуточных точек.

8.5. СУЩЕСТВОВАНИЕ И ЕДИНСТВЕННОСТЬ РЕШЕНИЙ СТОХАСТИЧЕСКИХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Пусть $a(t, x)$, $b(t, x)$ – измеримые, локально ограниченные функции на $\mathbb{R}_+ \times \mathbb{R}$ со значениями в \mathbb{R} , $\xi(t)$ – случайный непрерывный процесс, удовлетворяющий соотношению

$$\xi(t) = \xi(0) + \int_0^t a(s, \xi(s)) ds + \int_0^t b(s, \xi(s)) dw(s), \quad (8.24)$$

которое называется стохастическим дифференциальным уравнением (Ито). Интегралы понимаются как интегралы в среднем квадратичном. Вторым интегралом является интеграл Ито, $w(t)$ – стандартный винеровский процесс. В этом случае говорят, что для СП $\xi(t)$ существует дифференциал $d\xi(t)$ удовлетворяющий равенству

$$d\xi(t) = a(t, \xi(t)) dt + b(t, \xi(t)) dw(t).$$

Возникает задача: заданы функции $a(t, x)$, $b(t, x)$, и требуется найти функцию $\xi(t)$, удовлетворяющую (8.24). Значение $\xi(0)$ полагается заданным.

Теорема 8.2. Пусть функции $a(t, x)$, $b(t, x)$ удовлетворяют условиям:

1) для некоторого l и любых $t < T$ выполняются условия

$$|a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| \leq l |x - y|;$$

2) функции $a(t, 0)$, $b(t, 0)$ ограничены.

Тогда уравнение (8.24) имеет решение $\xi(t)$, и это решение единственно, т. е. если существует другое решение $\bar{\xi}(t)$ уравнения (8.24), то

$$P \left\{ \sup_t |\xi(t) - \bar{\xi}(t)| > 0 \right\} = 0.$$

Доказательство. Докажем вначале единственность. Допустим, что существуют два решения уравнения (8.24) $\xi(t)$ и $\bar{\xi}(t)$, для которых $\xi(0) = \bar{\xi}(0)$. Тогда

$$\bar{\xi}(t) - \xi(t) = \int_0^t (a(s, \bar{\xi}(s)) - a(s, \xi(s))) ds + \int_0^t (b(s, \bar{\xi}(s)) - b(s, \xi(s))) dw(s).$$

Пусть $c > 0$. Обозначим $\tau_c = \sup \{t : |\bar{\xi}(t) - \xi(t)| < c\}$, $t_c = \min(t, \tau_c)$.

$$\text{Тогда} \quad \bar{\xi}(t_c) - \xi(t_c) = \int_0^{t_c} (a(s, \bar{\xi}(s)) - a(s, \xi(s))) ds + \int_0^{t_c} (b(s, \bar{\xi}(s)) - b(s, \xi(s))) dw(s).$$

Используя неравенство Шварца и свойства интеграла Ито, получаем

$$\begin{aligned}
& \mathbf{E} \left\{ |\bar{\xi}(t_c) - \xi(t_c)|^2 \right\} \leq 2\mathbf{E} \left\{ \left(\int_0^{t_c} [a(s, \bar{\xi}(s)) - a(s, \xi(s))] ds \right)^2 \right\} + \\
& + 2\mathbf{E} \left\{ \left(\int_0^{t_c} [b(s, \bar{\xi}(s)) - b(s, \xi(s))] dw(s) \right)^2 \right\} \leq \\
& \leq 2t_c \mathbf{E} \left\{ \left(\int_0^{t_c} [a(s, \bar{\xi}(s)) - a(s, \xi(s))]^2 ds \right) \right\} + \\
& + 2 \left(\int_0^{t_c} \mathbf{E} \left\{ [b(s, \bar{\xi}(s)) - b(s, \xi(s))]^2 \right\} ds \right) \leq 2(t_c + 1) l^2 \int_0^{t_c} \mathbf{E} \left\{ |\bar{\xi}(t_c) - \xi(t_c)|^2 \right\} ds. \quad (8.25)
\end{aligned}$$

Таким образом, для каждого $t > 0$ существует константа $R = R(t)$ такая, что функция

$$g_c(t) = \mathbf{E} \left\{ |\bar{\xi}(t_c) - \xi(t_c)|^2 \right\} \quad \text{удовлетворяет неравенству} \quad g_c(t) \leq R \int_0^t g_c(s) ds.$$

Справедлива следующая лемма.

Лемма 8.2. Если функция $g(t)$ удовлетворяет неравенству $g(t) \leq C + R \int_0^t g(s) ds$,

то $g(t) \leq Ce^{Rt}$.

Доказательство. Поскольку

$$\left[\ln \left(C + R \int_0^t g(s) ds \right) \right]' = \frac{Rg(t)}{C + R \int_0^t g(s) ds} \leq R,$$

$$\text{то} \quad \ln \left(C + R \int_0^t g(s) ds \right) \leq \ln C + Rt. \quad \text{Отсюда} \quad C + R \int_0^t g(s) ds \leq Ce^{Rt}. \quad \square$$

Поскольку в нашем случае $C = 0$, то по лемме 8.2 получаем $g_c(t) = 0$ для всех $c > 0$ с вероятностью единица $\bar{\xi}(t_c) = \xi(t_c)$. Устремляя c к ∞ , получим для каждого $t > 0$

$$\bar{\xi}(t) = \xi(t), \quad (\text{п. н.}).$$

Из аксиом непрерывности получим

$$\mathbf{P} \{ \bar{\xi}(t) = \xi(t), t \in Q_+ \} = 1,$$

где Q_+ – множество неотрицательных рациональных чисел. А так как траектории СП $\bar{\xi}(t)$ и $\xi(t)$ непрерывны, то

$$\mathbf{P} \{ \bar{\xi}(t) = \xi(t), t \geq 0 \} = 1.$$

Единственность доказана.

Для доказательства существования решений используем метод последовательных приближений. Будем считать, что $\xi(0) = x \in \mathbb{R}$. Такое решение обозначим $\xi_x(t)$. Положим $\xi_x^{(0)}(t) = x$, а для $n > 0$

$$\xi_x^{(n)}(t) = x + \int_0^t a(s, \xi_x^{(n-1)}(s)) ds + \int_0^t b(s, \xi_x^{(n-1)}(s)) dw(s). \quad (8.26)$$

Если $\xi_x^{(n-1)}(t)$ – непрерывный неупреждающий процесс, то из условий теоремы следует, что $\xi_x^{(n)}(t)$ – непрерывный неупреждающий процесс. Поскольку $\xi_x^{(0)}(t) = x$ есть непрерывная неупреждающая функция, то последовательность $\xi_x^{(n)}(t)$ есть последовательность непрерывных неупреждающих функций.

Из условий теоремы следует, что для всех $T > 0$ существует константа K_T такая, что для всех $t \leq T$

$$|a(t, x)| \leq K_T(1 + |x|), |b(t, x)| \leq K_T(1 + |x|). \quad (8.27)$$

Из (8.27) следует

$$|\xi_x^{(n)}(t)|^2 \leq 3x^2 + 3 \left(\int_0^t a(s, \xi_x^{(n-1)}(s)) ds \right)^2 + 3 \left(\int_0^t b(s, \xi_x^{(n-1)}(s)) dw(s) \right)^2.$$

Отсюда, применяя (8.27) и неравенства, используемые в (8.26), получаем

$$|\xi_x^{(n)}(t)|^2 \leq 3x^2 + 3B_T + L_T \int_0^t \mathbf{E} \left\{ |\xi_x^{(n-1)}(s)|^2 \right\} ds, \quad (8.28)$$

где B_T и L_T зависят только от T и K_T .

Из этого неравенства следует, что функция $\mathbf{E} \left\{ |\xi_x^{(n)}(t)|^2 \right\}$ – локально ограниченная.

Более того, из (8.28) следует

$$\sup_{k \leq n} \mathbf{E} \left\{ |\xi_x^{(k)}(t)|^2 \right\} \leq 3x^2 + B_T + L_T \int_0^t \sup_{k \leq n} \mathbf{E} \left\{ |\xi_x^{(k)}(s)|^2 \right\} ds.$$

Применяя лемму 8.2, получим

$$\sup_{k \leq n} \mathbf{E} \left\{ |\xi_x^{(k)}(t)|^2 \right\} \leq (3x^2 + B_T) e^{TL_T}, t \leq T.$$

Покажем, что $\xi_x^{(n)}(t)$ сходится в среднем квадратичном. Имеем

$$\begin{aligned} \xi_x^{(n+1)}(t) - \xi_x^{(n)}(t) &= \int_0^t (a(s, \xi_x^{(n)}(s)) - a(s, \xi_x^{(n-1)}(s))) ds + \\ &+ \int_0^t (b(s, \xi_x^{(n)}(s)) - b(s, \xi_x^{(n-1)}(s))) dw(s). \end{aligned}$$

Так же, как и при доказательстве необходимости, для $t \leq T$ получаем

$$\mathbf{E} \left\{ |\xi_x^{(n+1)}(t) - \xi_x^{(n)}(t)|^2 \right\} \leq C_T \int_0^t \mathbf{E} \left\{ |\xi_x^{(n)}(s) - \xi_x^{(n-1)}(s)|^2 \right\} ds, \quad (8.29)$$

где C_T зависит только от T и K_T .

$$\text{Поскольку} \quad \mathbf{E} \left\{ \left| \xi_x^{(1)}(t) - \xi_x^{(0)}(t) \right|^2 \right\} \leq \mathbf{E} \left\{ \int_0^t a(s, x) ds + \int_0^t b(\Delta, x) dw(s) \right\} \leq R_T t$$

для $t \leq T$ и R_T зависит только от T , то по индукции из (8.29) получаем

$$\mathbf{E} \left\{ \left| \xi_x^{(n+1)}(t) - \xi_x^{(n)}(t) \right|^2 \right\} \leq R_T C_T^n \frac{t^{n+1}}{(n+1)!}.$$

Следовательно, ряд $\xi_x^{(0)}(t) + \sum_{n=0}^{\infty} (\xi_x^{(n+1)}(t) - \xi_x^{(n)}(t))$ сходится равномерно в среднем квадратичном. Его сумма $\xi_x(t) = \lim_{n \rightarrow \infty} \xi_x^{(n)}(t)$ удовлетворяет равенству (8.24). Покажем это. Так же, как и в доказательстве необходимости, получаем при $t \leq T$

$$\begin{aligned} & \mathbf{E} \left\{ \left(\int_0^t a(s, \xi_x(s)) ds + \int_0^t b(s, \xi_x(s)) dw(s) - \right. \right. \\ & \left. \left. - \int_0^t a(s, \xi_x^{(n)}(s)) ds - \int_0^t b(s, \xi_x^{(n)}(s)) dw(s) \right)^2 \right\} \leq K_T \int_0^t \mathbf{E} \left\{ \left| \xi_x(s) - \xi_x^{(n)}(s) \right|^2 \right\} ds. \end{aligned}$$

Тогда

$$\begin{aligned} & \mathbf{E} \left\{ \left(\xi_x(t) - x - \int_0^t a(s, \xi_x(s)) ds - \int_0^t b(s, \xi_x(s)) dw(s) \right)^2 \right\} = \\ & = \mathbf{E} \left\{ \left(\begin{aligned} & \xi_x(t) - \xi_x^{(n)}(t) + \int_0^t a(s, \xi_x^{(n-1)}(s)) ds - \int_0^t a(s, \xi_x(s)) ds + \\ & + \int_0^t b(s, \xi_x^{(n-1)}(s)) dw(s) - \int_0^t b(s, \xi_x(s)) dw(s) \end{aligned} \right)^2 \right\} \leq \\ & \leq 2\mathbf{E} \left\{ \left| \xi_x(t) - \xi_x^{(n)}(t) \right|^2 \right\} + K_T \int_0^t \mathbf{E} \left\{ \left| \xi_x(s) - \xi_x^{(n)}(s) \right|^2 \right\} ds \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

□

8.6. УПРАЖНЕНИЯ

1. Найти конечномерные распределения винеровского процесса.
2. Пусть $\xi^{(1)}(t)$ и $\xi^{(2)}(t)$, $t \geq 0$, — независимые случайные процессы, каждый из которых является процессом с независимыми приращениями. Доказать, что их сумма $\xi(t) = \xi^{(1)}(t) + \xi^{(2)}(t)$, $t \geq 0$, также является процессом с независимыми приращениями, $t \in \mathbb{R}$.
3. Пусть $\xi(t)$ — случайный процесс с независимыми приращениями. Доказать, что если для некоторых t_1 и t_2 и некоторой постоянной $\mathbf{P}\{\xi(t_1) - \xi(t_2) = a\} = 1$, то для любой пары u_1 и u_2 такой, что $t_1 \leq u_1 \leq u_2 \leq t_2$, существует постоянная b такая, что $\mathbf{P}\{\xi(u_1) - \xi(u_2) = b\} = 1$.

4. Пусть $\phi(t, z)$ – характеристическая функция однородного стохастически непрерывного процесса с независимыми приращениями $\xi(t)$. Доказать, что $\phi(t, z)$ непрерывна как функция t .
5. Пусть $\xi(t)$ – процесс с независимыми приращениями, $\phi(t, z)$ – его характеристическая функция. Доказать, что если $\phi(t, z)$ непрерывна по t в точке t_0 , то $\xi(t)$ стохастически непрерывен в точке t_0 .
6. Пусть ξ_1, \dots, ξ_n – независимые случайные величины, $t_1 < t_2 < \dots < t_n$ – точки из интервала $[a, b]$. Положим $\xi(t) = \sum_{t_k < t} \xi_k$. Доказать, что $\xi(t)$ – процесс с независимыми приращениями.
7. Пусть $\xi(t)$ – процесс с независимыми приращениями, $\phi(t, z)$ – его характеристическая функция. Доказать, что $z|\phi(t, z)|$ не возрастает как функция t .
8. Пусть $\xi(t)$ – однородный случайный процесс с независимыми приращениями, $\xi_0 = 0$, $\phi(t, z)$ – его характеристическая функция. Доказать, что для любых t и s $\phi(t + s, z) = \phi(t, z)\phi(s, z)$.
9. Пусть $\xi(t)$ – процесс с независимыми приращениями. Доказать, что если $\xi(t_0)$ имеет абсолютно непрерывное распределение при некотором t_0 , то $\xi(t)$ имеет абсолютно непрерывное распределение при любом $t \geq t_0$.
10. Пусть $\xi(t)$ – процесс с независимыми приращениями. Доказать, что функция $D\{\xi(t)\}$ не убывает по t .
11. Пусть $\xi(t)$, $a \leq t \leq b$ – однородный процесс с независимыми приращениями. Доказать, что $\xi(t)$ стохастически непрерывен всюду на $[a, b]$.
12. Пусть $\xi(t)$ – процесс с независимыми приращениями, η – некоторая случайная величина, определенная на том же вероятностном пространстве, что и $\xi(t)$. Будет ли процесс $\zeta t = \xi(t) + \eta$ процессом с независимыми приращениями?
13. Пусть $\xi(t)$, $t \geq 0$, – однородный процесс с независимыми приращениями, не равный почти наверное постоянной. Доказать, что $\xi(t)$ не является стохастически ограниченным.
14. Пусть $\xi(t)$, $t \geq 0$, – процесс с независимыми приращениями. Доказать, что если при некотором t_0 $P\{\xi(t_0) = \text{const}\} = 1$, то $P\{\xi(t) = \text{const}\} = 1$ для $t \leq t_0$.
15. Пусть $\xi(t)$ – процесс с независимыми приращениями. Доказать, что если функция $D\{\xi(t)\}$ непрерывна по t , то $\xi(t)$ стохастически непрерывен.
16. Пусть $\xi(t)$, $t \geq 0$, – однородный невырожденный процесс с независимыми приращениями. Доказать, что для любого $t > 0$ и любого $A > 0$ $P\{|\xi(t)| > A\} > 0$.
17. Пусть $w(t)$ – винеровский процесс. Найти совместную плотность распределения величин $w(u)$ и $w(v)$ $0 < u < v < 1$ при условии, что $w(1) = 0$.
18. Пусть $w(t)$ – винеровский процесс. Найти ковариацию величин $w(s)$ и $w(t)$, $s < t < 1$, при условии, что $w(1) = 0$.
19. Пусть $w(t)$ – винеровский процесс. Найти корреляционную функцию процесса $w^{(0)}(t) = w(t) - tw(1)$, рассматриваемого на отрезке $0 \leq t \leq 1$ (условный винеровский процесс).
20. Пусть $w^{(0)}(t)$, $0 \leq t \leq 1$ – условный винеровский процесс, определенный в преды-

- дущем упражнении. Доказать, что процесс $w(t) = (1+t)w^{(0)}\left(\frac{t}{1+t}\right)$, $t \geq 0$, – винеровский.
21. Пусть $w(t)$ – винеровский процесс. Доказать, что следующие процессы также винеровские: а) $w^{(1)}(t) = \begin{cases} 0, & t = 0, \\ tw\left(\frac{1}{t}\right), & t > 0; \end{cases}$ б) $w^{(2)}(t) = \sqrt{c}w\left(\frac{t}{c}\right)$, $t \geq 0$, $c = \text{const} > 0$.
22. Пусть $w^{(1)}(t)$ и $w^{(2)}(t)$ – независимые винеровские процессы. Доказать, что процесс $\frac{1}{\sqrt{2}}(w^{(1)}(t) + w^{(2)}(t))$, $t \geq 0$, также винеровский.
23. Пусть $w(t)$, $t \geq 0$, – винеровский процесс. Положим $w^{(0)}(t) = \begin{cases} w(t), & t \leq T, \\ 2w(T) - w(t), & t > T. \end{cases}$ Доказать, что $w^{(0)}(t)$ – винеровский процесс.
24. Доказать, что винеровский процесс не дифференцируем по вероятности.
25. Доказать, что винеровский процесс является марковским. Найти его переходную функцию.
26. Пусть $w(t)$ – винеровский процесс. Найти условную плотность величины $w(t)$, $t_1 < t < t_2$, при $w(t_1) = A$, $w(t_2) = B$.
27. Пусть $\tau(z)$, $z > 0$, – случайный момент времени, в который винеровский процесс $w(t)$ впервые достигает значения z . Найти плотность распределения $\tau(z)$. Показать, что математическое ожидание $\tau(z)$ бесконечно.
28. Пусть $\tau(z)$ – случайная величина, определенная в предыдущей задаче. Доказать, что композиция распределений случайных величин $\tau(z_1)$ и $\tau(z_2)$ совпадает с распределением случайной величины $\tau(z_1 + z_2)$.
29. Показать, что распределение случайной величины $\tau(z)$, определенной в упражнении 27, совпадает с распределением случайной величины $z^2\tau(1)$.
30. Поток машин, движущихся по шоссе в одном направлении, представляет собой пуассоновский поток с интенсивностью λ . Человек, «голосуя», пытается остановить первую попавшуюся машину. Найти закон распределения времени T , которое ему придется ждать, а также $E\{T\}$ и $D\{T\}$.
31. Пусть ξ и η – независимые неотрицательные случайные величины: $P\{\xi < x\} = 1 - e^{-ax}$, $P\{\eta < x\} = G(x)$. Доказать, что $P\{\xi < u + \eta/\xi \geq \eta\} = 1 - e^{-au}$, $u \geq 0$, в частности, для любого $t \geq 0$: $P\{\xi < u + t/\xi \geq \eta\} = 1 - e^{-au}$, $u \geq 0$. (Свойства отсутствия памяти у показательного распределения.)
32. Доказать, что случайный поток ν_t , полученный в результате наложения k независимых пуассоновских потоков $\nu_t^{(1)}, \dots, \nu_t^{(k)}$ с интенсивностями $\lambda_1, \dots, \lambda_k$ $\nu_t = \nu_t^{(1)} + \dots + \nu_t^{(k)}$, является пуассоновским с интенсивностью $\lambda = \lambda_1 + \dots + \lambda_k$.
33. Пусть задан пуассоновский поток с интенсивностью λ . Каждое требование этого потока с вероятностью p_i , $i = 1, \dots, k$, $\sum_{i=1}^k p_i = 1$ отнесем к i -му подпотoku независимо от остальных требований. Доказать, что i -й подпоток является пуассоновским с интенсивностью λp_i .

Глава 9

МАРКОВСКИЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ

9.1. ЦЕПИ МАРКОВА. ОПРЕДЕЛЕНИЯ И ОСНОВНЫЕ СВОЙСТВА

Будем рассматривать СП $\xi(t) \in D \subset \mathbb{R}$, $t \in T$. Значения СП $\xi(t)$ называются состояниями СП.

Определение 9.1. СП $\xi(t)$, $t \in T$, называется марковским процессом, если для любого натурального n , для любых точек $t_1 < t_2 < \dots < t_n < t$, $t_1, t_2, \dots, t_n, t \in T$, любых $x_1, \dots, x_n \in D$ и любого борелевского множества $B \in \mathcal{B}(\mathbb{R})$ верно равенство

$$P\{\xi(t) \in B | \xi(t_n) = x_n, \dots, \xi(t_1) = x_1\} = P\{\xi(t) \in B | \xi(t_n) = x_n\}. \quad (9.1)$$

Смысл этого соотношения состоит в следующем. Если момент t_n рассматривать как настоящий момент, а момент t – будущий момент времени, то соотношение (9.1) означает, что поведение процесса в будущем зависит только от того, в каком состоянии находится процесс в настоящий момент времени и не зависит от того, как вел себя процесс в прошлом.

Определение 9.2. СП $\xi(t)$, $t \in T$, $\xi(t) \in D \subset \mathbb{R}$, называется цепью Маркова, если он является марковским процессом и множество состояний D случайного процесса $\xi(t)$ – конечное или счетное. В дальнейшем будем считать, что $D = \{0, 1, 2, \dots\}$. Если множество T является конечным или счетным, то $\xi(t)$ называется цепью Маркова с дискретным временем. В этом случае полагают $T = \{0, 1, 2, \dots\}$. Если $T = [0, \infty]$ или интервал на числовой прямой, то СП $\xi(t)$ называется цепью Маркова с непрерывным временем.

В пп. 9.1–9.4 будем рассматривать цепи Маркова с непрерывным временем $t \in T = [0, \infty]$. Полученные результаты будут справедливы и для цепей Маркова с дискретным временем, только моменты времени нужно рассматривать как целочисленные величины.

Обозначим: $\mathcal{F}_{\leq t} = \sigma(\xi(s), s \leq t)$ – сигма-алгебра, порожденная множеством случайных величин $\{\xi(s) : s \leq t\}$; $\mathcal{F}_{=t} = \mathcal{F}_{\xi_t}$ – сигма-алгебра, порожденная случайной величиной $\xi(t)$. Тогда соотношение (9.1) будет эквивалентно соотношению ($t_n < t$):

$$P\{\xi(t) \in B | \mathcal{F}_{\leq t_n}\} = P(\xi(t) \in B | \mathcal{F}_{=t_n}). \quad (9.2)$$

Теорема 9.1. Для цепей Маркова соотношение (9.1) эквивалентно следующему соотношению: для любых $t_1 < t_2 < \dots < t_n < s_1 < \dots < s_m$, $t_i, s_j \in T$, $i = 1, \dots, n$, $j = 1, \dots, m$, $x_1, \dots, x_n \in D$ и любых борелевских множеств $B_1, \dots, B_m \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} P\{\xi(s_m) \in B_m, \dots, \xi(s_1) \in B_1 | \xi(t_n) = x_n, \dots, \xi(t_1) = x_1\} = \\ = P\{\xi(s_m) \in B_m, \dots, \xi(s_1) \in B_1 | \xi(t_n) = x_n\}. \end{aligned} \quad (9.3)$$

Доказательство. Обозначим случайные события: $C_k = \{\xi(s_k) \in B_k\}$, $A_n = \{\xi(t_n) = x_n, \dots, \xi(t_1) = x_1\}$, $A_{(n)} = \{\xi(t_n) = x_n\}$. Полагаем, что $P\{A_n\} \neq 0$. Преобразуем левую часть (9.3). Используя (9.1), получаем

$$\begin{aligned} P\{C_m C_{m-1} \dots C_1 | A_n\} &= P\{C_1 | A_{(n)}\} P\{C_2 | A_{(n)} C_1\} \dots P\{C_m | A_{(n)} C_1 \dots C_{m-1}\} = \\ &= P\{C_1 | A_{(n)}\} P\{C_2 | A_{(n)} C_2\} \dots P\{C_m | A_{(n)} C_1 \dots C_{m-1}\} = P\{C_m C_{m-1} \dots C_1 | A_{(n)}\}. \end{aligned}$$

□

Соотношение (9.3) равносильно соотношению

$$P\{\xi(s_m) \in B_m, \dots, \xi(s_1) \in B_1 | \mathcal{F}_{\leq t_n}\} = P\{\xi(s_m) \in B_m, \dots, \xi(s_1) \in B_1 | \mathcal{F}_{=t_n}\}, \quad (9.4)$$

где $t_n < s_1 < \dots < s_m$.

Обозначим: $P_i(t) = P\{\xi(t) = i\}$ – вероятность i -го состояния цепи Маркова в момент t ; $P_i = P_i(0)$ – вероятность начального состояния; $P_{ij}^{(s)}(t) = P\{\xi(t+s) = j | \xi(s) = i\}$ – переходная вероятность цепи Маркова, т. е. вероятность перехода $i \rightarrow j$ в момент времени s за t шагов.

Из формулы полной вероятности следует, что переходные вероятности $P_{ij}^{(s)}(t)$ удовлетворяют соотношению

$$P_{ij}^{(s)}(t+u) = \sum_k P_{ik}^{(s)}(t) P_{kj}^{(s+t)}(u),$$

которое называется *уравнением Колмогорова – Чепмена*.

Теорема 9.2. Конечномерные распределения цепи Маркова задаются вероятностями начальных состояний $\{P_i\}$ и переходными вероятностями $\{P_{ij}^{(s)}(t)\}$.

Доказательство. Результат следует из формулы умножения вероятностей и (9.1). Действительно, при $t_1 < t_2 < \dots < t_n$ для любых $i_1, \dots, i_n \in D$

$$P\{\xi(t_1) = i_1, \dots, \xi(t_n) = i_n\} = \sum_j P_j P_{ji_1}^{(0)}(t_1) P_{i_1 i_2}^{(t_1)}(t_2 - t_1) \dots P_{i_{n-1} i_n}(t_n - t_{n-1}).$$

□

Определение 9.3. Цепь Маркова называется однородной, если переходные вероятности $P_{ij}^{(s)}(t)$ не зависят от s .

Матрица $P(t) = (P_{ij}(t))$ называется матрицей переходных вероятностей. Матрица $P(t)$ при каждом t обладает следующими свойствами:

- 1) $0 \leq P_{ij}(t) \leq 1$, $i, j \in D$; $t \in T$;
- 2) $\sum_j P_{ij}(t) = 1$ для любого $i \in D$, $t \in T$.

Второе свойство следует из того факта, что за время t цепь Маркова с вероятностью единица из состояния i попадет в одно из состояний j . В матричном виде уравнение Колмогорова – Чепмена запишется в виде

$$P(t+u) = P(u)P(t).$$

Пользуясь формулой полной вероятности, получим равенство

$$P_j(t) = \sum_k P_k P_{kj}(t), \quad j \in D, \quad t \in T. \quad (9.5)$$

Теорема 9.3. Если цепь Маркова однородна, то для любых $t_1 < t_2 < \dots < t_n$, $s > 0$ и любых $i_1, i_2, \dots, i_n \in D$

$$\begin{aligned} P\{\xi(t_n + s) = i_n, \dots, \xi(t_1 + s) = i_1 | \xi(s) = i\} = \\ = P\{\xi(t_n) = i_n, \dots, \xi(t_1) = i_1 | \xi(0) = i\}. \end{aligned} \quad (9.6)$$

Доказательство. Используя формулу умножения и (9.1), имеем цепочку равенств

$$\begin{aligned} & P\{\xi(t_n + s) = i_n, \dots, \xi(t_1 + s) = i_1 | \xi(s) = i\} = \\ & = P\{\xi(t_1 + s) = i_1 | \xi(s) = i\} P\{\xi(t_2 + s) = i_2 | \xi(t_2 + s) = i_1\} \dots P\{\xi(t_n + s) = \\ & = i_n | \xi(t_{n-1} + s) = i_{n-1}\} = P\{\xi(t_1) = i_1 | \xi(0) = i\} P\{\xi(t_2) = i_2 | \xi(t_1) = i_1\} \dots P\{\xi(t_n) = \\ & = i_n | \xi(t_{n-1}) = i_{n-1}\} = P\{\xi(t_n) = i_n, \dots, \xi(t_1) = i_1 | \xi(0) = i\}. \end{aligned}$$

Для однородных цепей Маркова верна более общая теорема. \square

Теорема 9.4. Если цепь Маркова однородна, то для любых $t_1 < t_2 < \dots < t_n < s_1 < s_2 < \dots < s_m$, $s > 0$, любых борелевских множеств $B_1, \dots, B_m \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} & P\{\xi(s_m + s) \in B_m, \dots, \xi(s_1 + s) \in B_1 | \xi(t_n + s) = i_n, \dots, \xi(t_1 + s) = i_1\} = \\ & = P\{\xi(s_m) \in B_m, \dots, \xi(s_1) \in B_1 | \xi(t_n) = i_n\}. \end{aligned} \quad (9.7)$$

Доказательство следует из однородности цепи Маркова $\xi(t)$ и повторяет доказательство теоремы 9.1.

Определение 9.4. СВ $\tau \geq 0$ называется марковским моментом, если событие

$$\{\tau \leq t\} \in \mathcal{F}_t = \sigma(\xi(s), s \leq t).$$

Теорема 9.5. Если $\xi(t)$ – однородная, непрерывная справа цепь Маркова, τ – марковский момент, то для всех $s_1, \dots, s_m > 0$, $i_1, \dots, i_m \in D$ на множестве $(\tau < \infty)$ с вероятностью единица выполнено равенство

$$\begin{aligned} & P\{\xi(\tau + s_1) = i_1, \dots, \xi(\tau + s_m) = i_m | \xi(\tau) = i\} = \\ & = P\{\xi(s_1) = i_1, \dots, \xi(s_m) = i_m | \xi(0) = i\}. \end{aligned} \quad (9.8)$$

Равенство (9.8) называется строго марковским свойством.

Доказательство. Рассмотрим вначале случай, когда $\tau = t_1, t_2, \dots, t_k, \dots$. Обозначим

$$B_\tau = \{\xi(\tau + s_1) = i_1, \dots, \xi(\tau + s_m) = i_m\},$$

$$B_{t_k} = \{\xi(t_k + s_1) = i_1, \dots, \xi(t_k + s_m) = i_m\}.$$

Из определения условного математического ожидания следует, что соотношение (9.8) эквивалентно соотношению: для любого события $A \in F_{\xi(\tau)}$

$$E\{\mathbb{1}_A \mathbb{1}_{B_\tau}\} = E\{\mathbb{1}_A P\{B_0 | \xi(0) = \xi(\tau)\}\}. \quad (9.9)$$

Используя свойства условного математического ожидания и (9.6), получаем

$$\begin{aligned} E\{\mathbb{1}_A \mathbb{1}_{B_\tau}\} &= \sum_k E\{\mathbb{1}_A \mathbb{1}_{B_\tau} \mathbb{1}_{(\tau=t_k)}\} = \\ &= \sum_k E\{\mathbb{1}_A \mathbb{1}_{B_{t_k}} \mathbb{1}_{(\tau=t_k)}\} = \sum_k E\{\mathbb{1}_A \mathbb{1}_{(\tau=t_k)} P\{B_{t_k} | F_{\xi(t_k)}\}\} = \\ &= \sum_k E\{\mathbb{1}_A \mathbb{1}_{(\tau=t_k)} P\{B_{t_k} | \xi(t_k)\}\} = \sum_k E\{\mathbb{1}_A \mathbb{1}_{(\tau=t_k)} P\{B_0 | \xi(0) = \xi(t_k)\}\} = \\ &= \sum_k E\{\mathbb{1}_A \mathbb{1}_{(\tau=t_k)} P\{B_0 | \xi(0) = \xi(\tau)\}\} = E\{\mathbb{1}_A P\{B_0 | \xi(0) = \xi(\tau)\}\}. \end{aligned}$$

Рассмотрим общий случай.

Рассмотрим последовательность $\tau_n = k/n$, если $(k-1)/n < \tau \leq k/n$. Поскольку $\{\tau_n \leq t\} = \bigcup_{k \leq nt} \{\tau \leq k/n\} \in \mathcal{F}_t$, то τ_n – марковский момент, причем $\tau_n \geq \tau$ и при $n \rightarrow \infty$ $\tau_n \downarrow \tau$. Для моментов τ_n имеем

$$P\{B_{\tau_n} | \xi(\tau_n) = i\} = P\{B_0 | \xi(0) = i\}. \quad (9.10)$$

Поскольку случайные величины $\xi(t)$ являются дискретными и процесс $\xi(t)$ непрерывен справа, то при больших n $\xi(\tau_n) = \xi(\tau)$ и $B_{\tau_n} = B_\tau$ с вероятностью 1. Переходя в соотношении (9.10) к пределу при $n \rightarrow \infty$, получим (9.8). \square

9.2. СХОДИМОСТЬ ВЕРОЯТНОСТЕЙ ЦЕПИ МАРКОВА. СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ

Рассмотрим функцию $k(h) = 1 - \sup_{i,j} \frac{1}{2} \sum_n |P_{in}(h) - P_{jn}(h)|$, называемую коэффициентом эргодичности цепи Маркова. Из определения следует, что $0 \leq k(h) \leq 1$.

Теорема 9.6. Если для цепи Маркова $\xi(t)$, $t \in T$, существует $h > 0$ такое, что $k(h) > 0$, то для всех $t \in T$:

- 1) существуют пределы: $\lim_{t \rightarrow \infty} P_{ij}(t) = P_j^*$, не зависящие от i , и $\lim_{t \rightarrow \infty} P_j(t) = P_j^*$;
- 2) $|P_{ij}(t) - P_j^*| \leq (1 - k(h))^{\frac{t}{h}-1}$, $|P_j(t) - P_j^*| \leq (1 - k(h))^{\frac{t}{h}-1}$;
- 3) величины P_j^* удовлетворяют равенствам $P_j^* = \sum_n P_n^* P_{nj}(t)$;
- 4) если существуют $h > 0$, $j_0, \delta > 0$ такие, что $P_{ij_0}(h) \geq \delta > 0$ для всех i , то $k(h) \geq \delta$ и $\sum_j P_j^* = 1$.

Если величины P_j^* удовлетворяют равенствам $P_j^* = \sum_n P_n^* P_{nj}(t)$, $\sum_j P_j^* = 1$, то они называются стационарными вероятностями цепи Маркова.

Доказательство. Обозначим

$$\sum_n^+ a_n = \sum_{n: a_n > 0} a_n, \quad \sum_n^- a_n = \sum_{n: a_n < 0} a_n.$$

Тогда

$$\sum_n |a_n| = \sum_n^+ a_n + \sum_n^- a_n, \quad \sum_n a_n = \sum_n^+ a_n - \sum_n^- a_n.$$

Поскольку $\sum_n (P_{in}(h) - P_{jn}(h)) = 1 - 1 = 0$, то

$$\begin{aligned} \sum_n |P_{in}(h) - P_{jn}(h)| &= \sum_n^+ (P_{in}(h) - P_{jn}(h)) + \\ &+ \sum_n^- (P_{in}(h) - P_{jn}(h)) = 2 \sum_n^+ (P_{in}(h) - P_{jn}(h)). \end{aligned}$$

Тогда $k(h) = 1 - \sum_n^+ (P_{in}(h) - P_{jn}(h))$.

Обозначим

$$R_j(t) = \sup_i P_{ij}(t), \quad r_j(t) = \inf_i P_{ij}(t).$$

Покажем, что $R_j(t)$ убывает по t , а $r_j(t)$ возрастает по t . Пусть $t > 0$, $u > 0$. Тогда, используя уравнение Колмогорова – Чепмена, получаем

$$\begin{aligned} R_j(t+u) &= \sup_i P_{ij}(t+u) = \sup_i \sum_n P_{in}(t)P_{nj}(u) \leq R_j(u); \\ r_j(t+u) &= \inf_i P_{ij}(t+u) = \inf_i \sum_n P_{in}(t)P_{nj}(u) \geq r_j(u). \end{aligned}$$

Оценим разность:

$$\begin{aligned} R_j(t) - r_j(t) &= \sup_i P_{ij}(t) - \inf_i P_{ij}(t) = \sup_{i,l} (P_{ij}(t) - P_{lj}(t)) = \\ &= \sup_{i,l} \left(\sum_n P_{in}(h)P_{nj}(t-h) - \sum_n P_{ln}(h)P_{nj}(t-h) \right) = \\ &= \sup_{i,l} \left(\sum_n^+ (P_{in}(h) - P_{ln}(h))P_{nj}(t-h) - \sum_n^- (P_{in}(h) - P_{ln}(h))P_{nj}(t-h) \right) \leq \\ &\leq \sup_{i,l} \left(\sum_n^+ (P_{in}(h) - P_{ln}(h)) \right) (R_j(t-h) - r_j(t-h)) = \\ &= (1 - k(h))(R_j(t-h) - r_j(t-h)) \leq \dots \leq \\ &\leq (1 - k(h))^{\left[\frac{t}{h}\right]} \left(R_j\left(t - \left[\frac{t}{h}\right]h\right) - r_j\left(t - \left[\frac{t}{h}\right]h\right) \right) \leq (1 - k(h))^{\frac{t}{h}-1}. \end{aligned}$$

Поскольку $k(h) > 0$, $r_j(t) \leq P_{ij}(t) \leq R_j(t)$, $R_j(t) - r_j(t) \rightarrow 0$, то существует P_j^* такое, что

$$\lim_{t \rightarrow \infty} R_j(t) = P_j^* = \lim_{t \rightarrow \infty} r_j(t) = \lim_{t \rightarrow \infty} P_{ij}(t).$$

Из (9.5) и неравенств $r_j(t) \leq P_j(t) \leq R_j(t)$ получаем $P_j^* = \lim_{t \rightarrow \infty} P_j(t)$. Из предыдущих неравенств получаем

$$|P_{ij}(t) - P_j^*| \leq |R_j(t) - r_j(t)| \leq (1 - k(h))^{\frac{t}{h}-1}.$$

Аналогично

$$|P_j(t) - P_j^*| \leq |R_j(t) - r_j(t)| \leq (1 - k(h))^{\frac{t}{h}-1}.$$

Докажем теперь, что выполняется третье утверждение теоремы. Если число состояний цепи Маркова конечно, то, переходя в равенстве $P_{ij}(u+t) = \sum_n P_{in}(u)P_{nj}(t)$ к пределу при $u \rightarrow \infty$, получим $P_j^* = \sum_n P_n^* P_{nj}(t)$. Покажем, что это равенство будет верно, если число состояний цепи Маркова счетно.

Поскольку для любого N $P_{ij}(u+t) \geq \sum_{n \leq N} P_{in}(u)P_{nj}(t)$, то при $u \rightarrow \infty$ получаем $P_j^* \geq \sum_{n \leq N} P_n^* P_{nj}(t)$. Следовательно,

$$P_j^* \geq \sum_n P_n^* P_{nj}(t). \quad (9.11)$$

Предположим, что для некоторого j_0 $P_{j_0}^* > \sum_n P_n^* P_{nj_0}(t)$. Тогда, суммируя (9.11) по j , получаем противоречие: $\sum_j P_j^* > \sum_j \sum_n P_n^* P_{nj}(t) = \sum_n P_n^*$.

Докажем, наконец, четвертое утверждение. Поскольку для всех i $P_{ij_0} \geq \delta > 0$, то $r_{j_0}(h) \geq \delta > 0$ и, следовательно, $P_{j_0}^* > 0$. Тогда $\sum_j P_j^* > 0$. Докажем, что в условиях четвертого утверждения $k(h) \geq \delta$. Это следует из соотношения

$$1 - k(h) = \sup_{ij} \sum_n^+ (P_{in}(h) - P_{jn}(h)) \leq 1 - \delta.$$

Покажем, что $\sum_j P_j^* = 1$. Предположим противное: $\sum_j P_j^* < 1$. Возьмем вероятности начальных состояний $P_j = \frac{P_j^*}{\sum_n P_n^*}$. Имеем $P_j \geq 0$, $\sum_j P_j = 1$ и $P_j(t) = \sum_k P_k P_{kj}(t) = \frac{1}{\sum_n P_n^*} \sum_k P_k^* P_{kj}(t) = \frac{P_j^*}{\sum_n P_n^*} = P_j$. Поскольку $\{P_j^*\}$ не зависят от вероятностей начальных состояний, то из предыдущего равенства получаем $\lim_{t \rightarrow \infty} P_j(t) = P_j^* = P_j$. А это возможно, когда $\sum_j P_j^* = 1$. Теорема доказана. \square

9.3. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ЦЕПЕЙ МАРКОВА С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ

Пусть $\xi(t)$, $t \geq 0$, – однородная цепь Маркова с непрерывным временем и непрерывными справа траекториями. Обозначим $\tau(i)$ – время, в течение которого цепь Маркова будет находиться в состоянии i .

Используя свойства строгой марковости процесса $\xi(t)$ и однородности, получим

$$P\{\tau(i) > t + s\} = P\{\tau(i) > t\}P\{\tau(i) > t + s | \tau(i) > t\} = P\{\tau(i) > t\}P\{\tau(i) > s\}.$$

В силу стохастической непрерывности процесса $\xi(t)$ имеем

$$P\{\tau(i) > t\} = e^{-\lambda_i t}. \quad (9.12)$$

Обозначим π_{ij} – вероятность перехода цепи Маркова $\xi(t)$ из состояния i в состояние j . Из (9.12) следует, что вероятность более двух переходов за время Δt цепи Маркова будет $o(\Delta t)$. Тогда

$$P_{ij}(\Delta t) = P\{\tau(i) < \Delta t\}\pi_{ij} + o(\Delta t) = \lambda_j \pi_{ij} \Delta t + o(\Delta t) = \lambda_{ij} \Delta t + o(\Delta t),$$

где $\lambda_{ij} = \lambda_i \pi_{ij}$; λ_{ij} называется *интенсивностью перехода* процесса $\xi(t)$ из состояния i в состояние j . Считая, что

$$P_{ij}(0) = \delta_{ij}, \quad (9.13)$$

получаем

$$\lambda_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t} = P'_{ij}(0), \quad i \neq j. \quad (9.14)$$

Из (9.12) и (9.13) следует:

$$P'_{ii}(0) = \lim_{\Delta t \rightarrow 0} \frac{P_{ii}(\Delta t) - P_{ii}(0)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{e^{\lambda_i \Delta t} - 1}{\Delta t} = -\lambda_i. \quad (9.15)$$

Поскольку

$$\sum_j P_{ij}(\Delta t) = 1, \quad \text{то} \quad \sum_{j \neq i} \frac{P_{ij}(\Delta t)}{\Delta t} = \frac{1 - P_{ii}(\Delta t)}{\Delta t}. \quad (9.16)$$

Если число состояний цепи Маркова конечно, то, переходя в (9.16) к пределу при $\Delta t \rightarrow 0$, получаем $\sum_{j \neq i} \lambda_{ij} = \lambda_i$. Положим $\lambda_{ii} = -\lambda_i$. Тогда для всех i имеем

$$\sum_j \lambda_{ij} = 0. \quad (9.17)$$

Из уравнения Колмогорова – Чепмена находим $P_{ij}(t + \Delta t) = \sum_k P_{ik}(\Delta t) P_{kj}(t)$, поэтому

$$\frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} + P_{ij}(t) \frac{1 - P_{ii}(\Delta t)}{\Delta t} = \sum_{k \neq j} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t). \quad (9.18)$$

Если число состояний цепи Маркова конечно, то, переходя в (9.18) к пределу при $\Delta t \rightarrow 0$, получим

$$P'_{ij}(t) - \lambda_i P_{ij}(t) = \sum_{k \neq i} \lambda_{ik} P_{kj}(t), \quad \text{или} \quad P'_{ij}(t) = \sum_k \lambda_{ik} P_{kj}(t). \quad (9.19)$$

Систему уравнений (9.19) называют системой обратных дифференциальных уравнений Колмогорова для переходных вероятностей цепи Маркова.

В общем случае справедлива следующая теорема.

Теорема 9.7. Если для однородной, стохастически непрерывной, с непрерывными справа траекториями цепи Маркова $\xi(t)$ выполнено соотношение (9.17), то переходные вероятности $P_{ij}(t)$ удовлетворяют уравнениям (9.19).

Доказательство. Для конечной цепи Маркова теорема доказана. В общем случае из (9.18) получаем

$$\frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} + P_{ij}(t) \frac{1 - P_{ii}(\Delta t)}{\Delta t} \geq \sum_{\substack{k \neq i \\ k \leq N}} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t).$$

Отсюда следует

$$\lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} \geq \sum_k \lambda_{ik} P_{kj}(t). \quad (9.20)$$

С другой стороны,

$$\frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} + P_{ij}(t) \frac{1 - P_{ii}(\Delta t)}{\Delta t} \leq \sum_{\substack{k \neq i \\ k \leq N}} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t) + \sum_{k > N} \frac{P_{ik}(\Delta t)}{\Delta t}. \quad (9.21)$$

Поскольку $\sum_k P_{ik}(\Delta t) = 1$, то, используя (9.13), имеем при $N > i$

$$\sum_{k > N} \frac{P_{ik}(\Delta t)}{\Delta t} = \frac{1}{\Delta t} - \sum_{k \leq N} \frac{P_{ik}(t)}{\Delta t} = \frac{1 - P_{ii}(\Delta t)}{\Delta t} - \sum_{\substack{k \neq i \\ k \leq N}} \frac{P_{ik}(\Delta t)}{\Delta t} \xrightarrow{\Delta t \rightarrow 0} \sum_{k \leq N} \lambda_{ik} \xrightarrow{N \rightarrow \infty} 0.$$

Тогда, переходя в (9.21) к пределу при $\Delta t \rightarrow 0$, получаем

$$\lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} \leq \sum_k \lambda_{ik} P_{kj}(t). \quad (9.22)$$

Из (9.20) и (9.22) следует (9.19). \square

Запишем уравнение Колмогорова – Чепмена в виде $P_{ij}(t + \Delta t) = \sum_k P_{ik}(t) P_{kj}(\Delta t)$, поэтому

$$\frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} + P_{ij}(t) \frac{1 - P_{jj}(\Delta t)}{\Delta t} = \sum_{k \neq j} P_{ik}(t) \frac{P_{kj}(\Delta t)}{\Delta t}. \quad (9.23)$$

Если цепь Маркова $\xi(t)$ конечна, то, переходя в (9.23) к пределу при $\Delta t \rightarrow 0$, получаем

$$P'_{ij}(t) = \sum_k P_{ik}(t) \lambda_{kj}. \quad (9.24)$$

Система дифференциальных уравнений (9.24) называется системой прямых дифференциальных уравнений Колмогорова для переходных вероятностей цепи Маркова $\xi(t)$.

Теорема 9.8. Если для однородной, стохастически непрерывной, с непрерывными справа траекториями существует константа C такая, что для всех i и j $\lambda_{ij} \leq C$ и соотношение (9.14) выполняется равномерно по i и j , то справедливо (9.24).

Доказательство. Равенство (9.18) представим в виде

$$\begin{aligned} & \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} + P_{ij}(t) \frac{1 - P_{jj}(\Delta t)}{\Delta t} = \\ & = \sum_{k \neq j} P_{ik}(t) \left(\frac{P_{kj}(\Delta t)}{\Delta t} - \lambda_{kj} \right) + \sum_{k \neq j} P_{ik}(t) \lambda_{kj}. \end{aligned} \quad (9.25)$$

Из условий теоремы следует, что первое слагаемое в правой части (9.25) стремится к 0 при $\Delta t \rightarrow 0$, второе слагаемое конечно. Тогда, переходя в (9.25) к пределу при $\Delta t \rightarrow 0$, получим (9.24). \square

9.4. ВЕТВЯЩИЕСЯ ПРОЦЕССЫ С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ

Пусть в нулевой момент времени имеется одна частица. За время t она превращается в k частиц с вероятностью $P_k(t)$, $k = 0, 1, 2, \dots$. Затем каждая из k частиц независимо от предыстории и других частиц за время t превращается в l частиц с вероятностью $P_l(t)$, $l = 0, 1, 2, \dots$, и т. д. Каждая из обратившихся частиц независимо от предыстории и других частиц превращается в k частиц с вероятностями $P_k(t)$, $k = 0, 1, 2, \dots$.

Обозначим $\xi(t)$ – число частиц в момент времени t . Случайный процесс $\xi(t)$ является однородной цепью Маркова с непрерывным временем и называется ветвящимся процессом.

Предположим, что при $k \neq 1$ и $\Delta t \rightarrow 0$

$$P_k(\Delta t) = \lambda_k \Delta t + o(\Delta t). \quad (9.26)$$

При этом полагаем $P_1(\Delta t) = 1 - \lambda_1 \Delta t + o(\Delta t)$,

$$P_k(0) = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

Из (9.26) имеем $\lambda_k = P'_k(0)$, $k \neq 1$.

Обозначим $P_{kn}(t) = \mathbf{P}\{\xi(t) = n | \xi(0) = k\}$ – вероятность того, что за время t k частиц превратятся в n частиц. Положим $\lambda_1 = -\lambda$ и предположим, что $\sum_{k=0}^{\infty} \lambda_k = 0$.

Тогда из теоремы 9.7 следует, что вероятности $P_n(t) = P_{1n}(t)$ удовлетворяют дифференциальному уравнению Колмогорова

$$P'_n(t) = \sum_{k=0}^{\infty} \lambda_k P_{kn}(t), \quad n = 0, 1, \dots \quad (9.27)$$

Введем производящие функции:

$$F(t, z) = \sum_{n=0}^{\infty} P_n(t) z^n = \mathbf{E}\{z^{\xi(t)}\}, \quad F_k(t, z) = \sum_{n=0}^{\infty} P_{kn}(t) z^n = \mathbf{E}\{z^{\eta(t)}\},$$

где $\eta(t)$ – число частиц, в которые превращаются k частиц за время t . Умножая правую и левую части (9.27) на z^n и суммируя по n , получим

$$\frac{d}{dt} F(t, z) = \sum_{k=0}^{\infty} \lambda_k F_k(t, z). \quad (9.28)$$

Поскольку превращение частиц не зависит от других частиц, то

$$\eta(t) = \xi_1(t) + \dots + \xi_k(t),$$

где случайные величины $\xi_1(t), \dots, \xi_k(t)$ независимы и имеют то же распределение, что и $\xi(t)$. Тогда из свойств математического ожидания получим $F_k(t, z) = F^k(t, z)$. Подставляя это значение в (9.28), имеем

$$F'(t, z) = \sum_{k=0}^{\infty} \lambda_k F^k(t, z) \quad (9.29)$$

с начальным условием $F(0, z) = z$. Решение уравнения (9.29) равносильно решению уравнения

$$x' = \sum_k \lambda_k x^k \quad (9.30)$$

с начальным условием $x(0) = z$.

Обозначим $f(x) = \sum_{k=0}^{\infty} \lambda_k x^k$. Поскольку $f''(x) \geq 0$, то функция $f(x)$ является вогнутой; $f(1) = 0$, $f(0) = \lambda_0$. Возможно $f(x) = 0$ при $x = \alpha$, где $0 < \alpha < 1$, т. е. график функции $f(x)$ будет иметь вид, как представлено на рис. 9.1.

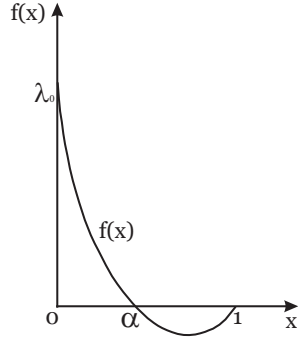
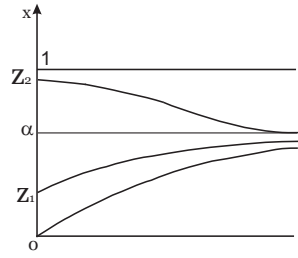
Траектории решений дифференциального уравнения (9.30) будут иметь вид, представленный на рис. 9.2.

Уравнение (9.30) будет иметь особые решения: $x(t) \equiv 1$, $x(t) \equiv \alpha$. Исследуем поведение решений уравнения (9.30) в точке $(0, 1)$. Вместо уравнения (9.30) рассмотрим эквивалентное ему дифференциальное уравнение для обратной функции $t = t(x)$ к функции

$$x = x(t) : \frac{dt}{dx} = \frac{1}{f(x)}.$$

Решение этого уравнения имеет вид

$$t = \int_x^z \frac{du}{f(u)}, \quad 0 \leq x \leq 1.$$

Рис. 9.1. График функции $f(x)$ Рис. 9.2. Графики интегральных кривых $x(t)$, $x(0) = z$

Рассмотрим интегральную кривую, проходящую через точку (t_0, x_0) :

$$t = t_0 + \int_{x_0}^x \frac{du}{f(u)}, \quad 0 \leq x < 1.$$

В случае, если для некоторой точки x_0 , $\alpha < x_0 < 1$, выполняется равенство

$$\int_{x_0}^1 \frac{dx}{f(x)} = -\infty, \quad (9.31)$$

всегда найдется значение $x_0 \leq z < 1$ такое, что

$$t(z) = t_0 + \int_{x_0}^z \frac{du}{f(u)} = 0,$$

т. е. все интегральные кривые пересекают ось $t = 0$ в некоторой точке $(0, z)$, где $0 \leq z < 1$. Следовательно, прямая $x(t) \equiv 1$ является единственной интегральной кривой, проходящей через точку $(0, 1)$.

Если

$$\int_{x_0}^1 \frac{dx}{f(x)} > -\infty, \quad (9.32)$$

то при достаточно большом $t_0 > 0$ интегральная кривая

$$t = t_0 + \int_{x_0}^x \frac{du}{f(u)}$$

проходит по прямой $x(t) \equiv 1$ до точки $(\tau, 1)$, где

$$\tau = t_0 + \int_{x_0}^1 \frac{dx}{f(x)}.$$

В этом случае через точку $(0, 1)$ проходит целое семейство интегральных кривых $x_\tau(t)$, каждая из которых отвечает своему значению $\tau \geq 0$. Среди них есть интегральная кривая $x_0(t)$, отвечающая значению $\tau = 0$ и лежащая ниже всех остальных интегральных кривых $x_\tau(t)$, т. е. $x_0(t) \leq x_\tau(t)$, $0 \leq \tau < \infty$.

Обозначим $\gamma = \min(t : \xi(t) = 0)$. Заметим, что события $\{\xi(t) = 0\} \sim (r < t) = (\xi(t) = 0)$ возрастают по t . Тогда по аксиоме непрерывности вероятностной меры

$$P\{\gamma < \infty\} = \lim_{t \rightarrow \infty} P\{\xi(t) = 0\} = \lim_{t \rightarrow \infty} P_0(t) = \lim_{t \rightarrow \infty} F(t, 0) = \alpha.$$

Следовательно, вероятность вырождения ветвящегося процесса (вероятность того, что настанет момент, когда не останется ни одной частицы) равна α .

Рассмотрим так называемое явление взрыва, когда образуется бесконечно много частиц. Вероятность того, что взрыв произойдет до момента t , есть

$$\begin{aligned} P\{\xi(t) = \infty\} &= 1 - P\{\xi(t) < \infty\} = 1 - \sum_{n=0}^{\infty} P\{\xi(t) = n\} = \\ &= 1 - \sum_{n=0}^{\infty} P_n(t) = 1 - \lim_{z \rightarrow 1} F(t, z). \end{aligned}$$

В случае выполнения соотношения (9.31) интегральная кривая $x(t) \equiv 1$ является единственной интегральной кривой, проходящей через точку $(0, 1)$. Следовательно, $\lim_{z \rightarrow 1} F(t, z) = 1$. Тогда вероятность взрыва будет равна 0.

В случае выполнения соотношения (9.32) $\lim_{z \rightarrow 1} F(t, z) = x_0(t)$, где функция $x_0(t)$ определена выше. В этом случае

$$P\{\xi(t) = \infty\} = 1 - x_0(t) > 0.$$

9.5. КЛАССИФИКАЦИЯ СОСТОЯНИЙ ЦЕПИ МАРКОВА С ДИСКРЕТНЫМ ВРЕМЕНЕМ

Будем рассматривать однородную цепь Маркова с дискретным временем $\xi(t)$, $t \in T$, $\xi(t) \in \mathbf{D}$, где $T = \{0, 1, 2, \dots\}$, $\mathbf{D} = \{0, 1, 2, \dots\}$.

Обозначим $P = P(1) = (P_{ij})$ – матрица переходных вероятностей за один шаг. Тогда из уравнения Колмогорова – Чепмена следует

$$P(n) = (P_{ij}(n)) = P^n.$$

Поскольку время t дискретно, то процесс $\xi(t)$ будем обозначать $\{\xi(n)\}$.

Определение 9.5. Состояние i цепи Маркова $\{\xi(n)\}$ называется *несущественным*, если есть такое состояние j и целое число $k_0 > 0$, что $P_{ij}(k_0) > 0$, а $P_{ji}(k) = 0$ для всех k . В противном случае состояние i называется *существенным*.

Удалим из множества D все несущественные состояния. Тогда оставшееся множество существенных состояний обладает тем свойством, что, попав в него, цепь Маркова никогда из него не выйдет. Поэтому основной интерес представляют существенные состояния.

Определение 9.6. Состояние j называется *достижимым из состояния i* , если существует $m \geq 0$, что $P_{ij}(m) > 0$ (обозначается $i \rightarrow j$). Состояния i и j называются *сообщающимися*, если $i \rightarrow j$ и $j \rightarrow i$ (обозначается $i \leftrightarrow j$).

Сообщающиеся состояния обладают следующим свойством транзитивности: если $i \leftrightarrow j$ и $j \leftrightarrow k$, то $i \leftrightarrow k$. Из этого свойства следует, что все сообщающиеся состояния разбиваются на непересекающиеся классы сообщающихся состояний.

Если класс S_i состоит из одного состояния i , то это состояние называется *поглощающим*.

Определение 9.7. Цепь Маркова, множество состояний которой образует один класс существенных состояний, называется *неприводимой*. В противном случае цепь называется *приводимой*.

Из сказанного выше следует, что цепь Маркова, начиная из некоторого состояния $\xi(0) = i_0$, попадает в некоторый момент в один из классов существенных сообщающихся состояний и в нем будет находиться все оставшееся время. Поэтому изучение цепи Маркова сводится к изучению классов сообщающихся между собой состояний.

В дальнейшем будем рассматривать один из классов сообщающихся между собой состояний. Это равносильно исследованию неприводимой цепи Маркова.

Определение 9.8. Состояние i называется *периодическим с периодом $d(i)$* , если $d(i)$ есть наибольший общий делитель чисел n таких, что $P_{ii}(n) > 0$.

Определение 9.9. Состояние i называется *нулевым*, если $P_{ii}(n) \rightarrow 0$ при $n \rightarrow \infty$, и *ненулевым* в противном случае.

Обозначим:

$$f_i(n) = P\{\xi(n) = i, \xi(n-1) \neq i, \dots, \xi(1) \neq i | \xi(0) = i\}. \quad (9.33)$$

$$F_i = \sum_{n=1}^{\infty} f_i(n), \quad (9.34)$$

где $f_i(n)$ – вероятность того, что цепь Маркова, выйдя из состояния i , впервые попадет в него на n -м шаге; $F_i(n)$ – вероятность того, что цепь Маркова, выйдя из состояния i , вновь когда-нибудь в него вернется.

Определение 9.10. Состояние i называется *возвратным*, если $F_i = 1$, и *невозвратным*, если $F_i < 1$.

Теорема 9.9. Состояние i *возвратно тогда и только тогда, когда расходится ряд*

$$P_i = \sum_{n=1}^{\infty} P_{ii}(n) = \infty. \quad (9.35)$$

Для невозвратного состояния i

$$F_i = \frac{P_i}{1 + P_i}. \quad (9.36)$$

Доказательство. По формуле полной вероятности имеем

$$P_{ii}(n) = f_i(1)P_{ii}(n-1) + f_i(2)P_{ii}(n-2) + \dots + f_i(n-1)P_{ii}(1) + f_i(n). \quad (9.37)$$

Введем производящие функции

$$P_i(z) = \sum_{n=1}^{\infty} z^n P_{ii}(n), \quad F_i(z) = \sum_{n=1}^{\infty} z^n f_i(n), \quad |z| \leq 1.$$

Умножая обе части (9.37) на z^n и суммируя по $n = 1, 2, \dots$, получаем $P_i(z) = (1 + P_i(z))F_i(z)$. Отсюда

$$F_i(z) = \frac{P_i(z)}{1 + P_i(z)}, \quad P_i(z) = \frac{F_i(z)}{1 - F_i(z)}. \quad (9.38)$$

Пусть $P_i = \infty$. Тогда $F_i \rightarrow 1$ при $z \rightarrow 1$, следовательно, $F_i(z) \rightarrow 1$ при $z \rightarrow 1$, т. е. $F_i = 1$.

Если $F_i = 1$, тогда $F_i(z) \rightarrow 1$ при $z \rightarrow 1$. Следовательно, из (9.38) получаем $P_i(z) \rightarrow \infty$ при $z \rightarrow 1$. Значит, $P_j = \infty$.

Если $P_i < \infty$, то при $z = 1$ из (9.38) получаем (9.36). \square

Теорема 9.10 (теорема солидарности). *В неприводимой цепи Маркова все состояния принадлежат одному типу: если хотя бы одно состояние возвратно, то все возвратны; если хотя бы одно периодически с периодом d , то и все периодически с периодом d .*

Доказательство. Пусть i и j – два различных состояния. В силу неприводимости цепи $\{\xi(n)\}$ существуют числа n и m такие, что $\alpha = P_{ij}(n) > 0$, $\beta = P_{ij}(m) > 0$. Из уравнения Колмогорова – Чепмена получаем

$$P_{ii}(n + m + k) = \sum_l \sum_s P_{il}(n) P_{ls}(k) P_{sj}(k).$$

Отсюда следует неравенство

$$P_{ii}(n + m + k) \geq P_{ij}(n) P_{jj}(k) P_{ij}(m) = \alpha \beta P_{jj}(k). \quad (9.39)$$

Следовательно,

$$\frac{1}{\alpha \beta} P_{ii}(k + n + m) \geq P_{jj}(k) \geq \alpha \beta P_{ii}(k - m - n). \quad (9.40)$$

Из неравенств (9.40) следует, что асимптотические свойства при $k \rightarrow \infty$ у $P_{ii}(k)$ и $P_{jj}(k)$ одинаковы: если состояние i – нулевое, то и состояние j – нулевое. Если i – возвратно, то и j – возвратно.

Пусть i – периодическое состояние с периодом $d(i)$. Полагая в (9.39) $k = 0$, получаем $P_{ii}(n + m) \geq \alpha \beta > 0$. Следовательно, $n + m$ делится на $d(i)$. Если k – такое, что $P_{jj}(k) > 0$, то k должно делиться на $d(j)$. При этом из (9.39) следует, что $n + m + k$ делится на $d(i)$. Поскольку $m + n$ делится на $d(i)$, то и k делится на $d(i)$. Отсюда получаем, что $d(i) \leq d(j)$. Аналогично, в силу симметрии, получаем $d(j) \leq d(i)$. Следовательно, $d(i) = d(j) = d$. \square

Теорема 9.11. *Если цепь Маркова $\{\xi(n)\}$ – неприводимая и периодическая с периодом d , то множество состояний цепи разбивается на d подклассов: D_0, D_1, \dots, D_{d-1} таких, что с вероятностью единица за один шаг цепь из класса D_k , $k < d - 1$, переходит в класс D_{k+1} , а из класса D_{d-1} – в класс D_0 .*

Доказательство. Выберем какое-нибудь состояние, например i , и построим классы D_0, D_1, \dots, D_{d-1} следующим образом: $j \in D_k$, если существует такое целое $l > 0$, что $P_{ij}(ld + k) > 0$.

Покажем, что построенные классы не пересекаются. Пусть $j \in D_k$ и для некоторого m $P_{ij}(m) > 0$. Покажем, что m можно представить в виде $m = l_1d + k$, где $l_1 > 0$ – целое число. Поскольку цепь $\{\xi(n)\}$ – неприводимая, то найдется такое $m_1 > 0$, что $P_{ji}(m_1) > 0$. Тогда из неравенств $P_{ii}(ld+k+m_1) \geq P_{ij}(ld+k)P_{ji}(m_1) > 0$, $P_{ii}(m_1+m) = P_{ij}(m)P_{ji}(m_1) > 0$ получаем, что $ld+k+m_1$ и $m+m_1$ делятся на d . Отсюда следует, что остаток от деления m на d равен k , т. е. $m = l_1d + k$.

Пусть $S \in D_k$. Покажем, что $\sum_{j \in D_{k+1}} P_{sj} = 1$, если $k < d-1$, и $\sum_{j \in D_0} P_{sj} = 1$, если $k = d-1$. Это равносильно тому, что $P_{sj} = 0$, если $S \in D_k$, а $j \notin D_{k+1}$, $k < d-1$;

$$S \in D_{d-1}, j \notin D_0. \quad (9.41)$$

Предположим, что (9.41) не выполнено, т. е. $P_{sj} > 0$, если $s \in D_k$, а $j \notin D_{k+1}$. Тогда по определению класса D_k $P_{is}(ld+k) > 0$. Следовательно, $P_{ij}(ld+k+1) \geq P_{is}(ld+k)P_{sj} > 0$, т. е. $j \in D_{k+1}$, что противоречит тому, что $j \notin D_{k+1}$. \square

Построенные классы состояний D_0, D_1, \dots, D_{d-1} периодической цепи $\{\xi(n)\}$ называются *циклическими подклассами*.

В случае $d = 1$ цепь Маркова называется непериодической.

9.6. ЭРГОДИЧЕСКИЕ ЦЕПИ МАРКОВА

Определение 9.11. Дискретная цепь Маркова $\{\xi(n)\}$, $n = 0, 1, 2, \dots$, называется эргодической, если для любых $i, j = 0, 1, \dots$ существуют независимые от i положительные пределы

$$\lim_{n \rightarrow \infty} P_{ij}(n) = P_{ij}^* > 0. \quad (9.42)$$

Лемма 9.1. Если наибольший общий делитель чисел $a_1 > 0, \dots, a_n > 0$ равен 1, то существует N такое, что для любого $k \geq N$

$$k = \lambda_1 a_1 + \dots + \lambda_n a_n,$$

где $\lambda_i > 0$ – целые числа.

Доказательство. Рассмотрим функцию $L(x) = x_1 a_1 + \dots + x_n a_n$, где (x_1, \dots, x_n) – вектор с целыми координатами. Пусть $d > 0$ – минимальное натуральное число, для которого существует такой вектор x^* , что $d = L(x^*)$. Покажем, что любое натуральное число, представимое в виде $L(x)$, делится на d . Предположим противное. Тогда существуют x, k и α , $0 < \alpha < d$, для которых $L(x) = kd + \alpha$. В силу линейности функции $L(x)$ получаем

$$L(x - kx^*) = kd + \alpha - kL(x^*) = kd + \alpha - kd = \alpha < d,$$

что противоречит минимальности d . Все числа a_1, \dots, a_n также являются значениями функции $L(x)$ и поэтому все делятся на d , а значит, по предположению $d = 1$.

Пусть $x^0 = (x_1^0, \dots, x_n^0)$ – вектор, для которого $\alpha(x^*) = 1$, k – любое натуральное число. Обозначим θ – остаток от деления числа k на число $a_1 + \dots + a_n$. Тогда $k = t(a_1 + \dots + a_n) + \theta = t(a_1 + \dots + a_n) + \theta L(x^0) = t(a_1 + \dots + a_n) + \theta(x_1^0 a_1 + \dots + x_n^0 a_n) = a_1(t + \theta x_1^0) + \dots + a_n(t + \theta x_n^0)$. При большом значении k t будет тоже велико. Следовательно, при большом k все $t + \theta x_i^0 > 0$, $i = 1, 2, \dots, n$. \square

В дальнейшем нам понадобится вспомогательное утверждение.

Теорема 9.12. Для того чтобы конечная цепь Маркова была эргодической, необходимо и достаточно, чтобы она была неприводимой и непериодической.

Доказательство. Необходимость. Поскольку число состояний цепи конечно и выполняется соотношение (9.42), то существует N такое, что $P_{ij}(n) > 0$ при $n \geq N$. Отсюда будет следовать неприводимость и непериодичность.

Достаточность. Пусть цепь Маркова неприводима и непериодична. Тогда для каждого состояния i найдутся числа $n_1^{(i)}, \dots, n_{k_i}^{(i)}$ такие, что $P_{ii}(n_l^{(i)}) > 0$, $l = 1, \dots, k_i$, и их наибольший общий делитель будет равен 1. Тогда по лемме 9.1 найдутся целые числа $l_1^{(i)}, \dots, l_{k_i}^{(i)}$ и число N_i такие, что для любого $k \geq N_i$: $k = l_1 n_1^{(i)} + \dots + l_{k_i} n_{k_i}^{(i)}$. Из уравнения Колмогорова – Чепмена получаем $P_{ii}(k) \geq \left(P_{ii}(n_1^{(i)})\right)^{l_1} \cdot \dots \cdot \left(P_{ii}(n_{k_i}^{(i)})\right)^{l_{k_i}} > 0$.

Положим $N^* = \max N_i$. Тогда для $k \geq N^*$ и всех i $P_{ii}(k) > 0$. Пусть $i \neq j$. В силу неприводимости цепи Маркова существует число n_{ij} такое, что $P_{ij}(n_{ij}) > 0$. Тогда из уравнения Колмогорова – Чепмена имеем: $P_{ij}(k + n_{ij}) \geq P_{ii}(k)P_{ij}(n_{ij}) > 0$ для всех $k \geq N^*$. Полагая $n = \max(N^*, \max_{ij} n_{ij})$, получаем $P_{ij}(k) > 0$ для всех $k \geq N$.

Тогда, используя теорему 9.6, получаем результат теоремы. \square

9.7. УПРАЖНЕНИЯ

1. Пусть ξ_0, ξ_1, \dots – последовательность независимых, одинаково распределенных целочисленных случайных величин. Доказать, что она образует цепь Маркова. Найти матрицу вероятностей перехода на n шагов.
2. Частица совершает случайное блуждание в плоскости по целочисленным точкам (i, j) таким, что $0 \leq i, j \leq n$. Из любой внутренней точки указанного квадрата частица с равными вероятностями, независимо от ее предыдущего движения, переходит в одну из соседних (по вертикали или горизонтали) точек. При выходе на границу квадрата частица далее:
 - а) движется по границе квадрата, детерминированного по часовой стрелке;
 - б) возвращается в ту точку, из которой она вышла на границу;
 - в) выбирает случайным образом направление на границе и движется по границе в выбранном направлении.
 Для каждого из указанных случаев определить, будет ли последовательность положений, занимаемых частицей, цепью Маркова.
3. В условиях предыдущего упражнения частица из каждой внутренней точки с равной вероятностью может переходить в одну из соседних (по горизонтали, вертикали или диагонали). Будет ли последовательность положений частицы цепью Маркова для каждого из трех указанных в предыдущем упражнении условий движения после выхода на границу?
4. В начальный момент времени в урне n_0 белых и m_0 черных шаров. Через каждую единицу времени из урны по схеме выбора без возвращения извлекается один шар. Пусть n_k – число белых, а m_k – число черных шаров в урне в момент времени k . Какие из указанных ниже последовательностей образуют цепь Маркова, а какие – нет:
 - а) n_k ; б) $n_k - m_k$; в) $n_k + m_k$; г) пара (n_k, m_k) ; д) $n_k - m_k + 1/(n_k + m_k + 2)$?
5. Пусть случайные величины ξ_0, \dots, ξ_n образуют цепь Маркова. Доказать, что случайные величины η_0, \dots, η_n , где $\eta_i = \xi_{n-i}$, также образуют цепь Маркова. Образуют ли цепь Маркова случайные величины ζ_0, \dots, ζ_n , где ζ_0, \dots, ζ_n – произвольная перестановка ξ_0, \dots, ξ_n ?

6. Пусть ξ_0, ξ_1, \dots – последовательность независимых случайных величин. Образует ли цепь Маркова последовательность $\xi_0 + \xi_1, \xi_1 + \xi_2, \xi_2 + \xi_3, \dots$?
7. Пусть ξ_0, ξ_1, \dots – последовательность случайных величин, образующих цепь Маркова. Будет ли цепью Маркова последовательность $\xi_0 + \xi_1, \xi_1 + \xi_2, \xi_2 + \xi_3, \xi_3 + \xi_4, \dots$?
8. Дана цепь Маркова с конечным числом состояний. Пусть ξ_i – состояние цепи на i -м шаге. Будет ли цепью Маркова последовательность η_0, η_1, \dots , где

$$\eta_1 = \begin{cases} 1, & \text{если } \xi_i = 1, \\ 0, & \text{если } \xi_i \neq 1. \end{cases}$$

9. Пусть ξ_0, ξ_1, \dots – последовательность независимых, одинаково распределенных случайных величин, принимающих значения -1 и $+1$ с вероятностями p и $q = 1 - p$ соответственно. Положим:

$$\text{а) } \eta_n = \xi_n \xi_{n+1}; \quad \text{б) } \eta_n = \max_{0 \leq i \leq n} \xi_i; \quad \text{в) } \eta_n = \prod_{i=0}^n \xi_i.$$

Будет ли последовательность η_0, η_1, \dots цепью Маркова?

10. Пусть ξ_0, ξ_1, \dots – последовательность независимых целочисленных случайных величин, причем

$$P\{\xi_n = k\} = p_k, \quad k = 0, \pm 1, \pm 2, \dots$$

Положим $\eta_n = \xi_0 + \dots + \xi_n$. Доказать, что последовательность η_0, η_1, \dots образует цепь Маркова. Найти соответствующую матрицу вероятностей перехода за один шаг.

11. Пусть ξ_0, ξ_1, \dots и η_0, η_1, \dots – две цепи Маркова. Будет ли цепью Маркова последовательность $\xi_0 + \eta_0, \xi_1 + \eta_1, \dots$?
12. Пусть ξ_0, ξ_1, \dots – независимые случайные величины с дискретным распределением, f_0, f_1, \dots – некоторые функции. Доказать, что последовательность случайных величин η_1, η_2, \dots , где $\eta_{k+1} = f_k(\eta_k, \xi_{k+1})$, образует цепь Маркова.
13. Матрица вероятностей перехода цепи Маркова с двумя состояниями имеет вид

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix},$$

$$0 \leq \alpha \leq 1, \quad 0 \leq \beta \leq 1.$$

Доказать, что матрица вероятностей перехода за n шагов

$$P^{(n)} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix},$$

и найти предел при $n \rightarrow \infty$.

14. Человек либо ходит на работу пешком, либо приезжает на автомобиле. Если он идет пешком, то вероятность, что на следующий день поедет на автомобиле, равна $0,6$. Если едет на автомобиле, то вероятность поехать на автомобиле на следующий день равна $0,7$. Составьте матрицу вероятностей переходов для его ежедневных способов попадания на работу. Если он идет пешком в понедельник, то какова вероятность, что он поедет на автомобиле:

- а) в среду;
 б) в среду и в пятницу;
 в) в пятницу;
 г) в следующий вторник;
 д) в понедельник шестью неделями позже?
 Ответить на перечисленные вопросы, если в понедельник он идет пешком или едет на автомобиле.

15. Мышь передвигается между двумя соседними комнатами A и B без какого-либо риска. Если она покинет комнату A через наружную дверь, то ее поймает кошка. Если покинет комнату B через наружную дверь – попадет в мышеловку. Исходное положение – мышь в комнате B . Каждую минуту она передвигается, не оставаясь в одной комнате: из комнаты A в комнату B – с вероятностью $3/4$, из комнаты B в комнату A – с вероятностью $7/8$. Найти:
 а) вероятность, что мышь поймана кошкой;
 б) вероятность, что мышь поймана мышеловкой;
 в) среднее значение и дисперсию числа минут, оставшихся у мыши до гибели.
16. Указать существенные и несущественные состояния цепи Маркова с матрицей вероятностей перехода за один шаг:

$$\text{а) } P = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}; \quad \text{б) } P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

17. Могут ли все состояния цепи Маркова быть несущественными, если:
 а) цепь Маркова имеет конечное число состояний;
 б) число состояний счетное?
18. Указать все пары сообщающихся состояний для цепи Маркова с матрицей вероятностей перехода за один шаг:

$$\text{а) } P = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}; \quad \text{б) } P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0,2 & 0 & 0,8 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

За сколько шагов из первого состояния можно попасть в третье?

19. Цепь Маркова имеет N состояний. Доказать, что:
 а) если j -е состояние достижимо из i -го ($i \neq j$), то оно может быть достигнуто меньше, чем за N шагов;
 б) если $P_{ii}^{(n)} > 0$ для некоторого n , то возвращение может произойти за N или менее шагов.

20. Будет ли цепь Маркова с матрицей вероятностей перехода за один шаг периодической, если:

$$\text{а) } P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}; \quad \text{б) } P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; \quad \text{в) } P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Для периодических цепей указать период d .

21. Выделить циклические подклассы C_0, C_1, \dots, C_{d-1} множества состояний периодических цепей Маркова в предыдущей задаче а), б). Для этого взять произвольное состояние k , выделить множество C_0 состояний, которые достижимы из k за d шагов, затем множество состояний C_1 , достижимых из состояния k за $d+1$ шаг, и т. д. Показать, что эволюция состояний марковской цепи происходит следующим образом:

$$C_0 \rightarrow C_1 \rightarrow C_2 \rightarrow \dots \rightarrow C_d \rightarrow C_0 \rightarrow C_1 \dots$$

22. Указать возвратные и невозвратные состояния цепи Маркова с матрицей вероятностей перехода за один шаг:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

23. Пусть цепь Маркова имеет периодические состояния. Доказать, что она не является эргодической.
24. Доказать, что цепь Маркова, имеющая по крайней мере два несообщающихся состояния, не является эргодической.
25. Доказать, что если цепь Маркова имеет по крайней мере одно несущественное состояние, то она не является эргодической.
26. Показать, что у неэргодической марковской цепи может существовать стационарное распределение, причем единственное.

27. Доказать, что все состояния цепи Маркова с матрицей переходных вероятностей P возвратные, если:

$$\text{а) } P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \text{б) } P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}; \quad \text{в) } P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix};$$

$$\text{г) } P = \begin{pmatrix} \frac{1}{n} & 0 & \frac{1}{n} & 0 & \dots \\ 0 & \frac{1}{n} & 0 & \frac{1}{n} & \dots \\ \frac{1}{n} & 0 & \frac{1}{n} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

28. Имеется цепь Маркова со счетным числом состояний и матрицей вероятностей перехода за один шаг:

$$P = \begin{pmatrix} p_1 & 1-p_1 & 0 & 0 & 0 & \dots \\ p_2 & 0 & 1-p_2 & 0 & 0 & \dots \\ p_3 & 0 & 0 & 1-p_3 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Доказать, что если ряд $\sum_{i=1}^{\infty} p_i$ сходится, то все состояния этой цепи возвратны, в противном случае – невозвратны.

29. Доказать, что для конечной цепи Маркова всегда существует стационарное распределение.
30. Пусть цепь Маркова имеет два состояния. Доказать, что имеет место один из трех случаев:
а) цепь эргодична; б) состояния не сообщаются; в) цепь периодическая.
31. Доказать, что любая цепь Маркова с конечным числом состояний имеет по крайней мере одно возвратное состояние.
32. Доказать, что для конечной цепи Маркова состояние возвратно тогда и только тогда, когда оно существенно. Показать, что это неверно для цепей со счетным числом состояний.
33. Эргодическая цепь Маркова с двумя состояниями имеет предельные вероятности p и $q = 1 - p$. Найти матрицу вероятностей перехода за один шаг.

34. Эргодичны ли цепи Маркова со следующими матрицами вероятностей перехода за один шаг:

$$\text{а) } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \text{б) } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \text{в) } \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}; \quad \text{г) } \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}; \quad \text{д) } \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix};$$

$$\text{е) } \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}; \quad \text{ж) } \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}?$$

35. Матрица вероятностей перехода P и начальное распределение $p^{(0)}$ цепи Маркова ξ_n имеют вид

$$P = \begin{pmatrix} \frac{3}{12} & \frac{2}{12} & \frac{1}{12} & \frac{3}{12} & \frac{1}{12} & \frac{2}{12} \\ \frac{1}{12} & \frac{1}{12} & \frac{3}{12} & \frac{1}{12} & \frac{4}{12} & \frac{2}{12} \\ 0 & 0 & \frac{3}{12} & \frac{3}{12} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix},$$

$$P^{(0)} = \left(\frac{1}{2} \frac{1}{2} 0 0 0 \right).$$

Найти:

- а) несущественные состояния;
 б) математическое ожидание времени τ до выхода из множества несущественных состояний;
 в) вероятности $P_i^{(\alpha)}$, $P_i^{(\beta)}$ попадания в классы состояний $\alpha = \{3; 4\}$, $\beta = \{5; 6\}$;
 г) предельное при $n \rightarrow \infty$ распределение n состояниям, т. е. величины $\pi_k = \lim_{n \rightarrow \infty} P\{\xi_n = k\}$.
36. Два автомобиля сдаются в аренду по одной и той же цене. Эти автомобили имеют следующие матрицы вероятностей перехода, соответствующие состояниям: 1 – работает хорошо и 2 – требует регулировки:

$$P_1 = \begin{pmatrix} 0,9 & 0,1 \\ 0,6 & 0,4 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0,8 & 0,2 \\ 0,7 & 0,3 \end{pmatrix}.$$

Какую из машин следует арендовать?

37. По двум урнам разложено N черных и N белых шаров так, что каждая урна содержит N шаров. Число черных шаров в первой урне в момент $n = 0, 1, 2, \dots$ обозначим ξ_n . В каждый целочисленный момент времени случайно выбирают по одному шару из каждой урны и меняют их местами. Показать, что ξ_n является цепью Маркова со стационарным распределением

$$\pi_k = (C_N^k)^2 / C_{2N}^N, \quad k = 0, 1, \dots, N.$$

38. Компания по прокату автомобилей выдает их на прокат в трех аэропортах: 1, 2 и 3. Клиенты возвращают автомобили в эти аэропорты в соответствии с матрицей вероятностей перехода за один шаг

$$P = \begin{pmatrix} 0,8 & 0,2 & 0 \\ 0,2 & 0 & 0,8 \\ 0,2 & 0,2 & 0,6 \end{pmatrix}.$$

1. Вычислить стационарные вероятности.
 2. Компания планирует построить ремонтную станцию в одном из аэропортов. Какой из них вы порекомендовали бы? Почему?
39. В городе M каждый житель имеет одну из трех профессий: A , B и C . Дети отцов, имеющих профессии A , B , C , наследуют профессии отцов с вероятностями $3/5$, $2/3$, $1/4$ соответственно и выбирают другую профессию из двух оставшихся с равными вероятностями. Найти:
- а) распределение по профессии в следующем поколении, если в данном поколении профессию A имеют 20 % жителей, B – 30 %, C – 50 %;
 - б) предельное распределение по профессиям, когда число поколений растет;
 - в) распределение по профессиям, не меняющееся при смене поколений.

40. Цепь Маркова имеет следующую матрицу вероятностей перехода за один шаг:

$$P = \begin{pmatrix} p_0 & p_1 & p_r & \dots & p_{m-1} \\ p_{m-1} & p_0 & p_1 & \dots & p_{m-2} \\ \dots & \dots & \dots & \dots & \dots \\ p_1 & p_2 & p_3 & \dots & p_0 \end{pmatrix},$$

где $0 < p_0 < 1$, $\sum_{i=0}^m p_i = 1$. Доказать, что

$$\lim_{n \rightarrow \infty} P\{\xi_n = i\} = \frac{1}{m}, \quad i = 1, 2, \dots, m.$$

41. Найти производящую функцию числа частиц в момент времени t для ветвящихся процессов с непрерывным временем и инфинитезимальными параметрами:

$$\text{а) } \lambda_0 = \lambda, \lambda_k = 0, k \geq 2; \quad \text{б) } \lambda_0 = 0, \lambda_2 = \lambda, \lambda_k = 0, k > 2.$$

42. В условиях предыдущего упражнения найти распределение вероятностей числа потомков частицы в момент времени t .

43. Найти производящую функцию числа частиц в момент времени t для ветвящегося процесса с непрерывным временем и производящей функцией инфинитезимальных параметров:

$$\text{а) } f(z) = \lambda_2 z^2 + \lambda_1 z - (\lambda_1 + \lambda_2); \quad \text{б) } f(z) = z^k - z, \quad k \geq 2; \quad \text{в) } f(z) = 1 - z - \sqrt{1 - z}.$$

44. В условиях предыдущего упражнения найти вероятность вырождения и вероятность взрыва.

Следующие упражнения рассматриваем для ветвящихся процессов с дискретным временем. Математическая модель этого процесса следующая: в нулевой момент времени имеется одна частица. В следующий момент времени эта частица превращается в k частиц первого поколения, $k = 0, 1, 2, \dots$. В следующий момент каждая частица независимо от других частиц превращается в k частиц второго поколения и т. д.

45. Пусть $\xi(t)$, $t = 0, 1, \dots$, – ветвящийся процесс, $\phi(z) = E\{z^{\xi}\} = \sum_{k=0}^{\infty} z^k P\{\xi = k\}$ – производящая функция числа потомков одной частицы, $\phi_n(z)$ – производящая функция числа частиц в n -м поколении. Показать, что

$$\phi_{n+1}(z) = \phi(\phi_n(z)), \quad n \geq 0.$$

46. Найти производящую функцию и распределение вероятностей числа частиц в n -м поколении, если производящая функция числа потомков одной частицы равна:

$$\text{а) } pz + 1 - p; \quad \text{б) } (1 - p)/(1 - pz); \quad \text{в) } 1 - p(1 - z)^{\alpha}, \quad 0 < \alpha < 1.$$

47. Пусть $\xi(t)$, $t = 0, 1, \dots$ – ветвящийся процесс, $\alpha_n = P\{\xi(n) = 0\}$. Доказать, что вероятность вырождения $\alpha = \lim_{n \rightarrow \infty} \alpha_n$ является корнем уравнения $\phi(z) = z$.

48. Доказать, что уравнение из предыдущего упражнения имеет два действительных корня при $\phi(z) \neq z$, причем один из корней равен 1.

49. Доказать, что в дискретном случае вероятность вырождения $\alpha < 1$, если $E\{\xi(1)\} > 1$ и $\alpha = 1$, если $E\{\xi(1)\} \leq 1$.

50. Найти вероятности выражения для ветвящихся процессов с производящей функцией числа потомков одной частицы:

$$\text{а) } (1 - p)(1 - pz); \quad \text{б) } 1 - p(1 - z)^{\alpha}, \quad 0 < \alpha < 1; \quad \text{в) } (1 + z + z^2 + z^3)/4.$$

51. Найти распределение времени вырождения τ для ветвящихся процессов с производящей функцией числа потомков одной частицы:

$$\text{а) } pz + 1 - p; \quad \text{б) } (1 - p)(1 - z)^{\alpha}, \quad 0 < \alpha < 1.$$

52. Найти производящую функцию общего числа частиц в первых n поколениях, если производящая функция числа потомков в одной частицы равна $pz + 1 - p$.

53. Пусть $\xi(0), \xi(1), \dots$ – ветвящийся процесс, $E\{\xi(1)\} = a$. Доказать, что

$$E\{\xi(n+k)/\xi(n)\} = a^k E\{\xi(n)\}.$$

54. Доказать, что если $a = \phi'(1) = E\{\xi(1)\}$ и $b = D\{\xi(1)\}$ конечны, то

$$D\{\xi(n)\} = a^{2n} D\{\xi(0)\} + \frac{ba^{n-1}(a^n - 1)}{a - 1} E\{\xi(0)\}.$$

Глава 10

МАРТИНГАЛЫ, ПОЛУМАРТИНГАЛЫ И МАРКОВСКИЕ МОМЕНТЫ

10.1. МАРТИНГАЛЫ, ПОЛУМАРТИНГАЛЫ И МАРКОВСКИЕ МОМЕНТЫ

Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) задана последовательность СВ $\{\xi_n\}$, $n \geq 0$ и последовательность σ -алгебр $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_n \subset \dots \subset \mathcal{F}$, \mathcal{F}_0 – тривиальная σ -алгебра. В дальнейшем будем предполагать, что случайные величины ξ_n F_n измеримы. Такую последовательность обозначают (ξ_n, \mathcal{F}_n) .

Определение 10.1. Последовательность (ξ_n, \mathcal{F}_n) называется мартингалом, если для всех n выполняются следующие условия:

$$\begin{aligned} 1) E\{|\xi_n|\} &< \infty, \\ 2) E\{\xi_{n+1}|\mathcal{F}_n\} &= \xi_n. \end{aligned} \tag{10.1}$$

Определение 10.2. Последовательность $\{\xi_n, \mathcal{F}_n\}$ называется субмартингалом, если для всех n выполняются следующие условия:

$$\begin{aligned} 1) E\{|\xi_n|\} &< \infty, \\ 2) E\{\xi_{n+1}|\mathcal{F}_n\} &\geq \xi_n. \end{aligned} \tag{10.2}$$

Определение 10.3. Последовательность (ξ_n, \mathcal{F}_n) называется супермартингалом, если для всех n выполняются следующие условия:

$$\begin{aligned} 1) E\{|\xi_n|\} &< \infty, \\ 2) E\{\xi_{n+1}|\mathcal{F}_n\} &\leq \xi_n. \end{aligned} \tag{10.3}$$

Замечание 10.1. Если $\mathcal{F}_n = \sigma(\xi_0, \dots, \xi_n)$ – σ -алгебра, порожденная случайными величинами ξ_0, \dots, ξ_n , и выполняются (10.1, 10.2, 10.3), то говорят, что сама последовательность образует мартингал, субмартингал и супермартингал соответственно. В дальнейшем будем считать, что для всех n $E\{|\xi_n|\} < \infty$.

Теорема 10.1. Соотношения (10.1, 10.2, 10.3) эквивалентны соотношениям

$$\int_A \xi_{n+1} dP = \int_A \xi_n dP, \tag{10.4}$$

$$\int_A \xi_{n+1} dP \geq \int_A \xi_n dP, \tag{10.5}$$

$$\int_A \xi_{n+1} dP \leq \int_A \xi_n dP \tag{10.6}$$

соответственно для любого события $A \in \mathcal{F}_n$.

Доказательство. По определению условного математического ожидания

$$\int_A \xi_{n+1} dP = \int_A E\{\xi_{n+1} | \mathcal{F}_n\} dP = \int_A \xi_n dP \quad \text{для всех } A \in \mathcal{F}_n.$$

Если верно (10.4), то из равенства

$$\int_A \xi_{n+1} dP = \int_A \xi_n dP$$

следует

$$\int_A E\{\xi_{n+1} | \mathcal{F}_n\} dP = \int_A \xi_n dP.$$

Поскольку подынтегральные случайные величины ξ_n измеримы, то они равны, т. е. верно (10.1). Из этих соотношений получаем, что из (10.1) следует (10.4). Эквивалентность соотношений (10.2) и (10.5), (10.3) и (10.6) доказывается аналогично. \square

Теорема 10.2. Соотношения (10.1), (10.2) и (10.3) эквивалентны соотношениям

$$1) E\{\xi_{n+1}\eta\} = E\{\xi_n\eta\}; \quad (10.7)$$

$$2) E\{\xi_{n+1}\eta\} \geq E\{\xi_n\eta\}; \quad (10.8)$$

$$3) E\{\xi_{n+1}\eta\} \leq E\{\xi_n\eta\} \quad (10.9)$$

соответственно для любой ограниченной случайной величины η \mathcal{F}_n -измеримой.

Доказательство. Теорема 10.1 следует из теоремы 10.2, если в качестве случайной величины η взять индикатор события A , т. е. $\eta = \mathbb{1}_A$. Обратное утверждение следует из свойств математического ожидания. Если верна теорема 10.1, то теорема 10.2 будет верна, если случайная величина η является индикатором события. В силу линейности математического ожидания теорема 10.2 верна для простых случайных величин η . Из теоремы о монотонной сходимости получим, что теорема 10.2 будет верна, если случайная величина η является неотрицательной. Представляя случайную величину η в виде $\eta = \eta^+ - \eta^-$, получим справедливость теоремы 10.2 для произвольных ограниченных \mathcal{F}_n измеримых случайных величин. \square

Теорема 10.3. Соотношения (10.1), (10.2), (10.3) эквивалентны соотношениям

$$E\{\xi_n | \mathcal{F}_m\} = \xi_m; \quad (10.10)$$

$$E\{\xi_n | \mathcal{F}_m\} \leq \xi_m; \quad (10.11)$$

$$E\{\xi_n | \mathcal{F}_m\} \geq \xi_m \quad (10.12)$$

соответственно для любых $m \leq n$.

Доказательство. Следует по индукции из равенства

$$E\{\xi_n | \mathcal{F}_m\} = E\{E\{\xi_n | \mathcal{F}_{n-1}\} | \mathcal{F}_m\} = E\{\xi_{n-1} | \mathcal{F}_m\}.$$

\square

Определение 10.4. Случайная величина $\tau = 0, 1, 2, \dots$ называется марковским моментом, если для всех n событие $(\tau = n) \in \mathcal{F}_n$. Если $P(\tau < \infty) = 1$, то случайная величина τ называется моментом остановки.

Пример 10.1. Пусть $\{\xi_n\}$ – последовательность случайных величин, τ – первый момент достижения уровня B , B – борелевское множество, $\tau = \min\{n : \xi_n \in B\}$. Тогда $(\tau = n) = \{\xi_0 \notin B, \xi_1 \notin B, \dots, \xi_{n-1} \notin B, \xi_n \in B\} \in \mathcal{F}_n$.

Теорема 10.4. Если τ и σ – марковские моменты, то $\tau + \sigma$, $\min(\tau, \sigma) = \tau \wedge \sigma$, $\max(\tau, \sigma) = \tau \vee \sigma$, $\tau = k$, где k постоянная, являются марковскими моментами.

Доказательство. Для каждого $n \geq 0$ имеем

$$\begin{aligned}(\tau + \sigma = n) &= \bigcup_{k=0}^n (\tau = k)(\sigma = n - k) \in \mathcal{F}_n; \\(\tau \wedge \sigma = n) &= (\tau = n) \bigcup_{k=0}^n (\sigma = k) + (\sigma = n) \bigcup_{k=0}^n (\tau = k); \\(\tau \vee \sigma = n) &= (\tau = n) \bigcup_{k=0}^n (\sigma = k) + (\sigma = n) \bigcup_{k=0}^n (\tau = k).\end{aligned}$$

Обозначим $\mathcal{F}_\tau = \{A \in \mathcal{F} : \text{для любого } n \geq 0 A(\tau = n) \in \mathcal{F}_n\}$. Если \mathcal{F}_n – совокупность событий, наблюдаемых до момента времени n , то \mathcal{F}_τ – совокупность событий, наблюдаемых до случайного момента времени τ . \square

Теорема 10.5. Совокупность событий \mathcal{F}_τ является σ -алгеброй.

Доказательство. Поскольку $\Omega(\tau = n) = (\tau = n) \in \mathcal{F}_n$, то $\Omega \in T_\tau$. Для любых n , $A_1, A_2 \in \mathcal{F}_\tau$ выполняются соотношения:

$$\begin{aligned}(A_1 + A_2)(\tau = n) &= A_1(\tau = n) + A_2(\tau = n) \in \mathcal{F}_n; \\A_1 A_2(\tau = n) &= A_1(\tau = n) A_2(\tau = n) \in \mathcal{F}_n; \\\bar{A}(\tau = n) &= (\tau = n) \overline{A(\tau = n)} \in \mathcal{F}_n,\end{aligned}$$

т. е. события $A_1 + A_2, A_1 A_2, \bar{A}_1, \bar{A}_2$ принадлежат \mathcal{F}_τ . Если $A_1, A_2, \dots \in \mathcal{F}_\tau$, то $\left(\bigcup_{n=1}^{\infty}\right)(\tau = n) = \bigcup_{n=1}^{\infty} A_n(\tau = n) \in \mathcal{F}_\tau$, т. е. $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_\tau$. \square

Теорема 10.6. Если (ξ_n, \mathcal{F}_n) – мартингал (субмартингал), то остановленная последовательность $(\xi_{n \wedge \tau}, \mathcal{F}_n)$ – мартингал (субмартингал).

Доказательство. Случайные величины $\xi_{n \wedge \tau} - \mathcal{F}_n$ -измеримы. Докажем, что для любого $A \in \mathcal{F}_n$

$$\int_A \xi_{n+1 \wedge \tau} dP = \int_A \xi_{n \wedge \tau} dP \text{ или } \int_A (\xi_{n+1 \wedge \tau} - \xi_{n \wedge \tau}) dP = 0.$$

Поскольку

$$\begin{aligned}\xi_{n \wedge \tau} &= \sum_{k=1}^n \xi_k \mathbb{I}_{(\tau=k)} + \xi_n \mathbb{I}_{(\tau > n)}, \\ \xi_{n+1 \wedge \tau} &= \sum_{k=0}^n \xi_{n+1 \wedge \tau} \mathbb{I}_{(\tau=k)} + \xi_{n+1 \wedge \tau} \mathbb{I}_{(\tau > n)} = \sum_{k=0}^n \xi_k \mathbb{I}_{(\tau > n)} + \xi_{n+1} \mathbb{I}_{(\tau > n)}.\end{aligned}$$

Тогда

$$\int_A (\xi_{n+1 \wedge \tau} - \xi_{n \wedge \tau}) dP = \int_A (\xi_{n+1} - \xi_n) \mathbb{I}_{(\tau > n)} dP = \int_{A(\tau > n)} (\xi_{n+1} - \xi_n) dP =$$

$$= \begin{cases} = 0, & \text{если } (\xi_n, \mathcal{F}_n) - \text{мартингал,} \\ \geq 0, & \text{если } (\xi_n, \mathcal{F}_n) - \text{субмартингал.} \end{cases}$$

□

Теорема 10.7 (о сохранении свойств мартингальности при замене времени на случайный момент). Пусть (ξ_n, \mathcal{F}_n) – мартингал (субмартингал), τ, σ – марковские моменты такие, что $0 \leq \tau \leq \sigma \leq N$. Тогда $E\{\xi_\sigma | \mathcal{F}_\tau\} = \xi_\tau$ ($E\{\xi_\sigma | \mathcal{F}_\tau\} \geq \xi_\tau$).

Доказательство. Рассмотрим вначале случай, когда $\sigma - \tau = 0$ или $\sigma - \tau = 1$. Тогда для любого события $A \in \mathcal{F}_\tau$

$$\begin{aligned} \int_A \xi_\tau dP &= \left[1 = \sum_{k=0}^N \mathbb{1}_{(\tau=k)} \right] = \sum_{k=0}^N \int_{A(\tau=k)} \xi_\tau dP = \sum_{k=0}^N \int_{A(\tau=k)} \xi_k dP = \\ &= \sum_{k=0}^N \left(\int_{A(\tau=k)(\sigma=k)} \xi_k dP + \int_{A(\tau=k)(\sigma>k)} \xi_k dP \right) = \\ &= \sum_{k=0}^N \left(\int_{A(\tau=k)(\sigma=k) \in \mathcal{F}_k} \xi_k dP + \int_{A(\tau=k)(\sigma>k)} \xi_{k+1} dP \right) = \sum_{k=0}^N \int_A \xi_\sigma I_{(\tau=k)} dP = \int_A \xi_\sigma dP. \end{aligned}$$

Общий случай. Рассмотрим последовательность марковских моментов

$$\tau_k = \min(\tau + k, \sigma), \quad \tau = \tau_0 \leq \tau_1 \leq \dots \leq \tau_N = \sigma.$$

Поскольку $\tau_{k+1} - \tau_k = 0, 1$ и $\mathcal{F}_{\tau_0} \subset \mathcal{F}_{\tau_1} \subset \dots \subset \mathcal{F}_{\tau_N}$, то

$$\begin{aligned} E\{\xi_\sigma | \mathcal{F}_\tau\} &= E\{\xi_{\tau_N} | \mathcal{F}_{\tau_0}\} = E\{E\{\xi_{\tau_N} | \mathcal{F}_{\tau_{N-1}}\} | \mathcal{F}_0\} = (\geq \text{ для субмартингалов}) = \\ &= E\{\xi_{\tau_{N-1}} | \mathcal{F}_{\tau_0}\} = \dots = \xi_{\tau_0} = \xi_\tau. \end{aligned}$$

□

Следствие 10.1. Если $\{\xi_n, \mathcal{F}_n\}$ – мартингал (субмартингал), $0 \leq \tau \leq \sigma \leq N$, то

$$E\{\xi_0\} = E\{\xi_\tau\} = E\{\xi_\sigma\} = E\{\xi_N\} \quad (E\{\xi_0\} \leq E\{\xi_\tau\} \leq E\{\xi_\sigma\} \leq E\{\xi_N\}).$$

10.2. НЕРАВЕНСТВА ДЛЯ СУБМАРТИНГАЛОВ

Теорема 10.8. Если (ξ_n, \mathcal{F}_n) – субмартингал, то справедливо

$$P\left\{\max_{n \leq N} \xi_n > \lambda\right\} \leq \frac{1}{\lambda} \int_{\max_{n \leq N} \xi_n > \lambda} \xi_N dP \leq \frac{1}{\lambda} E\{\xi_N^+\}.$$

Доказательство. Рассмотрим случайную величину $\tau = \min(n : \xi_n > \lambda)$. Если $\xi_n \leq \lambda$ для любого $n \leq N$, то $\tau = N$. Поскольку при всех n событие $(\tau = n) = (\xi_0 \leq \lambda, \dots, \xi_{n-1} \leq \lambda, \xi_n > \lambda) \in \mathcal{F}_n$, то τ – марковский момент. Поскольку $0 \leq \tau \leq N$, то из следствия 10.1 и свойств математического ожидания имеем

$$\begin{aligned} E\{\xi_N\} &\geq E\{\xi_\tau\} = E\{\xi_\tau\} \mathbb{1}_{\max_{n \leq N} \xi_n > \lambda} + E\{\xi_\tau\} \mathbb{1}_{\max_{n \leq N} \xi_n \leq \lambda} \geq \\ &\geq \lambda P\left\{\max_{n \leq N} \xi_n > \lambda\right\} + E\{\xi_N\} \mathbb{1}_{\max_{n \leq N} \xi_n \leq \lambda}, \end{aligned}$$

следовательно,

$$\lambda P \left\{ \max_{n \leq N} \xi_n > \lambda \right\} \leq E\{\xi_N\} \mathbb{1}_{\max_{n \leq N} \xi_n > \lambda} = \int_{\max_{n \leq N} \xi_n \leq \lambda} \xi_N dP.$$

□

Следствие 10.2 (неравенство Колмогорова). Если ξ_1, ξ_2, \dots – независимые случайные величины и для всех n $E\{\xi_n\} = 0$, $E\{\xi_n^2\} < \infty$, $S_n = \xi_1 + \dots + \xi_n$, то

$$P \left\{ \max_{n \leq N} |S_n| > \lambda \right\} \leq \frac{1}{\lambda^2} E\{S_N^2\} = \frac{1}{\lambda^2} \sum_{k=0}^N D\{\xi_k\}.$$

Доказательство. Поскольку последовательность (S_n, \mathcal{F}_n) – мартингал, то из неравенства Йенсена следует, что (S_n^2, \mathcal{F}_n) – субмартингал. Тогда по теореме 10.8

$$P \left\{ \max_{n \leq N} |S_n| > \lambda \right\} = P \left\{ \max_{n \leq N} |S_n|^2 > \lambda^2 \right\} \leq \frac{1}{\lambda^2} E\{S_N^{2+}\} = \frac{1}{\lambda^2} E\{S_N^2\}.$$

□

Лемма 10.1. Если $E\{\xi^p\} < \infty$, $\xi \geq 0$, $p \geq 1$, то $E\{\xi^p\} = p \int_0^\infty x^{p-1} P\{\xi \geq x\} dx$.

Доказательство. Поскольку

$$\int_0^\infty x^p dF_\xi(x) < \infty, \quad \text{то} \quad 0 < \int_y^\infty x^p dF_\xi(x) \xrightarrow{y \rightarrow \infty} 0.$$

Тогда

$$0 \leq \int_y^\infty x^p dF_\xi(x) \geq y^p \int_y^\infty dF_\xi(x) = y^p (1 - F_\xi(y)) = y^p P\{\xi \geq y\} \xrightarrow{y \rightarrow \infty} 0.$$

Интегрируя по частям, получаем

$$\begin{aligned} \int_0^\infty x^p dF_\xi(x) &= - \int_0^\infty x^p d(1 - F_\xi(x)) = \\ &= -x^p (1 - F_\xi(x))|_0^\infty + \int_0^\infty (1 - F_\xi(x)) dx^p = \int_0^\infty P\{\xi \geq x\} p x^{p-1} dx. \end{aligned}$$

□

Теорема 10.9. Если (ξ_n, F_n) – субмартингал, то:

1. при $p > 1$; $\xi_n \geq 0$ $E \left\{ \left(\max_{n \leq N} \xi_n \right)^p \right\} \leq \left(\frac{p}{p-1} \right)^p E\{\xi_N^p\};$
2. если $\xi_n \geq 0$, то $E \left\{ \max_{n \leq N} \xi_n \right\} \leq \frac{e}{e-1} (1 + E\{\xi_n \ln t \xi_n\}).$

Доказательство. Обозначим $\eta_N = \max_{n \leq N} \xi_n$. По теореме 10.8

$$P\{\eta_N > \lambda\} = 1/\lambda \int_{\eta_N > \lambda} \xi_N dP.$$

1. Будем считать, что $E\{\xi_N^p\} < \infty$. В этом случае $\eta_N^p \leq \left(\sum_{n=1}^N \xi_n\right)^p$ и

$$E\{\eta_N^p\} \leq N^{p-1} \sum_{n=1}^N E\{\xi_n^p\} \leq N^p E\{\xi_N^p\} < \infty.$$

Поскольку $\eta_N^p \geq \xi_N^p$, то в противном случае $E\{\eta_N^p\} = \infty$. Тогда, используя лемму 10.1 и теорему Фубини, получаем

$$\begin{aligned} E\{\eta_N^p\} &= p \int_0^\infty P\{\eta_N > x\} x^{p-1} dx \leq p \int_0^\infty x^{p-1} \frac{1}{x} \int_{\Omega} \xi_N \mathbb{1}_{(\eta_N > x)} dP dx = \\ &= p \int_{\Omega} \xi_N \int_0^\infty x^{p-2} \mathbb{1}_{(\eta_N > x)} dx dP = p \int_{\Omega} \xi_N \int_0^{\eta_N} x^{p-2} dx dP = \frac{p}{p-1} \int_{\Omega} \xi_N \eta_N^{p-1} dP = \\ &= \frac{p}{p-1} E\{\xi_N \eta_N^{p-1}\} \leq \frac{p}{p-1} (E\{\xi_N^p\})^{\frac{1}{p}} (E\{\eta_N^p\})^{\frac{p-1}{p}}. \end{aligned}$$

2. Используя лемму 10.1 и теорему Фубини, получим

$$\begin{aligned} E\{\eta_N - 1\} &= E\{\eta_N - 1\} \leq E\{\eta_N - 1\}^+ \leq \int_0^\infty P\{(\eta_N - 1)^+ > x\} dx = \\ &= \int_0^\infty P\{\eta_N - 1 > x\} dx = \int_0^\infty P\{\eta_N > 1 + x\} dx \leq \int_0^\infty \frac{1}{1+x} \int_{\Omega} \xi_N \mathbb{1}_{(\eta_N > 1+x)} dP dx = \\ &= \int_{\Omega} \xi_N \int_0^{(\eta_N - 1)^+} \frac{1}{1+x} dx dP = \int_{\Omega} \xi_N \ln(1+x)|_0^{(\eta_N - 1)^+} dP = \int_{\Omega} \xi_N \ln^+ \eta_N dP = E\{\xi_N\} \ln^+ \eta_N. \end{aligned}$$

Тогда

$$E\{\eta_N\} \leq 1 + E\{\xi_N \ln^+ \eta_N\}.$$

Воспользуемся неравенством для $a > 0, b > 0$:

$$a \ln^+ b \leq a \ln^+ a + b e^{-1}.$$

Тогда

$$E\{\eta_N\} \leq \frac{e}{e-1} (1 + E\{\xi_N \ln^+ \xi_N\}).$$

□

Пусть $-\infty < a < b < \infty$. Введем величины

$$\begin{aligned} \tau_1 &= \min(n : \xi_n \leq a); \quad \tau_2 = \min(n : \xi_n \geq b, n > \tau_1); \\ \tau_{2m-1} &= \min(n : n > \tau_{2m-2}, \xi_n \leq a); \quad \tau_{2m} = \min(n : n > \tau_{2m-1}, \xi_n \geq b) \dots \end{aligned}$$

Обозначим $\beta_n(a, b)$ – число пересечений уровней (a, b) снизу вверх последовательностью $\{\xi_n\}$ на промежутке $[0, n]$ $\beta_n(a, b) = \max(m : \tau_{2m} \leq n)$.

Теорема 10.10 (неравенство Дуба о среднем числе пересечения уровней субмартингалом). Если (ξ_n, \mathcal{F}_n) – субмартингал, то $E\{\beta_n(a, b)\} \leq \frac{E\{\xi_n^+\} + |a|}{b - a}$.

Доказательство. Из неравенств

$$\int_A \xi_{n+1}^+ dP \geq \int_{A(\xi_n > 0)} \xi_{n+1}^+ dP \geq \int_{A(\xi_n > 0)} \xi_{n+1} dP \geq \int_{A(\xi_n > 0)} \xi_n dP = \int_A \xi_n \mathbb{1}_{(\xi_n > 0)} dP = \int_A \xi_n^+ dP$$

для $A \in \mathcal{F}_n$ следует, что если (ξ_n, \mathcal{F}_n) – субмартингал, то (ξ_n^+, \mathcal{F}_n) – субмартингал. Если вместо субмартингала (ξ_n, \mathcal{F}_n) рассматривать последовательность $((\xi_n - a)^+, \mathcal{F}_n)$, то она тоже субмартингал. Заметим, что $\beta_n(a, b) = \beta_n^*(a, b - a)$, где $\beta_n^*(a, b - a)$ – число пересечений уровней $(0, b - a)$ субмартингалом $((\xi_n - a)^+, \mathcal{F}_n)$ на промежутке $[0, n]$.

Введем величины $\phi_i = \begin{cases} 1, & \tau_m < i \leq \tau_{m+1}, \text{ если } m \text{ нечетное,} \\ 0, & \tau_m < i \leq \tau_{m+1}, \text{ если } m \text{ четное.} \end{cases}$

Тогда

$$(b - a) \beta_n(a, b) \leq \sum_{i=1}^n \phi_i (\xi_i - \xi_{i-1}).$$

Покажем, что случайные величины ϕ_i \mathcal{F}_{i-1} -измеримы, т. е. что событие

$$(\phi_i = 1) \in \mathcal{F}_{i-1} :$$

$$(\phi_i = 1) = \bigcup_{k=1}^n (\tau_{2k-1} < i \leq \tau_{2k}) = \bigcup_{k=1}^n (\tau_{2k-1} < i; \tau_{2k} \geq i) \in \mathcal{F}_{i-1},$$

так как все τ_m – марковские моменты. Обозначим $\xi_n^* = (\xi_n - a)^+$. Тогда

$$\begin{aligned} (b - a) E\{\beta_n(a, b)\} &\leq \sum_{i=1}^n E\{\phi_i(\xi_i^* - \xi_{i-1}^*)\} = \\ &= \sum_{i=1}^n E\{E\{\phi_i(\xi_i^* - \xi_{i-1}^*) | \mathcal{F}_{i-1}\}\} = \sum_{i=1}^n E\{\phi_i(E\{\xi_i^* | \mathcal{F}_{i-1}\} - \xi_{i-1}^*)\} \leq \\ &\leq \sum_{i=1}^n E\{E\{\xi_i^* | \mathcal{F}_{i-1}\} - \xi_{i-1}^*\} = \sum_{i=1}^n (E\{\xi_i^*\} - E\{\xi_{i-1}^*\}) \leq E\{\xi_n^*\} \leq E\{\xi_n^+\} + |a|. \end{aligned}$$

□

10.3. ТЕОРЕМЫ О СХОДИМОСТИ

Теорема 10.11. Если (ξ_n, \mathcal{F}_n) – субмартингал и $\sup_n E\{\xi_n^+\} < \infty$, то с вероятностью 1 существует $\lim_{n \rightarrow \infty} \xi_n$.

Доказательство. От противного. Пусть предела не существует. Тогда

$$P\{\underline{\lim} \xi_n < \overline{\lim} \xi_n\} > 0. \quad (10.13)$$

Пусть $\omega \in (\underline{\lim} \xi_n(\omega) < \overline{\lim} \xi_n(\omega))$. Тогда найдутся рациональные числа $a = a(\omega)$ и $b = b(\omega)$ такие, что $\underline{\lim} \xi_n(\omega) \leq a < b \leq \overline{\lim} \xi_n(\omega)$. Покажем, что существуют постоянные $c, d \in Q$ такие, что $P\{\underline{\lim} \xi_n \leq c < d \leq \overline{\lim} \xi_n\} > 0$.

От противного. Пусть для любых $a, b \in Q$ $P\{\underline{\lim} \xi_n \leq a < b \leq \overline{\lim} \xi_n\} = 0$. Тогда $(\underline{\lim} \xi_n < \overline{\lim} \xi_n) = \bigcup_{(a,b) \in Q} (\underline{\lim} \xi_n \leq a < b \leq \overline{\lim} \xi_n)$, где объединение счетное. Тогда $P\{\underline{\lim} \xi_n < \overline{\lim} \xi_n\} \leq \sum_{a,b \in Q} P\{\underline{\lim} \xi_n \leq a < b \leq \overline{\lim} \xi_n\} = 0$. Получили противоречие с (10.13). Следовательно, если выполняется (10.13), то существуют $c, d \in Q$ такие, что $P\{\underline{\lim} \xi_n \leq c < d \leq \overline{\lim} \xi_n\} > 0$. Тогда

$$\beta_n(c, d) \xrightarrow{n \rightarrow \infty} \infty \quad (10.14)$$

на множестве ω ($\underline{\lim} \xi_n(\omega) < \overline{\lim} \xi_n(\omega)$). В силу (10.13) и (10.14)

$$E\{\beta_n(c, d)\} \xrightarrow{n \rightarrow \infty} \infty. \quad (10.15)$$

С другой стороны, по неравенству Дуба

$$E\{\beta_n(c, d)\} \leq \frac{E\{\xi_n^+\} + |c|}{d - c} \leq \frac{\sup E\{\xi_n^+\} + c}{d - c} < \infty.$$

Получили противоречие с (10.13). \square

Следствие 10.3. Если (ξ_n, \mathcal{F}_n) – субмартингал и $\xi_n \leq 0$, то существует $\lim_{n \rightarrow \infty} \xi_n$ почти наверное.

Доказательство. Следует из того, что $E\{\xi_n^+\} = 0$. \square

Следствие 10.4. Если (ξ_n, \mathcal{F}_n) – супермартингал и $\xi_n \geq 0$, то существует $\lim_{n \rightarrow \infty} \xi_n$ почти наверное.

Доказательство. Если (ξ_n, \mathcal{F}_n) – супермартингал, то $(-\xi_n, \mathcal{F}_n)$ – субмартингал и $E\{-\xi_n^+\} = 0$. \square

Замечание 10.2. Для мартингалов условие $\sup_n E\{\xi_n^+\} < \infty$ эквивалентно условию $\sup_n E\{|\xi_n|\} < \infty$.

Доказательство. Из равенства $|\xi_n| = \xi_n^+ + \xi_n^- = \xi_n^+ + \xi_n^+ + \xi_n^- - \xi_n^+ = 2\xi_n^+ - \xi_n$ следует

$$E\{|\xi_n|\} = 2E\{\xi_n^+\} - E\{\xi_n\} = 2E\{\xi_n^+\} - E\{\xi_1\}, \sup_n E\{|\xi_n|\} = 2 \sup_n E\{\xi_n^+\} - E\{\xi_1\}.$$

\square

Определение 10.5. Мартингал (ξ_n, \mathcal{F}_n) называется регулярным, если существует случайная величина η такая, что $E\{|\eta|\} < \infty$ и $\xi_n = E\{\eta | \mathcal{F}_n\}$.

Теорема 10.12. Следующие условия на мартингал (ξ_n, \mathcal{F}_n) эквивалентны:

1. (ξ_n, \mathcal{F}_n) – регулярный мартингал.
 2. Семейство $\{\xi_n\}$ – равномерно интегрируемо.
 3. $\xi_n \xrightarrow{L_1} \xi$ (сходится в L_1).
 4. $\xi_n \xrightarrow{п. н.} \xi_\infty$, причем $\xi_n = E\{\xi_\infty | \mathcal{F}_n\}$.
- Доказательство.** Покажем, что верно 1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 4) \Rightarrow 1).
1. Покажем, что из 1) следует 2). Тогда

Пусть (ξ_n, \mathcal{F}_n) – регулярный мартингал. Тогда

$$\xi_n = E\{\eta | \mathcal{F}_n\}, E\{|\eta|\} < \infty.$$

Надо показать, что

$$\sup_n \int_{(|\xi_n| > l)} |\xi_n| dP \xrightarrow{l \rightarrow \infty} 0 :$$

$$\begin{aligned}
\sup_n \int_{(|\xi_n|>l)} |\xi_n| dP &= \sup_n \int_{(|\xi_n|>l)} |E\{\eta|\mathcal{F}_n\}| dP \leq \sup_n \int_{(|\xi_n|>l)} E\{|\eta||\mathcal{F}_n\} dP = \\
&= \sup_n \int_{(|\xi_n|>l)} |\eta| dP \leq \sup_n \int_{\substack{(|\xi_n|>l) \\ (|\eta|\leq b)}} |\eta| dP + \sup_n \int_{\substack{(|\xi_n|>l) \\ (|\eta|>b)}} |\eta| dP \leq \\
&\leq b \sup_n P\{|\xi_n| > l\} + \sup_n \int_{\substack{(|\xi_n|>l) \\ (|\eta|>b)}} |\eta| dP \leq \frac{b}{l} \sup_n E\{|\xi_n|\} + \sup_n \int_{(|\eta|>b)} |\eta| dP \leq \\
&\leq \frac{b}{l} \sup_n E\{E\{|\eta||\mathcal{F}_n\}\} + \int_{(|\eta|>b)} |\eta| dP \leq \frac{b}{l} E\{|\eta|\} + \int_{(|\eta|>b)} |\eta| dP. \quad (10.16)
\end{aligned}$$

По теореме Лебега для любого $\epsilon < 0$ существует b_0 такое, что $\int_{(|\eta|>b_0)} |\eta| dP \leq \epsilon/2$. Тогда

можно подобрать l_0 такое, что правая часть (10.16) будет меньше ϵ при $l \geq l_0$.

2. Покажем, что из 2) следует 3).

Пусть последовательность $\{\xi_n\}$ равномерно интегрируема. Тогда $\sup_n E\{|\xi_n|\} < \infty$. Из теоремы 10.11 и замечания 10.2 будет следовать, что $\xi_n \xrightarrow{п. н.} \eta$ (η – некоторая случайная величина). Тогда из теорем для равномерно интегрируемых случайных величин следует, что $\xi_n \xrightarrow{L_1} \xi$.

3. Покажем, что из 3) следует 4).

Имеем $\xi_n \xrightarrow{L_1} \xi$. Докажем, что $\sup_n E\{|\xi_n|\} < \infty$:

$$\sup_n E\{|\xi_n|\} = \sup_n E\{|\xi_n - \xi + \xi|\} \leq \sup_n E\{|\xi_n - \xi|\} + E\{|\xi|\}.$$

Отсюда следует, что $\sup_n E\{|\xi_n|\} < \infty$. Из теоремы 10.12 и замечания 10.2 следует, что

$\xi_n \rightarrow \xi_\infty$ и $\xi = \xi_\infty$. Теперь покажем, что $\xi_n = E\{\xi_\infty | \mathcal{F}_n\}$. Поскольку $\xi_n \xrightarrow{L_1} \xi_\infty$, то для любого $A \in \mathcal{F}_m$ $\int_A \xi_n dP \xrightarrow{n \rightarrow \infty} \int_A \xi_\infty dP$. Пусть $m < n$ и $A \in \mathcal{F}_m$. Тогда

$$\int_A E\{\xi_n | \mathcal{F}_m\} dP = \int_A \xi_m dP.$$

Получим, что

$$\int_A \xi_m dP \xrightarrow{n \rightarrow \infty} \int_A \xi_\infty dP.$$

Поскольку левая часть от n не зависит, то:

$$\int_A \xi_m dP = \int_A \xi_\infty dP = \int_A E\{\xi_\infty | \mathcal{F}_m\} dP.$$

Поскольку это верно для любого $A \in \mathcal{F}_m$, то $\xi_m = E\{\xi_\infty | \mathcal{F}_m\}$.

4. Покажем, что из 4) следует 1).

Доказательство следует из 4), если в качестве η взять ξ_∞ . □

Теорема 10.13 (теорема Леви). Пусть $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_{n-1} \subset \mathcal{F}_n \subset \dots$ и случайная величина η такова, что $E\{\eta\} < \infty$.

Тогда $E\{\eta|\mathcal{F}_n\} \xrightarrow[n \rightarrow \infty]{\text{п. н.}} E\{\eta|\mathcal{F}_\infty\}$, где \mathcal{F}_∞ – σ -алгебра, порожденная $\bigcup_{i=0}^{\infty} \mathcal{F}_i$.

Доказательство. Положим $\xi_n = E\{\eta|\mathcal{F}_n\}$, тогда $\{\xi_n, \mathcal{F}_n\}$ – регулярный мартингал по определению. Тогда из предыдущей теоремы $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п. н.}} \xi_\infty$ и $\xi_n \xrightarrow{L_1} \xi_\infty$, следовательно, для любого $A \in \mathcal{F}$, $\int_A \xi_n dP \rightarrow \int_A \xi_\infty dP$. Возьмем $A \in \mathcal{F}_m$, $m \leq n$, тогда

$$\int_A \xi_m dP = \int_A \xi_\infty dP.$$

С другой стороны,

$$\int_A \xi_m dP = \int_A E\{\eta|\mathcal{F}_m\} dP = \int_A \eta dP.$$

Значит, для любого $A \in \mathcal{F}_m$

$$\int_A \eta dP = \int_A \xi_\infty dP. \quad (10.17)$$

Это верно и для любого $A \in \bigcup_m \mathcal{F}_m$, т. е. для любого $A \in \mathcal{F}_\infty$.

Поскольку ξ_∞ – \mathcal{F}_∞ -измерима, то для $A \in \mathcal{F}_\infty$ из (10.17) получаем

$$\int_A \eta dP = \int_A E\{\eta|\mathcal{F}_\infty\} dP = \int_A \xi_\infty dP.$$

Следовательно, $E\{\eta|\mathcal{F}_\infty\} = \xi_\infty = \lim_{n \rightarrow \infty} \xi_n = \lim_{n \rightarrow \infty} E\{\eta|\mathcal{F}_n\}$. \square

Следствие 10.5 (закон «0 или 1»). Пусть СВ η_1, η_2, \dots – независимые, одинаково распределенные случайные величины, $\mathcal{F}_1^n = \sigma(\eta_1, \dots, \eta_n)$, $\mathcal{F}_n^\infty = \sigma(\eta_n, \dots)$, $\mathcal{F}^\infty = \bigcap_n \mathcal{F}_n^\infty$.

Тогда для любого $A \in \mathcal{F}^\infty$ $P(A) = \begin{cases} 1, \\ 0. \end{cases}$

Доказательство. Рассмотрим последовательность $\mathcal{F}_1^n \subset \mathcal{F}_1^{n+1} \subset \dots$, $A \in \mathcal{F}^\infty$, тогда по теореме Леви $E\{\mathbb{1}_A|\mathcal{F}_1^n\} \rightarrow E\{\mathbb{1}_A|\mathcal{F}_1^\infty\} = \mathbb{1}_A$, так как $\mathcal{F}^\infty \subset \mathcal{F}_1^\infty$, с другой стороны, так как η_1, η_2, \dots независимы, тогда для любого n A и \mathcal{F}_1^n независимы. По свойству условного математического ожидания $E\{\mathbb{1}_A|\mathcal{F}_1^n\} = E\{\mathbb{1}_A\} = P(A)$, а значит, $P(A)$ либо 0, либо 1. \square

Примеры. Пусть ξ_1, ξ_2, \dots – независимые случайные величины. Тогда:

1. $P\{\text{существования } \lim \xi_n\} = \begin{cases} 0, \\ 1. \end{cases}$
2. $P\left\{\sum_{n=1}^{\infty} \xi_n < \infty\right\} = \begin{cases} 0, \\ 1. \end{cases}$
3. Пусть $S_n = \xi_1 + \dots + \xi_n$, $E\{\xi_n\} = 0$, $D\{\xi_n\} = \sigma^2 > 0$, $S_n^0 = \frac{S_n}{\sqrt{D\{S_n\}}}$.

Тогда $P\{\text{существования } \lim S_n^0\} = \begin{cases} 0, \\ 1. \end{cases}$

10.4. СХОДИМОСТЬ РЯДОВ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Теорема 10.14. Если случайные величины $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимы, $E\{\xi_k\} = 0$, и $\sum_{k=1}^{\infty} D\{\xi_k\} < \infty$, то последовательность $\{S_n\}$, $S_n = \sum_{k=1}^n \xi_k$, сходится почти наверное.

Доказательство. Поскольку последовательность $\{S_n\}$ образует мартингал, то достаточно доказать, что

$$\sup E\{|S_n|\} < \infty. \quad (10.18)$$

В условиях теоремы

$$D\{S_n\} = \sum_{k=1}^n D\{\xi_k\} \leq \sum_{k=1}^{\infty} D\{\xi_k\} < \infty, \text{ то } \sup_n D\{S_n\} < \infty,$$

отсюда

$$\sup_n E\{|S_n|\} \leq \sqrt{\sup_n E\{S_n^2\}} \leq \sqrt{\sup_n D\{S_n\}} < \infty.$$

□

Лемма 10.2. Пусть случайные величины ξ_1, ξ_2, \dots независимы, $E\{\xi_n\} = 0$, $n = 1, 2, 3, \dots$, $\mathcal{F}_n = \sigma(\xi_1, \xi_2, \dots, \xi_n)$, τ — конечный марковский момент относительно последовательности σ -алгебр, т. е. существует N такое, что $P\{\tau \leq N\} = 1$, $(\tau = n) \in \mathcal{F}_n$, тогда $E\left\{\sum_{k=1}^{\tau} \xi_k^2\right\} = E\left\{\sum_{k=1}^{\tau} D\{\xi_k\}\right\}$.

Доказательство. Используя равенство

$$\sum_{k=1}^N \xi_k^2 \mathbb{1}_{(\tau \geq k)} + 2 \sum_{l < j=2}^N \xi_l \xi_j \mathbb{1}_{(\tau \geq j)}$$

и независимость случайных величин ξ_k^2 и $\mathbb{1}_{(\tau \geq k)}$, а также ξ_l и $\mathbb{1}_{(\tau \geq j)}$ и ξ_j , получим

$$\begin{aligned} E\left\{\left(\sum_{k=1}^{\tau} \xi_k\right)^2\right\} &= \sum_{k=1}^N D\{\xi_k\} P\{\tau \geq k\} + 2 \sum_{l=j-1}^N E\{\xi_l \mathbb{1}_{(\tau \geq j)}\} E\{\xi_j\} = \\ &= \sum_{k=1}^N D\{\xi_k\} P\{\tau \geq k\} = E\sum_{k=1}^{\tau} D\{\xi_k\}. \end{aligned}$$

□

Лемма 10.3. Если существует $a > 0$ такое, что $P\left\{\sup_n |S_n| < a\right\} > 0$, $|\xi_k| \leq C$, $E\{\xi_k\} = 0$ для любого k , то $\sum_{k=1}^{\infty} D\{\xi_k\} < \infty$.

Доказательство. Рассмотрим марковский момент τ ,

$$\tau = (k : |S_1| \leq a, \dots, |S_{k-1}| \leq a, |S_k| > a) = \min(n : |S_n| > a), \quad S_n = \sum_{k=1}^n \xi_k,$$

$\tau = \infty$, если для любого n $|S_n| \leq a$.

$\tau_N = \min(\tau, N) = \tau \wedge N$ — ограниченный марковский момент.

По лемме 10.3

$$E \left\{ \sum_{k=1}^{\tau_N} \xi_k \right\}^2 = E \left\{ \sum_{k=1}^{\tau_N} D\{\xi_k\} \right\}. \quad (10.19)$$

Поскольку $|S_{\tau_N-1}| \leq a$, $|S_{\tau_N}| \leq a + c$, то из (10.19) имеем

$$E \left\{ \sum_{k=1}^{\tau_N} D\{\xi_k\} \right\} \leq (a + c)^2.$$

Поскольку $\sum_{k=1}^{\tau_N} D\{\xi_k\}$ возрастает по N и сходится к $\sum_{k=1}^{\tau} D\{\xi_k\}$, то

$$E \left\{ \sum_{k=1}^{\tau} D\{\xi_k\} \right\} \leq (a + c)^2.$$

Отсюда получаем

$$(a + c)^2 \geq E \left\{ \sum_{k=1}^{\tau} D\{\xi_k\} \right\} \geq P\{\tau = \infty\} \sum_{k=1}^{\infty} D\{\xi_k\}.$$

Поскольку по условию $P\{\tau = \infty\} > 0$, то получаем $\sum_{k=1}^{\infty} D\{\xi_k\} < \infty$. \square

Теорема 10.15 (теорема Колмогорова о трех рядах). Пусть ξ_1, ξ_2, \dots – независимые случайные величины. Тогда для сходимости ряда $\sum_{k=1}^{\infty} \xi_k < \infty$ необходимо, чтобы при любом c сходились ряды:

$$a) \sum_{k=1}^{\infty} P\{|\xi_k| > c\} < \infty;$$

$$б) \sum_{k=1}^{\infty} E\{\xi_k^c\} < \infty;$$

$$в) \sum_{k=1}^{\infty} D\{\xi_k^c\} < \infty, \quad \text{где } \xi_k^c = \begin{cases} \xi_k, & |\xi_k| \leq c, \\ 0, & |\xi_k| > c \end{cases}, \text{ и достаточно, чтобы при некотором } c \text{ сходились а)-в).}$$

Доказательство. Необходимость. Пусть $c > 0$ и $\sum_{k=1}^{\infty} \xi_k < \infty$. Рассмотрим $A_k = (|\xi_k| > c)$. Тогда по лемме Бореля – Кантелли $P\{\overline{\lim} A_k\} = 0$. Поскольку A_k – независимы, то из закона «0 или 1» следует $\sum_{k=1}^{\infty} P\{A_k\} < \infty$.

Рассмотрим $\{\bar{\xi}_k^c\}$ – последовательность независимых СВ, не зависящих от СВ ξ_k^c ; их распределения совпадают $\bar{\xi}_k^c \stackrel{d}{=} \xi_k^c$ для любого k .

Покажем, что существует такое a , что $P\left\{ \sum_{k=1}^{\infty} (\xi_k^c - \bar{\xi}_k^c) < a \right\} > 0$.

Предположим противное, что для всех a $P\left\{\sum_{k=1}^{\infty}(\xi_k^c - \bar{\xi}_k^c) < a\right\} = 0$. Тогда

$$1 = P\left\{\sum_{k=1}^{\infty}(\xi_k^c - \bar{\xi}_k^c) < \infty\right\} = \lim_{n \rightarrow \infty} P\left\{\sum_{k=1}^{\infty}(\xi_k^c - \bar{\xi}_k^c) < n\right\} = 0.$$

Противоречие. Используя лемму 10.3, получаем

$$\sum_{k=1}^{\infty} D\{\xi_k^c - \bar{\xi}_k^c\} = 2 \sum_{k=1}^{\infty} D\{\xi_k^c\} < \infty.$$

Отсюда

$$\sum_{k=1}^{\infty} D\{\xi_k^c\} < \infty.$$

Покажем, что сходится ряд б). Из равенства

$$\sum_{k=1}^{\infty} \xi_k = \sum_{k=1}^{\infty} (\xi_k - \xi_k^c) + \sum_{k=1}^{\infty} (\xi_k^c - E\{\xi_k^c\}) + \sum_{k=1}^{\infty} E\{\xi_k^c\} \quad (10.20)$$

и теоремы 10.15 следует, что сходится ряд

$$\sum_{k=1}^{\infty} (\xi_k^c - E\{\xi_k^c\}) < \infty.$$

Из сходимости рядов $\sum_{k=1}^{\infty} \xi_k < \infty$ и $\sum_{k=1}^{\infty} (\xi_k^c - E\{\xi_k^c\}) < \infty$ следует

$$\sum_{k=1}^{\infty} E\{\xi_k^c\} < \infty.$$

Необходимость доказана.

Достаточность. Пусть при некотором c сходятся а)–в). Тогда по теореме 10.15 сходится ряд

$$\sum_{k=1}^{\infty} (\xi_k^c - E\{\xi_k^c\}) < \infty.$$

Покажем, что сходится ряд $\sum_{k=1}^{\infty} (\xi_k - \xi_k^c)$. Поскольку $\sum_{k=1}^{\infty} P\{A_k\} < \infty$, то по лемме

Бореля – Кантелли $P\{\overline{\lim} A_k\} = 0$, или $P\{\underline{\lim} \bar{A}_k\} = 1$.

Тогда из определения нижнего предела следует, что для всех $\omega \in \underline{\lim} \bar{A}_k$ следует, что существует $n = n(\omega)$ такое, что при всех $k \geq n$ $|\xi_k| < c$. Тогда $\xi_k - \xi_k^c = 0$ при всех $k \geq n(\omega)$. Отсюда следует, что с вероятностью единица ряд $\sum_{k=1}^{\infty} (\xi_k - \xi_k^c) < \infty$. Поскольку

ряды в правой части (10.20) сходятся, то $\sum_{k=1}^{\infty} \xi_k < \infty$. \square

10.5. УПРАЖНЕНИЯ

1. Пусть ξ_0, ξ_1, \dots – последовательность независимых случайных величин с нулевыми математическими ожиданиями, $S_n = \xi_0 + \dots + \xi_n$. Доказать, что последовательность $\{S_n\}$ образует мартингал.
2. Пусть ξ_0, ξ_1, \dots – последовательность независимых случайных величин, $E\{\xi_n\} = 1$, $n = 0, 1, \dots$, $X_n = \prod_{i=0}^n \xi_i$. Доказать, что последовательность $\{X_n\}$ образует мартингал.
3. Пусть ξ – случайная величина с конечным математическим ожиданием, $\{\mathcal{F}_n\}_{n \geq 0}$ – неубывающая последовательность σ -алгебр. Положим $\xi_n = E\{\xi | \mathcal{F}_n\}$. Доказать, что последовательность $\{\xi_n, \mathcal{F}_n\}$ образует мартингал.
4. Пусть $\{\xi_n\}$ – мартингал, $E\{\xi_n^2\} < \infty$. Доказать, что $\{\xi_n^2\}$ – субмартингал.
5. Пусть $\{\xi_n\}$ – последовательность неотрицательных случайных величин, имеющих конечные математические ожидания $S_n = \xi_0 + \dots + \xi_n$. Доказать, что последовательность $\{S_n\}$ образует субмартингал.
6. Пусть $\{X_n, \mathcal{F}_n\}$ – мартингал, а $g(x)$ – выпуклая функция такая, что $E\{g(X_n)\} < \infty$, $n = 0, 1, \dots$. Доказать, что последовательность $\{g(X_n), \mathcal{F}_n\}$ образует субмартингал.
7. Пусть $\{\xi_n\}$ и $\{\eta_n\}$ – две последовательности случайных величин такие, что при каждом n существуют совместная плотность распределения случайных величин $\xi_1, \dots, \xi_n - f_n(x_1, \dots, x_n)$ и совместная плотность распределения случайных величин $\eta_1, \dots, \eta_n - g_n(x_1, \dots, x_n)$. Доказать, что последовательность

$$\zeta_n = \frac{g_n(\xi_1, \dots, \xi_n)}{f_n(\xi_1, \dots, \xi_n)},$$

образует мартингал.

8. Пусть τ и σ – марковские моменты. Доказать, что событие $(\tau < \sigma) \in \mathcal{F}_\tau \cap \mathcal{F}_\sigma$.
9. Пусть $\{\xi_n\}_{n \geq 0}$ – последовательность независимых, одинаково распределенных случайных величин, $P\{\xi_i = 0\} = P\{\xi_i = 2\} = 1/2$, $X_n = \prod_{i=1}^n \xi_i$. Показать, что не существует такой интегрируемой случайной величины ξ и неубывающего семейства σ -алгебр $\{\mathcal{F}_n\}$, что $X_n = E\{\xi | \mathcal{F}_n\}$.

ЧАСТЬ 3

МАТЕМАТИЧЕСКИЕ МЕТОДЫ
СТАТИСТИКИ

Глава 11

ТЕОРИЯ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Пусть проводится некоторый случайный эксперимент [9, 18, 30, 36], в котором измеряется $N \geq 1$ числовых величин $\xi_1, \dots, \xi_N \in \mathbb{R}^1$. Математической моделью такого случайного эксперимента является вероятностное пространство (Ω, \mathcal{F}, P) , на котором определен случайный N -вектор $\xi = (\xi_1, \dots, \xi_N)^T \in \mathbb{R}^N$ (« T » – символ транспонирования), имеющий некоторую N -мерную функцию распределения вероятностей

$$F(z; \theta) ::= P\{\xi_1 < z_1, \dots, \xi_N < z_N\}, \quad z = (z_1, \dots, z_N)^T \in \mathbb{R}^N,$$

где $\theta = (\theta_1, \dots, \theta_m)^T \in \Theta \subseteq \mathbb{R}^m$ – некоторый набор параметров (векторный параметр) распределения вероятностей.

Пусть осуществлено n независимых случайных экспериментов и зарегистрированы результаты: $x_1, x_2, \dots, x_n \in \mathbb{R}^N$ – первого, второго, \dots , n -го эксперимента. Построенное таким образом множество результатов экспериментов $X = \{x_1, \dots, x_n\}$ называется *случайной выборкой* объема n из N -мерного распределения вероятностей $F(\cdot; \theta)$. Число экспериментов n называется *объемом выборки*, результат i -го эксперимента $x_i = (x_{i1}, \dots, x_{iN})^T \in \mathbb{R}^N$ – i -м *выборочным значением*.

Далее выборку X иногда будем представлять в виде составного вектора: $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$. Математической моделью выборки в таком представлении является случайный nN -вектор $X \in \mathbb{R}^{nN}$, определенный на составном вероятностном пространстве $(\Omega^n, \mathcal{F}^n, P^n)$ и имеющий nN -мерную функцию распределения

$$\Psi(X; \theta) = \prod_{i=1}^n F(x_i; \theta), \quad X \in \mathbb{R}^{nN}, \quad \theta \in \Theta.$$

Отметим два важных свойства случайной выборки: независимость выборочных значений и их одинаковую распределенность.

Любое измеримое по Борелю функциональное преобразование выборки X :

$$t = t(X) = (t_1(X), \dots, t_K(X))^T \in \mathbb{R}^K, \quad X \in \mathbb{R}^{nN},$$

не зависящее от функции распределения $F(\cdot; \theta)$ или от ее параметра θ , называется *статистикой*.

В частности, при тождественном преобразовании $t(X) = X$ – выборка X также является статистикой.

Выводы (решения, оценки, суждения), основанные на статистиках $t = t(X)$, называются *статистическими выводами*.

Предмет математической статистики – это всевозможные статистические выводы. Математическая статистика – раздел математики, изучающий методы построения оптимальных по заданным критериям статистических выводов и анализа их свойств. *Математическая статистика* разделяется на два подраздела [8, 16]: теорию статистического оценивания параметров и теорию статистической проверки гипотез.

Пример 11.1. Пусть имеется некоторая коммерческая фирма, которая производит и реализует некоторый товар. Допустим, в установившемся режиме суточный доход этой фирмы $x \in \mathbb{R}^1$ является случайной величиной с некоторой функцией распределения вероятностей $F(z; \theta_0)$, $z \in \mathbb{R}^1$ ($N = 1$). В фирму обратилось рекламное агентство, которое предлагает увеличить доход фирмы θ_0 до $\theta > \theta_0$ за определенную плату.

Возникает вопрос: стоит ли это делать? Менеджер фирмы решает провести n пробных дней, в течение которых дается реклама и каждый день регистрируется объем продаж. В результате имеем выборку $X = \{x_1, \dots, x_n\}$, по которой решается задача оценивания среднего дохода θ , что относится к теории статистического оценивания параметров. Необходимо также проверить гипотезу $\theta > \theta_0$ об увеличении прибыли, что, в свою очередь, относится к теории статистической проверки гипотез.

11.1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Пусть наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с функцией распределения

$$F(z; \theta^0), \quad z \in \mathbb{R}^N, \quad \theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \Theta \subseteq \mathbb{R}^m.$$

Пространство \mathbb{R}^{nN} часто называют *выборочным пространством*: каждая точка в нем – выборка; Θ – *параметрическое пространство*. Предполагается, что функция распределения $F(\cdot; \theta^0)$ задана с точностью до вектора параметров $\theta^0 \in \Theta$. Задача статистического оценивания параметров заключается в том, чтобы по выборке X найти такое значение $\hat{\theta} = T(X) \in \Theta$, которое в определенном вероятностном смысле близко к θ^0 . Таким образом, возникает задача построения функционального преобразования $\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \Theta$, которое будем предполагать борелевским. Оценка $\hat{\theta} = T(X)$ задается статистикой $T(X)$ и поэтому называется *статистической оценкой*:

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_m \end{pmatrix} = T(X) = \begin{pmatrix} T_1(X) \\ \vdots \\ T_m(X) \end{pmatrix} \in \Theta \subseteq \mathbb{R}^m, \quad X \in \mathbb{R}^{nN},$$

где $\hat{\theta}_j = T_j(X)$, $j \in \{1, \dots, m\}$, – статистическая оценка j -го параметра.

Пример 11.2. Пусть $X = \{x_1, \dots, x_n\}$ – случайная выборка из одномерного ($N = 1$) нормального распределения $N_1(\theta_1^0, \theta_2^0)$:

$$F(z; \theta^0) = \Phi\left(\frac{z - \theta_1^0}{\sqrt{\theta_2^0}}\right), \quad z \in \mathbb{R}^1, \quad \theta^0 = (\theta_1^0, \theta_2^0)^T.$$

Требуется оценить два параметра: математическое ожидание θ_1^0 и дисперсию θ_2^0 .

Замечание 11.1. В теории вероятностей исходят из того, что модель точно задана, т. е. $F(\cdot; \theta^0)$ известна. В математической статистике вероятностная модель известна с точностью до значения вектора параметров θ^0 или задано семейство допустимых вероятностных моделей.

Замечание 11.2. В теории вероятностей принято случайные величины и их значения обозначать разными символами. Например, ξ – случайная величина, x – ее значение. В математической статистике и случайные величины, и их значения обозначаются одними и теми же символами.

Графически процесс статистического оценивания иллюстрируется на рис. 11.1.

Вероятностный смысл близости оценки $\hat{\theta} = T(X)$ к истинному значению θ^0 определяется следующими тремя требованиями, предъявляемыми к статистике $T(\cdot)$.

I. Состоятельность статистической оценки

Определение 11.1. *Статистическая оценка*

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \Theta$$

называется состоятельной, если, каково бы ни было истинное значение вектора параметров $\theta^0 \in \Theta$, при $n \rightarrow \infty$ имеет место сходимость оценки к истинному значению по вероятности:

$$\forall \theta^0 \in \Theta \quad \hat{\theta} P \theta^0, \quad n \rightarrow +\infty, \quad (11.1)$$

что эквивалентно соотношениям

$$\forall \theta^0 \in \Theta \quad \hat{\theta}_i P \theta_i^0, \quad i = 1, \dots, m, \quad n \rightarrow +\infty, \quad (11.2)$$

или

$$\forall \theta^0 \in \Theta, \quad \forall \epsilon > 0 \quad P_{\theta^0}\{|\hat{\theta} - \theta^0| > \epsilon\} \rightarrow 0, \quad n \rightarrow +\infty,$$

где

$$P_{\theta^0}(A) = \int_A d\Psi(X; \theta^0), \quad \Psi(X; \theta^0) = \prod_{i=1}^n F(x_i; \theta^0), \quad A \subseteq \mathcal{B}^{nN};$$

$P_{\theta^0}(\cdot)$ – вероятностная мера в выборочном пространстве, вычисленная при фиксированном истинном значении вектора параметров θ^0 ; \mathcal{B}^{nN} – борелевская σ -алгебра подмножеств из \mathbb{R}^{nN} [30].

Определение 11.2. Если в соотношениях (11.1) и (11.2) имеет место сходимость почти наверное, то говорят о сильной («строгой») состоятельности оценки:

$$\forall \theta^0 \in \Theta \quad \hat{\theta} \xrightarrow{\text{п. н.}} \theta^0, \quad n \rightarrow +\infty.$$

Разумеется, из сильной состоятельности следует состоятельность по вероятности.

II. Несмещенность статистической оценки

Определение 11.3. Смещением статистической оценки $\hat{\theta} = T(X)$ называется уклонение математического ожидания этой оценки от истинного значения вектора параметров θ^0 :

$$b(n; \theta^0) = (b_1, \dots, b_m)^T ::= E_{\theta^0}\{\hat{\theta}\} - \theta^0 = \int_{\mathbb{R}^{nN}} (T(X) - \theta^0) d\Psi(X; \theta^0).$$

Определение 11.4. Если, каковы бы ни были истинное значение θ^0 и объем выборки n , смещение равно нулю:

$$\forall n \geq 1, \quad \forall \theta^0 \in \Theta \quad b(n; \theta^0) = \mathbf{0}_m,$$

то оценка $\hat{\theta} = T(X)$ называется несмещенной. В противном случае – смещенной.

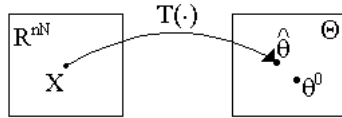


Рис. 11.1. Схема статистического оценивания

Определение 11.5. Если

$$\forall \theta^0 \in \Theta \quad \lim_{n \rightarrow \infty} b(n; \theta^0) = \mathbf{0}_m,$$

то оценка $\hat{\theta} = T(X)$ называется асимптотически несмещенной.

Несмещенность означает отсутствие систематической погрешности при оценивании параметров.

III. Точность статистической оценки

Определение 11.6. Матрицей вариаций статистической оценки $\hat{\theta} = T(X)$ называется $(m \times m)$ -матрица, определяемая следующим матричным соотношением:

$$V = V\{\hat{\theta}\} = (v_{jk})_{j,k=1}^m = \mathbf{E}_{\theta^0}\{(\hat{\theta} - \theta^0)(\hat{\theta} - \theta^0)^T\},$$

или покомпонентно $(j, k = 1, \dots, m)$:

$$v_{jk} = \mathbf{E}_{\theta^0}\{(\hat{\theta}_j - \theta_j^0)(\hat{\theta}_k - \theta_k^0)\} = \int_{\mathbb{R}^{nN}} (T_j(X) - \theta_j^0)(T_k(X) - \theta_k^0) d\Psi(X; \theta^0).$$

Свойства матрицы вариаций

- C1. Матрица вариаций симметрична: $V = V^T$.
- C2. Диагональный элемент $v_{jj} = \mathbf{E}_{\theta^0}\{(\hat{\theta}_j - \theta_j^0)^2\} \geq 0$, $j \in \{1, \dots, m\}$, – среднеквадратическая ошибка оценивания j -го параметра.
- C3. Пусть $\Sigma = \text{cov}\{\hat{\theta}, \hat{\theta}\} = \mathbf{E}_{\theta^0}\{(\hat{\theta} - \mathbf{E}_{\theta^0}\{\hat{\theta}\})(\hat{\theta} - \mathbf{E}_{\theta^0}\{\hat{\theta}\})^T\}$ – ковариационная матрица для $\hat{\theta}$. Тогда справедливо соотношение $V = \Sigma + b(n; \theta^0)b^T(n; \theta^0)$, где $b(n; \theta^0)$ – смещение оценки $\hat{\theta}$.
- C4. Если $\hat{\theta}$ – несмещенная оценка, то $V = \Sigma$.
- C5. Если $\text{tr}(V) \rightarrow 0$, то $\hat{\theta}$ – состоятельная оценка.

В качестве меры точности оценки $\hat{\theta}$ используются характеристики $\{v_{jj}, j = 1, \dots, m\}$, $\text{tr}(V)$, $|V|$. Чем меньше их значения, тем точнее оценка $\hat{\theta}$.

11.2. ВЫБОРОЧНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, ВЫБОРОЧНАЯ ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ, ВЫБОРОЧНЫЕ МОМЕНТЫ, ГИСТОГРАММА И ИХ СВОЙСТВА

Рассмотрим некоторые важнейшие элементарные статистики $T = T(X)$, используемые в дальнейшем для оценивания. Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ объема n из некоторого, для простоты, одномерного распределения вероятностей с функцией распределения $F(x) = \mathbf{P}\{x_j < x\}$, $x \in \mathbb{R}$ ($j \in \{1, \dots, n\}$).

Пусть определена характеристическая функция

$$g(t) = \mathbf{E}\{e^{itx_j}\} = \int_{-\infty}^{+\infty} e^{itx} dF(x), \quad t \in \mathbb{R}, \quad (11.3)$$

где i – мнимая единица, и для некоторого $k \geq 1$ существует начальный момент k -го порядка:

$$\alpha_k = \mathbf{E}\{x_j^k\} = \int_{-\infty}^{+\infty} x^k dF(x). \quad (11.4)$$

Примем обозначения:

$$I(z) = \begin{cases} 1, & z > 0; \\ 0, & z \leq 0; \end{cases} \quad - \text{единичная функция Хевисайда};$$

$$I_A(z) = \begin{cases} 1, & z \in A; \\ 0, & z \notin A; \end{cases} \quad - \text{индикатор множества } A.$$

Если $F(x)$ – абсолютно непрерывная функция распределения, то соответствующую ей плотность распределения обозначаем $f(x) = \frac{d}{dx}F(x)$.

Рассмотрим задачу построения по выборке X статистических оценок для $F(\cdot)$, $g(\cdot)$, $\{\alpha_k\}$ и $f(\cdot)$.

Определение 11.7. Выборочной функцией распределения (эмпирической функцией распределения), построенной по выборке X объема n , называется статистика

$$\hat{F}(x) ::= F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(x - x_j) = \frac{L_n(x)}{n}, \quad x \in \mathbb{R}, \quad (11.5)$$

где $L_n(x) = \sum_{j=1}^n \mathbb{I}(x - x_j)$ – число выборочных значений из $\{x_j\}_{j=1}^n$, для которых $x_j < x$.

Для построения графика выборочной функции распределения воспользуемся вариационным рядом выборки X : $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, где $x_{(j)}$ – j -я порядковая статистика, $j \in \{1, \dots, n\}$ (изучаемая в п. 11.3). График выборочной функции распределения представлен на рис. 11.2.

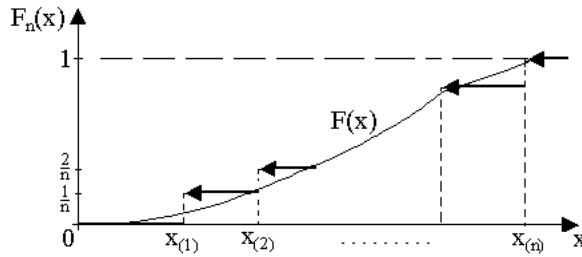


Рис. 11.2. Выборочная функция распределения

Определение 11.8. Выборочной характеристической функцией для выборки X называется комплекснозначная функция действительной переменной t :

$$\hat{g}(t) ::= g_n(t) = \frac{1}{n} \sum_{j=1}^n e^{itx_j}, \quad t \in \mathbb{R}. \quad (11.6)$$

Определение 11.9. Выборочным моментом k -го порядка для выборки X называется статистика

$$\hat{\alpha}_k ::= a_k = \frac{1}{n} \sum_{j=1}^n x_j^k. \quad (11.7)$$

Формально статистики (11.6) и (11.7) получаются подстановкой (11.5) в (11.3) и (11.4) с учетом того, что $\frac{d}{dz}I(z) ::= \delta(z)$ – обобщенная δ -функция Дирака, для которой

$$\int_{\mathbb{R}} G(z-b)\delta(z)dz = G(b), \quad b \in \mathbb{R},$$

где $G(\cdot)$ – произвольная функция, определенная на \mathbb{R} . Это проявление так называемого *подстановочного принципа* («plug-in») в статистике.

Теорема 11.1. *Выборочная функция распределения, выборочная характеристическая функция и выборочный момент k -го порядка являются сильно состоятельными, несмещенными оценками соответственно для $F(\cdot)$, $g(\cdot)$ и α_k . Причем для вариаций оценок справедливы следующие соотношения:*

$$V\{\hat{F}(x)\} = \mathbf{E}\{(\hat{F}(x) - F(x))^2\} = \frac{F(x)(1-F(x))}{n}, \quad V\{\hat{\alpha}_k\} = \frac{\alpha_{2k} - \alpha_k^2}{n}.$$

Доказательство. Для каждой из трех статистик доказательство проводится одинаково. Рассмотрим первую из статистик и по (11.5) представим ее в эквивалентном виде:

$$\hat{F}(x) = \frac{1}{n} \sum_{j=1}^n \xi_j, \quad \xi_j = \mathbb{I}(x - x_j),$$

где $\{\xi_j\}$ – случайные величины Бернулли, независимые в совокупности и одинаково распределенные, так как x – фиксированная величина; $\{x_j\}$ – выборочные значения, независимые в совокупности и одинаково распределенные с функцией распределения $F(\cdot)$. Здесь учтено, что независимость сохраняется при функциональном преобразовании. Вероятность «успеха»: $p = \mathbf{P}\{\xi_j = 1\} = \mathbf{E}\{\xi_j\} = \mathbf{P}\{x_j < x\} = F(x)$. Выполняются условия усиленного закона больших чисел [30], следовательно,

$$\frac{1}{n} \sum_{j=1}^n \xi_j \xrightarrow{\text{п. н.}} p = F(x), \quad n \rightarrow +\infty.$$

Далее $\mathbf{E}\{\hat{F}(x)\} = p = F(x)$, т. е. оценка является несмещенной. Вариация оценки

$$V\{\hat{F}(x)\} = D\{\hat{F}(x)\} = \frac{1}{n^2} \cdot n \cdot D\{\xi_1\} = \frac{p(1-p)}{n} = \frac{F(x)(1-F(x))}{n}.$$

□

Задание 11.1. Найти вариацию $V\{\hat{g}(t)\}$.

Теорема 11.2. *Выборочная функция распределения $\hat{F}(x)$ и выборочный момент α_k при $n \rightarrow +\infty$ являются асимптотически нормально распределенными статистиками, причем справедливы соотношения:*

1) если $0 < F(x) < 1$, то

$$\mathbf{P}\left\{\sqrt{n} \frac{\hat{F}(x) - F(x)}{\sqrt{F(x)(1-F(x))}} < z\right\} \rightarrow \Phi(z), \quad n \rightarrow +\infty, \quad z \in \mathbb{R},$$

где $\Phi(\cdot)$ – функция распределения вероятностей стандартного нормального закона $N_1(0, 1)$;

2) если момент порядка $2k$ ограничен: $\alpha_{2k} < +\infty$, и $\alpha_{2k} - \alpha_k^2 > 0$, то

$$P\left\{\sqrt{n}\frac{a_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} < z\right\} \rightarrow \Phi(z), \quad n \rightarrow +\infty, \quad z \in \mathbb{R}.$$

Доказательство. Первое утверждение основано на теореме Муавра – Лапласа, второе – на центральной предельной теореме (ЦПТ) Леви – Линдеберга (см. п. 5.8). \square

Задание 11.2. Исследовать статистику $g_n(t) = \text{Re}(g_n(t)) + i \cdot \text{Im}(g_n(t))$ на асимптотическую нормальность, применив многомерный аналог ЦПТ из п. 5.8 к $(\text{Re}(g_n(t)), \text{Im}(g_n(t)))^T \in \mathbb{R}^2$.

Рассмотрим теперь задачу оценивания плотности распределения $f(x) = \frac{d}{dx}F(x)$, $x \in \mathbb{R}$, по выборке $X = (x_1, \dots, x_n)^T$. Используя принцип «plug-in», построим подстановочную оценку $\tilde{f}(x) = \frac{d}{dx}F_n(x)$. Легко видеть, что эта оценка неконструктивна:

$$\tilde{f}(x) = \frac{1}{n} \sum_{j=1}^n \delta(x - x_j) = \begin{cases} 0, & x \notin \{x_1, \dots, x_n\}; \\ \text{не существует,} & x \in \{x_1, \dots, x_n\}. \end{cases}$$

Укажем другой способ статистического оценивания плотности $f(x)$.

Полагаем, что носителем плотности $f(\cdot)$ является $[x_-, x_+]$, т. е. плотность распределения $f(x)$ сосредоточена на отрезке $[x_-, x_+]$, вне его она равна нулю. Зададим некоторое натуральное число $K \geq 2$ и осуществим разбиение $[x_-, x_+]$ на K частей точками деления: $x_- = b_0 < b_1 < \dots < b_K = x_+$; k -й ячейкой гистограммы назовем числовой промежуток $\Gamma_k = [b_{k-1}, b_k)$, $k = 1, \dots, K$. Введем следующие обозначения: $\Delta_k = \mu(\Gamma_k) = b_k - b_{k-1}$ – «размер» (мера Лебега) k -й ячейки, $\nu_k = \sum_{j=1}^n \mathbb{I}_{\Gamma_k}(x_j)$ – число выборочных значений,

попавших в k -ю ячейку.

Определение 11.10. Гистограммой, или гистограммной оценкой плотности $f(x)$, называется статистика

$$\hat{f}(x) = \sum_{k=1}^K \frac{\nu_k}{n\Delta_k} \mathbb{I}_{\Gamma_k}(x), \quad x \in \mathbb{R}.$$

График гистограммы с наложенным графиком истинной плотности распределения приведен на рис. 11.3.

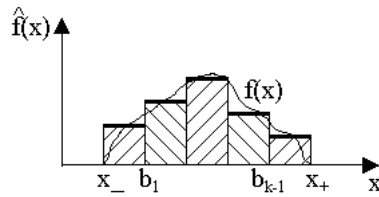


Рис. 11.3. Гистограмма

Заметим, что гистограмма, как и плотность распределения, удовлетворяет условию нормировки

$$\int_{-\infty}^{\infty} \hat{f}(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Отметим также, что гистограмма есть, вообще говоря, смещенная и несостоятельная оценка плотности распределения. Для ее состоятельности необходимо, чтобы при $n \rightarrow +\infty$ ячейки «измельчались» специальным образом:

$$K = K_n \rightarrow +\infty, \quad \max_{1 \leq k \leq K_n} \Delta_k \rightarrow 0.$$

11.3. ПОРЯДКОВЫЕ СТАТИСТИКИ, ИХ СВОЙСТВА И ПРИМЕНЕНИЕ В ТЕОРИИ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ

Пусть $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ — наблюдаемая случайная выборка объема n из некоторого абсолютно непрерывного распределения вероятностей с функцией распределения $F(x)$ и плотностью распределения $f(x) = \frac{d}{dx}F(x)$, $x \in \mathbb{R}$. Упорядочивая выборочные значения по возрастанию, построим *вариационный ряд*:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

где $x_{(j)}$ — j -е по величине выборочное значение, которое называется *j -й порядковой статистикой*. В частности, $x_{(1)} = \min_{1 \leq k \leq n} x_k$, $x_{(n)} = \max_{1 \leq k \leq n} x_k$ — минимальное и максимальное выборочные значения.

Найдем распределение вероятностей j -й порядковой статистики $\mathcal{L}\{x_{(j)}\}$. Введем в рассмотрение неполную бета-функцию:

$$B(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^y t^{a-1}(1-t)^{b-1} dt; \quad (11.8)$$

$$0 \leq y \leq 1, \quad a \geq 1, \quad b \geq 1; \quad B(1; a, b) \equiv 1.$$

Неполная бета-функция (11.8) связана с биномиальным распределением вероятностей:

$$B(p; a, n-a+1) = \sum_{k=a}^n C_n^k p^k (1-p)^{n-k}. \quad (11.9)$$

Теорема 11.3. Пусть $x_{(j)}$ — j -я порядковая статистика ($j \in \{1, \dots, n\}$), построенная по случайной выборке X объема n из некоторого распределения вероятностей с абсолютно непрерывной функцией распределения $F(x)$ и плотностью распределения $f(x) = \frac{d}{dx}F(x)$, тогда функция распределения и плотность распределения этой порядковой статистики задаются следующими соотношениями:

$$F_{x_{(j)}}(x) = B(F(x); j, n-j+1), \quad x \in \mathbb{R}; \quad (11.10)$$

$$p_{x_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} (F(x))^{j-1} (1-F(x))^{n-j} f(x). \quad (11.11)$$

Доказательство. Действуем по определению:

$$F_{x_{(j)}}(x) = P\{x_{(j)} < x\} = P\{L_n(x) \geq j\}, \quad (11.12)$$

где

$$L_n(x) = \sum_{i=1}^n \mathbb{I}(x - x_i) = \sum_{i=1}^n \xi_i. \quad (11.13)$$

В (11.13) $\xi_i = \mathbb{I}(x - x_i)$, $i = 1, \dots, n$, – независимые в совокупности, одинаково распределенные случайные величины Бернулли с вероятностью «успеха» $p = \mathbf{P}\{\xi_i = 1\} = \mathbf{P}\{x_i < x\} = F(x)$. Следовательно, $L_n(x)$ – сумма n случайных величин Бернулли – распределена по биномиальному закону $\mathcal{L}\{L_n(x)\} = Bi(n, F(x))$:

$$\mathbf{P}\{L_n(x) = k\} = C_n^k (F(x))^k (1 - F(x))^{n-k}, \quad k = 0, 1, \dots, n.$$

Далее по (11.12) и (11.9) имеем

$$\begin{aligned} F_{x_{(j)}}(x) &= \sum_{k=j}^n \mathbf{P}\{L_n(x) = k\} = \sum_{k=j}^n C_n^k (F(x))^k (1 - F(x))^{n-k} = \\ &= B(F(x); j, n - j + 1), \end{aligned}$$

что совпадает с (11.10).

Соотношение (11.11) получаем дифференцированием (11.10) с учетом (11.8). \square

Следствие 11.1. *Распределение вероятностей первой порядковой статистики $\mathcal{L}\{x_{(1)}\}$ задается соотношениями:*

$$F_{x_{(1)}}(x) = 1 - (1 - F(x))^n, \quad p_{x_{(1)}}(x) = n(1 - F(x))^{n-1} f(x).$$

Доказательство. Полагая в (11.10) и (11.11) $j = 1$ и используя условие нормировки для биномиального распределения, получаем требуемые результаты. \square

Следствие 11.2. *Распределение вероятностей n -й порядковой статистики $\mathcal{L}\{x_{(n)}\}$ задается соотношениями:*

$$F_{x_{(n)}}(x) = (F(x))^n, \quad p_{x_{(n)}}(x) = n(F(x))^{n-1} f(x).$$

Теорема 11.4 (о распределении r порядковых статистик). *Совместная плотность распределения вероятностей $r \geq 2$ порядковых статистик*

$$x_{(j_1)}, x_{(j_2)}, \dots, x_{(j_r)}, \quad 1 \leq j_1 < j_2 < \dots < j_r \leq n,$$

имеет вид

$$\begin{aligned} & p_{x_{(j_1)}, \dots, x_{(j_r)}}(z_1, \dots, z_r) = \\ &= \frac{n!}{(j_1 - 1)!(j_2 - j_1 - 1)! \dots (j_r - j_{r-1} - 1)!(n - j_r)!} \times \\ & \times F^{j_1-1}(z_1)(F(z_2) - F(z_1))^{j_2-j_1-1} \cdot \dots \cdot (1 - F(z_r))^{n-j_r} f(z_1) \cdot \dots \cdot f(z_r), \\ & -\infty < z_1 < z_2 < \dots < z_r < \infty. \end{aligned}$$

Следствие 11.3. *Совместная плотность распределения вероятностей всех n порядковых статистик имеет вид*

$$p_{x_{(1)}, \dots, x_{(n)}}(z_1, \dots, z_n) = n! f(z_1) \cdot \dots \cdot f(z_n), \quad z_1 < \dots < z_n.$$

Задание 11.3. Доказать следствие 11.2, теорему 11.4 и следствие 11.3.

Напомним понятие p -квантили, введенное в п. 3.11.

Определение 11.11. *Для некоторого p ($0 < p < 1$) p -квантилью функции распределения $F(\cdot)$ называется наименьший корень уравнения $F(\zeta) = p$, обозначаемый через ζ_p ($\zeta_p \in \mathbb{R}$).*

Определение 11.12. Выборочной p -квантилью, вычисленной по выборке X , называется статистика

$$Z_p = Z_p(X) = \begin{cases} x_{(k)}, & np = k - \text{целое}, \\ x_{(k+1)}, & k < np < k + 1. \end{cases} \quad (11.14)$$

Отметим, что из (11.14), (11.5) следует $F_n(Z_p) = p + O(1/n)$. Из данных определений и свойства состоятельности выборочной функции распределения имеем также состоятельность выборочной p -квантили: $Z_p \xrightarrow{P} t\zeta_p$, $n \rightarrow +\infty$.

Из теоремы 11.3 можно найти точный вид распределения $\mathcal{L}\{Z_p\}$. Для исследования асимптотических свойств выборочной p -квантили исследуем свойства $\mathcal{L}\{Z_p\}$ при $n \rightarrow \infty$.

Теорема 11.5 (о распределении выборочной квантили). Пусть выполнены условия теоремы 11.3, ζ_p — p -квантиль ($0 < p < 1$), плотность $f(x)$ — непрерывно дифференцируема в окрестности точки $x = \zeta_p$, причем $f(\zeta_p) > 0$, тогда выборочная p -квантиль Z_p при $n \rightarrow \infty$ распределена асимптотически нормально:

$$\mathcal{L}\left\{\sqrt{n}\frac{Z_p - \zeta_p}{\sqrt{p(1-p)}}f(\zeta_p)\right\} \rightarrow N_1(0, 1). \quad (11.15)$$

Доказательство. Введем в рассмотрение нормированную p -квантиль:

$$\eta_n = \sqrt{n}\frac{Z_p - \zeta_p}{\sqrt{p(1-p)}}f(\zeta_p), \quad (11.16)$$

тогда утверждение (11.15) можно переписать в эквивалентном виде:

$$F_{\eta_n}(y) = P\{\eta_n < y\} \rightarrow \Phi(y), \quad y \in \mathbb{R}.$$

Рассмотрим в (11.14) случай

$$np = k, \quad \frac{k}{n} = p. \quad (11.17)$$

Согласно (11.14) в этом случае

$$Z_p = x_{(k)}. \quad (11.18)$$

Заметим, что случай, когда $k < np < k + 1$ и, следовательно, $Z_p = x_{(k+1)}$, рассматривается аналогично.

Предполагая (11.17), используя (11.18) и (11.16), получаем

$$F_{\eta_n}(y) = P\left\{\sqrt{n}\frac{x_{(k)} - \zeta_p}{\sqrt{p(1-p)}}f(\zeta_p) < y\right\} = P\{x_{(k)} < y_n\}, \quad y \in \mathbb{R},$$

где

$$y_n = \zeta_p + y\frac{\sqrt{p(1-p)}}{\sqrt{n}f(\zeta_p)}. \quad (11.19)$$

Воспользуемся очевидным свойством выборочной функции распределения $F_n(\cdot)$:

$$\begin{aligned} F_{\eta_n}(y) &= P\{x_{(k)} < y_n\} = P\left\{F_n(y_n) \geq \frac{k}{n}\right\} = P\{F_n(y_n) \geq p\} = \\ &= P\left\{\sqrt{n}\frac{F_n(y_n) - F(y_n)}{\sqrt{F(y_n)(1-F(y_n))}} \geq \sqrt{n}\frac{p - F(y_n)}{\sqrt{F(y_n)(1-F(y_n))}}\right\}. \end{aligned} \quad (11.20)$$

Исследуем асимптотическое поведение ($n \rightarrow +\infty$) знаменателя в (11.20). Из (11.19) имеем

$$y_n = \zeta_p + O\left(\frac{1}{\sqrt{n}}\right), \quad y_n \rightarrow \zeta_p, \quad F(y_n) \rightarrow F(\zeta_p) = F(F^{-1}(p)) = p,$$

поэтому

$$\sqrt{F(y_n)(1-F(y_n))} \rightarrow \sqrt{p(1-p)}.$$

Проанализируем числитель в (11.20), пользуясь линейной формулой Тейлора:

$$\begin{aligned} \sqrt{n}(p - F(y_n)) &= \sqrt{n}\left(p - \left(p + f(\zeta_p)(y_n - \zeta_p) + O\left(\frac{1}{n}\right)\right)\right) = \\ &= -\sqrt{n}f(\zeta_p)\frac{y\sqrt{p(1-p)}}{\sqrt{n}f(\zeta_p)} + O\left(\frac{1}{\sqrt{n}}\right) = -y\sqrt{p(1-p)} + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Таким образом, переходя к пределу, из (11.20) в силу асимптотической нормальности $F_n(\cdot)$ получаем

$$F_n(y) \rightarrow 1 - \Phi\left(-\frac{y\sqrt{p(1-p)}}{\sqrt{p(1-p)}}\right) = \Phi(y), \quad n \rightarrow +\infty.$$

□

Определение 11.13. Если при $n \rightarrow \infty$ для некоторой статистической оценки $T_n = T_n(X)$ некоторого параметра θ выполняется предельное соотношение:

$$P\left\{\frac{T_n - \theta}{\sqrt{v_n}} < y\right\} \rightarrow \Phi(y), \quad y \in \mathbb{R},$$

где $v_n \rightarrow 0$, то принято говорить, что T_n является асимптотически нормальной оценкой параметра θ , а v_n называется асимптотической дисперсией этой оценки.

Следствие 11.4. Выборочная p -квантиль $Z_p = Z_p(X)$, определенная в (11.14), является состоятельной и асимптотически нормальной оценкой квантили $\zeta_p = F^{-1}(p)$, причем ее асимптотическая дисперсия

$$v_n = \frac{p(1-p)}{nf^2(\zeta_p)} \rightarrow 0, \quad n \rightarrow +\infty.$$

При $p = 1/2$ квантиль $m := \zeta_{1/2} = F^{-1}(1/2)$ называется медианой распределения вероятностей, а $Z_{1/2}$ — соответственно выборочной медианой.

Пользуясь теоремой 11.5, можно найти асимптотическое распределение для выборочной медианы, учитывая, что

$$\hat{m} := Z_{1/2} = \begin{cases} x_{(k)}, & n = 2k, \\ x_{(k+1)}, & n = 2k + 1. \end{cases} \quad (11.21)$$

Если $n = 2k$ четно, то существует еще одно определение выборочной медианы:

$$\hat{m} := \tilde{Z}_{1/2} = \frac{x_{(k)} + x_{(k+1)}}{2}. \quad (11.22)$$

При $n \rightarrow \infty$ свойства выборочных медиан (11.21) и (11.22) эквивалентны.

Следствие 11.5. Если $f(m) > 0$, то $Z_{1/2}$, $\tilde{Z}_{1/2}$ являются состоятельными, асимптотически несмещенными и асимптотически нормальными оценками истинного значения медианы $m = \zeta_{1/2}$ с асимптотической дисперсией

$$v_n = \frac{1}{4nf^2(\zeta_{1/2})}. \quad (11.23)$$

Следствие 11.6. Если $X = \{x_1, \dots, x_n\}$ – случайная выборка объема n из гауссовского распределения вероятностей $N_1(\theta, \sigma^2)$, то медиана совпадает с математическим ожиданием: $m = \theta$, а выборочная медиана $\hat{m} = Z_{1/2}$ является состоятельной и асимптотически нормальной оценкой истинного значения $m = \theta$ с асимптотической дисперсией

$$v_n = \frac{\pi \sigma^2}{2n}.$$

Доказательство. Этот результат непосредственно следует из п. 11.23 при подстановке $f(\theta) = 1/\sqrt{2\pi}\sigma$. \square

Задание 11.4. Сравнить по асимптотической дисперсии две статистические оценки параметра θ по выборке X объема n из $N_1(\theta, \sigma^2)$: выборочное среднее $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$ и выборочную медиану $\hat{m} = Z_{1/2}$.

Замечание 11.3. Выборочная медиана $\hat{m} = Z_{1/2}$ считается более предпочтительной, чем выборочное среднее \bar{x} , оценкой математического ожидания $\theta = E\{x_k\}$ по отношению к распределениям вероятностей с «хвостами», более «тяжелыми», чем у нормального распределения. Так, для распределения Лапласа с плотностью $f(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}}{\sigma}|x - \theta|\right)$, убывающей на «хвостах» (при $|x| \rightarrow +\infty$) «значительно» медленнее, чем плотность $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \theta)^2}{2\sigma^2}\right)$ нормального распределения $N_1(\theta, \sigma^2)$ с теми же значениями математического ожидания θ и дисперсии σ^2 , выборочная медиана оказывается оценкой максимального правдоподобия для θ (п. 11.7), а выборочное среднее является таковой в случае распределения $N_1(\theta, \sigma^2)$ (см. п. 11.7). Выборочная медиана обладает свойством робастности (см. п. 11.12).

Задание 11.5. После ознакомления с п. 11.7 показать, что выборочная медиана $\hat{m} = Z_{1/2}$ является оценкой максимального правдоподобия для математического ожидания распределения Лапласа.

11.4. РАНГОВЫЕ СТАТИСТИКИ И ИХ СВОЙСТВА

Пусть $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ – случайная выборка объема n из некоторого одномерного абсолютно непрерывного распределения вероятностей с функцией распределения $F(x) = P\{x_j < x\}$, $x \in \mathbb{R}$, и плотностью распределения $f(x) = \frac{d}{dx}F(x)$.

Для выборки-вектора $X \in \mathbb{R}^n$, у которого никакие две координаты не совпадают, обозначим $r_i = r_i(X)$ – число координат, не превосходящих x_i , т. е. номер x_i в вариационном ряду $x_{(1)} < x_{(2)} < \dots < x_{(n)}$.

Обозначим еще Π_n – множество $n!$ всевозможных перестановок из n элементов множества $\{1, 2, \dots, n\}$.

Определение 11.14. Статистика $r_i = r_i(X) : \mathbb{R}^n \rightarrow \{1, 2, \dots, n\}$ называется рангом i -го выборочного значения x_i . Вектор-столбец $r = (r_1, r_2, \dots, r_n)^T \in \Pi_n$ называется вектором рангов.

Заметим, что это определение рангов корректно, если только вероятность совпадения для любой пары выборочных значений равна нулю. Это свойство для выборки X выполняется, так как по предположению $F(\cdot)$ – абсолютно непрерывная функция распределения.

Лемма 11.1. Пусть $X \in \mathbb{R}^n$ – выборка из некоторого абсолютно непрерывного распределения, $r = (r_1, \dots, r_n)^T \in \Pi_n$ – ранговый вектор, а $X_{(\cdot)} = (x_{(1)}, \dots, x_{(n)})^T$ – вектор n порядковых статистик, тогда X и статистики $\{X_{(\cdot)}, r\}$ находятся во взаимно-однозначном соответствии.

Доказательство. Если выборка X задана, то $X_{(\cdot)}$ и r находятся единственным образом. Покажем, что справедливо и обратное. Действительно, если заданы $X_{(\cdot)}, r$, то выборка $X = (x_1, \dots, x_n)^T$ однозначно восстанавливается согласно соотношению

$$x_i = x_{(r_i)}, \quad i = 1, \dots, n. \quad (11.24)$$

□

Теорема 11.6. Если выборка $X = (x_1, \dots, x_n)^T$ есть случайный n -вектор с некоторой n -мерной плотностью $q(x_1, \dots, x_n)$, то статистика $X_{(\cdot)}$ имеет плотность

$$q_{(\cdot)}(y_{(1)}, \dots, y_{(n)}) = \sum_{J=(j_1, \dots, j_n)^T \in \Pi_n} q(y_{(j_1)}, \dots, y_{(j_n)}), \quad y_{(1)} < \dots < y_{(n)}. \quad (11.25)$$

Условное распределение рангового вектора r при фиксированном векторе порядковых статистик имеет следующий вид:

$$P\{r = (j_1, \dots, j_n)^T \mid X_{(\cdot)} = (y_{(1)}, \dots, y_{(n)})^T\} = \frac{q(y_{(j_1)}, \dots, y_{(j_n)})}{q_{(\cdot)}(y_{(1)}, \dots, y_{(n)})}, \quad J = (j_1, \dots, j_n)^T \in \Pi_n, \quad y_{(1)} < \dots < y_{(n)}. \quad (11.26)$$

Доказательство. Обозначим: \mathcal{A} – σ -алгебра борелевских подмножеств, порожденных n -векторами $(y_{(1)}, \dots, y_{(n)})^T$ с упорядоченными (в порядке возрастания) компонентами. Тогда, используя лемму 11.1, по свойству кратного интеграла для любого $A \in \mathcal{A}$ имеем выражение вероятностной меры:

$$\begin{aligned} P\{X_{(\cdot)} \in A\} &= \int_{X_{(\cdot)} \in A} q(y_1, \dots, y_n) dy_1 \dots dy_n = \sum_{J \in \Pi_n} \int_{\substack{r=J \\ X_{(\cdot)} \in A}} q(y_1, \dots, y_n) dy_1 \dots dy_n = \\ &= \sum_{J \in \Pi_n} \int_A q(y_{(j_1)}, \dots, y_{(j_n)}) dy_{(j_1)} \dots dy_{(j_n)} = \int_A q_{(\cdot)}(y_{(1)}, \dots, y_{(n)}) dy_{(1)} \dots dy_{(n)}. \end{aligned}$$

Здесь учтено, что при фиксированном r преобразование (11.24) линейно и имеет единственный якобиан. Из последнего соотношения следует, что $q_{(\cdot)}(\cdot)$ вида (11.25) является плотностью распределения вероятностей $X_{(\cdot)}$.

Рассуждая аналогично, для любых $J \in \Pi_n$, $A \in \mathcal{A}$ получаем

$$\begin{aligned} P\{r = J, X_{(\cdot)} \in A\} &= \int_{r=J, X_{(\cdot)} \in A} q(y_1, \dots, y_n) dy_1 \dots dy_n = \\ &= \int_A q(y_{(j_1)}, \dots, y_{(j_n)}) dy_{(j_1)} \dots dy_{(j_n)} = \\ &= \int_A \frac{q(y_{(j_1)}, \dots, y_{(j_n)})}{q_{(\cdot)}(y_{(1)}, \dots, y_{(n)})} q_{(\cdot)}(y_{(1)}, \dots, y_{(n)}) dy_{(1)} \dots dy_{(n)}. \end{aligned}$$

Отсюда и вытекает (11.26). □

Следствие 11.7. Если $X \in \mathbb{R}^n$ – случайная выборка объема n из некоторого абсолютно непрерывного распределения вероятностей с плотностью $f(x)$, $x \in \mathbb{R}$, то ранговый вектор $r = (r_1, \dots, r_n)^T \in \Pi_n$ и вектор порядковых статистик $X_{(\cdot)} = (x_{(1)}, \dots, x_{(n)})^T$ независимы. Ранговый вектор имеет равномерное распределение:

$$P\{r = J\} = \frac{1}{n!}, \quad J \in \Pi_n, \quad (11.27)$$

а плотность распределения $X_{(\cdot)}$ имеет вид (см. следствие 11.3):

$$q_{X_{(\cdot)}}(y_{(1)}, \dots, y_{(n)}) = n! f(y_{(1)}) \cdot \dots \cdot f(y_{(n)}), \quad (y_{(1)}, \dots, y_{(n)})^T \in \mathcal{A}. \quad (11.28)$$

Доказательство. По определению случайной выборки имеем

$$q(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i), \quad x_1, \dots, x_n \in \mathbb{R}.$$

Подставляя это выражение в (11.25), (11.26) и учитывая симметричность функции $q(\cdot)$, получаем (11.27), (11.28) и независимость $r, X_{(\cdot)}$. \square

Ранговые статистики широко используются в непараметрической статистике. Например, проверка случайности выборки (независимости выборочных значений и их одинаковой распределенности) сводится к проверке факта равномерности распределения вероятностей (11.27) рангового вектора.

11.5. НЕРАВЕНСТВО ИНФОРМАЦИИ. НЕРАВЕНСТВО КРАМЕРА – РАО. ЭФФЕКТИВНЫЕ ОЦЕНКИ

Как отмечалось ранее, оптимальную статистическую оценку $\hat{\theta} = T(X)$ для параметра $\theta \in \mathbb{R}$ целесообразно строить исходя из критерия минимума вариации оценки:

$$V\{T(\cdot)\} = E\{(T(X) - \theta)^2\} \rightarrow \min_{T(\cdot)}.$$

В этом параграфе выясним, чему равен минимум вариации: $V^* = \min_{T(\cdot)} V\{T(\cdot)\}$.

Если бы $V^* = 0$, то среднеквадратическая ошибка была бы нулевой, и оценка являлась бы точной: $\hat{\theta} \stackrel{\text{п. н.}}{=} \theta$. Но этот результат, оказывается, удается достичь лишь в вырожденных случаях. Далее будет показано, что вариация оценки ограничена снизу положительным числом:

$$V^* \geq \epsilon, \quad \epsilon = \epsilon(J_n) > 0,$$

где J_n – величина, определяющая количество информации, содержащейся в выборке X объема n о параметре θ .

11.5.1. Информационная матрица Фишера и ее свойства

Сформулируем задачу статистического оценивания параметров в общем виде: пусть наблюдается выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с плотностью распределения $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta = (\theta_1, \dots, \theta_m)^T \in \Theta \subseteq \mathbb{R}^m$ – неизвестное истинное значение векторного параметра, подлежащее оцениванию по X .

Обозначим:

$$p(X; \theta) = \prod_{l=1}^n p(x_l; \theta), \quad X \in \mathbb{R}^{nN}, \quad - \quad (11.29)$$

совместная плотность распределения выборки X , если параметр равен θ ;

$$\mathcal{P} = \{p(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\} \quad -$$

параметрическое семейство плотностей;

$$E_{\theta}\{\psi(x_l)\} = \int_{\mathbb{R}^N} \psi(x) p(x; \theta) dx \quad -$$

математическое ожидание, вычисленное при истинном значении параметра θ .

Аналогично:

$$E_{\theta}\{\Psi(X)\} = \int_{\mathbb{R}^{nN}} \Psi(X) p(X; \theta) dX \quad -$$

усреднение по распределению выборки при значении параметра θ ;

$$\nabla_{\theta} p(x; \theta) = \begin{pmatrix} \frac{\partial p(x; \theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial p(x; \theta)}{\partial \theta_m} \end{pmatrix} \in \mathbb{R}^m \quad - \text{градиент функции по } \theta \text{ в точке } x;$$

$$\nabla_{\theta}^2 p(x; \theta) = \left(\frac{\partial^2 p(x; \theta)}{\partial \theta_j \partial \theta_k} \right)_{j,k=1}^m \quad - (m \times m) \text{ — матрица вторых производных.}$$

Определение 11.15. Параметрическое семейство плотностей распределения вероятностей \mathcal{P} называется регулярным, если выполнены следующие условия регулярности.

R1. $p(x; \theta)$ — дважды дифференцируема по θ , $\forall \theta \in \Theta$, $\forall x \in \mathbb{R}^N$.

R2. Существуют конечные математические ожидания:

$$E_{\theta} \left\{ \left(\frac{\partial \ln p(x_l; \theta)}{\partial \theta_j} \right)^2 \right\} < +\infty, \quad E_{\theta} \left\{ \left| \frac{\partial^2 \ln p(x_l; \theta)}{\partial \theta_j \partial \theta_k} \right| \right\} < +\infty,$$

$$j, k = 1, \dots, m, \quad \theta \in \Theta, \quad l = 1, \dots, n.$$

R3. Допускается перестановка действий ∇_{θ}^j — дифференцирования j -го порядка по θ и интегрирования по $X \in \mathbb{R}^{nN}$:

$$\nabla_{\theta}^j \int_{\mathbb{R}^{nN}} \Psi(X) p(X; \theta) dX = \int_{\mathbb{R}^{nN}} \Psi(X) \nabla_{\theta}^j p(X; \theta) dX, \quad j = 1, 2.$$

Для формулировки четвертого условия регулярности введем некоторые дополнительные понятия.

Определение 11.16. Информационной матрицей Фишера для l -го выборочного значения x_l называется $(m \times m)$ -матрица $\mathcal{I} = (i_{jk})_{j,k=1}^m$, элементы которой вычисляются по формуле

$$i_{jk} = E_{\theta} \left\{ \frac{\partial \ln p(x_l; \theta)}{\partial \theta_j} \cdot \frac{\partial \ln p(x_l; \theta)}{\partial \theta_k} \right\}, \quad j, k = 1, \dots, m,$$

или в матричном виде:

$$\mathcal{J} = \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_i; \theta) (\nabla_\theta \ln p(x_i; \theta))^T \}. \quad (11.30)$$

Определение 11.17. Информационной матрицей Фишера для случайной выборки X объема n называется $(m \times m)$ -матрица

$$\mathcal{J}_n = \mathbf{E}_\theta \{ \nabla_\theta \ln p(X; \theta) (\nabla_\theta \ln p(X; \theta))^T \}. \quad (11.31)$$

Заметим, что математические ожидания в (11.30) и (11.31) существуют в силу условия **R2**.

R4. Информационная матрица Фишера невырождена:

$$|\mathcal{J}| \neq 0 \quad (|\mathcal{J}_n| \neq 0).$$

Следствие 11.8. Если выполнены условия регулярности **R1–R4**, то справедливо тождество $\mathbf{E}_\theta \{ \nabla_\theta \ln p(X; \theta) \} \equiv \mathbf{0}_m$.

Доказательство. В силу (11.31), условия **R3** при $j = 1$ и условия нормировки имеем

$$\begin{aligned} \mathbf{E}_\theta \{ \nabla_\theta \ln p(X; \theta) \} &= \mathbf{E}_\theta \left\{ \frac{1}{p(X; \theta)} \nabla_\theta p(X; \theta) \right\} = \\ &= \int_{\mathbb{R}^{nN}} \frac{1}{p(X; \theta)} \nabla_\theta p(X; \theta) p(X; \theta) dX = \int_{\mathbb{R}^{nN}} \nabla_\theta p(X; \theta) dX = \nabla_\theta \int_{\mathbb{R}^{nN}} p(X; \theta) dX = \nabla_\theta 1 = \mathbf{0}_m. \end{aligned}$$

□

Свойства информационной матрицы Фишера

C1. Информационная матрица Фишера – симметричная матрица: $\mathcal{J} = \mathcal{J}^T$ ($\mathcal{J}_n = \mathcal{J}_n^T$).

Доказательство. Следует из определений для информационных матриц (11.30), (11.31). □

C2. Для вычисления информационной матрицы Фишера могут быть использованы следующие формулы:

$$\mathcal{J} = \mathbf{E}_\theta \{ -\nabla_\theta^2 \ln p(x_i; \theta) \}; \quad (11.32)$$

$$\mathcal{J}_n = \mathbf{E}_\theta \{ -\nabla_\theta^2 \ln p(X; \theta) \}. \quad (11.33)$$

Доказательство. Докажем (11.32) (формула (11.33) доказывается аналогично). Пусть $l \in \{1, \dots, n\}$ – произвольный номер случайного выборочного наблюдения $x_l \in \mathbb{R}^N$. Вычислим и преобразуем правую часть (11.32) с использованием условий регулярности **R1–R4** и свойств математического ожидания $\mathbf{E}_\theta \{ \cdot \}$:

$$\begin{aligned} \mathbf{E}_\theta \{ -\nabla_\theta^2 \ln p(x_l; \theta) \} &= \mathbf{E}_\theta \left\{ -\nabla_\theta \left(\frac{1}{p(x_l; \theta)} \nabla_\theta p(x_l; \theta) \right) \right\} = \\ &= \mathbf{E}_\theta \left\{ \frac{1}{(p(x_l; \theta))^2} \nabla_\theta p(x_l; \theta) (\nabla_\theta p(x_l; \theta))^T \right\} - \mathbf{E}_\theta \left\{ \frac{1}{p(x_l; \theta)} \nabla_\theta^2 p(x_l; \theta) \right\} = \\ &= \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) (\nabla_\theta \ln p(x_l; \theta))^T \} - \int_{\mathbb{R}^N} \frac{1}{p(x; \theta)} \nabla_\theta^2 p(x; \theta) p(x; \theta) dx = \\ &= \mathcal{J} - \int_{\mathbb{R}^N} \nabla_\theta^2 p(x; \theta) dx = \mathcal{J} - \nabla_\theta^2 \int_{\mathbb{R}^N} p(x; \theta) dx = \mathcal{J} - \nabla_\theta^2 1 = \mathcal{J}. \end{aligned}$$

□

С3. Информационная матрица Фишера \mathcal{J}_n для всей выборки X объема n линейно зависит от n : $\mathcal{J}_n = n\mathcal{J}$.

Доказательство. Воспользуемся соотношениями (11.33) и (11.29):

$$\mathcal{J}_n = \mathbf{E}_\theta \left\{ -\nabla_\theta^2 \ln \prod_{l=1}^n p(x_l; \theta) \right\} = \sum_{l=1}^n \mathbf{E}_\theta \{ -\nabla_\theta^2 \ln p(x_l; \theta) \} = n\mathcal{J}.$$

□

С4. Информационная матрица Фишера \mathcal{J} есть ковариационная матрица для случайного вектора $\nabla_\theta \ln p(x_l; \theta) \in \mathbb{R}^m$.

Доказательство. По определению ковариационной матрицы имеем

$$\begin{aligned} \text{cov}\{\nabla_\theta \ln p(x_l; \theta), \nabla_\theta \ln p(x_l; \theta)\} &= \\ &= \mathbf{E}_\theta \{ (\nabla_\theta \ln p(x_l; \theta) - \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) \}) \times \\ &\quad \times (\nabla_\theta \ln p(x_l; \theta) - \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) \})^T \} = \\ &= \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) (\nabla_\theta \ln p(x_l; \theta))^T \} = \mathcal{J}, \end{aligned}$$

где учтено, что в силу следствия 11.8

$$\mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) \} \equiv \mathbf{0}_m.$$

□

С5. Информационная матрица Фишера \mathcal{J} в условиях регулярности **R1–R4** есть матрица положительно определенная: $\mathcal{J} \succ 0$.

Доказательство. В силу свойства С4 \mathcal{J} – ковариационная матрица, а любая ковариационная матрица является неотрицательно определенной: $\mathcal{J} \succeq 0$. Согласно условию регулярности **R4** $|\mathcal{J}| \neq 0$, поэтому $\mathcal{J} \succ 0$. □

11.5.2. Неравенство информации и его следствия. Эффективные оценки

Прежде чем определить понятие эффективности статистической оценки, получим так называемое неравенство информации.

Теорема 11.7 (неравенство информации). Пусть $\hat{\theta} = (\hat{\theta}_k)_{k=1}^m = T(X) = (T_k(X))_{k=1}^m \in \mathbb{R}^m$ – произвольная несмещенная статистическая оценка векторного параметра $\theta = (\theta_k)_{k=1}^m \in \mathbb{R}^m$ по случайной выборке X объема n из некоторого N -мерного распределения с плотностью $p(\cdot; \theta) \in \mathcal{P}$, $V = (v_{jk})_{j,k=1}^m = \mathbf{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\}$ – матрица вариаций оценки $\hat{\theta}$, а \mathcal{J}_n – информационная матрица Фишера для всей выборки X . Тогда если выполнены условия регулярности **R1–R4** для семейства \mathcal{P} , то $(m \times m)$ -матрица $\Delta = (\Delta_{jk})_{j,k=1}^m ::= V - \mathcal{J}_n^{-1}$ неотрицательно определена:

$$\Delta = V - \mathcal{J}_n^{-1} \succeq 0. \quad (11.34)$$

Доказательство. Введем в рассмотрение два вспомогательных случайных m -вектора:

$$U_1 = T(X) - \theta = \hat{\theta} - \theta \quad (11.35)$$

случайное уклонение оценки $\hat{\theta}$;

$$U_2 = \nabla_\theta \ln p(X; \theta). \quad (11.36)$$

Вычислим моменты первого и второго порядка для этих случайных векторов. В силу несмещенности оценки $\hat{\theta}$

$$\mathbf{E}_{\theta}\{U_1\} = \mathbf{0}_m, \quad (11.37)$$

а из следствия 11.8 имеем

$$\mathbf{E}_{\theta}\{U_2\} = \mathbf{0}_m. \quad (11.38)$$

Вычислим ковариации для векторов (11.35) и (11.36):

$$A_{ij} = \text{cov}\{U_i, U_j\} = \mathbf{E}_{\theta}\{U_i U_j^T\}, \quad i, j = 1, 2.$$

Согласно (11.35), (11.36) имеем

$$\begin{aligned} A_{11} &= V, \quad A_{22} = \mathcal{J}_n, \quad A_{21} = A_{12}^T, \\ A_{12} &= \mathbf{E}_{\theta}\{(T(X) - \theta)U_2^T\} = \mathbf{E}_{\theta}\{T(X)U_2^T\} - \theta \mathbf{E}_{\theta}\{U_2^T\} = \\ &= \mathbf{E}_{\theta}\{T(X)(\nabla_{\theta} \ln p(X; \theta))^T\} = \mathbf{E}_{\theta}\left\{T(X) \frac{1}{p(X; \theta)} (\nabla_{\theta} p(X; \theta))^T\right\} = \\ &= \int_{\mathbb{R}^{nN}} p(X; \theta) T(X) \frac{1}{p(X; \theta)} (\nabla_{\theta} p(X; \theta))^T dx = \\ &= \nabla_{\theta} \int_{\mathbb{R}^{nN}} T(X) p(X; \theta) dX = \nabla_{\theta} \theta = I_m, \end{aligned}$$

где I_m – единичная $(m \times m)$ -матрица.

При помощи линейного преобразования над случайными векторами U_1, U_2 построим случайный вектор

$$\Xi = (\xi_k)_{k=1}^m = U_1 - \mathcal{J}_n^{-1} U_2 \in \mathbb{R}^m. \quad (11.39)$$

Из (11.37), (11.38) и (11.39) очевидно, что $\mathbf{E}_{\theta}\{\Xi\} = \mathbf{0}_m$.

Вычислим ковариационную матрицу для случайного вектора Ξ :

$$\begin{aligned} \text{cov}\{\Xi, \Xi\} &= \mathbf{E}_{\theta}\{\Xi \Xi^T\} = \mathbf{E}_{\theta}\left\{(U_1 - \mathcal{J}_n^{-1} U_2)(U_1 - \mathcal{J}_n^{-1} U_2)^T\right\} = \\ &= A_{11} - A_{12} \mathcal{J}_n^{-1} - \mathcal{J}_n^{-1} A_{12}^T + \mathcal{J}_n^{-1} A_{22} \mathcal{J}_n^{-1} = \\ &= V - \mathcal{J}_n^{-1} - \mathcal{J}_n^{-1} + \mathcal{J}_n^{-1} \mathcal{J}_n \mathcal{J}_n^{-1} = V - \mathcal{J}_n^{-1} = \Delta. \end{aligned}$$

Таким образом, $\Delta = \text{cov}\{\Xi, \Xi\}$, а любая ковариационная матрица неотрицательно определена, следовательно, $\Delta \succeq 0$. \square

Следствие 11.9. Вариация $v_{kk} = \mathbf{E}\left\{\left(\hat{\theta}_k - \theta_k\right)^2\right\}$ оценки $\hat{\theta}_k = T_k(X)$ для параметра θ_k ограничена снизу положительным числом ($k \in \{1, \dots, m\}$):

$$v_{kk} \geq (\mathcal{J}_n^{-1})_{kk} > 0, \quad (11.40)$$

причем нижняя граница для вариации достигается тогда и только тогда, когда $\exists d_{k1}, \dots, d_{km} \in \mathbb{R}$ такие, что справедливо равенство

$$T_k(X) - \theta_k \stackrel{\text{п. н.}}{=} \sum_{j=1}^m d_{kj} \frac{\partial \ln p(X; \theta)}{\partial \theta_j}. \quad (11.41)$$

При этом если (11.41) выполняется, то

$$d_{kj} = (\mathcal{J}_n^{-1})_{kj}, \quad j = 1, \dots, m.$$

Доказательство. Из теоремы 11.7 следует:

$$\Delta_{kk} = (V - \mathcal{J}_n^{-1})_{kk} = v_{kk} - (\mathcal{J}_n^{-1})_{kk} \geq 0,$$

что приводит к (11.40). Неравенство (11.40) обращается в равенство тогда и только тогда, когда $\Delta_{kk} = \mathbf{E}\{\xi_k^2\} = 0$, а это, в свою очередь, равносильно тому, что $\xi_k = (T(X) - \theta - \mathcal{J}_n^{-1} \nabla_{\theta} \ln p(X; \theta))_k \stackrel{\text{п.н.}}{=} 0$, что и означает (11.41). \square

Следствие 11.10. *Определитель матрицы вариаций ограничен снизу положительным числом:*

$$|V| \geq \frac{1}{|\mathcal{J}_n|} > 0. \quad (11.42)$$

Доказательство. Воспользуемся (11.34): $\Delta = V - \mathcal{J}_n^{-1} \succeq 0$. Умножим Δ слева на матрицу $(\mathcal{J}_n^{1/2})^T$, а справа – на матрицу $\mathcal{J}_n^{1/2}$, где $\mathcal{J}_n^{1/2}$ – решение матричного уравнения $Y^T \mathcal{J}_n^{-1} Y = I_m$ относительно Y $\left(\mathcal{J}_n = \mathcal{J}_n^{1/2} (\mathcal{J}_n^{1/2})^T \right)$ [34, 11]. Для получившейся неотрицательно определенной матрицы определим спектр собственных значений $\{\lambda_1, \dots, \lambda_m\}$. Составим характеристическое уравнение:

$$\left| \left((\mathcal{J}_n^{1/2})^T V \mathcal{J}_n^{1/2} - I_m \right) - \lambda_j I_m \right| = 0, \quad \lambda_j \geq 0, \quad j = 1, \dots, m,$$

или, что эквивалентно:

$$\left| (\mathcal{J}_n^{1/2})^T V \mathcal{J}_n^{1/2} - \mu_j I_m \right| = 0, \quad \mu_j = 1 + \lambda_j \geq 1, \quad j = 1, \dots, m.$$

При этом $\{\mu_j\}_{j=1}^m$ – характеристические числа матрицы $(\mathcal{J}_n^{1/2})^T V \mathcal{J}_n^{1/2}$.

Воспользуемся свойством характеристических чисел матрицы и свойством определителя матрицы:

$$\left| (\mathcal{J}_n^{1/2})^T V \mathcal{J}_n^{1/2} \right| = \left| (\mathcal{J}_n^{1/2})^T \right| |V| |\mathcal{J}_n^{1/2}| = |\mathcal{J}_n^{1/2}| \left| (\mathcal{J}_n^{1/2})^T \right| |V| = |\mathcal{J}_n| |V| = \prod_{j=1}^m \mu_j \geq 1,$$

откуда и следует (11.42). \square

Определение 11.18. *Эффективностью несмещенной статистической оценки $\hat{\theta} = T(X)$ векторного параметра $\theta \in \mathbb{R}^m$ называется величина*

$$\mathcal{E}_n = \frac{\inf_{T(\cdot)} |V|}{|V|} = \frac{1}{|V| |\mathcal{J}_n|}.$$

Из (11.42) следует, что $0 \leq \mathcal{E}_n \leq 1$.

Определение 11.19. *Несмещенная оценка $\hat{\theta} = T(X)$ называется эффективной, если $\forall \theta \in \mathbb{R}^m, \forall n \geq 1$ эффективность $\mathcal{E}_n = 1$, иначе говоря, если неравенство (11.42) обращается в равенство.*

Следствие 11.11. *Если выполнены условия регулярности R1–R4 и $\hat{\theta} = T(X)$ – эффективная оценка, то ее матрица вариаций имеет вид*

$$V = \frac{1}{n} \mathcal{J}_n^{-1}.$$

Доказательство. Из хода доказательства следствия 11.10 заключаем:

$$\mu_j = 1 \Rightarrow \lambda_j = 0, \quad j = 1, \dots, m.$$

Следовательно, $\Delta = V - \mathcal{J}_n^{-1} = \mathbf{0}_m \Rightarrow V = \mathcal{J}_n^{-1}$. Отсюда, воспользовавшись свойством информационной матрицы Фишера: $\mathcal{J}_n = n\mathcal{J}$, получаем доказываемое. \square

Следствие 11.12 (критерий эффективности оценки). Для того чтобы в условиях регулярности **R1–R4** несмещенная оценка $\hat{\theta} = T(X)$ была эффективной, необходимо и достаточно, чтобы существовала такая положительно определенная матрица $\bar{D} \succ 0$, для которой выполняется соотношение

$$\nabla_{\theta} \ln p(X; \theta) \stackrel{n. n.}{=} \bar{D}(T(X) - \theta). \quad (11.43)$$

При этом если (11.43) выполняется, то $\bar{D} = J_n$ – информационная матрица Фишера для всей выборки.

Доказательство. Из определения эффективной оценки и следствия 11.9, а именно из системы соотношений (11.41), имеем

$$T(X) - \theta \stackrel{n. n.}{=} J_n^{-1} \nabla_{\theta} \ln p(X; \theta).$$

Умножая левую и правую части на J_n , получаем (11.43). \square

Следствие 11.13. В условиях регулярности **R1–R4** эффективная оценка существует.

Доказательство. В силу следствия 11.11:

$$V = \frac{1}{n} J_n^{-1} \rightarrow 0, \quad n \rightarrow +\infty \Rightarrow \hat{\theta} P \theta.$$

\square

11.5.3. Неравенство Крамера – Рао

Неравенство Крамера – Рао – частный случай неравенства информации, когда оценивается одномерный (скалярный) параметр $\theta \in \Theta \subseteq \mathbb{R}^1$, т. е. $m = 1$.

Полагая $m = 1$ в первых двух пунктах данного параграфа, упростим ранее введенные обозначения:

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^1 \quad -$$

некоторая несмещенная статистическая оценка параметра θ ;

$$V\{\hat{\theta}\} = E_{\theta}\{(\hat{\theta} - \theta)^2\} \geq 0 \quad -$$

вариация оценки $\hat{\theta}$;

$$J = - \int_{\mathbb{R}^N} \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} p(x; \theta) dx = \int_{\mathbb{R}^N} \left(\frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 p(x; \theta) dx > 0 \quad -$$

количество информации по Фишеру о параметре θ , содержащейся в отдельном выборочном значении $x_l \in \mathbb{R}^N$ ($l = 1, \dots, n$);

$$J_n = nJ \quad -$$

количество информации по Фишеру о параметре θ , содержащейся во всей выборке X объема n .

Теорема 11.8 (неравенство Крамера – Рао). Если плотность распределения $p(x; \theta)$ удовлетворяет ранее приведенным условиям регулярности **R1–R4**, то вариация любой несмещенной статистической оценки $\hat{\theta} = T(X)$ ограничена снизу положительным числом:

$$V\{\hat{\theta}\} \geq \frac{1}{J_n} = \frac{1}{nJ} > 0. \quad (11.44)$$

Доказательство. Положим $m = 1$ в теореме 11.7 либо в следствии 11.10. \square

Замечание 11.4. В условиях регулярности при конечном объеме выборки $n < +\infty$ невозможно оценить параметр θ безошибочно, т. е.

$$V\{\hat{\theta}\} \geq \epsilon > 0, \quad \epsilon = \frac{1}{nJ}.$$

Замечание 11.5. Увеличивая n , можно достичь наперед заданной точности оценивания, поскольку $\epsilon \rightarrow 0$ при $n \rightarrow \infty$.

Следствие 11.14. Для того чтобы несмещенная оценка $\hat{\theta} = T(X) \in \mathbb{R}^1$ была эффективной, т. е. чтобы неравенство Крамера – Рао (11.44) обращалось в равенство, необходимо и достаточно, чтобы существовало такое число $\bar{D} > 0$ такое, что справедливо соотношение

$$\frac{\partial \ln p(X; \theta)}{\partial \theta} \stackrel{\text{п. н.}}{=} \bar{D}(T(X) - \theta). \quad (11.45)$$

При этом если (11.45) выполняется, то $\bar{D} = J_n = nJ$.

Доказательство. Достаточно положить $m = 1$ в следствии 11.12. \square

11.6. МЕТОД МОМЕНТОВ. ОЦЕНКИ ПО МЕТОДУ МОМЕНТОВ И ИХ СВОЙСТВА

Метод моментов – простейший из методов построения статистических оценок, предложенный в 1900 г. английским статистиком К. Пирсоном.

Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ объема n из некоторого, для простоты, одномерного распределения вероятностей с функцией распределения $F(x; \theta^0)$, $x \in \mathbb{R}^1$, $\theta^0 \in \mathbb{R}^m$, где θ^0 – истинное значение вектора параметров, которое неизвестно. Задача заключается в том, чтобы по выборке X построить статистическую оценку $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T = T(X) = (T_1(X), \dots, T_m(X))^T \in \mathbb{R}^m$.

Будем предполагать, что функция распределения $F(\cdot; \theta)$ такова, что $\forall \theta = (\theta_1, \dots, \theta_m)^T \in \mathbb{R}^m$ существуют начальные моменты этого распределения до m -го порядка включительно. Пусть α_k – начальный момент k -го порядка:

$$\alpha_k = \alpha_k(\theta_1, \dots, \theta_m) ::= E_{\theta}\{x^k\} = \int_{-\infty}^{+\infty} x^k dF(x; \theta), \quad k = 1, \dots, m. \quad (11.46)$$

Будем дополнительно предполагать, что функции (11.46) задают взаимно-однозначное и непрерывное функциональное соответствие $\{\theta_1, \dots, \theta_m\} \leftrightarrow \{\alpha_1, \dots, \alpha_m\}$. Если в (11.46) истинные значения моментов известны: $\alpha_k^0 = \alpha_k(\theta_1^0, \dots, \theta_m^0)$, $k = 1, \dots, m$, то решая систему уравнений

$$\alpha_k(\theta_1, \dots, \theta_m) = \alpha_k^0, \quad k = 1, \dots, m,$$

относительно θ , однозначно определяем $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T$.

В качестве оценок начальных моментов будем использовать выборочные моменты (11.7):

$$\hat{\alpha}_k = a_k = \frac{1}{n} \sum_{l=1}^n x_l^k, \quad k = 1, \dots, m,$$

для которых ранее в теореме 11.1 была установлена сильная состоятельность:

$$a_k \xrightarrow{\text{п. н.}} \alpha_k, \quad n \rightarrow +\infty, \quad k = 1, \dots, m. \quad (11.47)$$

Кроме того, эти оценки – несмещенные и асимптотически нормальные.

Определение 11.20 (принцип метода моментов). В качестве оценки $\hat{\theta} = T(X)$ надлежит принимать такое значение $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T$, которое является решением системы из m уравнений

$$\alpha_k(\theta_1, \dots, \theta_m) = a_k, \quad k = 1, \dots, m. \quad (11.48)$$

Построенная таким образом оценка $\hat{\theta} = T(X)$ называется оценкой по методу моментов (ОММ).

Теорема 11.9. Пусть функция распределения $F(x; \theta)$, $x \in \mathbb{R}^1, \theta \in \mathbb{R}^m$, такова, что:

- 1) существуют начальные моменты $\alpha_k = \alpha_k(\theta)$, $k = 1, \dots, m$;
- 2) функциональное отображение $\alpha = \alpha(\theta) : \mathbb{R}^m \rightarrow \mathbb{R}^m$, задаваемое формулой (11.46), является взаимно-однозначным и непрерывным.

Тогда ОММ $\hat{\theta} = T(X)$, определяемая (11.48), является состоятельной оценкой, т. е. имеет место сходимость по вероятности

$$\hat{\theta} \xrightarrow{P} \theta^0, \quad n \rightarrow +\infty.$$

Доказательство. Результат теоремы вытекает из (11.47) и известного свойства сходимости по вероятности: если $\xi_n \xrightarrow{P} \xi$, $n \rightarrow +\infty$, и $f(\cdot)$ – непрерывная функция, то $f(\xi_n) \xrightarrow{P} f(\xi)$, $n \rightarrow +\infty$. \square

Замечание 11.6. Если $\nabla_{\theta} \alpha(\theta)|_{\theta=\theta^0} \neq 0_m$, то ОММ $\hat{\theta}$ – асимптотически нормальная оценка.

Замечание 11.7. Несмещенность и эффективность оценок по методу моментов в общем случае показать не удастся. Они не обладают высокой точностью и на практике используются как начальные приближения.

Задание 11.6. Наблюдается выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ из нормального распределения $N_1(\theta^0, \sigma^2)$, где математическое ожидание $\theta^0 \in \mathbb{R}^1$ неизвестно, а дисперсия $\sigma^2 > 0$ известна. Показать, что ОММ для θ^0 имеет вид

$$\hat{\theta} = T(X) = a_1 = \bar{x} = \frac{1}{n} \sum_{l=1}^n x_l.$$

11.7. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ И ЕГО СВОЙСТВА

11.7.1. Принцип максимального правдоподобия

Метод максимального правдоподобия был предложен в 1930 г. английским статистиком Р. Фишером. Этот метод является универсальным для построения статистических оценок.

Рассмотрим задачу статистического оценивания параметров в общем виде. Пусть наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с плотностью распределения $p(x; \theta^0)$, $x \in \mathbb{R}^N$, где $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \Theta \subseteq \mathbb{R}^m$ – неизвестное истинное значение вектора параметров. Обозначим: Θ^* – замыкание области Θ .

Определение 11.21. Функцией правдоподобия Фишера называется плотность распределения выборки X , причем X считается фиксированной, а вектор параметров $\theta \in \Theta^*$ – переменной величиной:

$$L = L(\theta) ::= p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) \geq 0, \quad \theta \in \Theta^*. \quad (11.49)$$

Логарифмической функцией правдоподобия Фишера называется функция

$$l = l(\theta) ::= \ln L(\theta) = \sum_{k=1}^n \ln p(x_k; \theta), \quad \theta \in \Theta^*. \quad (11.50)$$

В случае дискретного распределения вероятностей $P(z; \theta) = P_\theta\{x_k = z\}$, $z \in Z = \{z_1, \dots, z_M\}$, функция правдоподобия строится аналогично (11.49):

$$L(\theta) = \prod_{k=1}^n P(x_k; \theta), \quad \theta \in \Theta^*. \quad (11.51)$$

Выясним содержательный смысл функции правдоподобия. Из (11.51) видно, что $L(\theta)$ – это вероятность получить выборку X в предположении, что параметр равен θ . Эта величина рассматривается как степень правдоподобия зарегистрированной выборки X при данном θ (рис. 11.4) или степень правдоподобия θ при данной X .

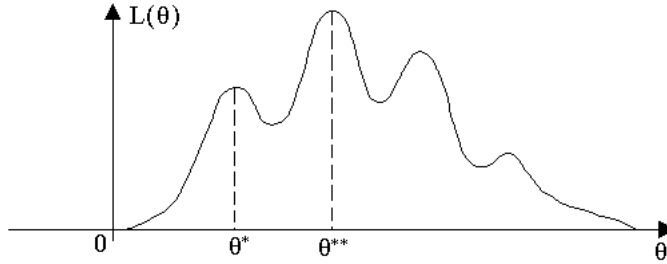


Рис. 11.4. Функция правдоподобия

Если $L(\theta^{**}) > L(\theta^*)$, то для данной выборки X значение параметра $\theta = \theta^{**}$ более правдоподобно, чем θ^* .

Пример 11.3. Менеджеру компьютерной фирмы предложили партию чипов. Поставщик утверждает, что доля годных чипов в этой партии составляет 95%, т. е. $\theta^0 = 0,95$. У менеджера есть предположение, что на самом деле $\theta^0 = 0,7$. Чтобы разрешить этот спор, решено провести следующий эксперимент: открывают контейнер, наудачу выбирают $n = 3$ коробки и проверяют чипы до обнаружения первого брака. Допустим, $x_1 = 4$, $x_2 = 6$, $x_3 = 3$ (x_k – число извлеченных небракованных чипов до первого обнаружения брака из k -й коробки). Что же более правдоподобно: предположение поставщика или менеджера?

Решение. Извлечение чипов – схема независимых испытаний Бернулли (см. п. 1.9). Случайная величина x_k – случайное число попыток до первой неудачи (обнаружения первого брака в данном случае) – имеет следующее распределение вероятностей (известное как геометрическое распределение):

$$P_\theta\{x_k = z\} = \theta^z(1 - \theta), \quad z = 0, 1, 2, \dots; \quad \theta \in \Theta^* = \{0,95; 0,7\}.$$

Построим функцию правдоподобия (11.51):

$$L(\theta) = \prod_{k=1}^3 (\theta^{x_k}(1 - \theta)) = (1 - \theta)^3 \theta^{13};$$

$$L(0,95) = 6,4 \cdot 10^{-5}, \quad L(0,7) = 2,6 \cdot 10^{-4}.$$

Таким образом, предположение менеджера более правдоподобно, чем утверждение поставщика.

Определение 11.22 (принцип максимума правдоподобия). В качестве статистической оценки параметра θ^0 надлежит использовать такое значение $\hat{\theta} = T(X)$, при котором функция правдоподобия (11.49) (или (11.51)) достигает максимума:

$$L(\hat{\theta}) = \max_{\theta \in \Theta^*} L(\theta), \quad \hat{\theta} = \arg \max_{\theta \in \Theta^*} L(\theta), \quad (11.52)$$

или, что эквивалентно, достигается максимальное значение логарифмической функции правдоподобия (11.50):

$$l(\hat{\theta}) = \max_{\theta \in \Theta^*} l(\theta), \quad \hat{\theta} = \arg \max_{\theta \in \Theta^*} l(\theta). \quad (11.53)$$

Построенная таким образом статистическая оценка $\hat{\theta} = T(X)$ называется оценкой максимального правдоподобия (ОМП, МП-оценкой).

11.7.2. Вычисление МП-оценки

В условиях регулярности **R1–R4** функция $l(\theta)$ дважды непрерывно дифференцируема по $\theta = (\theta_1, \dots, \theta_m)^T$, поэтому для нахождения максимума в (11.53) можно использовать необходимые и достаточные условия максимума. Необходимое условие имеет вид

$$\nabla_{\theta} l(\theta) = \mathbf{0}_m, \quad (11.54)$$

или покомпонентно:

$$\begin{cases} \frac{\partial l(\theta)}{\partial \theta_1} = 0, \\ \dots \\ \frac{\partial l(\theta)}{\partial \theta_m} = 0. \end{cases} \quad (11.55)$$

Определение 11.23. Система m уравнений (11.55) относительно m неизвестных $\theta_1, \dots, \theta_m$ называется системой уравнений правдоподобия.

Решая систему (11.55) (векторное уравнение (11.54)), находим стационарные точки, и чтобы среди них найти локальный максимум, необходимо проверить достаточное условие (отрицательную определенность матрицы вторых производных):

$$\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}} \prec 0. \quad (11.56)$$

Если локальных максимумов несколько, то глобальный максимум находится перебором значений функции $l(\cdot)$ в этих точках.

Пример 11.4 (продолжение примера 11.3). Построить оценку доли годных чипов в партии по методу максимального правдоподобия, предполагая, что $\theta \in \Theta = (0, 1)$, $\Theta^* = [0, 1]$.

Решение. Логарифмическая функция правдоподобия имеет вид

$$l(\theta) = \ln \prod_{k=1}^n (\theta^{x_k} (1 - \theta)) = n \ln(1 - \theta) + n\bar{x} \ln \theta,$$

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Воспользуемся (11.53), (11.54), (11.56), составим и решим уравнение правдоподобия ($m = 1$):

$$\nabla_{\theta} l(\theta) = -\frac{n}{1-\theta} + \frac{n\bar{x}}{\theta} = 0;$$

$$\hat{\theta} = \frac{\bar{x}}{\bar{x}+1} = \frac{\frac{13}{3}}{\frac{13}{3}+1} = \frac{13}{16} \approx 0,8.$$

Поскольку $\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}} < 0$, то $\hat{\theta}$ – оценка максимального правдоподобия.

Пример 11.5. Наблюдается случайная выборка $X = (x_1, \dots, x_n)^T$ объема n из нормального распределения $N_1(\theta^0, \sigma^2)$, где математическое ожидание θ^0 неизвестно, а дисперсия σ^2 известна. Построить МП-оценку для θ^0 и исследовать ее свойства.

Решение. Логарифмическая функция правдоподобия имеет вид

$$l(\theta) = \ln \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_k-\theta)^2}{2\sigma^2}} = n \ln \frac{1}{\sqrt{2\pi\sigma}} - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \theta)^2.$$

Составим и решим уравнение правдоподобия:

$$\nabla_{\theta} l(\theta) = \frac{2}{2\sigma^2} \sum_{k=1}^n (x_k - \theta) = 0;$$

$$\frac{1}{\sigma^2} n(\bar{x} - \theta) = 0;$$

$$\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Поскольку $\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}} = -n/\sigma^2 < 0$, то $\hat{\theta}$ – оценка максимального правдоподобия.

Таким образом, искомая МП-оценка построена. Проведем исследование ее свойств. Поскольку $\hat{\theta} = \bar{x}$ – выборочное среднее, т. е. выборочный момент первого порядка, то $\hat{\theta} \xrightarrow{п. н.} \theta^0$ при $n \rightarrow \infty$, и имеем сильную состоятельность для найденной оценки. Далее $E_{\theta^0}(\hat{\theta}) = \theta^0$, что означает ее несмещенность. Исследуем полученную оценку на эффективность, для этого воспользуемся критерием эффективности:

$$\frac{\partial \ln p(X; \theta)}{\partial \theta} = \nabla_{\theta} l(\theta) = \frac{n}{\sigma^2} (\hat{\theta} - \theta) = J_n(\hat{\theta} - \theta),$$

следовательно, $\hat{\theta}$ – эффективная оценка. Кроме того, $J_n = n/\sigma^2$ – количество информации по Фишеру во всей выборке о параметре θ . Поскольку $\hat{\theta}$ линейно зависит от гауссовских случайных величин $\{x_k\}_{k=1}^n$, то эта оценка имеет гауссовское распределение:

$$\mathcal{L}_{\theta^0}\{\hat{\theta}\} = N_1\left(\theta^0, \frac{\sigma^2}{n}\right).$$

11.7.3. Свойства МП-оценки

Приведем основные свойства МП-оценки.

С1. Инвариантность МП-оценки.

Пусть $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T$ – единственная МП-оценка для параметров $(\theta_1, \dots, \theta_m)^T$, а $(\tau_1, \dots, \tau_m)^T$ – некоторый набор m «новых» параметров, которые связаны со «старыми» параметрами некоторым взаимно-однозначным функциональным преобразованием:

$$\begin{aligned}\tau_1 &= f_1(\theta_1, \dots, \theta_m), \\ &\vdots \\ \tau_m &= f_m(\theta_1, \dots, \theta_m),\end{aligned}\tag{11.57}$$

тогда оценка максимального правдоподобия для $(\tau_1, \dots, \tau_m)^T$ также единственна и находится функциональным преобразованием оценки максимального правдоподобия «старых» параметров:

$$\begin{aligned}\hat{\tau}_1 &= f_1(\hat{\theta}_1, \dots, \hat{\theta}_m), \\ &\vdots \\ \hat{\tau}_m &= f_m(\hat{\theta}_1, \dots, \hat{\theta}_m).\end{aligned}$$

Доказательство. Это свойство доказывается с помощью замены переменных (11.57) в экстремальной задаче (11.52), (11.53). \square

С2. Связь МП-оценки с эффективной оценкой.

Пусть выполнены условия регулярности **R1–R4** и существует эффективная оценка $\hat{\theta}^* = T^*(X)$, тогда оценка максимального правдоподобия $\hat{\theta}$ совпадает с $\hat{\theta}^*$:

$$\hat{\theta} = \hat{\theta}^* = T^*(X).$$

Доказательство. Воспользуемся критерием эффективности оценки:

$$\nabla_{\theta} l(\theta) = J_n(\hat{\theta}^* - \theta).$$

С другой стороны, $\nabla_{\theta} l(\theta) = \mathbf{0}_m$ – система уравнений правдоподобия. Так как $J_n \succ 0$ и $\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}^*} \prec 0$, то единственная оценка максимального правдоподобия – это $\hat{\theta} = \hat{\theta}^*$. \square

Следствие 11.15. В условиях регулярности метод максимального правдоподобия является способом построения эффективных оценок.

С3. Сильная состоятельность МП-оценки.

Будем предполагать, что выполнены два дополнительных условия:

- а) $\Theta = \Theta^*$, т. е. параметрическое пространство – компакт;
- б) $\theta^0 \in \Theta$ – внутренняя точка параметрического пространства.

Определение 11.24. Информационной функцией Кульбака для случайного наблюдения $x_i \in \mathbb{R}^N$ с плотностью распределения $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta \in \Theta \subset \mathbb{R}^m$, называется функция 2m переменных:

$$J(\theta^0, \theta) ::= E_{\theta^0} \left\{ \ln \frac{p(x_i; \theta^0)}{p(x_i; \theta)} \right\} = \int_{\mathbb{R}^N} p(x; \theta^0) \ln \frac{p(x; \theta^0)}{p(x; \theta)} dx, \quad \theta^0, \theta \in \Theta.\tag{11.58}$$

Лемма 11.2. Справедливо неравенство

$$J(\theta^0, \theta) \geq 0, \quad \theta^0, \theta \in \Theta,\tag{11.59}$$

причем

$$J(\theta^0, \theta) = 0 \Leftrightarrow \int_{\mathbb{R}^N} |p(x; \theta) - p(x; \theta^0)| dx = 0.\tag{11.60}$$

Доказательство. Воспользуемся (11.58), неравенством Йенсена (см. п. 3.4) для выпуклой функции $y = g(x) = -\ln x$ и условием нормировки:

$$\begin{aligned} \mathcal{J}(\theta^0, \theta) &= E_{\theta^0} \left\{ -\ln \frac{p(x_l; \theta)}{p(x_l; \theta^0)} \right\} \geq -\ln E_{\theta^0} \left\{ \frac{p(x_l; \theta)}{p(x_l; \theta^0)} \right\} = \\ &= -\ln \int_{\mathbb{R}^N} p(x; \theta^0) \frac{p(x; \theta)}{p(x; \theta^0)} dx = -\ln 1 = 0, \end{aligned}$$

т. е. выполняется неравенство (11.59). В силу неравенства Йенсена

$$\mathcal{J}(\theta^0, \theta) = 0 \Leftrightarrow \frac{p(x_l; \theta)}{p(x_l; \theta^0)} \stackrel{\text{п. н.}}{=} 1,$$

что эквивалентно (11.60). □

Определение 11.25. Если

$$\int_{\mathbb{R}^N} |p(x; \theta) - p(x; \theta^0)| dx = 0 \Leftrightarrow \theta = \theta^0,$$

то значение параметра θ^0 называется идентифицируемым, в противном случае – неидентифицируемым.

Следствие 11.16. Если θ^0 – идентифицируемое значение параметра, то

$$\theta^0 = \arg \min_{\theta \in \Theta} \mathcal{J}(\theta^0, \theta). \quad (11.61)$$

Доказательство. Соотношение (11.61) следует из (11.59) и (11.60). □

Теорема 11.10 (о сильной состоятельности ОМП). Пусть θ^0 – идентифицируемое значение параметра – является внутренней точкой Θ , а функция плотности распределения вероятностей $p(x; \theta)$, $x \in \mathbb{R}^N$, непрерывна по θ и такова, что $\forall \theta \neq \theta^0$ существует $U_\theta \subset \Theta$ – окрестность точки θ , для которой

$$E_{\theta^0} \left\{ \inf_{\theta^* \in U_\theta} \ln \frac{p(x_l; \theta^0)}{p(x_l; \theta^*)} \right\} > -\infty. \quad (11.62)$$

Тогда МП-оценка $\hat{\theta}$ сильно состоятельна:

$$\hat{\theta} \xrightarrow{\text{п. н.}} \theta^0, \quad n \rightarrow \infty. \quad (11.63)$$

Доказательство. Представим оценку максимального правдоподобия в виде, эквивалентном (11.53):

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \left(\frac{1}{n} (l(\theta) - l(\theta^0)) \right) = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} (l(\theta^0) - l(\theta)) \right) = \arg \min_{\theta \in \Theta} \mathcal{J}_n(\theta^0, \theta),$$

где

$$\mathcal{J}_n(\theta^0, \theta) = \frac{1}{n} \sum_{k=1}^n \ln \frac{p(x_k; \theta^0)}{p(x_k; \theta)}.$$

Сравним этот результат с (11.61). По усиленному закону больших чисел имеем:

$$\mathcal{J}_n(\theta^0, \theta) \xrightarrow{\text{п. н.}} \mathcal{J}(\theta^0, \theta), \quad n \rightarrow \infty.$$

Из сходимости целевых функций, вообще говоря, не следует сходимость точек минимума (11.63). Для того чтобы это выполнялось, достаточно, чтобы имело место (11.62). \square

Следствие 11.17. *Оценка максимального правдоподобия состоятельна, т. е. $\hat{\theta} \xrightarrow{P} \theta^0$, $n \rightarrow \infty$.*

С4. Асимптотическая несмещенность и нормальность оценки максимального правдоподобия.

Теорема 11.11. *Пусть выполнены условия теоремы 11.10, а также условия регулярности **R1–R4**, причем условие **R1** выполняется в усиленном смысле: $p(x; \theta)$, $x \in \mathbb{R}^N$, — трижды дифференцируема по θ и существует окрестность $U_{\theta^0} \subset \Theta$ точки θ^0 такая, что*

$$\sup_{\theta \in U_{\theta^0}} E_{\theta^0} \left\{ \left| \frac{\partial^3 \ln p(x_i; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \right\} < +\infty, \quad i, j, k = 1, \dots, m.$$

Тогда при $n \rightarrow \infty$ оценка максимального правдоподобия $\hat{\theta}$ асимптотически нормально распределена с вектором математического ожидания θ^0 и ковариационной матрицей $J_n^{-1} = 1/n J^{-1}$, где $J \succ 0$ — информационная матрица Фишера для отдельного наблюдения. Другими словами:

$$\mathcal{L} \left\{ (J_n^{1/2})^T (\hat{\theta} - \theta^0) \right\} \rightarrow N_m(\mathbf{0}_m, I_m), \quad n \rightarrow +\infty,$$

где $J_n^{1/2}$ — решение матричного уравнения $Y^T J_n^{-1} Y = I_m$ по Y $\left(J_n = J_n^{1/2} (J_n^{1/2})^T \right)$.

Доказательство. Для простоты проведем доказательство в одномерном случае: $m = 1$, $\theta^0 \in \mathbb{R}^1$. Поскольку выполняются условия теоремы 11.10, то $\hat{\theta} \xrightarrow{p. n.} \theta^0$, $n \rightarrow \infty$, и, следовательно, $\exists \bar{n}$ такое, что $\forall n \geq \bar{n}$, $\hat{\theta} \in U_{\theta^0}$ с вероятностью 1. Дальнейший анализ проводим именно для таких значений $n \geq \bar{n}$. Запишем уравнение правдоподобия ($m = 1$):

$$\nabla_{\theta} l(\theta) = \frac{d}{d\theta} l(\theta) = l^{(1)}(\theta) = 0,$$

причем в силу **R1** функция $l^{(1)}(\theta)$ дважды непрерывно дифференцируема. Применим к левой части этого уравнения квадратичную формулу Тейлора в окрестности U_{θ^0} :

$$l^{(1)}(\theta^0) + (\hat{\theta} - \theta^0) l^{(2)}(\theta^0) + \frac{1}{2} (\hat{\theta} - \theta^0)^2 l^{(3)}(\bar{\theta}) = 0,$$

где $l^{(k)}(\theta)$ — k -я производная по θ , а $\bar{\theta} \in \Theta$ — «промежуточная» точка: $|\bar{\theta} - \theta^0| < |\hat{\theta} - \theta^0|$, поэтому $\bar{\theta} \xrightarrow{p. n.} \theta^0$ при $n \rightarrow \infty$.

Выразим случайное отклонение $\hat{\theta} - \theta^0$ из полученного выше уравнения следующим образом:

$$\hat{\theta} - \theta^0 = \frac{l^{(1)}(\theta^0)}{-l^{(2)}(\theta^0) - \frac{1}{2} (\hat{\theta} - \theta^0) l^{(3)}(\bar{\theta})}.$$

Умножим левую и правую части полученного соотношения на $\sqrt{n\mathcal{J}}$, а числитель и знаменатель правой части – на $1/n\mathcal{J}$:

$$\sqrt{n\mathcal{J}}(\hat{\theta} - \theta^0) = \frac{\frac{1}{\sqrt{n\mathcal{J}}}l^{(1)}(\theta^0)}{\frac{1}{n\mathcal{J}}\left(-l^{(2)}(\theta^0) - \frac{1}{2}(\hat{\theta} - \theta^0)l^{(3)}(\bar{\theta})\right)}. \quad (11.64)$$

Согласно усиленному закону больших чисел

$$-\frac{1}{n}l^{(2)}(\theta^0) = \frac{1}{n} \sum_{k=1}^n \left(-\frac{\partial^2 \ln p(x_k; \theta^0)}{(\partial \theta^0)^2} \right) \xrightarrow{\text{п. н.}} \mathcal{J}; \quad -\frac{1}{n}l^{(3)}(\bar{\theta}) \xrightarrow{\text{п. н.}} C,$$

где $|C| < +\infty$ – некоторая константа. Поскольку $\hat{\theta} - \theta^0 \xrightarrow{\text{п. н.}} 0$, то знаменатель сходится к 1 почти наверное.

Числитель в (11.64) при $n \rightarrow \infty$ распределен асимптотически нормально:

$$\mathcal{L}_{\theta^0} \left\{ \frac{1}{\sqrt{n\mathcal{J}}}l^{(1)}(\theta^0) \right\} \rightarrow N_1(0, 1),$$

в силу ЦПТ Леви – Линдберга (см. п. 5.5). \square

11.8. БАЙЕСОВСКИЙ МЕТОД СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Пусть в пространстве \mathbb{R}^N определено параметрическое семейство плотностей распределения вероятностей:

$$\mathcal{P} = \{p(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\},$$

где Θ – параметрическое пространство. Пусть $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \Theta$ – ненаблюдаемый случайный вектор параметров, определенный на вероятностном пространстве (Ω, \mathcal{F}, P) ; $\Pi(\theta)$, $\theta \in \Theta$, – априорная плотность распределения вероятностей случайного параметра θ^0 , которая предполагается известной. Наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из N -мерного распределения вероятностей с плотностью распределения $p(\cdot; \theta^0)$. Задача состоит в том, чтобы по выборке X построить статистическую оценку $\hat{\theta} = T(X)$ для θ^0 . Определим условную плотность распределения выборки при условии $\theta^0 = \theta$:

$$p(X|\theta) = p(X; \theta) = \prod_{l=1}^n p(x_l; \theta) = L(\theta).$$

Сформулированная модель статистического оценивания графически представлена на рис. 11.5.

Существуют два способа построения статистической оценки $\hat{\theta} = T(X)$.

Способ I. Применить метод максимального правдоподобия:

$$L(\theta) \rightarrow \max_{\theta \in \Theta}.$$

Недостатком этого способа является то, что он не учитывает априорную информацию о параметре θ^0 , заключенную в $\Pi(\cdot)$.

Способ II. Байесовский подход, учитывающий $\Pi(\cdot)$, который и рассматривается далее.

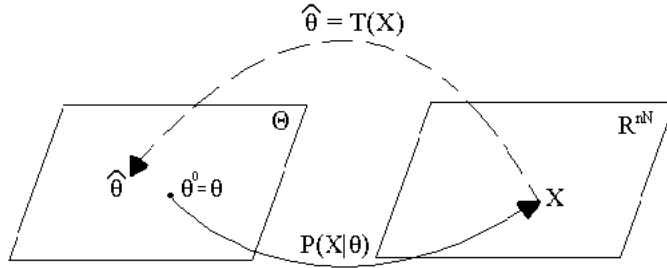


Рис. 11.5. Схема байесовского статистического оценивания

Определение 11.26. Функцией потерь (штрафа) в вышеформулированной задаче статистического оценивания параметров называется функция 2-х переменных:

$$w = w(\theta^*, \theta^{**}), \quad \theta^*, \theta^{**} \in \Theta,$$

где $w(\theta^*, \theta^{**}) \geq 0$ – величина потерь, которые несет статистик или лицо, принимающее решение, в ситуации, когда истинное значение параметра $\theta^0 = \theta^*$, а используется значение оценки $\hat{\theta} = \theta^{**}$.

Укажем две наиболее распространенные функции потерь.

Определение 11.27. Квадратичной функцией потерь называется функция

$$w = w(\theta^*, \theta^{**}) = c|\theta^* - \theta^{**}|^2, \quad c > 0. \quad (11.65)$$

Определение 11.28. (0 – 1)-функцией потерь называется функция

$$w = w(\theta^*, \theta^{**}) = 1 - \delta_{\theta^*, \theta^{**}} = \begin{cases} 0, & \theta^* = \theta^{**}, \\ 1, & \theta^* \neq \theta^{**}. \end{cases}$$

Пусть $\hat{\theta} = T(X)$ – некоторая статистическая оценка (случайная величина). Тогда ей соответствует $w = w(\theta^0, \hat{\theta})$ – случайная величина потерь.

Определение 11.29. Функционалом риска для оценки $\hat{\theta} = T(X)$ называется математическое ожидание случайной величины потерь:

$$r = r(T) = \mathbf{E} \left\{ w(\theta^0, \hat{\theta}) \right\} = \int_{\mathbb{R}^{nN}} \int_{\Theta} w(\theta, T(X)) p(X, \theta) d\theta dX \geq 0, \quad (11.66)$$

где $p(X, \theta) = p(X|\theta)\Pi(\theta) = p(X; \theta)\Pi(\theta) = L(\theta)\Pi(\theta)$ – совместная плотность распределения вероятностей выборки X и случайного параметра θ^0 .

Очевидно, что предпочтительнее та оценка $\hat{\theta} = T(X)$, для которой риск $r(T)$ меньше.

Определение 11.30 (байесовский принцип оптимальности). В качестве статистической оценки случайного параметра θ^0 надлежит использовать такую статистику $\hat{\theta} = T^*(X) : \mathbb{R}^{nN} \rightarrow \Theta$, для которой функционал риска (11.66) принимает минимальное значение:

$$r(T^*) = \min_{T(\cdot)} r(T). \quad (11.67)$$

Построенная таким образом статистическая оценка $\hat{\theta} = T^*(X)$ называется байесовской оценкой параметра θ^0 .

Теорема 11.12. Если функция потерь $w = w(\theta^*, \theta^{**})$ является квадратичной (определяется соотношением (11.65)), а априорная плотность $\Pi(\cdot)$ такова, что существует конечное априорное математическое ожидание (априорное среднее)

$$E\{\theta^0\} = \int_{\Theta} \theta \Pi(\theta) d\theta,$$

то байесовская оценка параметра θ^0 единственна и является апостериорным математическим ожиданием (или апостериорным средним):

$$\hat{\theta} = T^*(X) = E\{\theta^0|X\} = \int_{\Theta} \theta \Pi(\theta|X) d\theta, \quad (11.68)$$

где

$$\Pi(\theta|X) = \frac{p(X|\theta)\Pi(\theta)}{\int_{\Theta} p(X|\theta^*)\Pi(\theta^*)d\theta^*} \quad (11.69)$$

условная (апостериорная) плотность распределения случайного вектора θ^0 при условии наблюдения выборки X .

Доказательство. Запишем в явном виде экстремальную задачу (11.67) с учетом (11.65) и (11.66):

$$r(T) = c \int_{\mathbb{R}^{nN}} \int_{\Theta} |\theta - T(X)|^2 p(X|\theta) \Pi(\theta) d\theta dX \rightarrow \min_{T(\cdot)} \quad -$$

экстремальная задача, которая относится к задачам вариационного исчисления. Для простоты ее решение проведем при $m = 1$:

$$r(T) = c \int_{\mathbb{R}^{nN}} \int_{\Theta} (\theta - T(X))^2 p(X|\theta) \Pi(\theta) d\theta dX \rightarrow \min_{T(\cdot)}.$$

Необходимое условие минимума – это равенство нулю первой вариации функционала $r(T)$:

$$\delta r(T) = -2c \int_{\mathbb{R}^{nN}} \left(\int_{\Theta} (\theta - T(X)) p(X|\theta) \Pi(\theta) d\theta \right) \delta T(X) dX = 0,$$

где $\delta T(X)$ – произвольная вариация статистики $T(X)$. В силу этого произвола равенство нулю возможно лишь, если

$$\int_{\Theta} (\theta - T(X)) p(X|\theta) \Pi(\theta) d\theta = 0,$$

откуда и получаем (11.68), (11.69). Найденное решение соответствует минимуму, так как вторая вариация $\delta^2 r(T)|_{T(\cdot) \equiv T^*(\cdot)} > 0$. □

Теорема 11.13. В условиях теоремы 11.12 и условиях регулярности **R1–R4** байесовская оценка, определяемая (11.68), (11.69): $\hat{\theta} = T^*(X)$, при $n \rightarrow +\infty$ обладает теми же асимптотическими свойствами, что и оценка максимального правдоподобия, т. е. является сильно состоятельной, асимптотически несмещенной и асимптотически нормальной оценкой.

Замечание 11.8. Вычисления по формулам (11.68), (11.69) затруднительны из-за кратного интегрирования, часто – в бесконечных пределах.

Задание 11.7. Пусть $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ – случайная выборка объема n из $N_1(\theta^0, \sigma^2)$, где дисперсия $\sigma^2 > 0$ задана, а θ^0 – скалярная случайная величина с равномерным распределением на $[\alpha_1, \alpha_2]$, т. е. $\mathcal{L}\{\theta^0\} = R[\alpha_1, \alpha_2]$. Построить по (11.68), (11.69) байесовскую оценку для θ^0 .

11.9. ДОСТАТОЧНЫЕ СТАТИСТИКИ И ИХ СВОЙСТВА

Рассмотрим задачу статистического оценивания параметров в общем виде, акцентируя внимание на вычислительной сложности. Пусть имеется параметрическое семейство плотностей

$$\mathcal{P} = \{p(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\}$$

и наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого распределения вероятностей с плотностью $p(x; \theta^0)$, $x \in \mathbb{R}^N$, где $\theta^0 \in \Theta$ – неизвестное истинное значение вектора параметров. Необходимо построить статистическую оценку

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^m.$$

Проанализируем эту задачу с точки зрения ее вычислительной сложности. Необходимо построить m функций, каждая из которых зависит от Nn переменных. Упростим задачу следующим образом. Будем строить оценку в два этапа.

Этап 1. Сжатие данных – вырожденное функциональное преобразование выборки X :

$$t = t(X) = \begin{pmatrix} t_1(X) \\ \vdots \\ t_K(X) \end{pmatrix} : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K, \quad K < nN. \quad (11.70)$$

Этап 2. Построение оценки:

$$\hat{\theta} = \psi(t) = \begin{pmatrix} \psi_1(t) \\ \vdots \\ \psi_m(t) \end{pmatrix} : \mathbb{R}^K \rightarrow \mathbb{R}^m. \quad (11.71)$$

Статистика, которая позволяет осуществить сжатие данных $t = t(X)$ так, чтобы построенная на ее основе оценка $\hat{\theta}$ была так же точна, как и $\hat{\theta}$, называется *достаточной статистикой*.

Теперь перейдем к строгому математическому определению достаточной статистики. Для обобщения поставленной задачи статистического оценивания предположим, что θ^0 – случайный вектор с некоторой плотностью распределения вероятностей $\Pi(\theta)$, $\theta \in \Theta$. Определим условную плотность распределения вероятностей выборки:

$$p(X|\theta) = p(X; \theta) = \prod_{l=1}^n p(x_l; \theta).$$

Вычислим аналогично (11.69) по формуле Байеса апостериорную плотность распределения вероятностей случайного вектора θ^0 при условии, что наблюдается выборка X :

$$p(\theta|X) = \frac{p(X|\theta)\Pi(\theta)}{\int_{\Theta} p(X|\theta^*)\Pi(\theta^*)d\theta^*}. \quad (11.72)$$

По аналогии с (11.72) найдем условную плотность распределения θ^0 при условии наблюдения статистики $t = t(X)$:

$$q(\theta|t) = \frac{q(t|\theta)\Pi(\theta)}{\int_{\Theta} q(t|\theta^*)\Pi(\theta^*) d\theta^*}, \quad (11.73)$$

где $q(t|\theta)$ – условная плотность распределения случайной величины $t = t(X)$ при условии $\theta^0 = \theta$.

Определение 11.31. K -мерная статистика (11.70) называется достаточной статистикой для параметрического семейства \mathcal{P} , если условные плотности (11.72) и (11.73) совпадают:

$$q(\theta|t(X)) = p(\theta|X), \quad X \in \mathbb{R}^{nN}, \quad \theta \in \Theta. \quad (11.74)$$

При этом достаточная статистика с наименьшей размерностью K^* называется минимальной достаточной статистикой.

Замечание 11.9. Минимальная достаточная статистика наибольшим образом сжимает данные.

Замечание 11.10. Соотношение (11.74) означает, что статистика $t = t(X)$ несет столько же информации о параметре θ , сколько и вся выборка X .

Замечание 11.11. Если выборка X получена из дискретного распределения вероятностей, то это определение остается в силе, только вместо плотностей распределения необходимо везде использовать соответствующие вероятности.

Теорема 11.14 (критерий факторизации Неймана – Фишера). Для того чтобы статистика $t = t(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$ была достаточной статистикой для семейства плотностей распределения вероятностей \mathcal{P} , необходимо и достаточно, чтобы совместная плотность распределения выборки допускала следующую факторизацию, т. е. следующее разделение переменных ($X \in \mathbb{R}^{nN}$, $\theta \in \Theta$):

$$p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) = g(X)h(t(X); \theta), \quad (11.75)$$

где $g(X) \geq 0$ зависит только от выборки X , но не зависит от параметра θ , а $h(t(X); \theta) \geq 0$ зависит от θ , но зависимость от выборки X имеет место только через статистику $t(X)$.

Доказательство. Необходимость. Пусть $t = t(X)$ – достаточная статистика, тогда выполняется (11.74). Выразим условную плотность распределения вероятностей выборки с учетом (11.72)–(11.74):

$$p(X|\theta) = p(X; \theta) = \int_{\Theta} p(X; \theta^*)\Pi(\theta^*)d\theta^* \frac{q(t|\theta)}{\int_{\Theta} q(t|\theta^*)\Pi(\theta^*)d\theta^*}, \quad t = t(X).$$

Легко видеть, что первый сомножитель не зависит от θ , а второй может выступать в качестве $h(t(X); \theta)$, так что выполняется факторизация (11.75).

Достаточность. Пусть имеет место факторизация (11.75). Проверим выполнение (11.74). По (11.72) из (11.75) имеем

$$p(\theta|X) = \frac{g(X)h(t; \theta)\Pi(\theta)}{\int_{\Theta} g(X)h(t; \theta^*)\Pi(\theta^*)d\theta^*} = \frac{h(t; \theta)\Pi(\theta)}{\int_{\Theta} h(t; \theta^*)\Pi(\theta^*)d\theta^*}. \quad (11.76)$$

Вычислим апостериорную плотность $q(\theta|t)$, используя (11.73). Для этого сначала найдем $q(t|\theta)$. Воспользуемся методом функциональных преобразований случайных величин. Поскольку $K < nN$, то функциональное преобразование (11.70) – вырожденное. Введем вспомогательное преобразование: $\tau = \tau(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^{nN-K}$, причем функцию $\tau(\cdot)$ подберем таким образом, чтобы составное функциональное преобразование $X \rightarrow (t^T, \tau^T)^T$ было взаимно-однозначным, т. е. чтобы существовало единственное обратное преобразование $X = f(t, \tau)$ с якобианом преобразования $J(t, \tau) \neq 0$. Тогда по правилам функционального преобразования случайных величин имеем

$$q(t|\theta) = \int_{\mathbb{R}^{nN-K}} p(f(t, \tau); \theta) |J(t, \tau)| d\tau = h(t; \theta) \int_{\mathbb{R}^{nN-K}} g(f(t, \tau)) |J(t, \tau)| d\tau = h(t; \theta) B(t),$$

где

$$B(t) = \int_{\mathbb{R}^{nN-K}} g(f(t, \tau)) |J(t, \tau)| d\tau.$$

Подставляя это выражение в (11.73), получаем искомую плотность:

$$q(\theta|t) = \frac{B(t)h(t; \theta)\Pi(\theta)}{\int_{\Theta} B(t)h(t; \theta^*)\Pi(\theta^*)d\theta^*} = \frac{h(t; \theta)\Pi(\theta)}{\int_{\Theta} h(t; \theta^*)\Pi(\theta^*)d\theta^*}. \quad (11.77)$$

Сравнивая (11.76) и (11.77), видим, что они совпадают, т. е. выполняется (11.74), следовательно, $t = t(X)$ – достаточная статистика. \square

Задание 11.8. Пусть $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ – случайная выборка объема n из гауссовского распределения $N_1(\theta_1, \theta_2)$, $\theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2$. Построить достаточную статистику для θ .

Решение. Воспользуемся критерием факторизации Неймана – Фишера и выполним факторизацию (11.75).

1. Построим совместную плотность распределения выборки:

$$p(X; \theta) = \prod_{k=1}^n \left(\frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x_k - \theta_1)^2} \right).$$

2. Найдем достаточную статистику $t(X)$, осуществив разделение переменных X и θ с помощью эквивалентных преобразований:

$$\begin{aligned} p(X; \theta) &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \left(\frac{1}{\theta_2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\theta_2} \sum_{k=1}^n x_k^2 + \frac{\theta_1}{\theta_2} \sum_{k=1}^n x_k - \frac{n\theta_1^2}{2\theta_2}} = g(X)h(t(X); \theta); \\ g(X) &= \left(\frac{1}{\sqrt{2\pi}} \right)^n, \quad t(X) = (t_1(X), t_2(X))^T, \quad t_l(X) = \sum_{k=1}^n x_k^l, \quad l = 1, 2; \\ h(t; \theta) &= \left(\frac{1}{\theta_2} \right)^{\frac{n}{2}} e^{-\frac{t_2(X)}{2\theta_2} + \frac{\theta_1}{\theta_2} t_1(X) - \frac{n\theta_1^2}{2\theta_2}} \geq 0. \end{aligned}$$

Таким образом, найденная достаточная статистика является двумерной: $t(X) = (t_1(X), t_2(X))^T$ ($K = 2$).

Задание 11.9. Наблюдается выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ из $R[\theta_1, \theta_2]$. Показать, что достаточной статистикой является следующая двумерная статистика:

$$t = t(X) = (t_1(X), t_2(X))^T, \quad t_1(X) = \min_{1 \leq k \leq n} x_k = x_{(1)}, \quad t_2(X) = \max_{1 \leq k \leq n} x_k = x_{(n)}.$$

Свойства достаточной статистики

С1. Эквивалентная формулировка критерия факторизации.

Статистика $t = t(X)$ является достаточной для семейства \mathcal{P} тогда и только тогда, когда для $\forall \theta, \theta^* \in \Theta$, $\theta \neq \theta^*$, отношение правдоподобия зависит от выборки X только через статистику $t(X)$:

$$\frac{p(X; \theta)}{p(X; \theta^*)} = H(t(X); \theta, \theta^*), \quad X \in \mathbb{R}^{nN}. \quad (11.78)$$

Доказательство. Покажем, что (11.78) эквивалентно критерию факторизации (11.75). Выберем произвольное $\theta^* \in \Theta$ и зафиксируем его. Тогда из (11.78) следует:

$$p(X; \theta) = p(X; \theta^*) H(t(X); \theta, \theta^*) = g(X) h(t(X); \theta),$$

что эквивалентно (11.75). \square

С2. Свойство статистики $t = t(X)$ быть достаточной не зависит от вида априорной плотности $\Pi(\cdot)$.

Доказательство. Действительно, $\Pi(\cdot)$ не входит в критерий (11.75). \square

С3. Инвариантность достаточной статистики.

Достаточная статистика $t = t(X)$ единственна с точностью до произвольной биекции. Другими словами, если $\phi = \phi(t) : \mathbb{R}^K \rightarrow \mathbb{R}^K$ – некоторая биекция, то $\phi(t(X)) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$ также является достаточной статистикой.

Доказательство. Этот результат можно получить с помощью замены переменных в (11.75). \square

С4. Если семейство \mathcal{P} удовлетворяет условиям регулярности **R1–R4** и существует эффективная оценка $\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^m$, то статистика $T = T(X)$ является достаточной статистикой.

Доказательство. Воспользуемся критерием эффективности оценки:

$$\nabla_{\theta} \ln p(X; \theta) = J_n(\theta)(T(X) - \theta).$$

Проинтегрируем это векторное соотношение по θ . В результате получим следующее выражение для плотности:

$$p(X; \theta) = e^{A^T(\theta)T(X) + B(\theta) + C(X)}, \quad (11.79)$$

где $C(X)$ – функция выборки, появившаяся в результате интегрирования, а $A(\theta)$ и $B(\theta)$ должны удовлетворять следующим условиям:

$$\nabla_{\theta} A(\theta) = J_n(\theta), \quad \nabla_{\theta} B(\theta) = -J_n(\theta)\theta.$$

Сравнивая (11.75) и (11.79), заключаем, что в качестве $g(X)$ и $h(T(X); \theta)$ можно выбрать:

$$g(X) = e^{C(X)} \geq 0, \quad h(T(X); \theta) = e^{A^T(\theta)T(X) + B(\theta)} \geq 0.$$

Следовательно, в силу выполнения критерия факторизации $T = T(X)$ – достаточная статистика. \square

С5. Эквивалентное определение достаточной статистики.

Статистика $t = t(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$ является достаточной статистикой для семейства \mathcal{P} тогда и только тогда, когда условная плотность распределения выборки X при условии $t(X) = t^*$ не зависит от параметра θ , т. е.

$$p(X|t(X) = t^*; \theta) = \tilde{g}(X; t^*) = \text{invar}_{\theta}.$$

Задание 11.10. Доказать С5.

С6. Это свойство достаточной статистики сформулируем в виде теоремы.

Теорема 11.15 (Дармуа). Плотность распределения вероятностей $p(x; \theta)$, $x \in \mathbb{R}^N$, с m -мерным параметром $\theta \in \mathbb{R}^m$, допускающая m -мерную достаточную статистику $t = t(X) \in \mathbb{R}^m$, имеет вид

$$p(x; \theta) = \exp \left(\sum_{j=1}^m a_j(\theta) w_j(x) + b(\theta) + c(x) \right), \quad (11.80)$$

где $\{a_j(\cdot)\}, b(\cdot), c(\cdot), \{w_j(\cdot)\}$ – функции указанных выше переменных. При этом достаточная статистика имеет следующий аддитивный вид:

$$t = t(X) = (t_1(X), \dots, t_m(X))^T, \quad (11.81)$$

$$t_j(X) = \sum_{k=1}^n w_j(x_k), \quad j = 1, \dots, m.$$

Определение 11.32. Семейство плотностей вида (11.80) называется экспоненциальным семейством.

Следствие 11.18. Для экспоненциального семейства минимальная достаточная статистика является m -мерной и имеет вид (11.81).

С7. Сформулируем теорему, которая позволяет улучшать оценки с помощью достаточных статистик.

Теорема 11.16 (Колмогорова – Рао – Блекуэлла). Пусть $\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^1$ – некоторая несмещенная статистическая оценка параметра $\theta^0 \in \mathbb{R}^1$ по случайной выборке X , а $t = t(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$, $K < nN$, – достаточная статистика для семейства плотностей \mathcal{P} . Пусть далее с помощью условного математического ожидания определена борелевская функция (11.71):

$$\psi = \psi(t^*) ::= E\{T(X)|t(X) = t^*\}, \quad t^* \in \mathbb{R}^K, \quad \psi(\cdot) \in \mathbb{R}^1. \quad (11.82)$$

Тогда композиция функций

$$\tilde{\theta} = \psi(t(X)) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^1 \quad (11.83)$$

определяет несмещенную оценку параметра θ^0 , причем вариация этой оценки не превосходит вариации исходной оценки:

$$V\{\tilde{\theta}\} \leq V\{\hat{\theta}\}.$$

Доказательство. С учетом (11.82), (11.83) и определения условного математического ожидания представим оценку (11.83) в эквивалентном виде:

$$\tilde{\theta} = E\{T(X)|t(X)\}.$$

Проверим условие несмещенности с использованием формулы полного математического ожидания:

$$E\{\tilde{\theta}\} = E\{E\{T(X)|t(X)\}\} = E\{T(X)\} = E\{\hat{\theta}\} = \theta^0.$$

Сравним вариации оценок:

$$\begin{aligned} V\{\tilde{\theta}\} &= E\{(\tilde{\theta} - \theta^0)^2\} = E\{(E\{T(X) - \theta^0 | t(X)\})^2\} \leq \\ &\leq E\{E\{(T(X) - \theta^0)^2 | t(X)\}\} = E\{(T(X) - \theta^0)^2\} = V\{\hat{\theta}\}. \end{aligned}$$

Здесь использованы известное свойство из п. 1.4: $(E\{\xi\})^2 \leq E\{\xi^2\}$, и формула полного математического ожидания. \square

11.10. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И ЕГО СВОЙСТВА

11.10.1. Регрессионная модель.

Принцип метода наименьших квадратов

Метод наименьших квадратов (МНК) применяется для оценивания параметров стохастических зависимостей. Идея этого метода предложена немецким математиком К. Гауссом, а математические основы заложены русским математиком А. Марковым.

Рассмотрим следующую математическую модель. Пусть наблюдается случайная величина η , которая зависит от некоторой другой величины x . Зависимость задается в виде следующего стохастического уравнения:

$$\eta = f(x; \theta^0) + \epsilon, \quad (11.84)$$

где $f(x; \theta^0)$ – некоторая функция, вид которой известен; $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \Theta \subset \mathbb{R}^m$ – неизвестное истинное значение вектора параметров; $\eta = (\eta_1, \dots, \eta_N)^T \in \mathbb{R}^N$ – наблюдаемый случайный вектор, который часто называют вектором зависимых переменных, где η_i – i -я зависимая переменная ($i = 1, \dots, N$); $x = (x^1, \dots, x^M)^T \in \mathbb{R}^M$ – вектор-столбец M независимых переменных (предикторов, регрессоров). Относительно ненаблюдаемого случайного вектора ошибок наблюдений $\epsilon = (\epsilon_1, \dots, \epsilon_N)^T \in \mathbb{R}^N$, без потери общности, предполагаем:

$$E\{\epsilon\} = 0_N, \quad \Sigma = \text{cov}\{\epsilon, \epsilon\} = E\{\epsilon\epsilon^T\} \succ 0.$$

Определение 11.33. *Функция*

$$\bar{\eta} := E\{\eta\} = f(x; \theta^0), \quad x \in \mathbb{R}^M, \quad \theta^0 \in \Theta, \quad \bar{\eta} \in \mathbb{R}^N,$$

задающая зависимость математического ожидания $\bar{\eta} = E\{\eta\}$ от регрессоров x , называется функцией регрессии. При этом математическая модель наблюдений (11.84) называется регрессионной моделью наблюдений.

Определение 11.34. В зависимости от значений размерностей M и N существует следующая классификация регрессионных моделей.

1. Если $N > 1$, регрессионная модель (11.84) называется многомерной.
2. Если $N = 1$, $M > 1$, регрессионная модель (11.84) называется множественной.
3. Если $N = M = 1$, регрессионная модель (11.84) называется простой.

Определение 11.35. В зависимости от свойств функции $f(\cdot; \theta^0)$ существует следующая классификация регрессионных моделей: если функция $f(\cdot; \theta^0)$ линейна по параметру θ^0 , то регрессионная модель – линейная, в противном случае – нелинейная.

Пример 11.6. Рассмотрим коммерческую фирму, которая производит и продает некоторый товар. Доход фирмы за один месяц считаем случайной величиной $\eta \in \mathbb{R}^1$. Через $x \in \mathbb{R}^1$ обозначим расходы фирмы на рекламу товара за прошлый месяц. Предположим, что средний доход за месяц линейно зависит от x : $\bar{\eta} = E\{\eta\} = f(x; \theta^0) = \theta_1^0 + \theta_2^0 x$, $\theta^0 = (\theta_1^0, \theta_2^0)^T$. По приведенной классификации имеем простую линейную модель, а в результате экспериментов получаем множество точек, лежащих вблизи прямой $\bar{\eta} = \theta_1^0 + \theta_2^0 x$ (рис. 11.6).

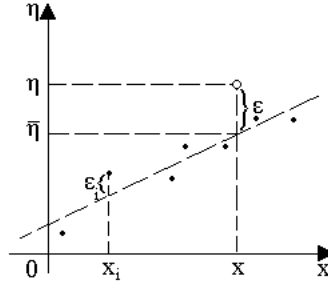


Рис. 11.6. Модель простой линейной регрессии

Сформулируем статистическую задачу. Пусть проведена серия n независимых случайных экспериментов вида (11.84). Для простоты предполагаем, что $N = 1$, т. е. рассматриваем множественную регрессионную модель; i -й случайный эксперимент заключается в следующем: фиксируется некоторое значение независимых переменных $x = x_i \in \mathbb{R}^M$, проводится эксперимент и регистрируется его результат $\eta = y_i \in \mathbb{R}^1$, причем между этими величинами имеет место соотношение вида (11.84), т. е.

$$y_i = f(x_i; \theta^0) + \epsilon_i, \quad i = 1, \dots, n. \quad (11.85)$$

Предполагается, что ошибки измерения $\{\epsilon_i\}$ независимы в совокупности и $E\{\epsilon_i\} = 0$, $D\{\epsilon_i\} = \sigma^2 < +\infty$. Задача заключается в том, чтобы по имеющимся наблюдениям $\{y_i\}$ и значениям факторов $\{x_i\}$ построить статистические оценки параметра θ^0 и дисперсии ошибок σ^2 .

Определение 11.36. Случайные эксперименты вида (11.85) называются активными, если экспериментатор имеет возможность выбирать значения факторов $x_1, \dots, x_n \in \mathbb{R}^M$ по своему усмотрению, в противном случае – пассивными.

Попытаемся решить поставленную задачу, пользуясь методом максимального правдоподобия. Для этого дополнительно предположим, что случайные ошибки наблюдения подчиняются нормальному закону распределения вероятностей:

$$\mathcal{L}\{\epsilon_i\} = N_1(0, \sigma^2), \quad i = 1, \dots, n. \quad (11.86)$$

Построим логарифмическую функцию правдоподобия. В силу формулы (11.85) и свойств нормального распределения из (11.86) имеем

$$\mathcal{L}\{y_i\} = N_1(f(x_i; \theta^0), \sigma^2), \quad i = 1, \dots, n.$$

Тогда логарифмическая функция правдоподобия

$$l(\theta) = \ln \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - f(x_i; \theta))^2}{2\sigma^2}} \right) = -\frac{1}{2}n \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} R^2(\theta), \quad (11.87)$$

где

$$R^2(\theta) = \sum_{i=1}^n (\delta y_i)^2 = \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \geq 0. \quad (11.88)$$

Определение 11.37. Величина $\delta y_i = y_i - f(x_i; \theta)$ называется невязкой (остатком) в i -м эксперименте ($i = 1, \dots, n$) при значении параметра θ , а функция $R^2(\theta)$, определяемая по (11.88), называется суммой квадратов невязок (остатков) при значении параметра θ .

Оценка максимального правдоподобия

$$\hat{\theta} = \arg \max_{\theta} l(\theta)$$

с учетом (11.87) имеет вид

$$\hat{\theta} = \arg \min_{\theta} R^2(\theta). \quad (11.89)$$

Определение 11.38 (принцип метода наименьших квадратов (МНК)).

Принцип метода наименьших квадратов состоит в том, что в качестве оценки $\hat{\theta}$ вектора параметров θ^0 принимается значение (11.89), которое минимизирует сумму квадратов остатков (11.88). Такая статистическая оценка $\hat{\theta}$ называется МНК-оценкой.

Укажем геометрический смысл МНК-оценки (рис. 11.7). В параметрическом семействе поверхностей $F = \{y = f(x; \theta), x \in \mathbb{R}^M : \theta \in \mathbb{R}^m\}$ с помощью (11.89) необходимо найти такую поверхность $y = f(x; \hat{\theta}), x \in \mathbb{R}^M$, для которой сумма квадратов отклонений (вдоль оси Y) экспериментальных точек $\{y_i\}$ от этой поверхности минимальна.

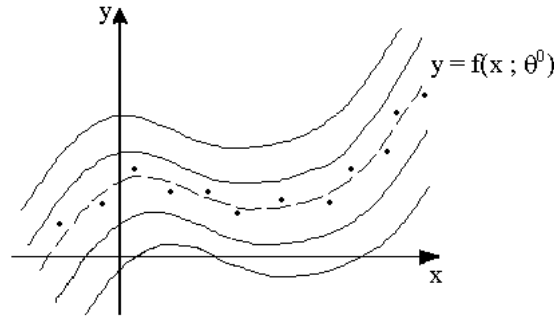


Рис. 11.7. Геометрический смысл МНК-оценки

Замечание 11.12. Соотношение (11.89) определяет МНК-оценку без ограничений, т. е. для $\theta^0 \in \mathbb{R}^m$. Если же $\theta^0 \in \Theta \subset \mathbb{R}^m$, то решается задача на условный минимум:

$$R^2(\theta) \rightarrow \min_{\theta \in \Theta}.$$

11.10.2. Оптимальные свойства МНК-оценки для линейной множественной регрессионной модели

Рассмотрим часто используемую на практике линейную множественную регрессионную модель.

Определение 11.39. Линейная множественная регрессионная модель является частным случаем модели (11.84):

$$\eta = \sum_{j=1}^m x_j \theta_j^0 + \epsilon, \quad (11.90)$$

где η – зависимая переменная; $x = (x^1, \dots, x^m)^T \in \mathbb{R}^m$ – m -вектор независимых переменных (факторов, предикторов, регрессоров), $m = M$.

Замечание 11.13. Модель (11.90) линейна не только по θ , но и без потери общности линейна по $\{x_j\}$, так как любая модель вида

$$\eta = \sum_{j=1}^m g_j(x) \theta_j^0 + \epsilon$$

приводится к (11.90) заменой:

$$x_j ::= g_j(x), \quad j = 1, \dots, m.$$

Модель наблюдений согласно (11.85), (11.90) примет вид

$$y_i = \sum_{j=1}^m x_{ij} \theta_j^0 + \epsilon_i, \quad i = 1, \dots, n, \quad (11.91)$$

где y_i – значение зависимой переменной в i -м эксперименте; x_{ij} – значение j -го фактора в i -м эксперименте ($x_i = (x_{i1}, \dots, x_{im})^T$); ϵ_i – случайная ошибка в i -м эксперименте.

Замечание 11.14. В п. 11.10.1 метод наименьших квадратов получен как частный случай метода максимального правдоподобия (в предположении гауссовости случайных ошибок $\{\epsilon_i\}$), но на практике метод наименьших квадратов используют и тогда, когда $\{\epsilon_i\}$ не являются гауссовскими случайными величинами.

С учетом замечания 11.14 в дальнейшем относительно $\{\epsilon_i\}$ предполагаем только следующее:

- 1) $\{\epsilon_i\}$ некоррелированы, т. е. $E\{\epsilon_i \epsilon_j\} = 0, i \neq j$;
- 2) $E\{\epsilon_i\} = 0$;
- 3) $D\{\epsilon_i\} = \sigma^2 < \infty$.

Удобно перейти к следующим матричным обозначениям:

$Y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ – вектор-столбец результатов наблюдений;

$\mathcal{E} = (\epsilon_1, \dots, \epsilon_n)^T \in \mathbb{R}^n$ – вектор-столбец ошибок наблюдений;

$X = (x_1 : \dots : x_n)^T$ – план эксперимента $((n \times m)$ -матрица значений факторов);

$\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \mathbb{R}^m$ – вектор-столбец оцениваемых параметров.

Тогда (11.91) можно переписать в матричном виде:

$$Y = X\theta^0 + \mathcal{E} \quad - \quad (11.92)$$

множественная линейная регрессионная модель, где относительно случайного вектора \mathcal{E} предполагается

$$E\{\mathcal{E}\} = \mathbf{0}_n, \quad \text{cov}\{\mathcal{E}, \mathcal{E}\} = \sigma^2 I_n. \quad (11.93)$$

Теорема 11.17. Пусть имеет место линейная множественная регрессионная модель (11.92), выполняются соотношения (11.93) и план эксперимента X – невырожденный:

$$|X^T X| \neq 0. \quad (11.94)$$

Тогда МНК-оценка $\hat{\theta}$ вектора параметров θ^0 единственна:

$$\hat{\theta} = C_0 Y, \quad C_0 = (X^T X)^{-1} X^T, \quad - \quad (11.95)$$

линейна по наблюдениям Y , несмещенная и имеет матрицу вариаций:

$$V = V\{\hat{\theta}\} = E_{\theta^0} \left\{ (\hat{\theta} - \theta^0) (\hat{\theta} - \theta^0)^T \right\} = \sigma^2 (X^T X)^{-1}.$$

Доказательство. Получим МНК-оценку из (11.89), (11.88) с учетом модели (11.92):

$$R^2(\theta) = (Y - X\theta)^T(Y - X\theta) = Y^TY - \theta^T X^T Y - Y^T X\theta + \theta^T X^T X\theta \rightarrow \min_{\theta}.$$

Необходимым условием минимума является

$$\nabla_{\theta} R^2(\theta) = \mathbf{0}_m. \quad (11.96)$$

Воспользуемся свойствами оператора ∇ . Для произвольной константы C , произвольного вектора b и матрицы A :

$$\nabla_{\theta}(C) = \mathbf{0}_m, \quad \nabla_{\theta}(b^T \theta) = \nabla_{\theta}(\theta^T b) = b, \quad \nabla_{\theta}(\theta^T A \theta) = 2A\theta.$$

Из (11.96) получим

$$\nabla_{\theta} R^2(\theta) = \mathbf{0}_m - 2X^T Y + 2X^T X\theta = \mathbf{0}_m.$$

Умножая на $(X^T X)^{-1}$ ($X^T X$ невырождена согласно (11.94)), имеем (11.95). Заметим, что

$$\nabla_{\theta}^2 R^2(\theta) = 2X^T X \succ 0.$$

Следовательно, найден единственный минимум.

Подставим (11.92) в (11.95):

$$\hat{\theta} = (X^T X)^{-1} X^T (X\theta^0 + \mathcal{E}) = \theta^0 + C_0 \mathcal{E}. \quad (11.97)$$

Отсюда найдем смещение оценки:

$$\mathbf{E}_{\theta^0} \{\hat{\theta} - \theta^0\} = \mathbf{E}_{\theta^0} \{C_0 \mathcal{E}\} = \mathbf{0}_m,$$

следовательно, оценка несмещенная. Аналогично вычислим матрицу вариаций:

$$V = V \left\{ \hat{\theta} \right\} = \mathbf{E}_{\theta^0} \{C_0 \mathcal{E} (C_0 \mathcal{E})^T\} = C_0 \mathbf{E}_{\theta^0} \{\mathcal{E} \mathcal{E}^T\} C_0^T = C_0 \sigma^2 I_n C_0^T = \sigma^2 (X^T X)^{-1}.$$

□

Обозначим: $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ – соответственно наименьшее и наибольшее характеристические числа неотрицательно определенной матрицы A .

Следствие 11.19. Если выполнено условие Эйкера:

$$\lambda_{\min}(X^T X) \rightarrow +\infty, \quad n \rightarrow +\infty, \quad (11.98)$$

то МНК-оценка $\hat{\theta}$, определяемая (11.95), является состоятельной оценкой:

$$\hat{\theta} \xrightarrow{P} \theta^0, \quad n \rightarrow +\infty. \quad (11.99)$$

Доказательство. Оценим следующую величину:

$$\begin{aligned} \mathbf{E}_{\theta^0} \left\{ \left| \hat{\theta} - \theta^0 \right|^2 \right\} &= \mathbf{E}_{\theta^0} \left\{ \left(\hat{\theta} - \theta^0 \right)^T \left(\hat{\theta} - \theta^0 \right) \right\} = \mathbf{E}_{\theta^0} \left\{ \text{tr} \left(\left(\hat{\theta} - \theta^0 \right) \left(\hat{\theta} - \theta^0 \right)^T \right) \right\} = \\ &= \text{tr} \left(V \left\{ \hat{\theta} \right\} \right) = \sigma^2 \text{tr} \left((X^T X)^{-1} \right) \leq \sigma^2 m \lambda_{\max} \left((X^T X)^{-1} \right). \end{aligned}$$

Известно свойство характеристических чисел матрицы [11, 34]: если выполняется (11.98), то $\lambda_{\max} \left((X^T X)^{-1} \right) \rightarrow 0$, следовательно:

$$\mathbf{E}_{\theta^0} \left\{ \left| \hat{\theta} - \theta^0 \right|^2 \right\} \rightarrow 0, \quad n \rightarrow +\infty,$$

а значит, выполняется (11.99). □

Замечание 11.15. Не при любом плане эксперимента X МНК-оценка является состоятельной.

Рассмотрим теперь задачу оценивания дисперсии σ^2 .

Определение 11.40. *Остаточной суммой квадратов называется статистика*

$$R_{\min}^2 = \sum_{i=1}^n \left(y_i - f(x_i; \hat{\theta}) \right)^2 = R^2(\hat{\theta}) \geq 0,$$

где $\hat{\theta}$ – МНК-оценка.

Учитывая линейную модель (11.92), легко получить

$$R_{\min}^2 = (Y - X\hat{\theta})^T (Y - X\hat{\theta}) \geq 0. \quad (11.100)$$

Теорема 11.18. *Если выполнены условия теоремы 11.17 и число наблюдений больше числа оцениваемых параметров: $n > m$, то несмещенной оценкой дисперсии σ^2 является статистика*

$$\hat{\sigma}^2 = \frac{1}{n-m} R_{\min}^2. \quad (11.101)$$

Доказательство. Используя теорему 11.17, имеем

$$Y - X\hat{\theta} = Y - X\theta^0 - XC_0\varepsilon = \varepsilon - XC_0\varepsilon = \left(I_n - X(X^T X)^{-1} X^T \right) \varepsilon.$$

Подставляя в (11.100), получаем

$$R_{\min}^2 = \varepsilon^T \left(I_n - X(X^T X)^{-1} X^T \right)^2 \varepsilon.$$

Установим несмещенность оценки (11.101), учитывая (11.93) и легко проверяемое свойство идемпотентности матрицы $I_n - X(X^T X)^{-1} X^T$ [11]:

$$\begin{aligned} E\{\hat{\sigma}^2\} &= \frac{1}{n-m} E\left\{ \text{tr} \left(\left(I_n - X(X^T X)^{-1} X^T \right)^2 \varepsilon \varepsilon^T \right) \right\} = \\ &= \frac{1}{n-m} \sigma^2 \text{tr} \left(I_n - X(X^T X)^{-1} X^T \right) = \frac{\sigma^2}{n-m} (n-m) = \sigma^2. \end{aligned}$$

□

Следствие 11.20. *Несмещенной оценкой для матрицы вариаций $V = V\{\hat{\theta}\}$ является следующая статистика:*

$$\hat{V} = \frac{R_{\min}^2}{n-m} (X^T X)^{-1}.$$

Теорема 11.19 (Гаусса – Маркова). *Если выполнены условия теоремы 11.17, то в семействе \mathcal{S} всевозможных линейных несмещенных статистических оценок вектора параметров $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \mathbb{R}^m$ МНК-оценка, определяемая теоремой 11.17: $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T = C_0 Y$, характеризуется наименьшими вариациями.*

Доказательство. Определим семейство \mathcal{S} , о котором говорится в теореме:

$$\mathcal{S} = \left\{ \tilde{\theta} = CY : C - \text{произвольная матрица, } E_{\theta^0} \{\tilde{\theta}\} = \theta^0, \forall \theta^0 \right\}.$$

Представим матрицу C в виде $C = C_0 + \Delta$, где $\Delta = (\Delta_{ij})$ – произвольная $(m \times n)$ -матрица. По теореме 11.17 для МНК-оценки имеем

$$\hat{\theta} = C_0 Y, \quad E_{\theta^0} \{\hat{\theta}\} = \theta^0, \quad \forall \theta^0;$$

$$V \left\{ \hat{\theta} \right\} = (v_{ij}) = \mathbf{E}_{\theta^0} \left\{ \left(\hat{\theta} - \theta^0 \right) \left(\hat{\theta} - \theta^0 \right)^T \right\} = \sigma^2 (X^T X)^{-1}.$$

Отсюда следует, что $\hat{\theta} \in \mathcal{S}$.

Получим ограничение на матрицу Δ , обусловленное несмещенностью МНК-оценки (11.95). Для любого $\theta^0 \in \mathbb{R}^m$ должно выполняться

$$\mathbf{E}_{\theta^0} \left\{ \tilde{\theta} \right\} = \mathbf{E}_{\theta^0} \left\{ (C_0 + \Delta) Y \right\} = \mathbf{E}_{\theta^0} \left\{ \hat{\theta} + \Delta Y \right\} = \theta^0 + \Delta \mathbf{E}_{\theta^0} \{ Y \} = \theta^0 + \Delta X \theta^0 = \theta^0.$$

Следовательно,

$$\Delta X = \mathbf{0},$$

и имеем эквивалентное представление для оценки $\tilde{\theta} \in \mathcal{S}$:

$$\tilde{\theta} = (C_0 + \Delta) Y = \hat{\theta} + \Delta (X \theta^0 + \varepsilon) = \hat{\theta} + \Delta \varepsilon.$$

С учетом этого вычислим матрицу вариаций:

$$\begin{aligned} \tilde{V} = (\tilde{v}_{ij}) &= V \left\{ \tilde{\theta} \right\} = \mathbf{E}_{\theta^0} \left\{ \left(\tilde{\theta} - \theta^0 \right) \left(\tilde{\theta} - \theta^0 \right)^T \right\} = \\ &= \mathbf{E}_{\theta^0} \left\{ \left(\left(\hat{\theta} - \theta^0 \right) + \Delta \varepsilon \right) \left(\left(\hat{\theta} - \theta^0 \right)^T + \varepsilon^T \Delta^T \right) \right\} = \\ &= V \left\{ \hat{\theta} \right\} + \Delta \mathbf{E} \left\{ \varepsilon \left(\hat{\theta} - \theta^0 \right)^T \right\} + \mathbf{E} \left\{ \left(\Delta \varepsilon \left(\hat{\theta} - \theta^0 \right)^T \right)^T \right\} + \Delta \mathbf{E}_{\theta^0} \left\{ \varepsilon \varepsilon^T \right\} \Delta^T. \end{aligned}$$

Учитывая (11.93) и представление (11.97) для МНК-оценки (11.95), полученное при доказательстве теоремы 11.17, имеем

$$\Delta \mathbf{E} \left\{ \varepsilon \left(\hat{\theta} - \theta^0 \right)^T \right\} = \Delta \mathbf{E}_{\theta^0} \left\{ \varepsilon \varepsilon^T C_0^T \right\} = \sigma^2 \Delta C_0^T = \sigma^2 \Delta X (X^T X)^{-1} X^T = \mathbf{0}.$$

В результате получим

$$\tilde{V} = (\tilde{v}_{ij}) = V + \sigma^2 \Delta \Delta^T,$$

откуда имеем выражения для диагональных элементов:

$$\tilde{v}_{ii} = v_{ii} + \sigma^2 \sum_{j=1}^n \Delta_{ij}^2 \geq v_{ii}, \quad i = 1, \dots, m,$$

причем равенство $\tilde{V} = V$ выполняется тогда и только тогда, когда $\Delta = \mathbf{0} \Leftrightarrow \tilde{\theta} = \hat{\theta}$.

□

11.10.3. Выравнивание экспериментальных данных прямой линией. Применение метода наименьших квадратов в эконометрике

Необходимо построить прямую $y = \hat{\theta}_1 + \hat{\theta}_2 x$ по методу наименьших квадратов для следующей линейной регрессионной модели:

$$y_i = \theta_1 + \theta_2 x_i + \epsilon_i, \quad i = 1, \dots, n.$$

Сведем эту модель к модели из п. 11.10.2:

$$x_{i1} := 1, \quad x_{i2} := x_i,$$

получим модель из п. 11.10.2, в которой $\theta = (\theta_1, \theta_2)^T$ – подлежащий оцениванию вектор параметров ($m = 2$).

Для вычисления оценки метода наименьших квадратов определим:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}; \quad (X^T X)^{-1} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1};$$

$$X^T Y = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n x_i y_i \right)^T; \quad \hat{\theta} = (X^T X)^{-1} X^T Y; \quad V\{\hat{\theta}\} = \sigma^2 (X^T X)^{-1}.$$

Пример 11.7 (оценивание производственной функции). Рассмотрим производственную функцию Кобба – Дугласа, часто используемую при моделировании экономики: $Y_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3}$, где i – номер года ($i = 1, 2, \dots, n, \dots$); Y_i – суммарная стоимость ВВП за i -й год; K_i – суммарная стоимость капиталовложений в i -м году; L_i – суммарная стоимость трудовых затрат в i -м году; $\beta_1, \beta_2, \beta_3 \geq 0$ – неизвестные параметры.

По данным $\{Y_i, K_i, L_i\}$ за n лет необходимо оценить параметры $\beta_1, \beta_2, \beta_3$. Приведем эту модель к линейной регрессионной модели с помощью логарифмирования: $y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3} + \epsilon_i$, $m = 3$, $\theta = (\theta_1, \theta_2, \theta_3)^T$, где $y_i = \ln Y_i$; $x_{i1} \equiv 1$; $x_{i2} = \ln K_i$; $x_{i3} = \ln L_i$; $\theta_1 = \ln \beta_1$; $\theta_2 = \beta_2$; $\theta_3 = \beta_3$; $\{\epsilon_i\}$ – случайные ошибки наблюдения.

Вектор оценок $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^T$ можно получить по (11.95) из теоремы 11.17.

После того как модель построена, можно построить прогноз ВВП в $(n+1)$ -м году:

$$Y_{n+1} = e^{\hat{\theta}_1} K_{n+1}^{\hat{\theta}_2} L_{n+1}^{\hat{\theta}_3},$$

где K_{n+1} , L_{n+1} – плановые значения капитальных вложений и трудовых затрат в $(n+1)$ -м году.

Заметим, что иногда применяется нелинейный метод наименьших квадратов:

$$\sum_{i=1}^n \left(Y_i - \beta_1 K_i^{\beta_2} L_i^{\beta_3} \right)^2 \rightarrow \min_{\beta_1, \beta_2, \beta_3}.$$

11.11. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ

11.11.1. Понятие об интервальном оценивании

Рассмотрим следующую задачу статистического оценивания параметров. Пусть в \mathbb{R}^N наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ из некоторого абсолютно непрерывного распределения вероятностей с некоторой плотностью $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta \in \mathbb{R}^1$ – неизвестное значение параметра. Примем следующие обозначения:

$$p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) \quad -$$

совместная плотность распределения вероятностей всей выборки;

$$P_{\theta}(A) = \int_A p(X; \theta) dX, \quad A \subseteq \mathcal{B}^{nN}, \quad -$$

вероятностная мера при условии, что параметр равен θ .

Рассмотрим задачу оценивания θ по X . Геометрический смысл статистического оценивания (см. рис. 11.1) состоит в том, что по выборке X пытаются построить точку $\hat{\theta}$, где $\hat{\theta} = T(X) \in \mathbb{R}^1$ – некоторая статистика. В силу того, что оценка интерпретируется точкой, такое оценивание называется *точечным*. В условиях регулярности известно следующее свойство точечных оценок (см. п. 11.5):

$$V\{\hat{\theta}\} = E_{\theta} \left\{ |\hat{\theta} - \theta|^2 \right\} \geq \frac{1}{nJ} > 0,$$

где $J > 0$ – количество информации по Фишеру.

Отсюда следует, что $P_{\theta} \{ \hat{\theta} = \theta \} = 0$ – точечная оценка с вероятностью единица ошибочна! Однако если возьмем какой-либо интервал $(\hat{\theta} - \delta, \hat{\theta} + \delta)$ длиной $2\delta > 0$, то

$$P_{\theta} \{ \hat{\theta} - \delta < \theta < \hat{\theta} + \delta \} > 0, \quad \forall \delta > 0.$$

На этом факте и основана идея *интервального оценивания*.

Определение 11.41. Пусть $0 < \epsilon < 1/2$. Случайный интервал $(\underline{\theta}, \bar{\theta}) \subset \mathbb{R}^1$, границы которого $\underline{\theta} = \underline{\theta}(X)$, $\bar{\theta} = \bar{\theta}(X)$ (причем $\underline{\theta}(X) < \bar{\theta}(X)$) являются статистиками, называется *доверительным интервалом для параметра θ* , если этот интервал «накрывает» истинное значение θ с наперед заданной достаточно большой вероятностью $1 - \epsilon$:

$$\forall \theta \in \mathbb{R} \quad P_{\theta} \{ \underline{\theta} < \theta < \bar{\theta} \} = 1 - \epsilon. \quad (11.102)$$

При этом величина $1 - \epsilon$ называется *доверительной вероятностью* (коэффициентом доверия, *доверительным уровнем*), а статистики $\underline{\theta}(X)$ и $\bar{\theta}(X)$ – соответственно *нижней и верхней доверительными границами*.

На практике доверительные вероятности обычно выбираются стандартным образом:

$$1 - \epsilon \in \{0,8; 0,85; 0,9; 0,95; 0,99; 0,995; 0,999\}.$$

Определение 11.42. Интервальной оценкой параметра θ по выборке X называется основанное на (11.102) вероятностное утверждение следующего вида:

$$\text{с вероятностью } 1 - \epsilon \quad \theta \in (\underline{\theta}, \bar{\theta}).$$

Замечание 11.16. Иногда доверительным уровнем называется следующая величина: $(1 - \epsilon)100\%$, т. е. доверительный уровень иногда измеряется в процентах.

Запишем условие (11.102) в эквивалентном виде:

$$\forall \theta \in \mathbb{R} \quad P_{\theta} \{ \theta \leq \underline{\theta} \} = \epsilon_1, \quad P_{\theta} \{ \theta \geq \bar{\theta} \} = \epsilon_2, \quad \epsilon_1 + \epsilon_2 = \epsilon.$$

В зависимости от соотношения между ϵ_1 и ϵ_2 известна следующая классификация типов доверительных интервалов.

Определение 11.43. Существуют три типа доверительных интервалов:

1) если $\epsilon_1 = \epsilon$, $\epsilon_2 = 0$, то интервальная оценка имеет вид: с вероятностью $1 - \epsilon$ $\theta > \underline{\theta}$, и доверительный интервал $(\underline{\theta}, \infty)$ называется *левосторонним доверительным интервалом*;

2) если $\epsilon_1 = 0$, $\epsilon_2 = \epsilon$, то интервальная оценка имеет вид: с вероятностью $1 - \epsilon$ $\theta < \bar{\theta}$, и доверительный интервал $(-\infty, \bar{\theta})$ называется правосторонним доверительным интервалом;

3) если $\epsilon_1 = \epsilon_2 = \epsilon/2$, то интервальная оценка принимает вид

$$\forall \theta \in \mathbb{R}, \quad P_{\theta} \{ \theta \leq \underline{\theta} \} = \frac{\epsilon}{2}, \quad P_{\theta} \{ \theta \geq \bar{\theta} \} = \frac{\epsilon}{2}; \quad (11.103)$$

при этом доверительный интервал $(\underline{\theta}, \bar{\theta})$ называется центральным доверительным интервалом (ЦДИ).

11.11.2. Методы построения доверительных интервалов

Рассмотрим три основных метода построения доверительных интервалов. Основное внимание будет уделено центральным доверительным интервалам, определяемым (11.103).

I. Метод обратной функции

Пусть некоторым образом построена состоятельная точечная оценка параметра θ : $\hat{\theta} = T(X)$. Предположим, что для нее вычислена функция распределения $F(z; \theta) = P_{\theta} \{ \hat{\theta} < z \}$, $z \in \mathbb{R}$, причем предполагаем, что $F(\cdot; \theta)$ монотонна по θ . Будем искать доверительные границы в виде

$$\underline{\theta} = \underline{\theta}(X) = f_1(\hat{\theta}), \quad \bar{\theta} = \bar{\theta}(X) = f_2(\hat{\theta}), \quad (11.104)$$

где $f_1(\cdot)$, $f_2(\cdot)$ – некоторые, пока не известные, монотонно возрастающие функции.

Подберем $f_1(\cdot)$, $f_2(\cdot)$ в (11.104) таким образом, чтобы выполнялись соотношения (11.103). Для первого из них имеем

$$\forall \theta \in \mathbb{R} \quad P_{\theta} \{ \theta \leq \underline{\theta} \} \equiv P_{\theta} \{ \theta \leq f_1(\hat{\theta}) \} = P_{\theta} \{ f_1^{-1}(\theta) \leq \hat{\theta} \} = 1 - F(\theta_1; \theta) = \frac{\epsilon}{2},$$

где $\theta_1 = f_1^{-1}(\theta)$, что эквивалентно $\theta = f_1(\theta_1)$.

Аналогично действуя со вторым соотношением, получаем $F(\theta_2 + 0; \theta) = \frac{\epsilon}{2}$, $\theta = f_2(\theta_2)$.

Таким образом, функции $f_1(\cdot)$, $f_2(\cdot)$ однозначно определяются следующими уравнениями:

$$F(\theta_1; \theta) = 1 - \frac{\epsilon}{2}, \quad \text{где } \theta = f_1(\theta_1); \quad (11.105)$$

$$F(\theta_2 + 0; \theta) = \frac{\epsilon}{2}, \quad \text{где } \theta = f_2(\theta_2). \quad (11.106)$$

Метод обратной функции порождает следующий алгоритм построения центрального доверительного интервала.

1. Строим состоятельную оценку $\hat{\theta} = T(X)$.
2. Находим функцию распределения этой оценки $F(\cdot; \theta)$.
3. Составляем и решаем по θ следующие два уравнения:

$$F(\hat{\theta}; \theta) = 1 - \frac{\epsilon}{2};$$

$$F(\hat{\theta} + 0; \theta) = \frac{\epsilon}{2}.$$

4. Определяем доверительные границы:

$\underline{\theta} = \underline{\theta}(X)$ – наименьший из корней, $\bar{\theta} = \bar{\theta}(X)$ – наибольший из корней.

Пример 11.8 (ЦДИ для среднего при известной дисперсии). Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ из нормального распределения $N(\theta, \sigma^2)$, где θ – неизвестное математическое ожидание, σ^2 – известная дисперсия. Необходимо построить 90% ЦДИ для θ ($1 - \epsilon = 0,9$).

Решение. Применим метод обратной функции.

1. Оценка $\hat{\theta} = \bar{x} = 1/n \sum_{k=1}^n x_k$ – выборочное среднее.

2. $\mathcal{L}\{\bar{x}\} = N_1(\theta, \sigma^2/n) \Rightarrow F(z; \theta) = \Phi((z - \theta)/\sigma/\sqrt{n})$, где $\Phi(\cdot)$ – функция распределения стандартного нормального закона.

3. Составляем и решаем уравнения ($g_\alpha = \Phi^{-1}(\alpha)$ – квантиль уровня α):

$$\begin{aligned}\Phi\left(\frac{\bar{x} - \theta}{\sigma/\sqrt{n}}\right) &= 1 - \frac{\epsilon}{2}, \quad \frac{\bar{x} - \theta}{\sigma/\sqrt{n}} = g_{1-\frac{\epsilon}{2}}; \\ \Phi\left(\frac{\bar{x} - \theta}{\sigma/\sqrt{n}}\right) &= \frac{\epsilon}{2}, \quad \frac{\bar{x} - \theta}{\sigma/\sqrt{n}} = g_{\frac{\epsilon}{2}} = -g_{1-\frac{\epsilon}{2}}.\end{aligned}$$

4. Вычисляя по таблицам [7]: $g_{0,95} \approx 1,65$, находим

$$\underline{\theta} = \bar{x} - 1,65 \frac{\sigma}{\sqrt{n}}, \quad \bar{\theta} = \bar{x} + 1,65 \frac{\sigma}{\sqrt{n}}.$$

Таким образом, с вероятностью 0,95: $\bar{x} - 1,65\sigma/\sqrt{n} < \theta < \bar{x} + 1,65\sigma/\sqrt{n}$.

Теорема 11.20. Пусть $\hat{\theta} = T(X)$ – некоторая состоятельная асимптотически несмещенная и асимптотически нормальная оценка параметра θ с асимптотической дисперсией $\sigma_n^2(\theta) \rightarrow 0$, $n \rightarrow \infty$, т. е. случайная величина $\tau = \tau(X; \theta) = (\hat{\theta} - \theta)/\sigma_n(\theta)$ при $n \rightarrow \infty$ распределена асимптотически нормально по закону $N_1(0, 1)$:

$$F_\tau(z) = P_\theta\{\tau < z\} \rightarrow \Phi(z), \quad z \in \mathbb{R}. \quad (11.107)$$

Тогда при $n \rightarrow \infty$ границы $(1 - \epsilon)100\%$ ЦДИ определяются из уравнений

$$\frac{\hat{\theta} - \theta}{\sigma_n(\theta)} = \pm g_{1-\frac{\epsilon}{2}}. \quad (11.108)$$

Доказательство. Применим метод обратной функции и воспользуемся асимптотикой (11.107). Составим уравнения (11.105) и (11.106), найдем

$$F(z; \theta) = P_\theta\{\hat{\theta} < z\} = P_\theta\left\{\frac{\hat{\theta} - \theta}{\sigma_n(\theta)} < \frac{z - \theta}{\sigma_n(\theta)}\right\} = F_\tau\left(\frac{z - \theta}{\sigma_n(\theta)}\right).$$

Учитывая (11.105), (11.106) и используя (11.107), получаем асимптотические соотношения:

$$\Phi\left(\frac{\hat{\theta} - \theta}{\sigma_n(\theta)}\right) = 1 - \frac{\epsilon}{2}; \quad \Phi\left(\frac{\hat{\theta} - \theta}{\sigma_n(\theta)}\right) = \frac{\epsilon}{2}.$$

Действуя дальше аналогично примеру 11.8, получаем (11.108). \square

Следствие 11.21. Если оценка $\hat{\theta}$ имеет нормальное распределение: $\mathcal{L}\{\hat{\theta}\} = N_1(\theta, \sigma_n^2(\theta))$, то границы, определяемые (11.108), являются точными при любом n .

Следствие 11.22. Если дисперсия $\sigma_n^2(\theta) = \sigma_n^2$ и не зависит от θ , то доверительные границы имеют явный вид:

$$\underline{\theta} = \hat{\theta} - g_{1-\frac{\epsilon}{2}} \sigma_n, \quad \bar{\theta} = \hat{\theta} + g_{1-\frac{\epsilon}{2}} \sigma_n.$$

Теорема 11.21 (интервальная оценка вероятности успеха). Пусть рассматривается схема n независимых испытаний Бернулли и $X = (x_1, \dots, x_n)^T \in \{0, 1\}^n$ – случайная выборка объема n из распределения $\text{Bi}(1, \theta)$, где $\theta \in (0, 1)$ – неизвестная вероятность успеха. Если $m = \sum_{i=1}^n x_i$ – зарегистрированное число успехов в n испытаниях, причем $m \in \{1, \dots, n-1\}$, то $(1 - \epsilon)100\%$ ЦДИ имеет границы $\underline{\theta}, \bar{\theta}$, являющиеся корнями следующих уравнений:

$$\sum_{i=m}^n C_n^i \theta^i (1 - \theta)^{n-i} = \frac{\epsilon}{2}; \quad (11.109)$$

$$\sum_{i=0}^m C_n^i \theta^i (1 - \theta)^{n-i} = \frac{\epsilon}{2}. \quad (11.110)$$

Доказательство. Применим метод обратной функции. Известна состоятельная оценка для θ : $\hat{\theta} = \frac{m}{n}$ – относительная частота, для которой

$$\mathcal{L}\{m\} = \text{Bi}(n, \theta), \quad F\left(\frac{k}{n}; \theta\right) = P_{\theta}\{m < k\} = \sum_{i=0}^{k-1} C_n^i \theta^i (1 - \theta)^{n-i}.$$

Составляем уравнение (11.105):

$$F\left(\frac{m}{n}; \theta\right) = \sum_{i=0}^{m-1} C_n^i \theta^i (1 - \theta)^{n-i} = 1 - \sum_{i=m}^n C_n^i \theta^i (1 - \theta)^{n-i} = 1 - \frac{\epsilon}{2},$$

откуда получаем (11.109).

Запишем уравнение (11.106):

$$F\left(\frac{m}{n} + 0; \theta\right) = \sum_{i=0}^m C_n^i \theta^i (1 - \theta)^{n-i} = \frac{\epsilon}{2},$$

что совпадает с (11.110). \square

Следствие 11.23. Если число успехов $m = 0$, то имеем правосторонний доверительный интервал с вероятностью $1 - \epsilon$: $0 < \theta < \bar{\theta}$, где $\bar{\theta} = 1 - \sqrt[n]{\epsilon}$.

Следствие 11.24. Если число успехов $m = n$, то имеем левосторонний доверительный интервал с вероятностью $1 - \epsilon$: $\underline{\theta} < \theta < 1$, где $\underline{\theta} = \sqrt[n]{\epsilon}$.

II. Метод стьюдентизации

Метод стьюдентизации был предложен английским статистиком В. Госсетом (псевдоним Student). Он обобщает метод обратной функции и состоит из следующих трех шагов.

1. Строится функция $\tau = \tau(X; \theta)$ такая, что выполняются два условия:

а) $\mathcal{L}\{\tau\} = \text{inv}\alpha_{\theta}$ – не зависит от параметра;

б) $\tau(\cdot; \theta)$ – непрерывна и монотонна по θ .

Например, в следствии 11.21 теоремы 11.20 была построена функция

$$\tau(X; \theta) = \frac{\hat{\theta} - \theta}{\sqrt{D_{\theta}\{\hat{\theta}\}}}.$$

Если $\hat{\theta}$ нормально распределена, то $\mathcal{L}\{\tau\} = N_1(0, 1)$, так что условия а) и б) выполняются.

2. По заданному доверительному уровню $1 - \epsilon$ выбираются два числа $g_-(\epsilon)$, $g_+(\epsilon)$ такие, что (рис. 11.8)

$$P_{\theta}\{g_-(\epsilon) < \tau < g_+(\epsilon)\} = 1 - \epsilon, \quad \forall \theta \in \mathbb{R}.$$

Когда числа $g_-(\epsilon)$, $g_+(\epsilon)$ находятся неоднозначно, используют еще одно дополнительное условие: $g_+(\epsilon) - g_-(\epsilon) \rightarrow \min$.

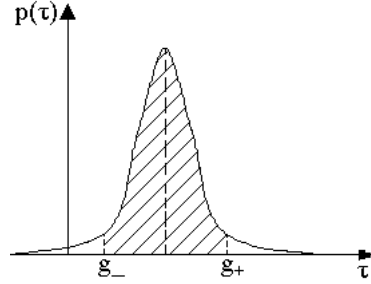


Рис. 11.8. Иллюстрация метода студентизации

Если τ имеет симметричное распределение вероятностей относительно $\tau = 0$, то $g_{\pm}(\epsilon) = \pm g(\epsilon)$. Например, если $\mathcal{L}\{\tau\} = N_1(0, 1)$, то $g(\epsilon) = \Phi^{-1}(1 - \epsilon/2)$.

3. Решаются два уравнения относительно θ :

$$\tau(X; \theta) = g_{\pm}(\epsilon).$$

Нижняя доверительная граница $\underline{\theta}_{\tau}$ равна наименьшему из корней, а верхняя доверительная граница $\bar{\theta}_{\tau}$ равна наибольшему из корней.

Пример 11.9 (ЦДИ для среднего при неизвестной дисперсии). Наблюдается случайная выборка $X = (x_1, \dots, x_n)^T$ из $N_1(\theta, \sigma^2)$; θ , σ^2 – неизвестные параметры. Построить ЦДИ для θ при доверительном уровне $1 - \epsilon$.

Решение. Воспользуемся методом студентизации.

1. Чтобы построить функцию $\tau = \tau(X; \theta)$, построим вначале несмещенные, состоятельные оценки для параметров θ , σ^2 :

$$\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 > 0.$$

Введем следующие нормированные случайные величины:

$$\xi_1 = \frac{\bar{x} - \theta}{\sigma/\sqrt{n}}, \quad \mathcal{L}\{\xi_1\} = N_1(0, 1);$$

$$\xi_2 = \frac{(n-1)s^2}{\sigma^2}, \quad \mathcal{L}\{\xi_2\} = \chi_{n-1}^2;$$

ξ_1 , ξ_2 – независимы.

Построим случайную величину:

$$\tau = \tau(X; \theta) = \frac{\xi_1}{\sqrt{\frac{\xi_2}{n-1}}}.$$

Известно, что $\mathcal{L}\{\tau\} = t_{n-1}$ – распределение вероятностей Стьюдента с $n-1$ степенью свободы; функция распределения в данном случае обозначается следующим образом:

$$P\{\tau < y\} = F_{t_{n-1}}(y), \quad y \in \mathbb{R}.$$

Выразим $\tau(X; \theta)$ через исходные статистики:

$$\tau(X; \theta) = \sqrt{n} \frac{\bar{x} - \theta}{s} \quad -$$

непрерывная и монотонная по θ функция.

2. Функция $F_{t_{n-1}}(y)$ есть функция распределения симметричной случайной величины:

$$F_{t_{n-1}}(-y) = 1 - F_{t_{n-1}}(y),$$

тогда получаем

$$P\{-g < \tau < g\} = F_{t_{n-1}}(g) - (1 - F_{t_{n-1}}(g)) = 2F_{t_{n-1}}(g) - 1 = 1 - \epsilon.$$

Из последнего равенства имеем $F_{t_{n-1}}(g) = 1 - \frac{\epsilon}{2}$, следовательно, $g(\epsilon) = F_{t_{n-1}}^{-1}\left(1 - \frac{\epsilon}{2}\right)$ – квантиль уровня $1 - \frac{\epsilon}{2}$ распределения Стьюдента с $n-1$ степенью свободы.

3. Из уравнений

$$\sqrt{n} \frac{\bar{x} - \theta}{s} = \pm F_{t_{n-1}}^{-1}\left(1 - \frac{\epsilon}{2}\right)$$

получаем

$$\underline{\theta}_\tau = \bar{x} - F_{t_{n-1}}^{-1}\left(1 - \frac{\epsilon}{2}\right) \frac{s}{\sqrt{n}};$$

$$\bar{\theta}_\tau = \bar{x} + F_{t_{n-1}}^{-1}\left(1 - \frac{\epsilon}{2}\right) \frac{s}{\sqrt{n}}.$$

Таким образом, ЦДИ $(\underline{\theta}_\tau, \bar{\theta}_\tau)$ построен. Его длина

$$\tilde{\Delta} = \bar{\theta}_\tau - \underline{\theta}_\tau = 2F_{t_{n-1}}^{-1}\left(1 - \frac{\epsilon}{2}\right) \frac{s}{\sqrt{n}}.$$

Задание 11.11. Сравнить полученный результат с примером 11.8. Как значение дисперсии σ^2 влияет на длину ЦДИ?

III. Метод построения асимптотически наикратчайших доверительных интервалов

Определение 11.44. Пусть с помощью некоторой функции $\tau = \tau(X; \theta)$ построен ЦДИ $(\underline{\theta}_\tau, \bar{\theta}_\tau)$ с доверительным уровнем $1 - \epsilon$. Средней длиной этого доверительного интервала называется число

$$\Delta = \Delta(\tau(\cdot)) ::= E\{\bar{\theta}_\tau - \underline{\theta}_\tau\} \geq 0.$$

Возникает задача построения наикратчайшего доверительного интервала, имеющего минимальную длину:

$$\Delta(\tau(\cdot)) \rightarrow \min_{\tau(\cdot)}.$$

Приведем асимптотический результат.

Теорема 11.22. Пусть в условиях регулярности рассматривается семейство центральных доверительных интервалов доверительного уровня $1 - \epsilon$ $(\underline{\theta}_\tau, \bar{\theta}_\tau)$, основанных на функции $\tau = \tau(X; \theta)$ следующего вида:

$$\tau(X; \theta) = \frac{1}{\sqrt{n D_\theta \{h(x_1; \theta)\}}} \sum_{k=1}^n h(x_k; \theta),$$

где $h(\cdot; \theta)$ – произвольная борелевская функция такая, что математическое ожидание $E_\theta \{h(x_1; \theta)\} = 0$, дисперсия $D_\theta \{h(x_1; \theta)\} < +\infty$; при этом доверительные границы определяются из уравнений

$$\tau(X; \theta) = \pm g_{1-\frac{\epsilon}{2}}, \quad g_\alpha = \Phi^{-1}(\alpha).$$

Тогда при $n \rightarrow \infty$ асимптотически наикратчайший центральный доверительный интервал определяется функцией

$$h(x; \theta) = h^*(x; \theta) = \ln p(x; \theta), \quad x \in \mathbb{R}^N,$$

где $p(\cdot; \theta)$ – плотность распределения вероятностей выборочных значений, а доверительные границы этого асимптотически наикратчайшего ЦДИ являются корнями уравнений

$$\frac{1}{\sqrt{n \mathcal{I}}} \frac{dl(\theta)}{d\theta} = \pm g_{1-\frac{\epsilon}{2}},$$

где

$$l(\theta) = \sum_{k=1}^n \ln p(x_k; \theta) \quad -$$

логарифмическая функция правдоподобия;

$$\mathcal{I} = E_\theta \left\{ -\frac{d^2 \ln p(x_k; \theta)}{d\theta^2} \right\} > 0 \quad -$$

количество информации по Фишеру.

Отметим, что одномерные доверительные интервалы могут быть обобщены на многомерный случай: $\theta \in \mathbb{R}^m$, $m \geq 1$.

Определение 11.45. Пусть $0 < \epsilon < 1$, тогда случайная область $V_\epsilon = V_\epsilon(X) \subset \mathbb{R}^m$, граница которой зависит от выборки X , называется доверительной областью для векторного параметра $\theta \in \mathbb{R}^m$, если эта область содержит θ с наперед заданной достаточно большой вероятностью $1 - \epsilon$, т. е.

$$\forall \theta \quad P_\theta \{\theta \in V_\epsilon(X)\} = 1 - \epsilon.$$

11.12. РОБАСТНОЕ СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

Реальные наблюдения, которые подвергаются статистическому анализу и на основе которых строятся точечные и интервальные оценки параметров, как правило, лишь весьма приближенно соответствуют той математической модели, на основе которой построена используемая процедура статистического оценивания. Это связано с тем, что на практике при регистрации наблюдений из выборки $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$, их сборе и хранении зачастую возникают достаточно грубые ошибки, вызванные сбоями регистрирующей аппаратуры, невнимательностью наблюдателя, а также несоответствием «природы» регистрируемых наблюдений предполагаемой (гипотетической) модели. В связи с этим возникает необходимость в статистических процедурах, которые не

должны терять свои оптимальные свойства в условиях отклонений от гипотетической модели. Такие статистические процедуры, по предложению американского статистика Дж. Тьюки, называются *устойчивыми* или *робастными* (от англ. *robust* – крепкий, стойкий, дюжий), и их разработка является одним из основных направлений в современной статистике [35, 28].

Напомним *гипотетическую модель*, принятую в теории статистического оценивания параметров (далее – модель M_o). Наблюдаемая случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n описывается некоторым N -мерным гипотетическим распределением вероятностей с плотностью $p(x; \theta^0)$, $x \in \mathbb{R}^N$, где $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \Theta \subseteq \mathbb{R}^m$ – неизвестное истинное значение вектора параметров.

Чаще всего на практике в качестве модели наблюдений с искажениями используется *модель Тьюки – Хьюбера* [35] (модель $M_{\epsilon+}^h$), согласно которой наблюдения x_1, \dots, x_n из выборки X объема n есть случайные N -векторы с искаженной (фактической) плотностью распределения вероятностей:

$$p^\epsilon(\cdot; \theta^0) \in \mathcal{P}(p(\cdot; \theta^0), \epsilon_+); \quad \mathcal{P}(p(\cdot; \theta^0), \epsilon_+) = \{p^\epsilon(x; \theta^0), x \in \mathbb{R}^N : \\ p^\epsilon(x; \theta^0) = (1 - \epsilon)p(x; \theta^0) + \epsilon h(x), \quad 0 \leq \epsilon \leq \epsilon_+ < 1\}, \quad (11.111)$$

где ϵ_+ – так называемый *уровень искажений* (при $\epsilon_+ = 0$ искажения отсутствуют и имеет место модель M_o), а ϵ – текущий уровень искажений, имеющий смысл вероятности появления аномальных наблюдений в выборке: при регистрации наблюдения $x \in \mathbb{R}^N$, согласно модели (11.111), оно с вероятностью $1 - \epsilon$ соответствует гипотетической модели M_o и описывается плотностью $p(\cdot; \theta^0)$, а с вероятностью ϵ является аномальным и определяется неизвестной плотностью «засоряющего» распределения $h(\cdot)$.

Другая интерпретация искажений Тьюки – Хьюбера (11.111) заключается в следующем: выборку X объема n рассматривают как неоднородную выборку, состоящую из двух подвыборок, первая из которых содержит «в среднем» $(1 - \epsilon) \cdot n$ наблюдений, имеющих гипотетическую плотность $p(\cdot; \theta^0)$, а вторая образована аномальными наблюдениями, описываемыми плотностью $h(\cdot)$, количество которых «в среднем» равно $\epsilon \cdot n$.

Искажения типа (11.111) обычно обусловлены появлением в выборке X резко выделяющихся аномальных наблюдений-«выбросов» (от англ. *outlier*), которые удалены от основной «массы» наблюдений и приводят к увеличению рассеяния и «утяжелению хвостов» вероятностного распределения наблюдений (рис. 11.9, $N = 2$).

В качестве гипотетической модели M_o обычно используется гауссовская модель. Не ограничивая общности, рассмотрим здесь случай одномерных наблюдений ($x \in \mathbb{R}$, $N = 1$) с гипотетической плотностью вида

$$p(x; \theta^0) = n_1(x|\theta^0, \sigma^2) = \Phi\left(\frac{x - \theta^0}{\sigma}\right), \quad (11.112)$$

где $\Phi(\cdot)$ – плотность распределения вероятностей стандартного нормального закона $N_1(0, 1)$; $\theta^0 \in \mathbb{R}$ – неизвестное истинное гипотетическое значение математического ожидания («параметр сдвига»); σ^2 – дисперсия в гипотетической модели ($\sigma > 0$ – «параметр масштаба», гипотетическое значение среднеквадратического отклонения).

В качестве искаженной модели, соответствующей M_o из (11.112), рассмотрим частный случай модели (11.111):

$$p^\epsilon(x; \theta^0) = (1 - \epsilon)\Phi\left(\frac{x - \theta^0}{\sigma}\right) + \epsilon\Phi\left(\frac{x - \theta^0}{k\sigma}\right), \quad 0 \leq \epsilon \leq \epsilon_+ < 1, \quad (11.113)$$

с плотностью «засоряющего» распределения $h(x) = n_1(x|\theta^0, k^2\sigma^2)$, дисперсия которого $k^2\sigma^2$ в k^2 раз больше дисперсии σ^2 гипотетического распределения ($k > 1$ – коэффициент увеличения среднеквадратического отклонения).

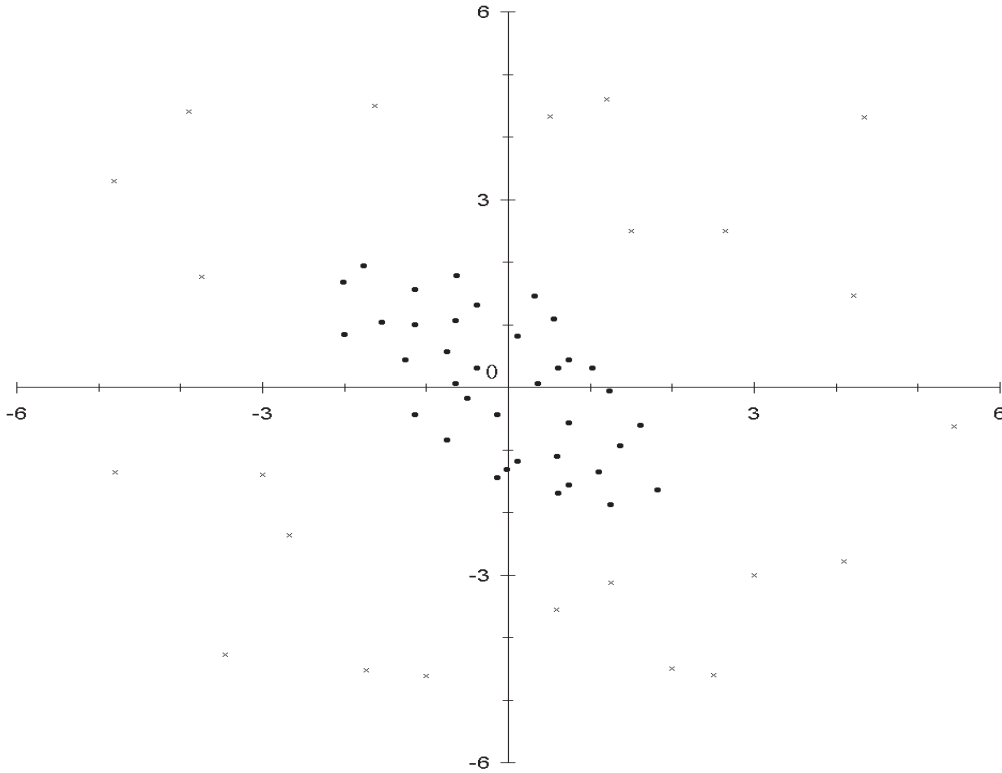


Рис. 11.9. Диаграмма рассеяния искаженной выборки:

● – гипотетические наблюдения; × – «выбросы»

В качестве показателей увеличения рассеяния и «утяжеления хвостов» для модели (11.112), (11.113) вычисляют

$$P = P\{|x_t - \theta_0| > 3\sigma\} = 2 \left(\Phi(-3) + \epsilon \left(\Phi\left(-\frac{3}{k}\right) - \Phi(-3) \right) \right) \quad (11.114)$$

вероятность появления «выброса» вне гипотетического интервала концентрации $[\theta^0 - 3\sigma, \theta^0 + 3\sigma]$ ($\Phi(\cdot)$ – функция распределения вероятностей стандартного нормального закона $N_1(0, 1)$);

$$\delta = \frac{D\{x_t\}}{\sigma^2} = 1 + (k^2 - 1)\epsilon \quad \text{– относительную дисперсию.} \quad (11.115)$$

Значения характеристик (11.114), (11.115) для гипотетической модели M_o из (11.112) (при $\epsilon = 0$ в (11.113)) равны соответственно: $P = 2\Phi(-3) \approx 0,0027$, $\delta = 1$, и увеличиваются с ростом текущего уровня искажений ϵ и дисперсии выбросов (увеличением ко-

эффицента k). Так, например, при $k = 5$ и $\epsilon = 0,1$ дисперсия наблюдений из выборки увеличивается в 3,4 раза ($\delta = 3,4$).

Одним из способов построения устойчивых оценок является подход, при котором для описания искажений в выборке используются распределения с «хвостами» более тяжелыми, чем у гипотетического распределения. Например, вместо модели (11.112), (11.113) для описания искаженной выборки используется распределение Лапласа, для которого МП-оценкой параметра «сдвига» является выборочная медиана, принадлежащая к классу так называемых *L-оценок*:

$$\hat{\theta}^L = \hat{\theta}^L(X) = \sum_{t=1}^n a_t g(x_{(t)}), \quad (11.116)$$

где $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ – вариационный ряд выборки $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$; $g(\cdot)$ – некоторая функция; $\{a_t\}_{t=1}^n$ – весовые коэффициенты.

Чаще всего используется частный случай (11.116):

$$\hat{\theta}^{L,\alpha} = \frac{1}{n-2q} \sum_{t=q+1}^{n-q} x_{(t)} \quad - \quad (11.117)$$

«усеченное» среднее, где q – наибольшее целое, не превосходящее $\alpha \cdot n$ ($0 < \alpha < 1/2$ задано). При $0 \leq \alpha < 1/n$ из (11.117) получаем обычное арифметическое среднее:

$$\hat{\theta}^{L,\alpha}|_{0 \leq \alpha < 1/n} = \frac{1}{n} \sum_{t=1}^n x_{(t)} = \bar{x},$$

а при $\alpha \rightarrow 1/2$ – выборочную медиану:

$$\hat{\theta}^{L,\alpha}|_{\alpha \rightarrow 1/2} = \begin{cases} x_{(l+1)}, & n = 2l + 1, \\ \frac{x_{(l)} + x_{(l+1)}}{2}, & n = 2l. \end{cases}$$

Теперь вернемся к модели искажений Тьюки – Хьюбера общего вида (11.111). В качестве робастных статистических оценок параметра $\theta^0 \in \Theta \subseteq \mathbb{R}^m$ швейцарский статистик П. Хьюбер предложил использовать так называемые *M-оценки* $\hat{\theta}^M = \hat{\theta}^M(X)$, определяемые по выборке $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ как решение экстремальной задачи (Θ^* – замыкание Θ):

$$\sum_{t=1}^n \phi(x_t; \theta) \rightarrow \min_{\theta \in \Theta^*}, \quad (11.118)$$

где $\phi(\cdot; \theta)$ – некоторая функция, определяющая конкретный тип оценок и их точность.

Очевидно, что *M-оценка* с функцией $\phi(\cdot; \theta) \equiv -\ln p(\cdot; \theta)$ – обычная оценка максимального правдоподобия (МП-оценка) из п. 11.7, построенная для гипотетической модели M_o (когда искажения наблюдений в выборке отсутствуют). Более того, оценки вида (11.118) в условиях гипотетической модели M_o при определенных условиях на функции $\phi(\cdot; \theta)$ обобщают МП-оценки.

Определение 11.46. *Оценкой минимального контраста (ОМК, МК-оценкой) называется оценка, определяемая из соотношения (11.118) с функцией $\phi(\cdot; \theta) \equiv \phi^0(\cdot; \theta)$, называемой функцией контраста и удовлетворяющей соотношению*

$$F(\theta^0, \theta^0) < F(\theta, \theta^0), \quad \forall \theta \in \Theta, \quad \theta \in \Theta^*, \quad \theta \neq \theta^0, \quad (11.119)$$

где

$$F(\theta, \theta^0) = \int_{\mathbb{R}^N} \Phi^0(x; \theta) p(x; \theta^0) dx -$$

математическое ожидание от $\Phi^0(\cdot; \theta)$, вычисленное по гипотетической плотности $p(\cdot; \theta^0)$.

В условиях гипотетической модели (при отсутствии искажений) МК-оценка (11.118), (11.119) при определенных условиях регулярности обладает почти всеми свойствами МП-оценки (за исключением разве что эффективности), например она строго состоятельна.

При предполагаемом наличии искажений Тьюки – Хьюбера (модель $M_{\epsilon_+}^h$ из (11.111)), основываясь на принципе «усечения» Хьюбера [35, 28], в качестве робастной M -оценки (11.118) можно предложить «усеченную» МК-оценку $\hat{\theta}^{M,C} = \hat{\theta}^{M,C}(X)$ с функцией $\Phi(\cdot; \theta) \equiv \Phi^C(\cdot; \theta)$ вида

$$\Phi^C(x; \theta) = \begin{cases} \Phi^0(x; \theta), & \Phi^0(x; \theta) < C; \\ C, & \Phi^0(x; \theta) \geq C, \end{cases} \quad (11.120)$$

где $\Phi^0(\cdot; \theta)$ – гипотетическая функция контраста из (11.119), а C – параметр «усечения»: $C \in (-\infty; +\infty]$, позволяющий ограничить влияние аномальных наблюдений-«выбросов» на статистическую оценку (при $C = +\infty$ получаем классическую ОМК (11.118), (11.119)).

Для выбора значения параметра «усечения» C в (11.120) можно использовать так называемый минимаксный подход Хьюбера [35, 28], который состоит в следующем.

В качестве оптимальной функции $\Phi^*(\cdot; \theta) \in \Phi$ в (11.118) надлежит выбирать ту, которая на семействе функций $\Phi(\cdot; \theta) \in \Phi$ минимизирует «наихудшее» по модели $M_{\epsilon_+}^h$ ($p^\epsilon(\cdot; \theta^0) \in \mathcal{P}(p(\cdot; \theta^0), \epsilon_+)$) значение некоторого показателя погрешности оценки (например, асимптотической дисперсии $V = V(\Phi(\cdot; \theta), p^\epsilon(\cdot; \theta^0))$):

$$\Phi^*(\cdot; \theta) = \arg \min_{\Phi(\cdot; \theta) \in \Phi} \sup_{p^\epsilon(\cdot; \theta^0) \in \mathcal{P}(p(\cdot; \theta^0), \epsilon_+)} V(\Phi(\cdot; \theta), p^\epsilon(\cdot; \theta^0)). \quad (11.121)$$

Из условия типа (11.121) для семейства «усеченных» функций (11.120) находим оптимальное по Хьюберу значение параметра «усечения» C^* . Так, для параметра «сдвига» гауссовской гипотетической плотности (11.112) при наличии искажений типа (11.111) (считая $\epsilon = \epsilon_+$ – зафиксированным, «параметр масштаба» σ – известным, а плотности «засоряющих» распределений – симметричными: $h(x) = h(-x)$, $x \in \mathbb{R}$), исходя из принципа минимакса асимптотической дисперсии (11.121) на классе всех произвольных функций $\Phi(\cdot; \theta)$ в (11.118) (включающем в себя и «усеченные» функции) Хьюбер получил «усеченную» робастную оценку [35, 28], определяемую из следующего уравнения относительно θ :

$$\sum_{t=1}^n \psi_b \left(\frac{x_t - \theta}{\sigma} \right) = 0,$$

где

$$\psi_b(z) = \max\{-b, \min\{z, b\}\}, \quad z \in \mathbb{R},$$

а значение параметра «усечения» $b = b(\epsilon_+) > 0$ является решением уравнения

$$2\Phi(b) - 1 + \frac{2\phi(b)}{b} = \frac{1}{1 - \epsilon_+}.$$

В «предельных» случаях при $\epsilon_+ \rightarrow 0$ ($b \rightarrow +\infty$) имеем выборочное среднее, а при $\epsilon_+ \rightarrow 1$ ($b \rightarrow 0$) – выборочную медиану.

11.13. УПРАЖНЕНИЯ

- Пусть выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей $N_1(a, \sigma^2)$. Вычислить математическое ожидание статистик $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$ и $s^2 = 1(n-1) \sum_{i=1}^n (x_i - \bar{x})^2$, где $\bar{x} = \sum_{i=1}^n x_i / n$ – выборочное среднее.
- Выборка $X = (x_1, \dots, x_n)$ получена из экспоненциального распределения вероятностей с параметром $\lambda > 0$. Статистика $Y = (y_1, \dots, y_n)$ построена следующим образом: $y_i = 1 - e^{-\lambda x_i}$, $i = 1, \dots, n$. Найти распределение вероятностей элементов статистики Y .
- Пусть выборка $X = (x_1, \dots, x_n)$ получена из распределения вероятностей с известной функцией распределения $F(x)$, $x \in \mathbb{R}$. Найти функцию распределения вероятностей первой порядковой статистики.
- Выборка $X = (x_1, \dots, x_n)$ получена из равномерного распределения вероятностей на отрезке $[0, b]$, $b > 0$. Вычислить математическое ожидание и дисперсию n -й порядковой статистики.
- Пусть выборка $X = (x_1, \dots, x_n)$ получена из распределения вероятностей с функцией распределения $F(x)$, $x \in \mathbb{R}$. Найти совместную функцию распределения $x_{(1)}$ и $x_{(n)}$.
- Дана выборка $X = (3, 0, 2, 4, 3, 6, 0, 3, 1)$. Построить по этой выборке эмпирическую функцию распределения $\hat{F}(x)$, $x \in \mathbb{R}$, и вычислить $\hat{F}(3, 5) - \hat{F}(3)$.
- Пусть $\hat{F}(x)$ – эмпирическая функция распределения, построенная по выборке $X = (x_1, \dots, x_n)$. Пусть $a \in \mathbb{R}$, $a > 0$. Является ли эмпирической функцией распределения функция $\hat{F}(ax)$? В случае положительного ответа привести выборку, которой она соответствует.
- Пусть $\hat{F}(x)$ – эмпирическая функция распределения, построенная по выборке $X = (x_1, \dots, x_n)$. Существует ли выборка, для которой $\hat{F}(x^3)$ является эмпирической функцией распределения? В случае положительного ответа указать такую выборку.
- Пусть выборка $X = (x_1, \dots, x_n)$ получена из распределения вероятностей с заданными дисперсией σ^2 и четвертым центральным моментом μ_4 . По выборке X построена оценка дисперсии – статистика $\hat{\sigma}^2$. Вычислить ее дисперсию.
- Выборка $X = (x_1, \dots, x_n)$ получена из распределения вероятностей с заданным начальным моментом α_k порядка k , $k \in \mathbb{N}$. По выборке X построен выборочный момент a_k порядка k . Вычислить математическое ожидание a_k .
- Выборка $X = (x_1, \dots, x_n)$ получена из распределения вероятностей с заданными моментами до порядка $k+s$, $s \in \mathbb{N}$, включительно: $\alpha_1, \dots, \alpha_{k+s}$. Вычислить ковариацию выборочных моментов a_k и a_s , построенных по выборке X .
- Выполнить упражнение 9 в случае, когда выборка X получена из нормального распределения вероятностей с заданной дисперсией σ^2 .
- Пусть $\hat{\theta}$ – несмещенная оценка неизвестного параметра θ , имеющая конечную положительную дисперсию $D\{\hat{\theta}\}$. Является ли статистика $(\hat{\theta})^2$ несмещенной оценкой

для θ^2 ? В случае отрицательного ответа указать смещение.

14. Известно, что $E\{T_n\} = \theta + b_n$, $D\{T_n\} = \delta_n$, где $b_n \xrightarrow{n \rightarrow \infty} 0$, $\delta_n \xrightarrow{n \rightarrow \infty} 0$. Доказать, что T_n является состоятельной оценкой параметра θ . Что можно сказать о несмещенности этой оценки?
15. По выборке $X = (x_1, \dots, x_n)$ оценивается параметр θ равномерного распределения вероятностей на отрезке $[0, \theta]$. Являются ли статистики $\hat{\theta}_1 = \frac{n+1}{n}x_{(n)}$, $\hat{\theta}_2 = (n+1)x_{(1)}$ несмещенными оценками? Какую из них предпочтительнее использовать и почему?
16. Пусть выборка $X = (x_1, \dots, x_n)$ получена из показательного распределения вероятностей с параметром $\lambda > 0$. Является ли оценка $\hat{\lambda} = (\bar{x})^{-1}$ несмещенной? Найти смещение в случае отрицательного ответа. Является ли оценка состоятельной?
17. Выборка $X = (x_1, \dots, x_n)$ получена из равномерного распределения вероятностей на отрезке $[0, \theta]$. Найти минимальную достаточную статистику для параметра θ .
18. Пусть выборка $X = (x_1, \dots, x_n)$ получена из равномерного распределения вероятностей на отрезке $[\theta_1, \theta_2]$. Будут ли достаточными статистиками для параметра $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ следующие статистики: а) \bar{x} ; б) $x_{(n)}$; в) $\begin{pmatrix} x_{(1)} \\ x_{(n)} \end{pmatrix}$?
19. Выборка $X = (x_1, \dots, x_n)$ получена из равномерного распределения вероятностей на отрезке $[-\theta, \theta]$. Найти минимальную достаточную статистику для параметра θ .
20. Выборка $X = (x_1, \dots, x_n)$ получена из показательного (экспоненциального) распределения вероятностей с параметром $\lambda > 0$. Найти минимальную достаточную статистику для этого параметра.
21. Выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей $N_1(\mu, \sigma^2)$. Вычислить количество информации по Фишеру, содержащееся в одном наблюдении о неизвестном параметре, если: а) параметр μ неизвестен, значение σ^2 известно; б) дисперсия σ^2 неизвестна, значение μ известно.
22. Выборка $X = (x_1, \dots, x_n)$ получена из биномиального распределения вероятностей с параметрами N и p ; значение $p \in (0, 1)$ известно. Доказать, что статистика \bar{x} – минимальная достаточная статистика.
23. Пусть построена эффективная оценка $t = t(X)$ по выборке X в классе оценок со смещением $\alpha \cdot \theta$, α – заданная постоянная. Построить эффективную оценку в классе несмещенных оценок.
24. Построить оценку максимального правдоподобия для неизвестного значения дисперсии σ^2 нормального распределения вероятностей $N_1(\mu, \sigma^2)$, из которого получена выборка $X = (x_1, \dots, x_n)$, если значение μ известно.
25. Выборка $X = (x_1, \dots, x_n)$ получена из равномерного распределения вероятностей на отрезке $[0, \theta]$; значение параметра $\theta > 0$ неизвестно. Построить оценку максимального правдоподобия параметра θ .
26. Выборка $X = (x_1, \dots, x_n)$ получена из показательного распределения вероятностей с неизвестным значением параметра $\lambda > 0$. Построить для λ оценку максимального правдоподобия.

27. Выборка $X = (x_1, \dots, x_n)$ представляет собой результаты n независимых испытаний Бернулли с неизвестной вероятностью «успеха» p . Построить для нее оценку максимального правдоподобия.
28. Построить оценку максимального правдоподобия параметра $\lambda > 0$ распределения Пуассона по выборке $X = (x_1, \dots, x_n)$ из этого распределения вероятностей.
29. Выборка $X = (x_1, \dots, x_n)$ получена из равномерного распределения вероятностей на отрезке $[0, \theta]$. Построить оценку неизвестного значения параметра θ по методу моментов. Исследовать построенную оценку на несмещенность и состоятельность. Является ли построенная оценка эффективной?
30. В схеме независимых испытаний Бернулли с неизвестной вероятностью «успеха» произведено n испытаний, в которых «успех» наблюдался m_n раз. Построить оценку вероятности θ по методу моментов.
31. Выборка $X = (x_1, \dots, x_n)$ получена из распределения Пуассона с неизвестным значением параметра $\lambda > 0$. Построить оценку по методу моментов параметра λ и исследовать ее на несмещенность, состоятельность и эффективность.
32. Случайное число ξ столкновений частиц, получающихся при распаде ядра урана в результате бомбардировки его нейтронами, с молекулами газа в камере Вильсона имеет распределение вероятностей

$$P\{\xi = x\} = \frac{1}{2} \left(e^{-\theta_1} \frac{\theta_1^x}{x!} + e^{-\theta_2} \frac{\theta_2^x}{x!} \right), \quad x = 0, 1, 2, \dots, \quad 0 < \theta_1 < \theta_2,$$

т. е. так называемое «двойное» распределение Пуассона. Методом моментов оценить вектор неизвестных значений параметров $\theta = (\theta_1, \theta_2)$.

33. Выборка $X = (x_1, \dots, x_n)$ получена из показательного (экспоненциального) распределения вероятностей с параметром $1/\sqrt{\lambda}$. Построить оценку параметра $\lambda > 0$ по методу моментов.
34. Выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей при неизвестном значении математического ожидания θ , значение дисперсии σ^2 задано. Построить асимптотически наикратчайший центральный доверительный интервал уровня 0,9 для параметра θ .
35. Выборка $X = (x_1, \dots, x_n)$ получена из экспоненциального распределения вероятностей при неизвестном значении параметра $\lambda > 0$. Построить асимптотически наикратчайший ЦДИ уровня $1 - \epsilon$ для этого параметра.
36. В трех наудачу выбранных единицах товара обнаружена только одна бракованная. Построить 95 % доверительный интервал для вероятности p события, состоящего в том, что наудачу выбранная единица товара окажется годной.
37. При выборочной проверке 10 компаний, занимающихся разработкой программного обеспечения, налоговая инспекция не выявила ни одного нарушения налогового законодательства. Построить 80 %, 90 % и 99 % доверительные интервалы для доли нарушителей налогового законодательства, считая, что для описания числа нарушителей адекватна биномиальная модель.

Глава 12

ТЕОРИЯ СТАТИСТИЧЕСКОЙ ПРОВЕРКИ ГИПОТЕЗ

12.1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ СТАТИСТИЧЕСКОЙ ПРОВЕРКИ ГИПОТЕЗ

Рассмотрим следующую статистическую задачу. Пусть имеется случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathcal{X} \subseteq \mathbb{R}^{nN}$ из некоторого N -мерного распределения вероятностей $P_\theta(\cdot)$, заданного на измеримом пространстве (Ω, \mathcal{F}) , где $\theta \in \Theta \subseteq \mathbb{R}^m$ – неизвестное истинное значение векторного параметра; Θ – параметрическое пространство; $\mathcal{X} \subseteq \mathbb{R}^{nN}$ – выборочное пространство. Задано некоторое натуральное число $K \geq 2$ и определено некоторое разбиение параметрического пространства Θ на K областей:

$$\Theta = \bigcup_{k=0}^{K-1} \Theta_k, \quad \Theta_k \cap \Theta_l = \emptyset, \quad k \neq l.$$

Определение 12.1. Утверждение о неизвестном значении параметра θ , состоящее в том, что $\theta \in \Theta_k$, называется k -й гипотезой:

$$H_k : \theta \in \Theta_k, \quad k = 0, \dots, K-1.$$

Определение 12.2. Гипотеза H_k называется простой, если мощность $|\Theta_k| = 1$, в противном случае – сложной (составной).

Определение 12.3. Пусть на самом деле значение параметра $\theta \in \Theta_\nu$, где $\nu \in \{0, \dots, K-1\}$, тогда говорят, что гипотеза H_ν истинна.

Задача статистической проверки гипотез H_0, \dots, H_{K-1} состоит в том, чтобы по наблюдаемой выборке X неким оптимальным образом оценить номер ν истинной гипотезы: $d = d(X) = k$ – выносим решение в пользу гипотезы H_k ($d = d(X)$ – статистическая оценка для ν). Возможно K решений ($k \in \{0, 1, \dots, K-1\}$). Множество возможных решений обозначим

$$D = \{0, 1, \dots, K-1\}, \quad |D| = K,$$

и назовем пространством решений.

Определение 12.4. Решающим правилом (решающей функцией, критерием, тестом) в вышеформулированной задаче статистической проверки гипотез называется функциональное отображение выборочного пространства \mathcal{X} в пространство решений D :

$$\mathcal{X} \xrightarrow{d(\cdot)} D. \quad (12.1)$$

Схема принятия решений в задаче проверки гипотез представлена на рис. 12.1. Существуют два основных типа решающих правил (РП) (12.1).

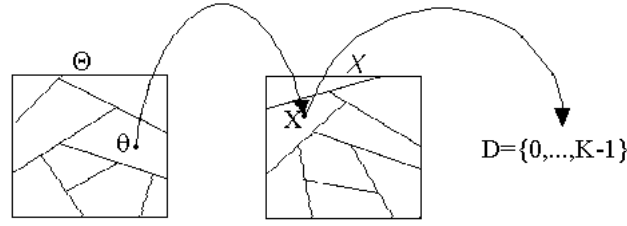


Рис. 12.1. Иллюстрация задачи проверки гипотез

Определение 12.5. Нерандомизированным РП называется отображение (12.1) следующего вида:

$$d = d(X) = \begin{cases} 0, & X \in \mathcal{X}_0, \\ \vdots & \\ K-1, & X \in \mathcal{X}_{K-1}, \end{cases}$$

где $\{\mathcal{X}_0, \dots, \mathcal{X}_{K-1}\}$ – некоторое детерминированное борелевское разбиение выборочного пространства:

$$\mathcal{X} = \bigcup_{k=0}^{K-1} \mathcal{X}_k, \quad \mathcal{X}_k \cap \mathcal{X}_l = \emptyset, \quad k \neq l.$$

При этом, если выборка X фиксирована, то решение $d = d(X)$ неслучайно.

Определение 12.6. Рандомизированным РП называется случайное отображение (12.1) следующего вида:

$$d = d(X, \omega), \quad \omega \in \Omega, \quad X \in \mathcal{X}, \quad d \in D,$$

причем если выборка X фиксирована, то решение $d = d(X, \omega)$ является дискретной случайной величиной с множеством значений D и некоторым дискретным распределением вероятностей:

$$\phi_i = \phi_i(X) = P\{d = i | X\}, \quad i \in D.$$

При этом борелевские функции $\phi_i = \phi_i(X)$, $i \in D$, удовлетворяют следующим ограничениям:

$$0 \leq \phi_i(X) \leq 1, \quad i \in D; \quad \sum_{i \in D} \phi_i(X) = 1, \quad X \in \mathcal{X},$$

и называются критическими функциями.

Укажем пошаговый алгоритм принятия решения с помощью рандомизированного решающего правила.

1. По выборке X вычисляем значения критических функций: $\phi_i = \phi_i(X)$, $i \in D$, и определяем дискретное распределение вероятностей $\{\phi_0(X), \phi_1(X), \dots, \phi_{K-1}(X)\}$.

2. Проводим случайный эксперимент (жребий) со множеством исходов D и дискретным распределением вероятностей, найденным на шаге 1.

3. Регистрируем исход k этого жребия и принимаем решение $d = k$.

Нерандомизированное решающее правило есть частный случай рандомизированного решающего правила, если критические функции принимают одно из двух возможных значений:

$$\phi_i(X) \in \{0, 1\}, \quad X \in \mathcal{X}; \quad \mathcal{X}_i = \{X : \phi_i(X) = 1\}, \quad i \in D.$$

12.2. РЕШАЮЩЕЕ ПРАВИЛО НЕЙМАНА – ПИРСОНА

Рассмотрим задачу проверки двух гипотез ($K = 2$). Пусть имеется случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$, образованная наблюдениями с плотностью распределения вероятностей $p(x; \theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta$, где θ – неизвестное истинное значение параметра. Обозначим:

$$p(X; \theta) = \prod_{l=1}^n p(x_l; \theta) \quad -$$

совместная плотность распределения всей выборки X ;

$$P_\theta(A) = \int_A p(X; \theta) dX, \quad A \in \mathcal{B}^{nN}, \quad -$$

вероятностная мера, определенная при значении параметра θ .

Пространство параметров Θ разбито на две непересекающиеся области:

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset.$$

Определены две гипотезы:

$$\begin{aligned} H_0 &: \theta \in \Theta_0, \\ H_1 &: \theta \in \Theta_1. \end{aligned}$$

Поскольку $H_1 = \overline{H_0}$, то H_1 называется *альтернативой*, а H_0 – *нулевой гипотезой*.

Задача заключается в построении решающего правила (теста) для проверки гипотез H_0, H_1 по выборке X . Построим РП общего вида – рандомизированное решающее правило:

$$d = d(X, \omega), \quad X \in \mathbb{R}^{nN}, \quad \omega \in \Omega, \quad d \in D = \{0, 1\}; \quad (12.2)$$

$$P\{d(X, \omega) = 1|X\} = \phi(X), \quad P\{d(X, \omega) = 0|X\} = 1 - \phi(X), \quad (12.3)$$

где $0 \leq \phi(X) \leq 1$ – некоторая произвольная критическая функция.

Существует бесконечно много РП $d(\cdot)$, различающихся критическими функциями $\phi(\cdot)$. Как найти оптимальную критическую функцию $\phi_*(\cdot)$? Сформулируем критерий оптимальности.

Определение 12.7. Условимся говорить, что при принятии решений имеет место ошибка I рода, если на самом деле верна гипотеза H_0 ($\nu = 0$), а принято решение $d = 1$ в пользу H_1 . При этом вероятностью ошибки I рода называется число

$$\begin{aligned} \alpha &= \alpha(\theta) ::= P_\theta\{d(X, \omega) = 1\} = \\ &= E_\theta\{\phi(X)\} = \int_{\mathbb{R}^{nN}} \phi(X) p(X; \theta) dX, \quad \theta \in \Theta_0. \end{aligned}$$

Определение 12.8. Принято говорить, что имеет место ошибка II рода, если на самом деле верна гипотеза H_1 ($\nu = 1$), а принято решение $d = 0$ в пользу H_0 . При этом вероятностью ошибки II рода называется число

$$\begin{aligned} \beta &= \beta(\theta) ::= P_\theta\{d(X, \omega) = 0\} = E_\theta\{1 - \phi(X)\} = \\ &= 1 - \int_{\mathbb{R}^{nN}} \phi(X) p(X; \theta) dX, \quad \theta \in \Theta_1. \end{aligned}$$

Определение 12.9. Мощностью решающего правила $d = d(X, \omega)$ называется вероятность правильного принятия альтернативы H_1 ($\gamma = d = 1$):

$$w = w(\theta) = \int_{\mathbb{R}^{nN}} \phi(X) p(X; \theta) dX, \quad \theta \in \Theta_1.$$

Отметим, что $w(\theta) = 1 - \beta(\theta)$.

Было бы целесообразно найти функцию $\phi_*(\cdot)$ так, что $\alpha = \beta = 0$. Очевидно, что в условиях регулярности это невозможно (как и в п. 11.5 невозможно было построить статистическую оценку с нулевой вариацией $V = 0$). Как поступать в этой ситуации?

Определение 12.10 (принцип оптимальности Неймана – Пирсона). Критическую функцию $\phi(X)$ в рандомизированном решающем правиле (12.2), (12.3) надлежит выбирать таким образом, чтобы вероятность ошибки I рода не превосходила некоторого наперед заданного числа $\epsilon \in (0, 1)$, а вероятность ошибки II рода была минимальной, т. е. выбирать как решение следующей экстремальной задачи:

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \epsilon, \quad \sup_{\theta \in \Theta_1} \beta(\theta) \rightarrow \min_{0 \leq \phi(\cdot) \leq 1},$$

что эквивалентно

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \epsilon, \quad \inf_{\theta \in \Theta_1} w(\theta) \rightarrow \max_{0 \leq \phi(\cdot) \leq 1}. \quad (12.4)$$

При этом оптимальная критическая функция $\phi_*(X)$ и соответствующее ей решающее правило $d_*(X, \omega)$ называются соответственно критической функцией Неймана – Пирсона и решающим правилом Неймана – Пирсона, величина $\sup_{\theta \in \Theta_0} \alpha(\theta)$ – размером теста, а ϵ – уровнем значимости теста.

Отметим, что экстремальная задача (12.4) – задача вариационного исчисления. Нейман и Пирсон решили эту задачу лишь для случая, когда H_0, H_1 – простые гипотезы: $\Theta = \{\theta_0, \theta_1\}$, а $\Theta_0 = \{\theta_0\}$ и $\Theta_1 = \{\theta_1\}$ – одноточечные множества.

Упростим обозначения:

$$H_i : \theta = \theta_i, \quad i = 0, 1; \quad \theta_0, \theta_1 \in \Theta, \quad \theta_0 \neq \theta_1;$$

$$p_i(X) = p(X; \theta_i), \quad X \in \mathbb{R}^{nN};$$

$$P_i(B) = P_{\theta_i}(B), \quad Q(B) = \frac{1}{2}(P_0(B) + P_1(B)), \quad B \in \mathcal{B}^{nN};$$

$$E_i\{\cdot\} \equiv E_{\theta_i}\{\cdot\}.$$

Определение 12.11. Статистикой отношения правдоподобия для проверки простых гипотез H_0, H_1 называется статистика

$$L = L(X) = \frac{p_1(X)}{p_0(X)} \geq 0.$$

Теорема 12.1 (Неймана – Пирсона). Для вышеформулированной задачи проверки простых гипотез H_0, H_1 справедливы следующие три утверждения:

1) для любого наперед заданного ϵ , $0 < \epsilon < 1$, найдутся такие постоянные $c_* > 0$, $\kappa_* \in [0, 1]$, что решающее правило (тест) Неймана – Пирсона $d = d_*(X, \omega)$ с критической функцией

$$\phi_*(X) = \begin{cases} 0, & L(X) < c_*; \\ \varkappa_*, & L(X) = c_*; \\ 1, & L(X) > c_*, \end{cases} \quad (12.5)$$

имеет размер, в точности равный ϵ ;

2) тест Неймана – Пирсона $d = d_*(X, \omega)$, определяемый (12.5), имеет наибольшую мощность w_* среди всех тестов, размер которых не превосходит ϵ ;

3) тест $d = d_*(X, \omega)$ – единственный с точностью до множества B нулевой меры: $Q(B) = 0$.

Доказательство. Запишем экстремальную задачу (12.4) с учетом определений 12.7, 12.9 в явном виде:

$$\begin{aligned} \alpha &= \int \phi(X) p_0(X) dX \leq \epsilon, \\ w &= \int_{\mathbb{R}^{nN}} \phi(X) p_1(X) dX \rightarrow \max_{\phi(\cdot)}. \end{aligned} \quad (12.6)$$

Доказательство разобьем на три части.

1. Введем в рассмотрение функцию распределения статистики $L = L(X)$ при верной гипотезе H_0 :

$$F_0(y) = P_0\{L(X) < y\}, \quad y \geq 0.$$

Считая произвольными c_* и \varkappa_* , вычислим для теста (12.5) вероятность ошибки I рода:

$$\begin{aligned} \alpha &= E_0\{\phi(X)\} = \varkappa_* P_0\{L = c_*\} + P_0\{L > c_*\} = \\ &= \varkappa_*(F_0(c_* + 0) - F_0(c_*)) + (1 - F_0(c_* + 0)). \end{aligned} \quad (12.7)$$

В этом соотношении подберем произвольные константы c_* , \varkappa_* таким образом, чтобы $\alpha = \alpha_* = \epsilon$. Для этого выберем c_* так, чтобы

$$F_0(c_*) \leq 1 - \epsilon \leq F_0(c_* + 0).$$

Возможны два случая:

а) c_* – точка непрерывности функции $F_0(\cdot)$; тогда

$$c_* = F_0^{-1}(1 - \epsilon),$$

и из (12.7) имеем

$$\alpha_* = \epsilon \quad \forall \varkappa_*;$$

б) c_* – точка разрыва; в этом случае выберем

$$\varkappa_* = \frac{F_0(c_* + 0) - (1 - \epsilon)}{F_0(c_* + 0) - F_0(c_*)}.$$

Подставляя это значение в (12.7), получаем

$$\alpha_* = \epsilon.$$

2. Для критической функции $\phi_*(\cdot)$ имеем

$$\alpha_* = \epsilon, \quad w_* = \int_{\mathbb{R}^{nN}} \phi_*(X) p_1(X) dX. \quad (12.8)$$

Выберем произвольную критическую функцию $\tilde{\Phi}(X)$, для которой обозначим

$$\tilde{\alpha} = \int_{\mathbb{R}^{nN}} \tilde{\Phi}(X) p_0(X) dX \leq \epsilon, \quad \tilde{w} = \int_{\mathbb{R}^{nN}} \tilde{\Phi}(X) p_1(X) dX.$$

Покажем, что: $\tilde{w} \leq w_*$. Легко проверить поточечно, используя (12.5), что

$$\Delta = \Delta(X) ::= (p_1(X) - c_* p_0(X))(\Phi_*(X) - \tilde{\Phi}(X)) \geq 0, \quad \forall X \in \mathbb{R}^{nN}.$$

Отсюда с учетом (12.8) получаем

$$\int_{\mathbb{R}^{nN}} \Delta(X) dX = w_* - c_* \alpha_* - \tilde{w} + c_* \tilde{\alpha} = w_* - \tilde{w} - c_*(\epsilon - \tilde{\alpha}) \geq 0.$$

Следовательно,

$$w_* \geq \tilde{w} + c_*(\epsilon - \tilde{\alpha}) \geq \tilde{w}.$$

3. В экстремальной задаче (12.6) значения интегралов не изменятся, если значение $\Phi_*(X)$ изменить на множестве B : $Q(B) = 0$. □

Следствие 12.1. Если функция распределения $F_0(\cdot)$ статистики отношения правдоподобия $L = L(X)$ при гипотезе H_0 — непрерывная функция, то РП Неймана — Пирсона является нерандомизированным и имеет следующий вид:

$$d = d_*(X) = \begin{cases} 0, & L(X) < c_*, \\ 1, & L(X) \geq c_*, \end{cases}$$

где $c_* = F_0^{-1}(1 - \epsilon)$.

Доказательство. Обратимся к первому пункту доказанной выше теоремы 12.1. Если $F_0(\cdot)$ непрерывна, то \varkappa_* можно выбирать произвольно на отрезке $[0, 1]$, поэтому выберем $\varkappa_* = 1$, тогда согласно (12.5) $\Phi_*(X) \in \{0; 1\}$, т. е. жребий исчезает и имеет место детерминированное (нерандомизированное) решающее правило. □

12.3. БАЙЕСОВСКОЕ РЕШАЮЩЕЕ ПРАВИЛО

Предположим, что параметр θ — случайная величина, принимающая одно из двух возможных значений:

$$\theta \in \Theta = \{\theta_0, \theta_1\}; \quad P\{\theta = \theta_i\} = \Pi_i, \quad 0 < \Pi_i < 1, \quad i = 0, 1; \quad \Pi_0 + \Pi_1 = 1.$$

Наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого распределения вероятностей с условной плотностью $p(x|\theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta$.

Обозначим

$$p_i(X) = \prod_{j=1}^n p(x_j|\theta_i), \quad i = 0, 1, \quad -$$

условная плотность распределения выборки X при условии, что $\theta = \theta_i$.

Истинное значение θ неизвестно, и определены две простые гипотезы:

$$H_i : \theta = \theta_i, \quad i = 0, 1.$$

Задача заключается в построении теста для проверки H_0, H_1 по выборке X .

Эту задачу можно решить с помощью теста Неймана – Пирсона, но в таком случае игнорируется информация о $\{\Pi_i\}$. Поэтому рассмотрим другой подход (см. для сравнения п. 11.8).

Построим рандомизированное решающее правило:

$$\begin{aligned} d &= d(X, \omega) \in D = \{0, 1\}, \quad X \in \mathbb{R}^N, \quad \omega \in \Omega; \\ P\{d(X, \omega) = 1 | X\} &= \phi(X), \quad P\{d(X, \omega) = 0 | X\} = 1 - \phi(X), \end{aligned} \quad (12.9)$$

где $\phi(X)$ – произвольная критическая функция ($0 \leq \phi(X) \leq 1$).

Обозначим: $\nu = \nu(\omega) \in \{0, 1\}$ – случайная величина Бернулли – номер истинной гипотезы H_ν . В силу случайности θ

$$P(H_i) = P\{\theta = \theta_i\} = \Pi_i, \quad i = 0, 1,$$

поэтому Π_i принято называть *априорной вероятностью* i -й гипотезы.

Определение 12.12. *Функцией потерь в рассматриваемой задаче проверки двух гипотез H_0, H_1 называется функция двух переменных:*

$$w = w(i, j) \geq 0, \quad i, j \in D = \{0, 1\},$$

где $w(i, j)$ – величина потерь, которые несет статистик в ситуации, когда на самом деле $\nu = i$ (верна H_i), а принято решение $d = j$ в пользу гипотезы H_j .

Определение 12.13. *Принято говорить, что имеет место $(0 - 1)$ -функция потерь, если*

$$w(i, j) = 1 - \delta_{ij} = \begin{cases} 0, & i = j; \\ 1, & i \neq j. \end{cases}$$

Функцию потерь удобно задавать в виде матрицы потерь: $W = (w_{ij})$, $w_{ij} = w(i, j)$. В случае $(0 - 1)$ -матрицы потерь имеем

$$W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Определение 12.14. *Функционалом риска называется математическое ожидание случайных потерь (средние потери)*

$$r = r(\phi(\cdot)) = E\{w(\nu, d(X, \omega))\} \geq 0. \quad (12.10)$$

Определение 12.15 (байесовский принцип оптимальности). *Критическую функцию $\phi(\cdot)$ в рандомизированном решающем правиле (12.9) надлежит выбирать таким образом, чтобы функционал риска (12.10) достигал минимального значения:*

$$r(\phi^*(\cdot)) = \inf_{\phi(\cdot)} r(\phi(\cdot)). \quad (12.11)$$

При этом критическая функция $\phi^*(\cdot)$, определяемая (12.11), называется *байесовской критической функцией*, а соответствующее решающее правило $d^*(X, \omega)$, определяемое (12.9), – *байесовским решающим правилом (БРП)*.

Теорема 12.2. *Пусть в сформулированной выше задаче проверки простых гипотез H_0, H_1 функция потерь имеет следующий вид:*

$$w(i, j) = \begin{cases} 0, & i = j, \\ w_0, & i = 0, \quad j = 1, \\ w_1, & i = 1, \quad j = 0, \end{cases} \quad (12.12)$$

где $w_0 > 0$, $w_1 > 0$ – некоторые заданные величины.

Тогда байесовская критическая функция задается соотношением ($X \in \mathbb{R}^{nN}$):

$$\Phi^*(X) = \begin{cases} 0, & L(X) < C^*, \\ \varkappa^*, & L(X) = C^*, \\ 1, & L(X) > C^*, \end{cases} \quad (12.13)$$

где

$$L(X) = \frac{p_1(X)}{p_0(X)} \geq 0, \quad C^* = \frac{\Pi_0 w_0}{\Pi_1 w_1} \geq 0, \quad \varkappa^* \in [0, 1]. \quad (12.14)$$

Доказательство. Прежде всего заметим, что в силу произвола в выборе \varkappa^* байесовская критическая функция неоднозначна. Сформулируем экстремальную задачу (12.11) в явном виде и решим ее. Из (12.10) и (12.12) имеем

$$\begin{aligned} r &= r(\Phi(\cdot)) = \\ &= 0 \cdot P\{d(X, \omega) = \nu\} + w_0 \cdot P\{\nu = 0, d = 1\} + w_1 \cdot P\{\nu = 1, d = 0\} = \\ &= w_0 \cdot P\{\nu = 0\} \cdot P\{d = 1|H_0\} + w_1 \cdot P\{H_1\} \cdot P\{d = 0|H_1\} = \\ &= w_0 \cdot \Pi_0 \cdot \alpha + w_1 \cdot \Pi_1 \cdot \beta. \end{aligned}$$

Используя интегральные представления вероятностей ошибок α , β из п. 12.2, приходим к задаче:

$$\begin{aligned} r &= r(\Phi(\cdot)) = \\ &= \Pi_0 w_0 \int_{\mathbb{R}^{nN}} \Phi(X) p_0(X) dX + \Pi_1 w_1 \int_{\mathbb{R}^{nN}} (1 - \Phi(X)) p_1(X) dX = \\ &= \Pi_1 w_1 - \int_{\mathbb{R}^{nN}} \Phi(X) G(X) dX \rightarrow \min_{\Phi(\cdot)}, \end{aligned} \quad (12.15)$$

где $G(X) = \Pi_1 w_1 p_1(X) - \Pi_0 w_0 p_0(X)$.

Из (12.15) получаем эквивалентную экстремальную задачу:

$$\int_{\mathbb{R}^{nN}} \Phi(X) G(X) dX \rightarrow \max_{\Phi(\cdot)},$$

являющуюся линейной задачей вариационного исчисления, причем ограничения носят точечный, локальный характер: $0 \leq \Phi(X) \leq 1$, $X \in \mathbb{R}^{nN}$, поэтому решение очевидно:

$$\Phi^*(X) = \begin{cases} 0, & G(X) < 0, \\ \varkappa^*, & G(X) = 0, \\ 1, & G(X) > 0. \end{cases}$$

С учетом обозначений (12.14) последнее соотношение эквивалентно (12.13). Найдём минимум риска, который при этом достигается:

$$r^* = r(\Phi^*) = \Pi_1 w_1 - \int_{G(X) > 0} G(X) dX.$$

□

Следствие 12.2. Среди байесовских решающих правил (12.13) существует нерандомизированное решающее правило:

$$d = d^*(X) = \begin{cases} 0, & L(X) < C^*, \\ 1, & L(X) \geq C^*. \end{cases}$$

Доказательство. Для доказательства достаточно выбрать $\varkappa^* = 1$. \square

Следствие 12.3. Если имеет место $(0-1)$ -функция потерь, т. е. $w_0 = w_1 = 1$, и гипотезы H_0, H_1 равновероятны $P_0 = P_1 = 1/2$, то БРП имеет вид

$$d = d^*(X) = \begin{cases} 0, & p_0(X) > p_1(X), \\ 1, & p_1(X) \geq p_0(X). \end{cases} \quad (12.16)$$

Доказательство. Из формулы (12.14) имеем $C^* = 1, \varkappa^* = 1$. \square

Заметим в заключение, что решающее правило (12.16) часто называют *тестом максимального правдоподобия*.

12.4. ПРОВЕРКА ГИПОТЕЗ О РАСПРЕДЕЛЕНИИ ВЕРОЯТНОСТЕЙ. КРИТЕРИИ СОГЛАСИЯ

Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ из одномерного распределения вероятностей с неизвестной функцией распределения $F(x)$, $x \in \mathbb{R}^1$. Пусть $F_0(x)$, $x \in \mathbb{R}^1$, – некоторая гипотетическая (предполагаемая) функция распределения вероятностей. В связи с этим определены две гипотезы:

$$\begin{aligned} H_0 &: F(\cdot) \equiv F_0(\cdot); \\ H_1 = \overline{H_0} &: F(\cdot) \neq F_0(\cdot). \end{aligned}$$

Нулевая гипотеза H_0 – простая, альтернатива H_1 – сложная гипотеза.

Определение 12.16. Гипотеза H_0 означает, что выборка X согласуется с распределением, задаваемым гипотетической функцией распределения вероятностей $F_0(\cdot)$, и поэтому называется гипотезой согласия; решающее правило $d = d(X)$ для проверки гипотез H_0, H_1 – критерием согласия.

12.4.1. χ^2 -Критерий согласия Пирсона

Воспользуемся той же схемой группировки выборки X , что и при построении гистограммы (см. п. 11.2).

Разобьем числовую прямую на $K \geq 2$ ячеек точками деления $-\infty = b_0 < b_1 < b_2 < \dots < b_K = +\infty$:

$$\Gamma_k = [b_{k-1}, b_k), \quad k = 1, 2, \dots, K.$$

Вычислим два набора величин:

1) теоретические вероятности попадания случайного наблюдения x_i в ячейки $\{\Gamma_k\}$ при условии, что верна H_0 ($k = 1, \dots, K$):

$$p_k^0 = P_{H_0}\{x_i \in \Gamma_k\} = P_{H_0}\{b_{k-1} \leq x_i < b_k\} = F_0(b_k) - F_0(b_{k-1}) > 0;$$

$$\sum_{k=1}^K p_k^0 = 1;$$

2) относительные и абсолютные частоты попадания в ячейки, вычисленные по выборке X :

$$\hat{p}_k = \frac{\mathbf{v}_k}{n}, \quad \mathbf{v}_k = \sum_{i=1}^n \mathbb{I}_{\Gamma_k}(x_i), \quad k = 1, \dots, K; \quad \sum_{k=1}^K \mathbf{v}_k = n, \quad \sum_{k=1}^K \hat{p}_k = 1.$$

Определение 12.17. χ^2 -Статистикой Пирсона называется статистика

$$\chi^2 = \sum_{k=1}^K \frac{(\mathbf{v}_k - np_k^0)^2}{np_k^0} = n \sum_{k=1}^K \frac{(\hat{p}_k - p_k^0)^2}{p_k^0} \geq 0. \quad (12.17)$$

Из (12.17) видно, что χ^2 – это с точностью до множителя n взвешенная с весами $\{1/p_k^0\}$ сумма квадратов отклонений частот $\{\hat{p}_k\}$ от теоретических значений $\{p_k^0\}$.

Лемма 12.1. Если верна гипотеза H_0 , то при $n \rightarrow \infty$ имеет место сходимость к χ^2 -распределению с $K-1$ степенью свободы:

$$\mathcal{L}_{H_0}\{\chi^2\} \rightarrow \chi_{K-1}^2,$$

или

$$P_{H_0}\{\chi^2 < y\} \rightarrow F_{\chi_{K-1}^2}(y), \quad n \rightarrow +\infty, \quad y \geq 0, \quad (12.18)$$

где $F_{\chi_{K-1}^2}(\cdot)$ – функция χ^2 -распределения с $K-1$ степенью свободы.

Доказательство. Обозначим нормированные частоты, входящие в (12.17):

$$\mathbf{v}_k^*(n) = \frac{\mathbf{v}_k - np_k^0}{\sqrt{n}}, \quad k = 1, \dots, K; \quad \sum_{k=1}^K \mathbf{v}_k^*(n) \equiv 0. \quad (12.19)$$

Определим $(K-1)$ -вектор нормированных частот:

$$\mathbf{v}^*(n) = (\mathbf{v}_1^*(n), \dots, \mathbf{v}_{K-1}^*(n))^T \in \mathbb{R}^{K-1}.$$

Заметим, что в силу условия нормировки в (12.19) $\mathbf{v}_K^*(n)$ линейно зависит от $\mathbf{v}^*(n)$:

$$\mathbf{v}_K^*(n) = - \sum_{l=1}^{K-1} \mathbf{v}_l^*(n). \quad (12.20)$$

В асимптотике $n \rightarrow +\infty$ воспользуемся центральной предельной теоремой для $\mathbf{v}^*(n)$ (многомерный аналог теоремы Муавра – Лапласа [25]):

$$\mathcal{L}_{H_0}\{\mathbf{v}^*(n)\} \rightarrow N_{K-1}(\mathbf{0}_{K-1}, \Sigma), \quad (12.21)$$

где $\Sigma = (\sigma_{ij})$ – асимптотическая ковариационная матрица, для элементов которой справедлива формула $(i, j = 1, \dots, K-1)$

$$\sigma_{ij} = p_i^0(\delta_{ij} - p_j^0).$$

Непосредственной проверкой можно убедиться, что в этом случае для обратной матрицы справедливо выражение

$$\Sigma^{-1} = (\bar{\sigma}_{ij}), \quad \bar{\sigma}_{ij} = \frac{1}{p_K^0} + \frac{\delta_{ij}}{p_i^0}. \quad (12.22)$$

Из (12.17), (12.19)–(12.22) имеем представление для χ^2 -статистики в виде квадратичной формы:

$$\chi^2 = \sum_{k=1}^{K-1} \frac{(\mathbf{v}_k^*(n))^2}{p_k^0} + \frac{\left(\sum_{l=1}^{K-1} \mathbf{v}_l^*(n)\right)^2}{p_K^0} = \sum_{k,l=1}^{K-1} \mathbf{v}_k^*(n) \mathbf{v}_l^*(n) \bar{\sigma}_{kl} = (\mathbf{v}^*(n))^T \Sigma^{-1} \mathbf{v}^*(n).$$

Введем в рассмотрение случайный вектор

$$\xi(n) = (\xi_k(n)) = (\Sigma^{1/2})^{-1} \mathbf{v}^*(n) \in \mathbb{R}^{K-1},$$

где матрица $\Sigma^{1/2}$ – решение по Y матричного уравнения $Y^T \Sigma^{-1} Y = I_{K-1}$ ($\Sigma = \Sigma^{1/2} (\Sigma^{1/2})^T$). Тогда χ^2 -статистика представима в виде суммы квадратов:

$$\chi^2 = \xi^T(n) \xi(n) = \sum_{k=1}^{K-1} \xi_k^2(n),$$

причем в силу (12.21) $\mathcal{L}_{H_0}\{\xi(n)\} \rightarrow N_{K-1}(\mathbf{0}_{K-1}, I_{K-1})$. Отсюда по определению χ^2 -распределения получаем (12.18). \square

Определение 12.18. χ^2 -Критерий согласия Пирсона имеет вид

$$d = d(X) = \begin{cases} 0, & \chi^2 < F_{\chi_{K-1}^2}^{-1}(1 - \epsilon), \\ 1, & \chi^2 \geq F_{\chi_{K-1}^2}^{-1}(1 - \epsilon), \end{cases} \quad (12.23)$$

где $F_{\chi_{K-1}^2}^{-1}(1 - \epsilon)$ – квантиль χ^2 -распределения с $K - 1$ степенью свободы уровня $1 - \epsilon$, $0 < \epsilon < 1$.

Теорема 12.3. При $n \rightarrow \infty$ асимптотический размер теста Пирсона (12.23) совпадает с наперед заданным уровнем значимости ϵ :

$$\alpha = P_{H_0}\{d(X) = 1\} \rightarrow \epsilon, \quad n \rightarrow +\infty.$$

Доказательство. Основано на лемме 12.1 и состоит в вычислении α с использованием (12.18). \square

Замечание 12.1. χ^2 -Критерий согласия Пирсона справедлив и для N -мерных ($N > 1$) функций распределения $F(x)$, $x \in \mathbb{R}^N$.

Замечание 12.2. χ^2 -Критерий Пирсона допускает обобщение для случая сложных гипотез согласия:

$$H_0 : F(\cdot) \in F_0,$$

где

$$F_0 = \{F_0(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\} -$$

некоторое параметрическое семейство функций распределения с m -мерным параметром θ . Появляется дополнительное ограничение: $K > m + 1$, и критерий согласия имеет вид

$$d = d^0(X) = \begin{cases} 0, & \inf_{\theta} \chi^2(\theta) < F_{\chi_{K-m-1}^2}^{-1}(1 - \epsilon), \\ 1, & \inf_{\theta} \chi^2(\theta) \geq F_{\chi_{K-m-1}^2}^{-1}(1 - \epsilon), \end{cases}$$

где

$$\chi^2(\theta) = \sum_{k=1}^K \frac{(\nu_k - np_k(\theta))^2}{np_k(\theta)} \geq 0, \quad p_k(\theta) = P_\theta\{x_i \in \Gamma_k\}, \quad k = 1, \dots, K.$$

Заметим также, что статистика

$$\tilde{\theta} = \arg \min_{\theta} \chi^2(\theta)$$

называется χ^2 -оценкой параметра θ .

12.4.2. Критерий Колмогорова

Будем предполагать, что гипотеза согласия $H_0 : F(\cdot) \equiv F_0(\cdot)$ – простая, причем функция $F_0(\cdot)$ непрерывна. Построим по наблюдаемой выборке X выборочную функцию распределения, изученную в п. 11.2 (рис. 12.2):

$$\hat{F}(x) = F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x - x_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x - x_{(i)}), \quad x \in \mathbb{R}^1,$$

где $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ – вариационный ряд выборки.

Как было установлено в п. 11.2, выборочная функция распределения – сильно состоятельная оценка истинной функции распределения:

$$F_n(x) \xrightarrow{п. н.} F(x), \quad n \rightarrow +\infty, \quad x \in \mathbb{R}^1.$$

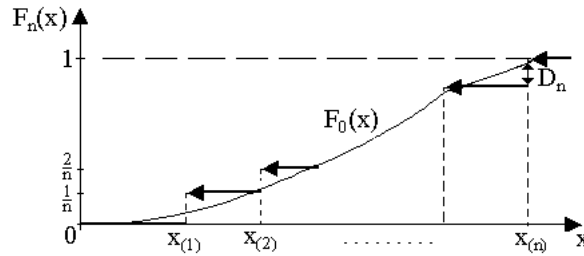


Рис. 12.2. Иллюстрация критерия Колмогорова

Определение 12.19. Расстоянием Колмогорова между выборочной функцией распределения $F_n(x)$ и гипотетической функцией распределения $F_0(x)$ называется величина

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)| \quad (0 \leq D_n \leq 1). \quad (12.24)$$

Лемма 12.2. Если верна гипотеза H_0 , то случайные величины $u_1 = F_0(x_1)$, $u_2 = F_0(x_2)$, \dots , $u_n = F_0(x_n)$ независимы и одинаково распределены со стандартным равномерным распределением:

$$\mathcal{L}_{H_0}\{u_i\} = \mathcal{R}[0, 1], \quad i = 1, \dots, n.$$

Доказательство. Независимость $\{u_i\}$ вытекает из независимости $\{x_i\}$ и соответствующей теоремы о борелевских функциональных преобразованиях. Вычислим функцию распределения для случайной величины u_i при условии, что верна гипотеза H_0 ($0 \leq y \leq 1$):

$$F_{u_i}(y) ::= P_{H_0}\{u_i < y\} = P_{H_0}\{x_i < F_0^{-1}(y)\} = F_0(F_0^{-1}(y)) = y.$$

Это и есть функция распределения стандартного равномерного закона $\mathcal{R}[0, 1]$. \square

Теорема 12.4. Если верна нулевая гипотеза H_0 , то распределение вероятностей статистики Колмогорова D_n , определяемой формулой (12.24), не зависит от гипотетической функции распределения $F_0(\cdot)$:

$$\mathcal{L}_{H_0}\{D_n\} = \mathcal{L}\{D_n^*\}, \quad D_n^* = \sup_{0 \leq u \leq 1} |\Psi_n(u) - u|;$$

$$\Psi_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(u - u_i),$$

где $\{u_i\}$ – независимые в совокупности случайные величины, имеющие стандартное равномерное распределение: $\mathcal{L}\{u_i\} = \mathcal{R}[0, 1]$, $i = 1, \dots, n$.

Доказательство. Вытекает из леммы 12.2 и замены переменных $x = F_0^{-1}(u)$, $u = F_0(x)$ в экстремальной задаче (12.24). \square

Теорема 12.5. Если верна гипотеза H_0 , то при бесконечном увеличении объема выборки n ($n \rightarrow +\infty$) функция распределения нормированной статистики $\sqrt{n}D_n$ сходится к распределению Колмогорова:

$$P_{H_0}\{\sqrt{n}D_n < z\} \rightarrow K(z) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 z^2}, \quad z \geq 0.$$

Определение 12.20. Пусть ϵ – некоторое число: $0 < \epsilon < 1$, а $K^{-1}(1 - \epsilon)$ – квантиль распределения Колмогорова уровня $1 - \epsilon$. Тогда тест

$$d = d(X) = \begin{cases} 0, & \sqrt{n}D_n < K^{-1}(1 - \epsilon), \\ 1, & \sqrt{n}D_n \geq K^{-1}(1 - \epsilon) \end{cases} \quad (12.25)$$

называется критерием согласия Колмогорова.

Замечание 12.3. Распределение статистики D_n не зависит от гипотетической функции распределения $F_0(\cdot)$, в результате чего имеет место универсальность теста Колмогорова.

Следствие 12.4. При $n \rightarrow \infty$ асимптотический размер теста Колмогорова (12.25) совпадает с наперед заданным уровнем значимости $\epsilon \in (0, 1)$.

Доказательство. Основывается на соотношениях (12.25), теореме 12.5 и заключается в проверке следующего факта: $\alpha \rightarrow \epsilon$, где α – вероятность ошибки I рода. \square

12.5. КРИТЕРИЙ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ ДЛЯ ПРОВЕРКИ СЛОЖНЫХ ГИПОТЕЗ

В п. 12.4 были рассмотрены методы проверки специальных сложных гипотез – гипотез согласия. Рассмотрим теперь универсальный критерий проверки сложных гипотез – критерий отношения правдоподобия, суть которого состоит в следующем.

Пусть наблюдается случайная выборка $X \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с плотностью $p(x; \theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta \subseteq \mathbb{R}^m$. Определены две, в общем случае сложные, гипотезы:

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \subset \Theta; \\ H_1 &: \theta \in \Theta_1 = \Theta \setminus \Theta_0. \end{aligned}$$

Аналогично критерию Неймана – Пирсона введем в рассмотрение статистику отношения правдоподобия:

$$\tilde{\Lambda}_n = \tilde{\Lambda}_n(X; \Theta_0) = \frac{\sup_{\theta \in \Theta_1} p(X; \theta)}{\sup_{\theta \in \Theta_0} p(X; \theta)}, \quad (12.26)$$

где

$$p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) = L(\theta) \quad -$$

функция правдоподобия. Часто рассматривают эквивалентные статистики:

$$\Lambda_n = \Lambda_n(X; \Theta_0) = \frac{\sup_{\theta \in \Theta} p(X; \theta)}{\sup_{\theta \in \Theta_0} p(X; \theta)} \geq 1; \quad (12.27)$$

$$\lambda_n = \lambda_n(X; \Theta_0) = \frac{1}{\Lambda_n(X; \Theta_0)} = \frac{\sup_{\theta \in \Theta_0} p(X; \theta)}{\sup_{\theta \in \Theta} p(X; \theta)} \in [0, 1].$$

Статистики (12.27) взаимно-однозначно функционально связаны между собой и статистикой отношения правдоподобия (12.26):

$$\Lambda_n = \max\{1, \tilde{\Lambda}_n\}, \quad \lambda_n = \min\left\{1, \frac{1}{\tilde{\Lambda}_n}\right\} \in [0, 1].$$

Определение 12.21. Критерием отношения правдоподобия (КОП) в задаче проверки сложных гипотез H_0, H_1 по выборке X объема n называется следующий статистический критерий ($C \in [0, 1]$):

$$d = d(X) = \begin{cases} 0, & \lambda_n(X; \Theta_0) > C, \\ 1, & \lambda_n(X; \Theta_0) \leq C, \end{cases}$$

или в эквивалентном виде

$$d = d(X) = \begin{cases} 0, & -2 \ln \lambda_n(X; \Theta_0) < \delta, \\ 1, & -2 \ln \lambda_n(X; \Theta_0) \geq \delta, \end{cases} \quad (12.28)$$

где критическое (пороговое) значение $\delta \geq 0$ выбирается так, чтобы критерий имел наперед заданный уровень значимости ϵ :

$$\begin{aligned} P_{\theta}\{d(X) = 1\} &= \int_{\lambda_n(X; \Theta_0) \leq C} p(X; \theta) dX = \\ &= P_{\theta}\{-2 \ln \lambda_n \geq \delta\} \leq \epsilon, \quad \forall \theta \in \Theta_0. \end{aligned} \quad (12.29)$$

Как было показано в п. 12.2 (теорема 12.1, известная как фундаментальная лемма Неймана – Пирсона), для простых гипотез КОП оптимален при любом n . Для сложных гипотез это, вообще говоря, не так. Тем не менее КОП широко применяется на практике. Оказывается, что он асимптотически оптимален при $n \rightarrow \infty$.

Рассмотрим вначале случай простой гипотезы H_0 и сложной альтернативы H_1 .

Теорема 12.6. Пусть определены простая гипотеза $H_0 : \theta = \theta_0$, где $\theta_0 = (\theta_{0i}) \in \Theta$ – фиксированная точка m -мерной области Θ , и сложная альтернатива $H_1 : \theta \neq \theta_0$, тогда при выполнении условий регулярности, принятых в теории оценок максимального правдоподобия (см. п. 11.7), асимптотический ($n \rightarrow \infty$) размер КОП (12.28), в котором $\delta = F_{\chi_m^2}^{-1}(1 - \epsilon)$ – квантиль уровня $1 - \epsilon$ для χ^2 -распределения с m степенями свободы, совпадает с ϵ :

$$P_{\theta_0}\{d(X) = 1\} = P_{\theta_0}\{-2 \ln \lambda_n(X; \theta_0) \geq \delta\} \rightarrow \epsilon. \quad (12.30)$$

Доказательство. В силу (12.29) для доказательства (12.30) достаточно показать сходимости к χ^2 -распределению:

$$\mathcal{L}_{H_0}\{-2 \ln \lambda_n(X; \theta_0)\} \rightarrow \chi_m^2, \quad n \rightarrow \infty,$$

где из (12.27) следует

$$-2 \ln \lambda_n(X; \theta_0) = 2 \left(l(\hat{\theta}) - l(\theta_0) \right); \quad (12.31)$$

$l(\theta) = \ln p(X; \theta)$ – логарифмическая функция правдоподобия; $\hat{\theta} = \arg \max_{\theta \in \Theta} p(X; \theta)$ – ОМП, которая в условиях регулярности обладает свойством сильной состоятельности:

$$\hat{\theta} \xrightarrow{p. n.} \theta^0, \quad n \rightarrow \infty.$$

Следовательно, $\forall \epsilon > 0 \exists \bar{n} = \bar{n}(\epsilon)$, что $\forall n \geq \bar{n}$ с вероятностью единица: $|\hat{\theta} - \theta_0| \leq \epsilon$. В силу условий регулярности применим к $l(\theta_0)$ в (12.31) квадратичную формулу Тейлора с остаточным членом в форме Лагранжа (учитывая, что $\nabla_{\theta} l(\theta)|_{\theta=\hat{\theta}} = \mathbf{0}_m$):

$$-2 \ln \lambda_n(X; \theta_0) = n(\hat{\theta} - \theta_0)^T \left(-\frac{1}{n} \nabla_{\theta^*}^2 l(\theta^*) \right) (\hat{\theta} - \theta_0).$$

Здесь θ^* – промежуточная точка: $|\theta^* - \theta_0| < |\hat{\theta} - \theta_0|$, следовательно, $\theta^* \xrightarrow{p. n.} \theta_0$. В силу условий регулярности выполняется усиленный закон больших чисел:

$$-\frac{1}{n} \nabla_{\theta^*}^2 l(\theta^*) = \frac{1}{n} \sum_{k=1}^n (-\nabla_{\theta^*}^2 \ln p(x_k; \theta^*)) \xrightarrow{p. n.} E_{\theta_0} \{-\nabla_{\theta_0}^2 \ln p(x_k; \theta_0)\} = \mathcal{J}(\theta_0),$$

где $\mathcal{J} = \mathcal{J}(\theta_0)$ – информационная матрица Фишера для выборочного значения. Из свойств ОМП

$$\mathcal{L}_{H_0} \left\{ \sqrt{n} \left(\hat{\theta} - \theta_0 \right) \right\} \rightarrow N_m(\mathbf{0}_m, \mathcal{J}^{-1}(\theta_0)), \quad n \rightarrow \infty,$$

и по свойствам линейного преобразования многомерного нормального распределения (см. п. 14.1)

$$\mathcal{L}_{H_0} \left\{ \sqrt{n} \left(\mathcal{J}^{\frac{1}{2}}(\theta_0) \right)^T \left(\hat{\theta} - \theta_0 \right) \right\} = \mathcal{L}\{\xi_n\} \rightarrow N_m(\mathbf{0}_m, I_m) = \mathcal{L}\{\xi\}.$$

В результате имеем

$$\mathcal{L}_{H_0}\{-2 \ln \lambda_n(X; \theta_0)\} \rightarrow \mathcal{L}\{\xi^T \xi\} = \chi_m^2, \quad n \rightarrow \infty.$$

□

Замечание 12.4. Если H_0 является сложной гипотезой вида

$$H_0 : \theta \in \Theta_0, \quad \Theta_0 = \{\theta : h_j(\theta) = 0, \quad j = 1, \dots, m-s\},$$

где $1 \leq s \leq m-1$, $\{h_j(\cdot)\}_{j=0}^{m-s}$ – фиксированные непрерывные функции, то результат теоремы 12.6 остается в силе, только

$$\delta = F_{\chi_{m-s}^2}^{-1}(1 - \epsilon).$$

12.6. ПОНЯТИЕ О ПОСЛЕДОВАТЕЛЬНОМ АНАЛИЗЕ ВАЛЬДА

Ранее рассмотренные статистические критерии Неймана – Пирсона, Байеса, χ -квадрат, Колмогорова, отношения правдоподобия характеризуются общим свойством: объем n наблюдаемой выборки $X_n = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ фиксируется априори. Такая ситуация возникает при обработке ранее полученных (архивных) экспериментальных данных в так называемых пассивных экспериментах. Однако на практике часто встречаются ситуации, когда проверку гипотез необходимо осуществлять в ходе активного эксперимента и решать вопрос о его продолжении или прекращении (в режиме «on-line»). В 1950 г. американский статистик А. Вальд выдвинул идею о том, что априорно фиксировать объем выборки n необязательно. Целесообразно определять необходимый объем выборки в ходе самого эксперимента, так что сам объем выборки становится случайной величиной, зависящей от экспериментальных данных: $n = n(X)$. Методы проверки гипотез, основанные на этой идее, получили название *последовательного анализа*. Рассмотрим простейшую модель последовательного анализа Вальда.

Пусть определена случайная последовательность наблюдений $x_1, x_2, \dots, x_n, \dots \in \mathbb{R}^N$ – независимых в совокупности, одинаково распределенных случайных векторов, имеющих плотность $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta \in \Theta = \{\theta_0, \theta_1\}$ – неизвестное истинное значение параметра, принимающего одно из двух различных возможных значений. Определены две простые гипотезы:

$$H_0 : \theta = \theta_0;$$

$$H_1 : \theta = \theta_1.$$

Обозначим: $X_m = (x_1^T, \dots, x_m^T)^T \in \mathbb{R}^{mN}$ – составная выборка, полученная при проведении m экспериментов ($m = 1, 2, \dots$); $d_m \in D = \{0, 1, 2\}$ – возможное решение после проведения m экспериментов.

Решение $d_m = 0(1)$ означает, что по результатам m экспериментов X_m принимается гипотеза $H_0(H_1)$, а решение $d_m = 2$ – что для вынесения решения об истинности H_0 или H_1 данных, содержащихся в X_m , недостаточно, – и необходимо осуществить $(m + 1)$ -й эксперимент.

Последовательный анализ характеризуется схемой наблюдения и принятия решений, представленной на рис. 12.3.

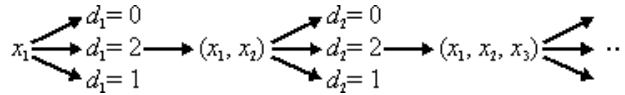


Рис. 12.3. Схема последовательного анализа Вальда

Процесс увеличения числа экспериментов m продолжается до тех пор, пока для некоторого случайного n не будет принято «окончательное» решение $d_n \in \{0, 1\}$. Задать последовательное решающее правило (*последовательный критерий*) – значит задать последовательность решающих правил:

$$d_m = d_m(X_m), \quad X_m \in \mathbb{R}^{mN}, \quad d_m \in \{0, 1, 2\}, \quad m = 1, 2, \dots$$

Точность последовательного критерия характеризуют условные вероятности ошибок I и II рода:

$$\alpha = P_{\theta_0}\{d_n = 1\}, \quad \beta = P_{\theta_1}\{d_n = 0\},$$

а также условные математические ожидания объемов выборок (средние длительности экспериментов):

$$W_i = E_{\theta_i}\{n\}, \quad i = 0, 1.$$

Определение 12.22 (принцип оптимальности Вальда). Последовательное решающее правило $\{d_m = d_m(X_m)\}$ надлежит выбирать таким образом, чтобы вероятности ошибок I и II рода α и β не превосходили соответственно наперед заданных достаточно малых уровней α_0 и β_0 , а условные средние объемы выборок W_0, W_1 были бы минимальны.

Этому принципу, оказывается, удовлетворяет последовательный критерий отношения правдоподобия Вальда следующего вида:

$$d_m = d_m(X_m) = \begin{cases} 0, & L_m(X_m) < B, \\ 2, & B \leq L_m(X_m) < A, \\ 1, & L_m(X_m) \geq A, \end{cases} \quad (12.32)$$

где

$$L_m(X_m) = \frac{p(X_m; \theta_1)}{p(X_m; \theta_0)} - \quad (12.33)$$

статистика отношения правдоподобия;

$$p(X_m; \theta_i) = \prod_{j=1}^m p(x_j; \theta_i), \quad i = 0, 1, \quad - \quad (12.34)$$

плотность распределения случайной выборки X_m при $\theta = \theta_i$ ($i = 0, 1$); $A, B \in \mathbb{R}$ ($B < A$) – некоторые, пока не определенные числовые параметры правила, называемые соответственно *верхним порогом* и *нижним порогом* и выбираемые из принципа оптимальности Вальда.

Решающее правило Вальда (12.32)–(12.34) допускает эквивалентное представление, удобное для дальнейшего анализа:

$$d_m = d_m(Z_m) = \begin{cases} 0, & Z_m < \ln B, \\ 2, & \ln B \leq Z_m < \ln A, \\ 1, & Z_m \geq \ln A, \end{cases} \quad (12.35)$$

где

$$Z_m = \sum_{j=1}^m z_j, \quad z_j = \ln \frac{p(x_j; \theta_1)}{p(x_j; \theta_0)}.$$

Теорема 12.7. Пусть конечны условные дисперсии:

$$0 < \sigma_i^2 = D_{\theta_i} \left\{ \ln \frac{p(x_j; \theta_1)}{p(x_j; \theta_0)} \right\} < \infty, \quad i = 0, 1, \text{ и } |\ln A| < \infty, |\ln B| < \infty,$$

тогда последовательный критерий Вальда (12.35) представляет собой шаговый процесс, заканчивающийся с вероятностью единица принятием одной из гипотез H_0 или H_1 .

Доказательство. Пусть на самом деле верна гипотеза H_i ($i \in \{0, 1\}$). Тогда по условию теоремы существует условное математическое ожидание, определяемое с учетом (12.35):

$$\mu_i = E_{\theta_i}\{z_j\} = \int_{\mathbb{R}^N} p(x; \theta_i) \ln \frac{p(x; \theta_1)}{p(x; \theta_0)} dx,$$

причем в силу свойства функции Кульбака из п. 11.7 $\mu_1 > 0$, $\mu_0 < 0$. К сумме Z_m при $m \rightarrow \infty$ применима ЦПТ Леви – Линдеберга из п. 5.5:

$$P_{\theta_i} \left\{ \frac{Z_m - m\mu_i}{\sqrt{m}\sigma_i} < y \right\} \rightarrow \Phi(y), \quad y \in \mathbb{R}, \quad (12.36)$$

где $\Phi(\cdot)$ – функция распределения вероятностей закона $N_1(0, 1)$.

Исследуем асимптотику для вероятности \bar{Q}_m «незавершения шагового процесса принятия решений за m шагов». Согласно (12.35)

$$\bar{Q}_m = P_{\theta_i} \{ \ln B \leq Z_1 < \ln A, \ln B \leq Z_2 < \ln A, \dots, \ln B \leq Z_m < \ln A \}.$$

В силу свойства вероятности

$$P \left\{ \bigcap_{k=1}^m C_k \right\} \leq P\{C_l\}, \quad l \in \{1, \dots, m\},$$

имеем

$$\bar{Q}_m \leq P_{\theta_i} \{ \ln B \leq Z_m < \ln A \},$$

и вероятность Q_m «завершения шагового процесса за m шагов» оценивается с учетом (12.36) снизу следующим образом:

$$\begin{aligned} Q_m &= 1 - \bar{Q}_m \geq 1 - P_{\theta_i} \{ \ln B \leq Z_m < \ln A \} = \\ &= 1 - P_{\theta_i} \left\{ \frac{\ln B - m\mu_i}{\sqrt{m}\sigma_i} \leq \frac{Z_m - m\mu_i}{\sqrt{m}\sigma_i} < \frac{\ln A - m\mu_i}{\sqrt{m}\sigma_i} \right\} = \\ &= 1 - P_{\theta_i} \left\{ \frac{1}{\sqrt{m}} \cdot \frac{\ln B}{\sigma_i} - \sqrt{m} \cdot \frac{\mu_i}{\sigma_i} \leq Z_{mi}^0 < \frac{1}{\sqrt{m}} \cdot \frac{\ln A}{\sigma_i} - \sqrt{m} \frac{\mu_i}{\sigma_i} \right\} \rightarrow 1. \end{aligned}$$

□

Важным вопросом является выбор пороговых значений A, B в (12.35). Можно показать, что если выбирать их по принципу Вальда, то

$$A \leq \frac{1 - \beta_0}{\alpha_0}, \quad B \geq \frac{\beta_0}{1 - \alpha_0}.$$

При этом если берется знак «=», то $\alpha + \beta \leq \alpha_0 + \beta_0$. Заметим, что точные выражения $A = A(\alpha_0, \beta_0)$, $B = B(\alpha_0, \beta_0)$ получить не удастся. Условные средние объемы выборок правила Вальда при таком упрощенном выборе A, B определяются следующими соотношениями:

$$\begin{aligned} W_0 &= \frac{(1 - \alpha_0) \ln \frac{\beta_0}{1 - \alpha_0} + \alpha_0 \ln \frac{1 - \beta_0}{\alpha_0}}{\mu_0}; \\ W_1 &= \frac{\beta_0 \ln \frac{\beta_0}{1 - \alpha_0} + (1 - \beta_0) \ln \frac{1 - \beta_0}{\alpha_0}}{\mu_1}. \end{aligned}$$

12.7. УПРАЖНЕНИЯ

1. Выборка $X = (x_1, \dots, x_n)$ получена из распределения Пуассона с параметром λ , относительно которого имеются две простые гипотезы: $H_0 : \lambda = 1$, $H_1 : \lambda = 3$. Для проверки этих гипотез используется следующий статистический критерий: принимается H_0 при $x_{(n)} \leq 1$, в противном случае принимается H_1 . Указать минимальный объем выборки, при котором мощность этого критерия превышает заданное значение γ .
2. Выборка $X = (x_1, \dots, x_n)$ получена из неизвестного дискретного распределения вероятностей, относительно которого имеются две гипотезы: H_0 состоит в том, что наблюдались значения случайной величины, принимающей значения 0 и 1 с вероятностями p_0 и p_1 соответственно, $p_0 + p_1 = 1$; альтернатива H_1 состоит в том, что наблюдаемая случайная величина принимает значения 1 и 2 с вероятностями q_1 и q_2 , $q_1 + q_2 = 1$; $p_1, p_2, q_1, q_2 > 0$. Для статистической проверки гипотез H_0 , H_1 используется следующий критерий: принимается гипотеза H_1 при $x_i \geq 1$, $i = 1, \dots, n$; в противном случае принимается нулевая гипотеза. Вычислить вероятности ошибок первого и второго рода такого критерия.
3. Выполнить упражнение 2 в случае, когда используется следующий статистический критерий: принимается гипотеза H_1 , если хотя бы один из элементов выборки равен 2, иначе принимается гипотеза H_0 .
4. Выборка $X = (x_1, \dots, x_n)$ получена из распределения Пуассона при неизвестном значении параметра θ . Относительно этого значения имеются две гипотезы: $H_0 : \theta \leq \theta_0$ и $H_1 : \theta > \theta_0$. Построить равномерно наиболее мощный критерий размера ϵ для статистической проверки этих гипотез.
5. Выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей $N_1(a, \theta)$; значение дисперсии θ неизвестно. Используя статистику $1/n \sum_{i=1}^n (x_i - a)^2$, построить равномерно наиболее мощный критерий размера ϵ для проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta \neq \theta_0$.
6. Выборка $X = (x_1, \dots, x_n)$ получена из показательного распределения вероятностей с параметром λ , значение которого неизвестно. Используя статистику \bar{x} , построить при больших значениях n равномерно наиболее мощный критерий размера ϵ для проверки гипотезы $H_0 : \lambda = \lambda_0$ против альтернативы $H_1 : \lambda > \lambda_0$.
7. Выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей с неизвестным значением математического ожидания a и единичной дисперсией, а выборка $Y = (y_1, \dots, y_m)$ – из распределения $N_1(b, 1)$ при неизвестном значении $b \leq a$. Выборки X и Y независимы. Проверяется гипотеза о равенстве математических ожиданий a и b . Принимается гипотеза $H_0 : a = b$, если $\sqrt{nm/(n+m)}(\bar{x} - \bar{y}) \leq c$, иначе принимается альтернатива $H_1 : a > b$; $c > 0$ – заданный параметр критерия. Найти размер данного критерия в зависимости от c .
8. Наблюдается число разладок оборудования на одном из производственных участков за смену. За каждый час смены наблюдалось следующее число разладок: 16, 17, 19, 16, 24, 19, 17, 16 (всего 144 разладки). При уровне значимости 0,05 проверить гипотезу H_0 о том, что число разладок подчиняется соответствующему равномерному распределению вероятностей.

9. Выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей $N_1(a, \theta)$, где значение параметра a известно, дисперсия – неизвестный параметр. Построить критерий Неймана – Пирсона для проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta = \theta_1$, $\theta_1 > \theta_0 > 0$, имеющий вероятность ошибки первого рода α . Вычислить для построенного критерия вероятность ошибки второго рода.
10. Выборка $X = (x_1, \dots, x_n)$ получена из нормального распределения вероятностей $N_1(\theta, \sigma^2)$ при неизвестных значениях параметров θ, σ^2 . Относительно параметра θ имеются две простые гипотезы: $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$, $\theta_1 > \theta_0$. Считая априорные вероятности гипотез равными, построить байесовское решающее правило проверки гипотез H_0, H_1 .
11. Выборка $X = (x_1, \dots, x_n)$ получена из геометрического распределения вероятностей с параметром p , где параметр p может принимать лишь значения $1/2$ и $1/4$ с вероятностями $1/3$ и $2/3$ соответственно. Построить байесовское решающее правило проверки двух простых гипотез $H_0 : p = 1/2$, $H_1 : p = 1/4$.
12. Выборка $X = (x_1, \dots, x_n)$ получена из биномиального распределения вероятностей с параметрами N и p , где p может принимать лишь значения $1/3$ и $2/3$ с априорными вероятностями $1/5$ и $4/5$ соответственно, а значение параметра N известно и фиксировано. Построить байесовское решающее правило проверки нулевой гипотезы $H_0 : p = 1/3$ против альтернативы $H_1 : p = 2/3$.
13. Рассматривается задача контроля качества больших партий выпускаемой предприятием продукции. Партия считается годной, если доля дефектных изделий $\theta = 0,01$ (гипотеза H_0), и бракованной, если доля дефектных изделий $\theta = \theta_1 > 0,01$ (гипотеза H_1). Построить последовательный критерий отношения вероятностей Вальда для проверки изделий, если заданы значения вероятностей ошибок I и II рода: $\alpha = 0,01$, $\beta = 0,001$.

ЧАСТЬ 4

СТАТИСТИЧЕСКИЙ АНАЛИЗ МНОГОМЕРНЫХ ДАННЫХ

Глава 13

РАЗВЕДОЧНЫЙ АНАЛИЗ И СЖАТИЕ ДАННЫХ

13.1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ И ЕГО ОСНОВНЫЕ ЭТАПЫ

Пусть исследуется сложная система (объект или явление), относительно которой по имеющейся информации необходимо сделать научно обоснованные выводы.

Определение 13.1. *Прикладная статистика – это научная дисциплина, основанная на методах теории вероятностей и математической статистики, целью которой является формирование выводов и принятие решений относительно сложной системы посредством статистического анализа данных о ней.*

Статистический анализ данных можно условно разделить на следующие этапы.

1. Исследование сложной системы и сбор априорной информации.
2. Формирование предположений относительно системы и построение ее математической модели.
3. Планирование и проведение экспериментов с системой и сбор статистических данных.
4. Первичный (разведочный) анализ и сжатие данных.
5. Подбор методов и непосредственное проведение статистического анализа.
6. Анализ полученных результатов и проверка их адекватности.
7. Принятие решения об удовлетворительности полученных результатов или необходимости возврата на предыдущие этапы.

Статистический анализ данных обычно проводится с использованием компьютеров и различных пакетов прикладных программ [32], в основу которых положены теоретически обоснованные статистические методы.

13.2. ВЫБОРКА ИЗ МНОГОМЕРНЫХ НАБЛЮДЕНИЙ КАК МОДЕЛЬ СТАТИСТИЧЕСКИХ ДАННЫХ

Адекватной математической моделью статистических данных является выборка $X = \{x_1, \dots, x_n\}$ объема $n \geq 1$, где $x_t = (x_{t1}, \dots, x_{tN})^T \in \mathbb{R}^N$ – случайный N -вектор, определенный на вероятностном пространстве (Ω, \mathcal{F}, P) и являющийся наблюдением над объектом $\omega_t \in \Omega$ ($t = 1, \dots, n$) или наблюдением над одним и тем же объектом $\omega \in \Omega$ в различные моменты времени $t = 1, \dots, n$. Здесь предполагается, что с течением времени свойства объектов не меняются, т. е. они статичны. Случай, когда это не так, будет рассмотрен в части 5.

Обычно используются две основные модели данных.

Одновыборочная модель. Предполагается, что все наблюдения x_1, \dots, x_n из выборки X являются независимыми в совокупности и одинаково распределенными случайными векторами с некоторой функцией распределения вероятностей $F(x)$, $x \in \mathbb{R}^N$.

Многовыборочная модель. В этой модели исходная выборка X объема $n \geq 2$ считается неоднородной и представляется в виде совокупности из L подвыборок ($2 \leq L \leq n$):

$$X = \bigcup_{i=1}^L X^{(i)}; \quad X^{(i)} \cap X^{(j)} = \emptyset, \quad i \neq j, \quad i, j = 1, \dots, L.$$

Каждая подвыборка $X^{(i)}$ состоит из n_i наблюдений с одной и той же функцией распределения $F_i(x)$, $x \in \mathbb{R}^N$ ($i = 1, \dots, L$; $n_1 + \dots + n_L = n$).

По характеру компонент (*признаков*), из которых образованы векторы-наблюдения $x_t = (x_{t1}, \dots, x_{tN})^T$ ($t = 1, \dots, n$), выделяют следующие модели данных:

- а) *непрерывные*, когда все N признаков имеют совместное абсолютно-непрерывное распределение вероятностей с некоторой плотностью $f(x)$, $x \in \mathbb{R}^N$, которая однозначно определяет их совместную функцию распределения $F(x)$, $x \in \mathbb{R}^N$ [30];
- б) *дискретные*, когда наблюдения являются дискретными случайными векторами;
- в) *комбинированные*, когда часть признаков – непрерывные случайные величины, а остальные – дискретные.

13.3. РАЗВЕДОЧНЫЙ АНАЛИЗ ДАННЫХ

Пусть имеется выборка $X = \{x_1, \dots, x_n\}$ объема n и предполагается, что образующие ее наблюдения $\{x_t \in \mathbb{R}^N\}_{t=1}^n$ – независимые в совокупности, одинаково распределенные случайные N -векторы с функцией распределения $F(x)$, $x \in \mathbb{R}^N$.

Статистический анализ данных обычно начинается с *разведочного анализа*, при котором, как правило, исследуются одномерные (маргинальные) распределения вероятностей каждого признака по отдельности. Поэтому далее в этом параграфе, не ограничивая общности, выборку будем предполагать образованной из одномерных наблюдений ($N = 1$): $x_t \in \mathbb{R}$, $t = 1, \dots, n$, и считать, что она соответствует наблюдениям над каким-либо признаком из исходной выборки. Функция распределения в этом случае также одномерная: $F(x)$, $x \in \mathbb{R}$, и при абсолютно непрерывном распределении ей соответствует плотность $f(x) = \frac{d}{dx}F(x)$, $x \in \mathbb{R}$.

При разведочном анализе обычно вычисляют следующие статистические характеристики (см. также пп. 11.2–11.4).

I. Характеристики положения

1. Вариационный ряд:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

где

$$x_- = x_{(1)} = \min_{1 \leq t \leq n} x_t, \quad x_+ = x_{(n)} = \max_{1 \leq t \leq n} x_t \quad -$$

соответственно *минимальное* и *максимальное значения*.

2. Выборочное среднее:

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad -$$

статистическая оценка математического ожидания $\mu = E\{x_t\}$.

3. *Выборочная медиана:*

$$\hat{m} = \begin{cases} x_{(k+1)}, & n = 2k + 1, \\ \frac{x_{(k)} + x_{(k+1)}}{2}, & n = 2k, \end{cases} -$$

статистическая оценка медианы m : $F(m) = 1/2$.

II. Характеристики рассеяния

1. *Выборочная дисперсия:*

$$s^2 = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})^2 -$$

оценка истинного значения дисперсии $\sigma^2 = D\{x_t\} = E\{(x_t - \mu)^2\}$ (при этом s – оценка среднеквадратического отклонения $\sigma = \sqrt{D\{x_t\}} = \sqrt{E\{(x_t - \mu)^2\}}$).

2. *Размах:*

$$x_+ - x_-.$$

3. *Интервал концентрации:*

$$[\bar{x} - 3s, \bar{x} + 3s] \quad \left(P\{x_t \in [\mu - 3\sigma, \mu + 3\sigma]\} \geq \frac{8}{9} \right).$$

III. Характеристики формы

1. *Выборочная асимметрия:*

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^3}{s^3} -$$

оценка коэффициента асимметрии $\beta_1 = E\{(x_t - \mu)^3\}/\sigma^3$.

2. *Выборочный эксцесс:*

$$\hat{\beta}_2 = \frac{\frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^4}{s^4} - 3 -$$

оценка истинного значения эксцесса $\beta_2 = E\{(x_t - \mu)^4\}/\sigma^4 - 3$.

Последние две характеристики имеют особое значение при сравнении распределения статистических данных с нормальным распределением $N_1(\mu, \sigma^2)$.

Коэффициент асимметрии показывает, насколько распределение анализируемых данных отличается от симметричного относительно математического ожидания μ распределения вероятностей в ту или иную сторону:

если $\beta_1 < 0$, то $P\{x_t \leq \mu\} < 1/2$;

если $\beta_1 > 0$, то $P\{x_t \leq \mu\} > 1/2$.

По значению эксцесса можно судить о том, насколько «хвосты» распределения «тяжелее» ($\beta_2 > 0$) или «легче» ($\beta_2 < 0$) «хвостов» нормального распределения.

В случае нормального распределения ($\mathcal{L}\{x_t\} = N_1(\mu, \sigma^2)$) обе эти характеристики равны нулю ($\beta_1 = \beta_2 = 0$).

Далее в разведочном анализе пытаются «подогнать» распределение вероятностей, что сводится к проверке гипотез согласия (см. п. 12.4):

$$\begin{aligned} H_0 & : F(\cdot) \equiv F_0(\cdot); \\ H_1 = \overline{H_0} & : F(\cdot) \neq F_0(\cdot), \end{aligned}$$

где $F_0(\cdot)$ – предполагаемая (гипотетическая) функция распределения. Причем если она задана параметрически (например, предполагается нормальное распределение $N_1(\mu, \sigma^2)$), то в качестве неизвестных параметров в $F_0(\cdot)$ используются их статистические оценки (для нормального распределения это будут выборочные среднее \bar{x} и дисперсия s^2).

Для графической иллюстрации точности «подгонки» строят выборочную функцию распределения $\hat{F}(\cdot)$ и гистограмму $\hat{f}(\cdot)$ (см. п. 11.2) с «наложенными» на них соответственно «подогнанной» функцией распределения $F_0(\cdot)$ (см. рис. 12.2) и плотностью $f_0(x) = \frac{d}{dx}F_0(x)$ (см. рис. 11.3).

Замечание 13.1. На практике, как правило, «подгоняют» данные нормальным распределением. Если это не удается, то предпринимают различные попытки нормализовать данные. В случае положительных выборочных значений самая распространенная из таких попыток – прологарифмировать наблюдения: $x_t := \ln(x_t)$, $t = 1, \dots, n$. Если среди выборочных значений встречаются отрицательные либо нулевые ($x_- \leq 0$), то это преобразование обобщают: $x_t := \ln(x_t + c)$, $t = 1, \dots, n$, где $c > |x_-|$ – некоторая положительная величина.

13.4. ПРОБЛЕМА СЖАТИЯ ДАННЫХ. МЕТОД ГЛАВНЫХ КОМПОНЕНТ

Пусть в пространстве \mathbb{R}^N зарегистрирована случайная выборка $X = \{x_1, \dots, x_n\}$ объема n : $x_t \in \mathbb{R}^N$, $t = 1, \dots, n$.

Как правило, при сборе данных стараются определить максимальное количество признаков (характеристик) и размерность N пространства наблюдений оказывается достаточно большой.

Возникает проблема *сжатия данных*. Необходимо преобразовать исходную выборку $X = \{x_1, \dots, x_n\}$ в сжатую выборку $Y = \{y_1, \dots, y_n\}$: $y_t \in \mathbb{R}^m$, $t = 1, \dots, n$, с меньшим числом признаков m ($m < N$), несущих информацию об исходных наблюдениях с минимальными (наперед заданными) потерями.

Формально решение задачи сжатия данных сводится к поиску борелевского преобразования:

$$y = B(x) : \mathbb{R}^N \rightarrow \mathbb{R}^m, \quad (13.1)$$

переводящего исходное наблюдение $x = (\tilde{x}_1, \dots, \tilde{x}_N)^T \in \mathbb{R}^N$ в сжатое наблюдение $y = (\tilde{y}_1, \dots, \tilde{y}_m)^T \in \mathbb{R}^m$. Выборка Y получается из выборки X преобразованием каждого из n наблюдений: $y_t = B(x_t)$, $t = 1, \dots, n$.

Преобразование в (13.1), вообще говоря, нелинейное. Однако на практике обычно используют линейные преобразования:

$$y = Bx : \mathbb{R}^N \rightarrow \mathbb{R}^m,$$

где B – $(m \times N)$ -матрица.

Проблема заключается в выборе матрицы B . Одним из способов ее решения является *метод главных компонент*, наиболее часто используемый на практике для сжатия данных.

Согласно методу главных компонент случайный N -вектор-наблюдение $x = (\tilde{x}_1, \dots, \tilde{x}_N)^T \in \mathbb{R}^N$ с ковариационной $(N \times N)$ -матрицей

$$\Sigma = \text{cov}\{x, x\} = \mathbf{E}\{(x - \mu)(x - \mu)^T\},$$

$\mu = E\{x\} \in \mathbb{R}^N$ – вектор математического ожидания, подвергается следующему линейному преобразованию ($y = (\tilde{y}_1(x), \dots, \tilde{y}_N(x))^T$):

$$\tilde{y}_k = \tilde{y}_k(x) = \Psi_k^T x, \quad k = 1, \dots, N, \quad (13.2)$$

где $\{\Psi_k\}_{k=1}^N$ являются ортонормированными собственными векторами ковариационной матрицы Σ , удовлетворяют соотношениям

$$\Sigma \Psi_k = \lambda_k \Psi_k, \quad k = 1, \dots, N; \quad (13.3)$$

$$\Psi_k^T \Psi_l = \delta_{kl}, \quad k, l = 1, \dots, N, \quad (13.4)$$

и соответствуют упорядоченным по убыванию собственным числам $\{\lambda_k\}_{k=1}^N$ матрицы Σ : $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$.

Полученные таким образом случайные величины $\tilde{y}_1, \dots, \tilde{y}_N$ называют *главными компонентами* для исходного наблюдения $x = (\tilde{x}_1, \dots, \tilde{x}_N)^T$. Исследуем вероятностные свойства главных компонент.

Теорема 13.1. *Главные компоненты $\tilde{y}_1, \dots, \tilde{y}_N$, удовлетворяющие (13.2)–(13.4), некоррелированы*

$$\text{cov}\{\tilde{y}_k, \tilde{y}_j\} = 0, \quad k \neq j, \quad k, j = 1, \dots, N,$$

а их дисперсии равны соответствующим собственным числам

$$D\{\tilde{y}_k\} = \lambda_k, \quad k = 1, \dots, N. \quad (13.5)$$

Доказательство. С учетом (13.2)–(13.4) вычислим ковариацию ($k, j = 1, \dots, N$):

$$\begin{aligned} \text{cov}\{\tilde{y}_k, \tilde{y}_j\} &= E\{(\tilde{y}_k - E\{\tilde{y}_k\})(\tilde{y}_j - E\{\tilde{y}_j\})\} = \\ &= \Psi_k^T E\{(x - \mu)(x - \mu)^T\} \Psi_j = \Psi_k^T \Sigma \Psi_j = \lambda_j \Psi_k^T \Psi_j = \delta_{kj} \lambda_j. \end{aligned}$$

□

Следствие 13.1. *Суммарная дисперсия исходных признаков равна суммарной дисперсии главных компонент:*

$$\sum_{k=1}^N D\{\tilde{x}_k\} = \sum_{k=1}^N D\{\tilde{y}_k\} = \sum_{k=1}^N \lambda_k. \quad (13.6)$$

Доказательство. Очевидно и следует из (13.5) и известного свойства матрицы [34, 11]:

$$\sum_{k=1}^N D\{\tilde{x}_k\} = \text{tr}(\Sigma) = \sum_{k=1}^N \lambda_k.$$

□

Соотношения (13.5), (13.6) позволяют предложить критерий выбора *информативных признаков* в пространстве главных компонент, который состоит в том, что признаки, имеющие малые дисперсии, отбрасываются, а рассматриваются лишь m ($m \leq N$) первых главных компонент из $\tilde{y}_1, \dots, \tilde{y}_m, \dots, \tilde{y}_N$ ($\lambda_1 \leq \dots \leq \lambda_m \leq \dots \leq \lambda_N$).

Число m обычно определяется по наперед заданной малой величине $\epsilon \in [0, 1)$:

$$m = m(\epsilon) = \min\{k : 1 - \kappa(k) \leq \epsilon, \quad k = 1, \dots, N\}; \quad (13.7)$$

$$\kappa(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^N \lambda_j},$$

где $0 < \varkappa(k) \leq 1$ в (13.7) – относительная доля суммарной дисперсии первых k главных компонент. Чем ближе $\varkappa(k)$ к единице (а $(1 - \varkappa(k))$ – к нулю), тем меньше потери информации при сжатии данных ($\varkappa(N) = 1$).

Замечание 13.2. На практике обычно истинное значение ковариационной матрицы Σ неизвестно, и вместо нее в методе главных компонент используется оценка по исходной выборке $X = \{x_1, \dots, x_n\}$:

$$S = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T, \quad \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t,$$

вероятностные свойства которой будут исследованы в следующих главах.

Замечание 13.3. Метод главных компонент применяется не только для сжатия данных. Он позволяет декоррелировать их – строить выборки из наблюдений с некоррелированными признаками. С его помощью также производят визуализацию многомерных данных: на плоскости ($m = 2$) и в трехмерном пространстве ($m = 3$). При этом учитывается информация, заключенная во всех исходных признаках, чего нельзя сказать о диаграммах рассеяния, на которых исходные признаки отображаются попарно (на плоскости) и тройками (в трехмерном пространстве).

Замечание 13.4. Недостаток метода главных компонент состоит в том, что главные компоненты, по сравнению с исходными признаками, не имеют на практике физической интерпретации. С этой точки зрения лучше напрямую выбирать информативные признаки из множества исходных признаков и не производить их функциональных преобразований, т. е. осуществлять так называемый прямой отбор информативных признаков.

Глава 14

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И ОЦЕНИВАНИЕ ЕГО ПАРАМЕТРОВ

14.1. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И ЕГО ОСНОВНЫЕ СВОЙСТВА

Пусть при проведении разведочного анализа установлено, что некоторые компоненты имеют нормальные распределения вероятностей. Следующее предположение состоит в том, чтобы считать, что эти компоненты имеют совместное многомерное нормальное распределение.

Определение 14.1. Если компоненты непрерывного случайного N -вектора $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$ имеют конечные моменты второго порядка: $E\{(\tilde{x}_j)^2\} < +\infty$, $j = 1, \dots, N$, а его плотность распределения вероятностей задается соотношением

$$f(x) = n_N(x|\mu, \Sigma) = (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad (14.1)$$

где $\mu = E\{x\}$ – N -вектор математического ожидания, а

$$\Sigma = E\{(x - \mu)(x - \mu)^T\} \quad -$$

невыврожденная ковариационная $(N \times N)$ -матрица ($|\Sigma| \neq 0$), то говорят, что $x \in \mathbb{R}^N$ имеет невырожденное многомерное нормальное распределение, и кратко обозначают этот факт: $\mathcal{L}\{x\} = N_N(\mu, \Sigma)$.

Замечание 14.1. В случае вырожденной ковариационной матрицы Σ ($|\Sigma| = 0$) иногда также определяют вырожденное многомерное нормальное распределение, которое, однако, не описывается плотностью (14.1). Считается, что случайный N -вектор $x \in \mathbb{R}^N$ имеет вырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| = 0$), если он может быть представлен в виде $x = By + b$, где случайный p -вектор $y \in \mathbb{R}^p$ ($p < N$) имеет невырожденное p -мерное нормальное распределение, $(p \times N)$ -матрица линейного преобразования B – полный ранг: $\text{rank}(B) = p$; p -вектор сдвига $b = \mu - BE\{y\}$.

Согласно (14.1) многомерное нормальное распределение однозначно определяется своими параметрами – вектором математического ожидания $\mu = (\tilde{\mu}_j)_{j=1}^N$:

$$\tilde{\mu}_j = E\{\tilde{x}_j\} \quad -$$

математическое ожидание j -й компоненты, и ковариационной матрицей $\Sigma = (\sigma_{jk})_{j,k=1}^N$, где

$$\sigma_{jk} = \sigma_{kj} = \text{cov}\{\tilde{x}_j, \tilde{x}_k\} = E\{(\tilde{x}_j - \tilde{\mu}_j)(\tilde{x}_k - \tilde{\mu}_k)\} \quad -$$

ковариация j -й и k -й компонент, а

$$\sigma_{jj} = D\{\tilde{x}_j\} = E\{(\tilde{x}_j - \tilde{\mu}_j)^2\} \quad -$$

дисперсия j -й компоненты ($j, k = 1, \dots, N$).

Все характеристики многомерного нормального распределения $N_N(\mu, \Sigma)$ выражаются через его параметры μ и Σ , в том числе:

а) коэффициент корреляции:

$$\rho_{jk} = \frac{\text{cov}\{\tilde{x}_j, \tilde{x}_k\}}{\sqrt{D\{\tilde{x}_j\}D\{\tilde{x}_k\}}} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}} \in [-1, 1], \quad j, k = 1, \dots, N;$$

б) центральные моменты третьего порядка:

$$E\{(\tilde{x}_i - \tilde{\mu}_i)(\tilde{x}_j - \tilde{\mu}_j)(\tilde{x}_k - \tilde{\mu}_k)\} = 0, \quad i, j, k = 1, \dots, N;$$

в) центральные моменты четвертого порядка:

$$E\{(\tilde{x}_i - \tilde{\mu}_i)(\tilde{x}_j - \tilde{\mu}_j)(\tilde{x}_k - \tilde{\mu}_k)(\tilde{x}_l - \tilde{\mu}_l)\} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{kj};$$

$$E\{(\tilde{x}_i - \tilde{\mu}_i)^4\} = 3\sigma_{ii}^2, \quad i, j, k, l = 1, \dots, N.$$

Для того чтобы исследовать другие свойства многомерного нормального распределения, вычислим его характеристическую функцию, однозначно определяющую распределение вероятностей.

Теорема 14.1. Пусть случайный N -вектор $x \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), тогда его характеристическая функция $\phi_x(\mathbf{t}) = E\{e^{it^T x}\}$, $\mathbf{t} \in \mathbb{R}^N$, имеет вид

$$\phi_x(\mathbf{t}) = e^{it^T \mu - \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}}, \quad (14.2)$$

где i – мнимая единица.

Доказательство. Воспользуемся видом плотности многомерного нормального распределения (14.1) и запишем характеристическую функцию:

$$\begin{aligned} \phi_x(\mathbf{t}) &= E\{e^{it^T x}\} = \int_{\mathbb{R}^N} e^{it^T x} n_N(x|\mu, \Sigma) dx = \\ &= \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \int_{\mathbb{R}^N} e^{it^T x - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx. \end{aligned}$$

Выделим полный квадрат в выражении, стоящем под экспонентой:

$$\begin{aligned} it^T x - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) &= \\ &= -\frac{1}{2}(x - (\mu + i\Sigma \mathbf{t}) + i\Sigma \mathbf{t})^T \Sigma^{-1}(x - (\mu + i\Sigma \mathbf{t}) + i\Sigma \mathbf{t}) + it^T x = \\ &= -\frac{1}{2}(x - (\mu + i\Sigma \mathbf{t}))^T \Sigma^{-1}(x - (\mu + i\Sigma \mathbf{t})) + it^T \mu - \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}. \end{aligned}$$

Продолжим вычисление характеристической функции и с учетом условия нормировки получим

$$\begin{aligned} \phi_x(\mathbf{t}) &= \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \int_{\mathbb{R}^N} e^{-\frac{1}{2}(x - (\mu + i\Sigma \mathbf{t}))^T \Sigma^{-1}(x - (\mu + i\Sigma \mathbf{t})) + it^T \mu - \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}} dx = \\ &= e^{it^T \mu - \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}} \int_{\mathbb{R}^N} n_N(x|\mu + i\Sigma \mathbf{t}, \Sigma) dx = e^{it^T \mu - \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}}, \end{aligned}$$

что совпадает с (14.2). \square

Характеристическая функция позволяет исследовать поведение нормально распределенных наблюдений при линейных преобразованиях и определить маргинальные распределения вероятностей выбранных компонент.

Теорема 14.2. Пусть $y = Cx + b \in \mathbb{R}^m$ – линейное преобразование нормального случайного N -вектора $x \in \mathbb{R}^N$ ($m \leq N$) с невырожденным нормальным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), где $b \in \mathbb{R}^m$; $(m \times N)$ -матрица линейного преобразования C имеет полный ранг: $\text{rank}(C) = m$, тогда случайный вектор $y \in \mathbb{R}^m$ имеет невырожденное m -мерное нормальное распределение: $\mathcal{L}\{y\} = N_m(C\mu + b, C\Sigma C^T)$ ($|C\Sigma C^T| \neq 0$).

Доказательство. Воспользуемся выражением (14.2) для характеристической функции $\phi_x(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^N$, исходного случайного вектора $x \in \mathbb{R}^N$ и найдем характеристическую функцию $\phi_y(\tau)$, $\tau \in \mathbb{R}^m$, для его линейного преобразования $y = Cx + b \in \mathbb{R}^m$:

$$\begin{aligned}\phi_y(\tau) &= \mathbf{E}\left\{e^{i\tau^T y}\right\} = \mathbf{E}\left\{e^{i\tau^T (Cx+b)}\right\} = \mathbf{E}\left\{e^{i(C^T \tau)^T x}\right\} e^{i\tau^T b} = \phi_x(C^T \tau) e^{i\tau^T b} = \\ &= e^{i(C^T \tau)^T \mu - \frac{1}{2}(C^T \tau)^T \Sigma C^T \tau} e^{i\tau^T b} = e^{i\tau^T (C\mu + b) - \frac{1}{2}\tau^T (C\Sigma C^T)\tau}.\end{aligned}$$

Выражение в правой части полученного соотношения является характеристической функцией многомерного нормального распределения с m -вектором математического ожидания $C\mu + b$ и ковариационной $(m \times m)$ -матрицей $C\Sigma C^T$, которая в силу невырожденности исходной ковариационной $(N \times N)$ -матрицы Σ и полного ранга матрицы C ($\text{rank}(C) = m$) имеет ранг m и, следовательно, невырождена. \square

Следствие 14.1. Пусть случайный N -вектор $x = (\tilde{x}_1, \dots, \tilde{x}_N)^T \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$) и пусть $x^* = (\tilde{x}_{j_1}, \dots, \tilde{x}_{j_m})^T \in \mathbb{R}^m$ – m -вектор, образованный из каких-либо $m \leq N$ его компонент ($j_k \neq j_l \in \{1, \dots, N\}$, $k \neq l$, $k, l = 1, \dots, m$). Тогда случайный m -вектор x^* имеет невырожденное m -мерное маргинальное нормальное распределение с математическим ожиданием $\mu^* = (\mu_{j_k})_{k=1}^m \in \mathbb{R}^m$ и ковариационной $(m \times m)$ -матрицей $\Sigma^* = (\sigma_{j_k j_l})_{k,l=1}^m$ ($|\Sigma^*| \neq 0$), образованными из соответствующих компонент исходных математического ожидания $\mu = (\mu_k)_{k=1}^N \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы $\Sigma = (\sigma_{kl})_{k,l=1}^N$: $\mathcal{L}\{x^*\} = N_m(\mu^*, \Sigma^*)$.

Доказательство. Представим m -вектор $x^* = (\tilde{x}_{j_1}, \dots, \tilde{x}_{j_m})^T$ как линейное преобразование исходного N -вектора $x \in \mathbb{R}^N$: $x^* = Cx$, где $(m \times N)$ -матрица C на пересечении k -й строки и j_k -го столбца содержит единицы ($j_k \in \{1, \dots, N\}$, $k = 1, \dots, m$), а остальные ее элементы равны нулю. Очевидно, что $\text{rank}(C) = m$, и из теоремы 14.2 с учетом вида матрицы C получаем доказываемое. \square

Следствие 14.2. Компоненты случайного N -вектора $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$, распределенного по невырожденному многомерному нормальному закону $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), имеют одномерные маргинальные нормальные распределения: $\mathcal{L}\{\tilde{x}_j\} = N_1(\mu_j, \sigma_{jj})$ ($\sigma_{jj} > 0$), $j = 1, \dots, N$.

Замечание 14.2. Если компоненты случайного вектора имеют одномерные нормальные распределения, то их совместное распределение, вообще говоря, не является многомерным нормальным. Оно будет таковым в случае их независимости (при этом некоррелированности компонент не достаточно!).

Следствие 14.3. Пусть случайный вектор $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$, распределенный по многомерному нормальному закону $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), разбит на вектор-блоки: $x = ((x^{(1)})^T; \dots; (x^{(K)})^T)^T$. Тогда случайные векторы $\{x^{(k)}\}_{k=1}^K$ независимы тогда и только тогда, когда они некоррелированы.

Доказательство. По следствию 14.1 случайные векторы $\{x^{(k)}\}_{k=1}^K$ имеют нормальные распределения: $\mathcal{L}\{x^{(k)}\} = N_{N(k)}(\mu^{(k)}, \Sigma^{(k)})$ ($|\Sigma^{(k)}| \neq 0$), $k = 1, \dots, K$, являющиеся маргинальными по отношению к распределению исходного вектора $x \in \mathbb{R}^N$ ($N^{(1)} + \dots + N^{(K)} = N$). Но по известному критерию независимости [30]

$\{x^{(k)}\}_{k=1}^K$ независимы тогда и только тогда, когда их совместная плотность распределения представима в виде произведения их плотностей:

$$n_N(x|\mu, \Sigma) = \prod_{k=1}^K n_{N^{(k)}}(x^{(k)}|\mu^{(k)}, \Sigma^{(k)}), \quad x = \left((x^{(1)})^T \dots (x^{(K)})^T \right)^T \in \mathbb{R}^N.$$

Исходя из вида (14.1) плотности многомерного нормального распределения, заключаем, что последнее соотношение выполняется тогда и только тогда, когда случайные векторы $\{x^{(k)}\}_{k=1}^K$ некоррелированы: $\Sigma = \text{diag}\{\Sigma^{(1)}, \dots, \Sigma^{(K)}\}$. \square

Следствие 14.4. Пусть случайные векторы $\{x^{(k)} \in \mathbb{R}^{N^{(k)}}\}_{k=1}^K$ независимы в совокупности и имеют невырожденные нормальные распределения:

$\mathcal{L}\{x^{(k)}\} = N_{N^{(k)}}(\mu^{(k)}, \Sigma^{(k)})$ ($|\Sigma^{(k)}| \neq 0$), $k = 1, \dots, K$, тогда составной вектор $x = \left((x^{(1)})^T \dots (x^{(K)})^T \right)^T \in \mathbb{R}^N$, $N = \sum_{k=1}^K N^{(k)}$, также имеет невырожденное нормальное распределение: $\mathcal{L}\{x\} = N_N(\mu, \Sigma)$, с вектором математического ожидания $\mu = \left((\mu^{(1)})^T \dots (\mu^{(K)})^T \right)^T$ и ковариационной матрицей $\Sigma = \text{diag}\{\Sigma^{(1)}, \dots, \Sigma^{(K)}\}$ ($|\Sigma| \neq 0$).

С невырожденным многомерным нормальным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$) в силу вида его плотности (14.1) связывают метрику Махаланобиса:

$$\rho(x, z) = \sqrt{(x - z)^T \Sigma^{-1} (x - z)}, \quad x, z \in \mathbb{R}^N,$$

которая позволяет построить в \mathbb{R}^N так называемый эллипсоид рассеяния с центром в математическом ожидании $\mu \in \mathbb{R}^N$:

$$V_r = \{x : \rho(x, \mu) \leq r, \quad r > 0\}.$$

Воспользуемся известным в матричном анализе [34, 11] представлением для невырожденной ковариационной $(N \times N)$ -матрицы Σ , которая при этом в силу известного свойства неотрицательной определенности ковариационной матрицы положительно определена ($\Sigma \succ 0$):

$$\Sigma = \Sigma^{1/2} (\Sigma^{1/2})^T,$$

где невырожденная $(N \times N)$ -матрица $\Sigma^{1/2}$ определяется неоднозначно и является решением по Y матричного уравнения $Y^T \Sigma^{-1} Y = I_N$. В частности, можно выбрать $\Sigma^{1/2} = \left(\sqrt{\lambda_1} \Psi_1 \dots \sqrt{\lambda_N} \Psi_N \right)$, где $\{\lambda_k, \Psi_k\}_{k=1}^N$ – собственные числа и соответствующие им ортонормированные собственные векторы матрицы Σ , определенные в (13.3), (13.4).

Введем в рассмотрение случайный N -вектор $y = (\Sigma^{1/2})^{-1} (x - \mu)$. Согласно теореме 14.1, $\mathcal{L}\{y\} = N_N(\mathbf{0}_N, I_N)$, и случайная величина $\rho^2(x, \mu) = y^T y$ является суммой квадратов N независимых в совокупности стандартных нормальных случайных величин и имеет χ^2 -распределение с N степенями свободы. Поэтому вероятность попадания случайного вектора x в эллипсоид рассеяния V_r радиуса r равна: $P\{x \in V_r\} = F_{\chi_N^2}(r^2)$. Этот факт позволяет по наперед заданному уровню вероятности $P\{x \in V_r\} = 1 - \alpha$, где $\alpha \in (0, 1)$ – мало, определить радиус эллипсоида рассеяния:

$$r = r(\alpha) = \sqrt{F_{\chi_N^2}^{-1}(1 - \alpha)},$$

где $F_{\chi_N^2}^{-1}(1 - \alpha)$ – квантиль уровня $1 - \alpha$ от χ^2 -распределения с N степенями свободы.

Эллипсоид рассеяния V_r , $r = r(\alpha)$, является областью концентрации наблюдений вокруг своего математического ожидания, а область в \mathbb{R}^N вне эллипсоида рассеяния содержит наблюдения, описываемые «хвостом» распределения вероятностей: вероятность попадания в эту область мала и равна α .

14.2. ОЦЕНКИ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ПАРАМЕТРОВ МНОГОМЕРНОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ. ВЫБОРОЧНОЕ СРЕДНЕЕ И ВЫБОРОЧНАЯ КОВАРИАЦИОННАЯ МАТРИЦА

Пусть имеется случайная выборка $X = \{x_1, \dots, x_n\}$ объема n из невырожденного N -мерного нормального распределения $N_N(\mu, \Sigma)$ с неизвестными истинными значениями параметров: математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы Σ ($|\Sigma| \neq 0$). Воспользуемся методом максимального правдоподобия и построим их статистические оценки $\hat{\mu}$ и $\hat{\Sigma}$ по выборке X .

Теорема 14.3. Пусть наблюдения $x_1, \dots, x_n \in \mathbb{R}^N$, образующие выборку $X = \{x_t\}_{t=1}^n$ объема $n > N$, являются независимыми в совокупности, одинаково распределенными случайными N -векторами с невырожденным нормальным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), тогда единственными оценками максимального правдоподобия (МП-оценками) для вектора математического ожидания μ и ковариационной матрицы Σ являются соответственно выборочное среднее:

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t, \quad (14.3)$$

и выборочная ковариационная матрица:

$$\hat{\Sigma} = \frac{1}{n} A, \quad A = \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T. \quad (14.4)$$

Доказательство. Воспользуемся видом (14.1) плотности $n_N(\cdot | \mu, \Sigma)$ многомерного нормального распределения $N_N(\mu, \Sigma)$ и запишем логарифмическую функцию правдоподобия по выборке $X = \{x_t\}_{t=1}^n$ для неизвестных значений параметров μ и Σ :

$$\begin{aligned} l(\mu, \Sigma) &= \ln \prod_{t=1}^n n_N(x_t | \mu, \Sigma) = \\ &= -\frac{nN}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^n (x_t - \mu)^T \Sigma^{-1} (x_t - \mu) = \\ &= -\frac{nN}{2} \ln(2\pi) + \frac{n}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \sum_{t=1}^n \text{tr}(\Sigma^{-1} (x_t - \mu)(x_t - \mu)^T) = \\ &= -\frac{nN}{2} \ln(2\pi) + \frac{n}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \text{tr} \left(\Sigma^{-1} \sum_{t=1}^n (x_t - \mu)(x_t - \mu)^T \right), \end{aligned}$$

где использованы следующие свойства следа матрицы $\text{tr}(\cdot)$: а) $c = \text{tr}(c)$, $c \in \mathbb{R}$; б) $\text{tr}(BC) = \text{tr}(CB)$; в) $\text{tr}(B + D) = \text{tr}(B) + \text{tr}(D)$.

Учтем очевидное тождество

$$\sum_{t=1}^n (x_t - \bar{x}) \equiv \mathbf{0}_N \quad (14.5)$$

и преобразуем выражение

$$\begin{aligned} \sum_{t=1}^n (x_t - \mu)(x_t - \mu)^T &= \sum_{t=1}^n (x_t - \bar{x} + \bar{x} - \mu)(x_t - \bar{x} + \bar{x} - \mu)^T = \\ &= A + n(\bar{x} - \mu)(\bar{x} - \mu)^T, \end{aligned}$$

где \bar{x} и A определены соответственно в (14.3) и (14.4).

Подставим правую часть последнего соотношения в выражение для логарифмической функции правдоподобия, воспользуемся приведенными выше свойствами следа матрицы и получим

$$\begin{aligned} l(\mu, \Sigma) &= -\frac{nN}{2} \ln(2\pi) + \frac{n}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \text{tr}(\Sigma^{-1}A) - \\ &\quad - \frac{n}{2} (\bar{x} - \mu)^T \Sigma^{-1} (\bar{x} - \mu). \end{aligned} \quad (14.6)$$

Из вида (14.6) логарифмической функции правдоподобия $l(\mu, \Sigma)$ заключаем, что максимум по $\mu \in \mathbb{R}^N$ в ней достигается лишь на выборочном среднем $\hat{\mu} = \bar{x}$, которое и является МП-оценкой для вектора математического ожидания μ .

Из (14.6) также видно, что вместо МП-оценки $\hat{\Sigma}$ для ковариационной матрицы Σ целесообразно строить МП-оценку $\widehat{\Sigma^{-1}}$ для обратной ковариационной матрицы Σ^{-1} . В силу взаимно-однозначного соответствия между Σ и Σ^{-1} МП-оценкой для Σ будет $\hat{\Sigma} = (\widehat{\Sigma^{-1}})^{-1}$.

Найдем МП-оценку для $\Sigma^{-1} = (\sigma_{kj}^*)_{k,j=1}^N$ из решения задачи:

$$l^*(\Sigma^{-1}) = l(\bar{x}, \Sigma) = -\frac{nN}{2} \ln(2\pi) + \frac{n}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \text{tr}(\Sigma^{-1}A) \rightarrow \max_{\Sigma^{-1}}.$$

Вычислим частные производные от $l^*(\Sigma^{-1})$ по σ_{kj}^* , $k, j = 1, \dots, N$, используя поэлементное представление матрицы $A = (a_{kj})_{k,j=1}^N$, разложение определителя по строке и известное в матричном анализе представление Крамера для элементов обратной матрицы [34, 11]:

$$\frac{\partial l^*(\Sigma^{-1})}{\partial \sigma_{kj}^*} = \frac{n}{2} \frac{1}{|\Sigma^{-1}|} \frac{\partial |\Sigma^{-1}|}{\partial \sigma_{kj}^*} - \frac{1}{2} \frac{\partial}{\partial \sigma_{kj}^*} \sum_{i,l=1}^N \sigma_{il}^* a_{li} = \frac{n}{2} \sigma_{kj} - \frac{1}{2} a_{jk}.$$

Согласно необходимому условию максимума полученные выражения для производных приравняем к нулю и получаем МП-оценку $\hat{\Sigma} = (\hat{\sigma}_{kj})_{k,j=1}^N$ для ковариационной матрицы $\Sigma = (\sigma_{kj})_{k,j=1}^N$: $\hat{\sigma}_{kj} = a_{jk}/n$, $j, k = 1, \dots, N$, откуда в силу симметричности матрицы A : $a_{jk} = a_{kj}$, $j, k = 1, \dots, N$, окончательно имеем $\hat{\Sigma} = \frac{A}{n}$, что совпадает с (14.4). Условие $n > N$ обеспечивает невырожденность полученной оценки ковариационной матрицы: $P\{|A| = 0\} = 0$. \square

14.3. ВЕРОЯТНОСТНЫЕ СВОЙСТВА ВЫБОРОЧНОГО СРЕДНЕГО И ВЫБОРОЧНОЙ КОВАРИАЦИОННОЙ МАТРИЦЫ. НЕСМЕЩЕННАЯ ВЫБОРОЧНАЯ КОВАРИАЦИОННАЯ МАТРИЦА

Пусть по выборке $X = \{x_1, \dots, x_n\}$ объема $n > N$ из независимых в совокупности многомерных наблюдений с невырожденным N -мерным нормальным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$) в качестве оценок вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы Σ используются МП-оценки: $\hat{\mu} = \bar{x}$ – выборочное среднее и $\hat{\Sigma} = \frac{A}{n}$ – выборочная ковариационная матрица, полученные в теореме 14.3. Как МП-оценки они обладают всеми их свойствами: сильная состоятельность, асимптотическая несмещенность и асимптотическая нормальность. Однако в силу специфики многомерного нормального распределения они имеют ряд дополнительных свойств. Сначала исследуем их распределения вероятностей.

Определение 14.2. Говорят, что случайная $(N \times N)$ -матрица A имеет распределение Уишарта $W_N(\Sigma, m)$, если она распределена как матрица $A = \sum_{j=1}^m z_j z_j^T$, где случайные N -векторы $\{z_j\}_{j=1}^m$ независимы в совокупности и одинаково распределены по невырожденному нормальному закону $N_N(\mathbf{0}_N, \Sigma)$ ($|\Sigma| \neq 0$).

Распределение Уишарта обладает следующими очевидными свойствами.

- C1. Если случайные матрицы $\{A_k\}_{k=1}^K$ независимы в совокупности и имеют распределения Уишарта: $\mathcal{L}\{A_k\} = W_N(\Sigma, m_k)$, $k = 1, \dots, K$, то их сумма также имеет распределение Уишарта: $\mathcal{L}\left\{\sum_{k=1}^K A_k\right\} = W_N\left(\Sigma, \sum_{k=1}^K m_k\right)$.
- C2. Если случайная матрица A имеет распределение Уишарта $W_N(\Sigma, m)$, то $E\{A\} = m\Sigma$, и $\mathcal{L}\{cA\} = W_N(c\Sigma, m)$, где $c > 0$ – константа.

Теорема 14.4. В условиях теоремы 14.3 МП-оценки $\hat{\mu} = \bar{x}$ и $\hat{\Sigma} = \frac{A}{n}$, определенные в (14.3) и (14.4), независимы и имеют следующие распределения вероятностей:

$$\mathcal{L}\{\bar{x}\} = N_N\left(\mu, \frac{1}{n}\Sigma\right), \quad \mathcal{L}\{n\hat{\Sigma}\} = \mathcal{L}\{A\} = W_N(\Sigma, n-1). \quad (14.7)$$

Доказательство. Построим ортогональное преобразование выборки $X = \{x_1, \dots, x_n\}$. Через $C = (c_{ij})_{i,j=1}^n$ обозначим специальную ортогональную $(n \times n)$ -матрицу, обладающую свойством $CC^T = C^T C = I_n$ ($C^{-1} = C^T$) [34, 11]. Наложим на матрицу $C = (c_{ij})_{i,j=1}^n$ ограничение, положив элементы последней строки в ней равными $\frac{1}{\sqrt{n}}$: $c_{nj} = \frac{1}{\sqrt{n}}$, $j = 1, \dots, n$. Преобразуем выборку $X = \{x_t\}_{t=1}^n$ в $Z = \{z_t\}_{t=1}^n$:

$$z_t = \sum_{j=1}^n c_{tj} x_j, \quad t = 1, \dots, n. \quad (14.8)$$

Случайные векторы $\{z_t\}_{t=1}^n$ являются линейными преобразованиями нормально распределенных, независимых в совокупности случайных векторов $\{x_t\}_{t=1}^n$, имеющих по следствию 14.4 совместное нормальное распределение. По теореме 14.2 $\{z_t\}_{t=1}^n$ также имеют совместное нормальное распределение и по следствию 14.1 нормально распределены: $\mathcal{L}\{z_t\} = N_N(v_t, \Sigma_t)$, $t = 1, \dots, n$. Найдем параметры этих распределений.

Для математических ожиданий имеем

$$\begin{aligned}\mathbf{v}_t &= \mathbf{E}\{z_t\} = \mathbf{E}\left\{\sum_{j=1}^n c_{tj}x_j\right\} = \sum_{j=1}^n c_{tj}\boldsymbol{\mu} = \\ &= \sum_{j=1}^n c_{tj}\frac{1}{\sqrt{n}}\sqrt{n}\boldsymbol{\mu} = \sum_{j=1}^n c_{tj}c_{nj}\sqrt{n}\boldsymbol{\mu} = \delta_{tn}\sqrt{n}\boldsymbol{\mu}, \quad t = 1, \dots, n.\end{aligned}$$

Вместо вычисления ковариационных матриц $\{\Sigma_t\}_{t=1}^n$ решим более общую задачу – найдем ковариации векторов $\{z_t\}_{t=1}^n$:

$$\begin{aligned}\text{cov}\{z_t, z_l\} &= \mathbf{E}\{(z_t - \mathbf{v}_t)(z_l - \mathbf{v}_l)^T\} = \\ &= \mathbf{E}\left\{\left(\sum_{j=1}^n c_{tj}x_j - \sum_{j=1}^n c_{tj}\boldsymbol{\mu}\right)\left(\sum_{i=1}^n c_{li}x_i - \sum_{i=1}^n c_{li}\boldsymbol{\mu}\right)^T\right\} = \\ &= \sum_{j,i=1}^n c_{tj}\mathbf{E}\{(x_j - \boldsymbol{\mu})(x_i - \boldsymbol{\mu})^T\}c_{li} = \\ &= \sum_{j,i=1}^n c_{tj}\delta_{ji}\Sigma c_{li} = \sum_{j=1}^n c_{tj}c_{lj}\Sigma = \delta_{tl}\Sigma, \quad t, l = 1, \dots, n,\end{aligned}$$

где учтено, что $\{x_i\}_{i=1}^n$ независимы в совокупности, и использовано свойство ортогональной матрицы: $CC^T = I_n$.

Таким образом, случайные векторы $\{z_t\}_{t=1}^n$ имеют совместное нормальное распределение, некоррелированы, имеют следующие нормальные распределения:

$$\mathcal{L}\{z_t\} = N_N(\mathbf{0}_N, \Sigma), \quad t = 1, \dots, n-1, \quad \mathcal{L}\{z_n\} = N_N(\sqrt{n}\boldsymbol{\mu}, \Sigma), \quad (14.9)$$

и по следствию 14.3 независимы в совокупности.

Выразим теперь оценки \bar{x} и $\hat{\Sigma}$ через $\{z_t\}_{t=1}^n$.

Из (14.8) и (14.3) получим

$$z_n = \sum_{j=1}^n c_{nj}x_j = \frac{1}{\sqrt{n}} \sum_{j=1}^n x_j = \sqrt{n}\bar{x},$$

и выборочное среднее \bar{x} допускает представление

$$\bar{x} = \frac{1}{\sqrt{n}}z_n, \quad (14.10)$$

является линейным преобразованием z_n и имеет нормальное распределение $N_N\left(\boldsymbol{\mu}, \frac{\Sigma}{n}\right)$, что доказывает первую часть (14.7).

С учетом тождества (14.5) запишем выборочную ковариационную матрицу $\hat{\Sigma}$ из (14.4):

$$\hat{\Sigma} = \frac{1}{n}A = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T = \frac{1}{n} \left(\sum_{t=1}^n x_t x_t^T - n\bar{x}\bar{x}^T \right).$$

Воспользовавшись вспомогательным соотношением

$$\begin{aligned} \sum_{t=1}^n z_t z_t^T &= \sum_{t=1}^n \sum_{j=1}^n c_{tj} x_j \left(\sum_{i=1}^n c_{ti} x_i \right)^T = \\ &= \sum_{j,i=1}^n \sum_{t=1}^n c_{tj} c_{ti} x_j x_i^T = \sum_{j,i=1}^n \delta_{ji} x_j x_i^T = \sum_{t=1}^n x_t x_t^T, \end{aligned}$$

и доказанным выше представлением (14.10), получим

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n-1} z_t z_t^T. \quad (14.11)$$

Согласно (14.9) и определению распределения Уишарта: $\mathcal{L}\{n\hat{\Sigma}\} = \mathcal{L}\{A\} = W_N(\Sigma, n-1)$, и (14.7) окончательно доказано.

Независимость статистических оценок \bar{x} и $\hat{\Sigma}$ непосредственно следует из представлений (14.10) и (14.11), согласно которым \bar{x} и $\hat{\Sigma}$ не содержат общих случайных векторов из $\{z_t\}_{t=1}^n$: \bar{x} определяется z_n , а $\hat{\Sigma}$ строится на основе $\{z_t\}_{t=1}^{n-1}$. \square

Теперь исследуем другие вероятностные свойства оценок \bar{x} и $\hat{\Sigma}$.

Теорема 14.5. В условиях теоремы 14.3 выборочное среднее \bar{x} из (14.3) является несмещенной оценкой для математического ожидания μ : $E\{\bar{x}\} = \mu$, $n \geq 1$, а выборочная ковариационная матрица $\hat{\Sigma} = \frac{A}{n}$ — асимптотически несмещенной: $E\{\hat{\Sigma}\} \rightarrow \Sigma$, $n \rightarrow +\infty$.

Несмещенной оценкой ковариационной матрицы является

$$S = \frac{1}{n-1} A, \quad (14.12)$$

где матрица $A = (a_{kl})_{k,l=1}^N$ определена в (14.4).

Все оценки \bar{x} , $\hat{\Sigma}$ и S сильно состоятельны:

$$\bar{x} \xrightarrow{p. n.} \mu, \quad \hat{\Sigma} \xrightarrow{p. n.} \Sigma, \quad S \xrightarrow{p. n.} \Sigma, \quad n \rightarrow +\infty,$$

выборочное среднее \bar{x} имеет нормальное распределение: $\mathcal{L}\{\bar{x}\} = N_N\left(\mu, \frac{1}{n}\Sigma\right)$, $n \geq 1$, а элементы выборочной ковариационной матрицы $\hat{\Sigma} = (\hat{\sigma}_{kl})_{k,l=1}^N$ и несмещенной выборочной ковариационной матрицы $S = (s_{kl})_{k,l=1}^N$ — совместные асимптотически нормальные распределения с ковариациями, определяемыми элементами истинной ковариационной матрицы $\Sigma = (\sigma_{kl})_{k,l=1}^N$:

$$\begin{aligned} \lim_{n \rightarrow +\infty} n \cdot \text{cov}\{\hat{\sigma}_{ij}, \hat{\sigma}_{kl}\} &= \lim_{n \rightarrow +\infty} n \cdot \text{cov}\{s_{ij}, s_{kl}\} = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{kj}, \\ i, j, k, l &= 1, \dots, N. \end{aligned} \quad (14.13)$$

Доказательство. Справедливость большинства утверждений теоремы следует из свойств МП-оценок, которыми являются $\hat{\mu} = \bar{x}$ и $\hat{\Sigma}$. Однако докажем их все, основываясь на представлениях для $\hat{\mu} = \bar{x}$ и $\hat{\Sigma}$, полученных при доказательстве теоремы 14.4.

Несмещенность и нормальность выборочного среднего \bar{x} ($n \geq 1$) уже доказаны в теореме 14.4.

С учетом (14.11), (14.9) вычислим математическое ожидание выборочной ковариационной матрицы $\hat{\Sigma}$:

$$\mathbf{E}\{\hat{\Sigma}\} = \mathbf{E}\left\{\frac{1}{n} \sum_{t=1}^{n-1} z_t z_t^T\right\} = \frac{n-1}{n} \mathbf{E}\{z_t z_t^T\} = \frac{n-1}{n} \Sigma \rightarrow \Sigma, \quad n \rightarrow +\infty,$$

и асимптотическая несмещенность $\hat{\Sigma}$ доказана. С учетом последнего соотношения для оценки $S = (1 - n^{-1})\hat{\Sigma}$ из (14.12) имеем $\mathbf{E}\{S\} = \Sigma$, $n \geq 1$, что означает ее несмещенность.

Сильная состоятельность оценок \bar{x} , $\hat{\Sigma}$ и S следует из усиленного закона больших чисел для независимых в совокупности, одинаково распределенных случайных векторов ($n \rightarrow +\infty$):

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \xrightarrow{\text{п. н.}} \mu;$$

$$\hat{\Sigma} = \frac{n-1}{n} \frac{1}{n-1} \sum_{t=1}^{n-1} z_t z_t^T \xrightarrow{\text{п. н.}} \Sigma; \quad S = \frac{1}{n-1} \sum_{t=1}^{n-1} z_t z_t^T \xrightarrow{\text{п. н.}} \Sigma,$$

где $\{x_t\}_{t=1}^n$ — независимы в совокупности и одинаково распределены с математическим ожиданием $\mu = \mathbf{E}\{x_t\}$, а $\{z_t z_t^T\}_{t=1}^{n-1}$ — независимые в совокупности, одинаково распределенные случайные матрицы с $\mathbf{E}\{z_t z_t^T\} = \Sigma$.

Доказательство последнего утверждения теоремы основано на использовании центральной предельной теоремы для одинаково распределенных случайных векторов [4]: пусть $v_1, \dots, v_m, v_{m+1}, \dots \in \mathbb{R}^q$ — последовательность независимых в совокупности, одинаково распределенных случайных векторов с математическим ожиданием $\mu_v = \mathbf{E}\{v_t\}$ и ковариационной матрицей $\Sigma_v = \text{cov}\{v_t, v_t\}$, $t = 1, \dots, m, m+1, \dots$, тогда

$$\mathcal{L}\left\{\frac{1}{\sqrt{m}} \sum_{t=1}^m (v_t - \mu_v)\right\} \rightarrow N_q(\mathbf{0}_q, \Sigma_v), \quad m \rightarrow +\infty.$$

Согласно (14.11): $n\hat{\Sigma} = (n-1)S = A = \sum_{t=1}^{n-1} z_t z_t^T$, где случайные N -векторы $z_t = (z_{t1}, \dots, z_{tN})$, $t = 1, \dots, n-1$, независимы в совокупности и согласно (14.9) распределены по закону $N_N(\mathbf{0}_N, \Sigma)$. Поэтому, будучи составленными в вектор, нормированные элементы

$$\left\{\frac{a_{kl} - (n-1)\sigma_{kl}}{\sqrt{n-1}}\right\}_{k \geq l=1}^N$$

матрицы $A = (a_{kl})_{k,l=1}^N$: $a_{kl} = \sum_{t=1}^{n-1} z_{tk} z_{tl}$ (учтено, что $\mathbf{E}\left\{\frac{a_{kl}}{n-1}\right\} = \sigma_{kl}$ и $a_{kl} = a_{lk}$), имеют асимптотически нормальное ($n \rightarrow +\infty$) совместное $N(N+1)/2$ -мерное распределение с нулевыми математическими ожиданиями и ковариациями:

$$\begin{aligned} & \text{cov}\left\{\frac{a_{ij} - (n-1)\sigma_{ij}}{\sqrt{n-1}}, \frac{a_{kl} - (n-1)\sigma_{kl}}{\sqrt{n-1}}\right\} = \\ & = \text{cov}\{z_{ti} z_{tj}, z_{tk} z_{tl}\} = \mathbf{E}\{z_{ti} z_{tj} z_{tk} z_{tl}\} - \mathbf{E}\{z_{ti} z_{tj}\} \mathbf{E}\{z_{tk} z_{tl}\} = \\ & = \sigma_{ij} \sigma_{kl} + \sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{kj} - \sigma_{ij} \sigma_{kl} = \sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{kj}, \quad i, j, k, l = 1, \dots, N, \end{aligned}$$

где использована известная формула для моментов четвертого порядка многомерного нормального распределения из п. 14.1.

Очевидное соотношение $(i, j, k, l = 1, \dots, N)$

$$\begin{aligned} \lim_{n \rightarrow +\infty} n \cdot \text{cov}\{\hat{\sigma}_{ij}, \hat{\sigma}_{kl}\} &= \lim_{n \rightarrow +\infty} n \cdot \text{cov}\{s_{ij}, s_{kl}\} = \\ &= \lim_{n \rightarrow +\infty} \text{cov}\left\{\frac{a_{ij} - (n-1)\sigma_{ij}}{\sqrt{n-1}}, \frac{a_{kl} - (n-1)\sigma_{kl}}{\sqrt{n-1}}\right\} \end{aligned}$$

завершает доказательство. \square

Замечание 14.3. В теореме 14.4 было установлено, что матрица A из (14.4), на основе которой строятся оценки $\hat{\Sigma} = \frac{A}{n}$ и $S = \frac{A}{n-1}$ для ковариационной матрицы Σ , имеет распределение Уишарта: $\mathcal{L}\{A\} = W_N(\Sigma, n-1)$, и по свойству C2 распределения Уишарта оценки $\hat{\Sigma}$ и S также имеют это распределение: $\mathcal{L}\{\hat{\Sigma}\} = W_N\left(\frac{\Sigma}{n}, n-1\right)$, $\mathcal{L}\{S\} = W_N\left(\frac{\Sigma}{n-1}, n-1\right)$, и при этом не требуется асимптотика $n \rightarrow +\infty$. Однако, несмотря на то что для плотности распределения Уишарта получено аналитическое представление [4], удобнее использовать более простое асимптотическое ($n \rightarrow +\infty$) нормальное распределение для элементов $\hat{\Sigma}$ и S , полученное в теореме 14.5.

Построенные в данной главе выборочное среднее $\bar{x} = (\bar{x}_k)_{k=1}^N$, выборочная ковариационная матрица $\hat{\Sigma} = (\hat{\sigma}_{kl})_{k,l=1}^N$: $\hat{\sigma}_{kl} = \frac{a_{kl}}{n}$ и ее несмещенный аналог $S = (s_{kl})_{k,l=1}^N$: $s_{kl} = \frac{a_{kl}}{n-1}$ легко вычисляются покомпонентно по выборке $X = \{x_t\}_{t=1}^n$ объема n , где $x_t = (x_{t1}, \dots, x_{tN})^T \in \mathbb{R}^N$, $t = 1, \dots, n$:

$$\bar{x}_k = \frac{1}{n} \sum_{t=1}^n x_{tk}, \quad a_{kl} = \sum_{t=1}^n (x_{tk} - \bar{x}_k)(x_{tl} - \bar{x}_l), \quad k, l = 1, \dots, N, \quad (14.14)$$

и широко используются в многомерном статистическом анализе (см. пример в конце п. 15.4). Не предполагая нормальности наблюдений из выборки, их также можно рассматривать как оценки для вектора математического ожидания и ковариационной матрицы по многомерному аналогу метода моментов (п. 11.6). Однако при этом они как статистические оценки теряют ряд своих свойств. Например, в этом случае затруднительно установить их распределение вероятностей.

14.4. УПРАЖНЕНИЯ

1. Доказать соотношения для центральных моментов третьего и четвертого порядков многомерного нормального распределения, приведенные в пунктах б) и в) (п. 14.1).
2. Привести алгоритм вычисления матрицы $\Sigma^{1/2}$ для невырожденной ковариационной матрицы Σ ($\Sigma \succ 0$): $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^T$.
3. Пусть случайный вектор $x = (\tilde{x}_1, \tilde{x}_2)^T$ имеет двухмерное нормальное распределение

$$N_2(\mu, \Sigma), \quad \mu = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Показать, что его плотность может быть вычислена по формуле

$$p(x) = n_2(x|\mu, \Sigma) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \times \\ \times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(\tilde{x}_1 - \tilde{\mu}_1)^2}{\sigma_{11}} - 2\rho\frac{(\tilde{x}_1 - \tilde{\mu}_1)(\tilde{x}_2 - \tilde{\mu}_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(\tilde{x}_2 - \tilde{\mu}_2)^2}{\sigma_{22}}\right)\right),$$

где $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$ – коэффициент корреляции между \tilde{x}_1 и \tilde{x}_2 .

4. Пусть $\{\tilde{x}_i\}_{i=1}^N$ независимы и $\tilde{x}_i \sim N_1(\beta + \gamma z_i, \sigma^2)$, где z_i – заданные числа ($i=1, \dots, N$) и $\sum_{i=1}^N z_i = 0$.

а) Найти совместное распределение $x = (\tilde{x}_1, \dots, \tilde{x}_N)^T$.

- б) Найти совместное распределение $\bar{x} = \frac{\sum_{i=1}^N \tilde{x}_i}{N}$ и $g = \frac{\sum_{i=1}^N \tilde{x}_i z_i}{\sum_{j=1}^N z_j^2}$.

5. Пусть $(x_i, y_i)^T$ ($i = 1, 2, 3$) независимы и одинаково распределены:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right).$$

а) Найти совместное распределение всех этих шести случайных величин.

б) Найти совместное распределение $\sum_{i=1}^3 \frac{x_i}{3}, \sum_{i=1}^3 \frac{y_i}{3}$.

6. Пусть $x = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)^T \sim N_3(\mu, \Sigma)$,

$$\mu = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 2 \end{pmatrix}.$$

Найти ковариационную матрицу случайного вектора $y = (\tilde{y}_1, \tilde{y}_2)^T \in \mathbb{R}^2$, где $\tilde{y}_1 = \tilde{x}_1 + \tilde{x}_2$, $\tilde{y}_2 = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3$.

Глава 15

ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ

15.1. ИССЛЕДОВАНИЕ ПАРНОЙ ЗАВИСИМОСТИ ПРИЗНАКОВ. ВЫБОРОЧНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ. Z-СТАТИСТИКА ФИШЕРА И ПРОВЕРКА ГИПОТЕЗ О ЗНАЧЕНИИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Пусть в пространстве из $N \geq 1$ признаков наблюдается выборка $X = \{x_1, \dots, x_n\}$ объема n , образованная наблюдениями с вектором математического ожидания $\mu = E\{x_t\} \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицей $\Sigma = (\sigma_{kl})_{k,l=1}^N = \text{cov}\{x_t, x_t\}$ ($t = 1, \dots, n$). Элементы ковариационной матрицы σ_{kl} , $k, l = 1, \dots, N$, описывают *парные зависимости* признаков между собой. Однако на практике для исследования парной зависимости признаков чаще используется коэффициент корреляции, называемый еще *парной корреляцией* ($k, l = 1, \dots, N$):

$$\rho_{kl} = \frac{\sigma_{kl}}{\sqrt{\sigma_{kk}\sigma_{ll}}}, \quad (15.1)$$

и его статистическая оценка по выборке X – *выборочный коэффициент корреляции*:

$$r_{kl} = \frac{a_{kl}}{\sqrt{a_{kk}a_{ll}}}, \quad (15.2)$$

получаемый из (15.1) подстановкой вместо неизвестных элементов ковариационной матрицы $\Sigma = (\sigma_{kl})_{k,l=1}^N$ соответствующих элементов выборочной ковариационной матрицы $\hat{\Sigma} = (\hat{\sigma}_{kl})_{k,l=1}^N$: $\hat{\sigma}_{kl} = a_{kl}/n$, или несмещенной выборочной ковариационной матрицы $S = (s_{kl})_{k,l=1}^N$: $s_{kl} = a_{kl}/(n-1)$, где статистики a_{kl} , $k, l = 1, \dots, N$, определены в (14.14).

Коэффициент корреляции ρ_{kl} и его оценка r_{kl} обладают рядом удобных для проведения исследований простых свойств ($k, l = 1, \dots, N$).

C1. $-1 \leq \rho_{kl} \leq 1$, $-1 \leq r_{kl} \leq 1$.

C2. ρ_{kl} и r_{kl} инвариантны относительно масштабного преобразования признаков в выборке, т. е. не изменяют своих значений при умножении признаков в наблюдениях $x_t = (x_{t1}, \dots, x_{tN})^T$ на соответствующие константы: $x_t := (b_1x_{t1}, \dots, b_Nx_{tN})^T$ ($t = 1, \dots, n$).

C3. Если k -й и l -й признаки независимы, то $\rho_{kl} = 0$ (обратное, вообще говоря, неверно). Если $|\rho_{kl}| = 1$, то k -й и l -й признаки x_{tk} и x_{tl} связаны линейной зависимостью: $x_{tk} \stackrel{\text{п. н.}}{=} b_{kl}x_{tl} + d_{kl}$, где b_{kl} и d_{kl} – константы (верно и обратное).

Предположим теперь, что наблюдения из выборки $X = \{x_1, \dots, x_n\}$ независимы в совокупности и имеют невырожденное многомерное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Исследуем дополнительные свойства выборочного коэффициента корреляции (15.2), который в этом случае является оценкой максимального правдоподобия для коэффициента корреляции (15.1) (поскольку (15.2) – функциональное преобразование выборочной ковариационной матрицы, являющейся оценкой максимального правдоподобия).

Нам понадобится следующая лемма [4].

Лемма 15.1. Пусть $U_1, \dots, U_m \in \mathbb{R}^M$ – асимптотически нормальная, сходящаяся по вероятности последовательность M -мерных случайных векторов ($m \rightarrow +\infty$): $U_m \xrightarrow{P} b$, $\mathcal{L}\{\sqrt{m}(U_m - b)\} \rightarrow N_M(\mathbf{0}_M, W)$, где $b \in \mathbb{R}^M$ – некоторый M -вектор; а $W = W^T \succ 0$ – положительно определенная симметричная $(M \times M)$ -матрица.

Пусть функция $\Phi = \Phi(u) : \mathbb{R}^M \rightarrow \mathbb{R}^1$ имеет в окрестности точки $u = b$ первую и вторую производные, и $\Phi_b = \nabla_u \Phi(u)|_{u=b} \neq \mathbf{0}_M$, тогда предельным распределением при $m \rightarrow +\infty$ для $\sqrt{m}(\Phi(U_m) - \Phi_b)$ будет $N_1(0, \Phi_b^T W \Phi_b)$.

Доказательство. Основано на разложении функции $\Phi = \Phi(u)$ в точке $u = b$ в ряд Тейлора с остаточным членом в форме Лагранжа: $\Phi(u) = \Phi(b) + (\nabla_u \Phi(u)|_{u=b})^T (u - b) + O(|u - b|^2)$, и на вычислении соответствующих моментов. \square

Теорема 15.1. В условиях теоремы 14.3 выборочный коэффициент корреляции r_{kl} из (15.2) имеет асимптотически нормальное распределение:

$$\mathcal{L}\{\sqrt{n-1}(r_{kl} - \rho_{kl})\} \rightarrow N_1(0, (1 - \rho_{kl}^2)^2), \quad n \rightarrow +\infty \quad (15.3)$$

при условии, что соответствующее истинное абсолютное значение коэффициента корреляции не равно единице: $|\rho_{kl}| \neq 1$ ($k, l = 1, \dots, N$).

Доказательство. Воспользуемся леммой 15.1 и определим последовательность векторов $\{U_m = (U_{m1}, U_{m2}, U_{m3})^T \in \mathbb{R}^3\}$ ($m = n - 1$), где

$$U_m = \begin{pmatrix} \frac{s_{kk}(m)}{\sigma_{kk}} \\ \frac{s_{kl}(m)}{\sqrt{\sigma_{kk}\sigma_{ll}}} \\ \frac{s_{ll}(m)}{\sigma_{ll}} \end{pmatrix} = \begin{pmatrix} \frac{1}{m} a_{kk}(m) \\ \sigma_{kk} \\ \frac{1}{m} a_{kl}(m) \\ \sqrt{\sigma_{kk}\sigma_{ll}} \\ \frac{1}{m} a_{ll}(m) \\ \sigma_{ll} \end{pmatrix},$$

а $a_{kl} = a_{kl}(m)$, $k, l = 1, \dots, N$, определены в (14.14). Из сильной состоятельности несмещенной выборочной ковариационной матрицы $S = (s_{kl})_{k,l=1}^N$: $s_{kl} = s_{kl}(m) = a_{kl}(m)/m$, доказанной в теореме 14.5, заключаем, что последовательность векторов $\{U_m\}$ также сходится почти наверное, а тем более, по вероятности, при $m \rightarrow +\infty$ ($n \rightarrow +\infty$) к вектору $b = (1, \rho_{kl}, 1)^T$. Согласно той же теореме и доказательству к ней в силу асимптотической нормальности элементов матрицы $S = (s_{kl})_{k,l=1}^N$ из соотношений (14.13) имеем: $\mathcal{L}\{\sqrt{m}(U_m - b)\} \rightarrow N_3(\mathbf{0}_3, W)$, $m \rightarrow +\infty$, где

$$W = \begin{pmatrix} 2 & 2\rho_{kl} & 2\rho_{kl}^2 \\ 2\rho_{kl} & 1 + \rho_{kl}^2 & 2\rho_{kl} \\ 2\rho_{kl}^2 & 2\rho_{kl} & 2 \end{pmatrix}.$$

Выразим выборочный коэффициент корреляции $r_{kl} = r_{kl}(m)$ из (15.2) через вектор $U_m = (U_{m1}, U_{m2}, U_{m3})^T$: $r_{kl} = r_{kl}(m) = \Phi(U_m) = \frac{U_{m2}}{\sqrt{U_{m1}U_{m3}}}$.

Вычислим производные от функции $\Phi(u) = u_2/\sqrt{u_1 u_3}$ по $u = (u_1, u_2, u_3)^T$ в точке $u = b = (1, \rho_{kl}, 1)^T$:

$$\Phi_b = \begin{pmatrix} -\frac{1}{2} u_1^{-\frac{3}{2}} u_2 u_3^{-\frac{1}{2}} \\ u_1^{-\frac{1}{2}} u_3^{-\frac{1}{2}} \\ -\frac{1}{2} u_1^{-\frac{1}{2}} u_2 u_3^{-\frac{3}{2}} \end{pmatrix} \Big|_{u=b} = \begin{pmatrix} -\frac{1}{2} \rho_{kl} \\ 1 \\ -\frac{1}{2} \rho_{kl} \end{pmatrix} \neq \mathbf{0}_3.$$

Согласно лемме 15.1 подставим полученные выражения для W и Φ_b в предельное распределение $N_1(0, \Phi_b^T W \Phi_b)$ и получим доказываемое. \square

Следствие 15.1. Пусть $\Psi = \Psi(v) : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ — дважды дифференцируема в точке $v = \rho_{kl}$ и $\frac{d}{dv}\Psi(v)|_{v=\rho_{kl}} \neq 0$, тогда в условиях теоремы 15.1 для выборочного коэффициента корреляции r_{kl} выполняется ($n \rightarrow +\infty$):

$$\mathcal{L}\{\sqrt{n-1}(\Psi(r_{kl}) - \Psi(\rho_{kl}))\} \rightarrow N_1\left(0, \left(\frac{d}{dv}\Psi(v)\Big|_{v=\rho_{kl}}\right)^2 (1 - \rho_{kl}^2)^2\right). \quad (15.4)$$

Доказательство следует из результата теоремы 15.1 и леммы 15.1. \square

Замечание 15.1. В соотношениях (15.3) и (15.4) вместо $n-1$ можно использовать n (поскольку $n \rightarrow +\infty$), однако считается, что $n-1$ «точнее».

Выберем преобразование $\Psi = \Psi(v) : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ так, чтобы дисперсия предельного нормального распределения в (15.4) была единичной:

$$\left(\frac{d}{dv}\Psi(v)\Big|_{v=\rho_{kl}}\right)^2 (1 - \rho_{kl}^2)^2 = 1.$$

Получим дифференциальное уравнение:

$$\frac{d}{dv}\Psi(v)\Big|_{v=\rho_{kl}} = \frac{1}{1 - \rho_{kl}^2} = \frac{1}{(1 - \rho_{kl})(1 + \rho_{kl})} = \frac{1}{2} \left(\frac{1}{1 + \rho_{kl}} + \frac{1}{1 - \rho_{kl}} \right),$$

решая которое, найдем $\Psi(v) = \frac{1}{2} \ln \frac{1+v}{1-v}$.

Определение 15.1. Пусть r_{kl} — выборочный коэффициент корреляции (15.2), тогда статистика

$$Z = Z(r_{kl}) = \Psi(r_{kl}) = \frac{1}{2} \ln \frac{1 + r_{kl}}{1 - r_{kl}} \quad (15.5)$$

называется *Z-Статистикой Фишера*, а преобразование $\Psi(\cdot)$ — *Z-преобразованием Фишера*.

Следствие 15.2. В условиях теоремы 15.1 случайная величина $\sqrt{n-1}(Z(r_{kl}) - Z(\rho_{kl}))$ имеет при $n \rightarrow +\infty$ стандартное нормальное распределение $N_1(0, 1)$.

Z -статистика Фишера (15.5) имеет большое прикладное значение — с ее помощью проверяются гипотезы о значении коэффициента корреляции:

$$\begin{aligned} H_0 & : \rho_{kl} = \rho_{kl}^o; \\ H_1 = \overline{H_0} & : \rho_{kl} \neq \rho_{kl}^o, \end{aligned}$$

где ρ_{kl}^o — предполагаемое значение коэффициента корреляции.

Воспользуемся следствием 15.2 и получим при $n \rightarrow +\infty$ следующий критерий для проверки гипотез H_0, H_1 :

$$\begin{cases} H_0 & : |\sqrt{n-1}(Z(r_{kl}) - Z(\rho_{kl}^o))| \leq \Delta; \\ H_1 = \overline{H_0} & : |\sqrt{n-1}(Z(r_{kl}) - Z(\rho_{kl}^o))| > \Delta, \end{cases} \quad (15.6)$$

где порог критерия Δ определяется по наперед заданному малому значению уровня значимости $\alpha = P\{H_1|H_0\} = 1 - P\{H_0|H_0\} \in (0, 1)$ (вероятность принять гипотезу H_1 при условии, что верна гипотеза H_0) с учетом того, что при верной H_0 статистика $\sqrt{n-1}(Z(r_{kl}) - Z(\rho_{kl}^o))$ критерия (15.6) имеет стандартное нормальное распределение:

$$\Delta = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad (15.7)$$

квантиль уровня $1 - \alpha/2$ стандартного нормального закона $N_1(0, 1)$.

Замечание 15.2. Построенный критерий (15.6), (15.7) позволяет также определить доверительный интервал (п. 11.11) уровня $1 - \alpha$ для ρ_{kl} :

$$P \left\{ \text{th} \left(Z(r_{kl}) - \frac{\Delta}{\sqrt{n-1}} \right) < \rho_{kl} < \text{th} \left(Z(r_{kl}) + \frac{\Delta}{\sqrt{n-1}} \right) \right\} = 1 - \alpha,$$

где $\Psi^{-1}(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}} = \text{th}(v)$ – преобразование, обратное к преобразованию Фишера $\Psi(v) = \frac{1}{2} \ln \frac{1+v}{1-v}$.

Замечание 15.3. На практике чаще всего проверяются гипотезы:

$$\begin{aligned} H_0 &: \rho_{kl} = 0; \\ H_1 = \overline{H_0} &: \rho_{kl} \neq 0. \end{aligned}$$

Если принимается H_0 , то k -й и l -й признаки ($k \neq l = 1, \dots, N$) считаются некоррелированными, а в силу предположения их совместной нормальности – и независимыми. Поэтому гипотезу H_0 иногда называют гипотезой независимости. Проверяются гипотезы H_0, H_1 при помощи критерия (15.6), (15.7) ($\rho_{kl}^0 = 0$ и $Z(\rho_{kl}^0) = Z(0) = 0$).

В случае выборки «малого объема» (асимптотика $n \rightarrow +\infty$ не предполагается) для проверки гипотез H_0, H_1 считается целесообразным использование другого критерия:

$$\begin{cases} H_0 &: \sqrt{n-2} \frac{|r_{kl}|}{\sqrt{1-r_{kl}^2}} \leq \Delta; \\ H_1 = \overline{H_0} &: \sqrt{n-2} \frac{|r_{kl}|}{\sqrt{1-r_{kl}^2}} > \Delta, \end{cases}$$

где относительно статистики критерия в п. 15.4 (теорема 15.5, соотношение (15.20)) будет установлено, что при истинной гипотезе H_0 она является модулем случайной величины, имеющей t -распределение Стьюдента с $n - 2$ степенями свободы и порог критерия $\Delta = F_{t_{n-2}}^{-1}(1 - \alpha/2)$ – соответствующая квантиль уровня $1 - \alpha/2$ ($\alpha = P\{H_1|H_0\} \in (0, 1)$). Однако данный критерий «более чувствителен» к отклонениям распределения вероятностей выборочных значений компонент от нормального закона, и при «достаточно большом» объеме выборки n лучше использовать критерий (15.6), (15.7), основанный на Z -статистике Фишера (15.5).

15.2. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ: УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ И ЧАСТНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Пусть случайный N -вектор-наблюдение $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Иногда необходимо исследовать зависимость определенного подмножества из m ($m < N$) его признаков между собой при фиксированных (выбранных) значениях остальных $N - m$ признаков.

Не ограничивая общности, будем предполагать, что исследуемые m признаков расположены в начале N -вектора-наблюдения $x \in \mathbb{R}^N$, что порождает следующее разбиение его на два подвектора:

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix};$$

$$x^{(1)} = (\tilde{x}_1, \dots, \tilde{x}_m)^T \in \mathbb{R}^m, \quad x^{(2)} = (\tilde{x}_{m+1}, \dots, \tilde{x}_N)^T \in \mathbb{R}^{N-m},$$

а соответствующего ему N -вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы Σ – на соответствующие блоки:

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma_{21} = \Sigma_{12}^T. \quad (15.8)$$

Вычислим условное распределение вероятностей случайного m -вектора $x^{(1)}$ при фиксированном значении компонент $(N - m)$ -вектора $x^{(2)}$.

Теорема 15.2. Пусть случайный N -вектор $x \in \mathbb{R}^N$, имеющий невырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), разбит на два подвектора: $x = \left((x^{(1)})^T : (x^{(2)})^T \right)^T$, где $x^{(1)} \in \mathbb{R}^m$, $x^{(2)} \in \mathbb{R}^{N-m}$ ($m < N$), тогда условное распределение случайного вектора $x^{(1)}$ при фиксированном значении вектора $x^{(2)}$ – m -мерное нормальное со следующими значениями параметров:

$$\begin{aligned} L\{x^{(1)}|x^{(2)}\} &= N_m(\mu_{1|2}, \Sigma_{11|2}); \\ \mu_{1|2} &= \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}), \quad \Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T. \end{aligned} \quad (15.9)$$

Доказательство. Перейдем от исходного случайного N -вектора-наблюдения $x = \left((x^{(1)})^T : (x^{(2)})^T \right)^T \in \mathbb{R}^N$ к вспомогательному вектору $y = \left((y^{(1)})^T : (y^{(2)})^T \right)^T \in \mathbb{R}^N$, используя невырожденное линейное функциональное преобразование $y = y(x) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ следующего вида:

$$y^{(1)} = x^{(1)} + Bx^{(2)}, \quad y^{(2)} = x^{(2)},$$

где $(m \times (N - m))$ -матрица преобразования B выбирается так, чтобы случайные векторы $y^{(1)}$ и $y^{(2)}$ были некоррелированными: $\text{cov}\{y^{(1)}, y^{(2)}\} = \mathbf{0}_{m \times (N-m)}$ – нулевая $(m \times (N - m))$ -матрица.

Чтобы выбрать матрицу B , вычислим ковариацию между $y^{(1)}$ и $y^{(2)}$, предварительно введя в рассмотрение центрированные случайные векторы $\dot{x}^{(j)} = x^{(j)} - \mu^{(j)}$ ($j = 1, 2$):

$$\begin{aligned} \text{cov}\{y^{(1)}, y^{(2)}\} &= \mathbf{E}\left\{ (y^{(1)} - \mathbf{E}\{y^{(1)}\}) (y^{(2)} - \mathbf{E}\{y^{(2)}\})^T \right\} = \\ &= \mathbf{E}\left\{ (\dot{x}^{(1)} + B\dot{x}^{(2)}) (\dot{x}^{(2)})^T \right\} = \Sigma_{12} + B\Sigma_{22}. \end{aligned}$$

Приравнявая правую часть полученного соотношения к $\mathbf{0}_{m \times (N-m)}$, получаем $B = -\Sigma_{12}\Sigma_{22}^{-1}$, где в силу невырожденности исходной матрицы Σ ее диагональный блок Σ_{22} также невырожден.

Отметим, что случайный вектор $y = y(x) = \left((y^{(1)})^T : (y^{(2)})^T \right)^T$ как линейное преобразование нормального случайного вектора x также имеет нормальное распределение, а его подвекторы $y^{(1)}$ и $y^{(2)}$ – маргинальные нормальные распределения: $N_m(\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11|2})$ и $N_{N-m}(\mu^{(2)}, \Sigma_{22})$ соответственно, где $\Sigma_{11|2}$ определена в (15.9) и является ковариационной матрицей вектора $y^{(1)} = x^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}x^{(2)}$. Из некоррелированности $y^{(1)}$ и $y^{(2)}$ в силу их совместного нормального распределения следует их независимость. Поэтому плотность распределения вероятностей вектора y равна произведению плотностей распределения векторов $y^{(1)}$ и $y^{(2)}$:

$$\tilde{p}(y) = n_m(y^{(1)}|\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11|2}) n_{N-m}(y^{(2)}|\mu^{(2)}, \Sigma_{22}).$$

Запишем плотность распределения вектора $x = x(y)$ как преобразования y в x :

$$x^{(1)} = y^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}y^{(2)}, \quad x^{(2)} = y^{(2)}.$$

Отметим, что преобразования здесь имеют единичный якобиан, и из известного соотношения для плотности при невырожденном преобразовании [30] имеем

$$\begin{aligned} p(x) &= \tilde{p}(y)|_{y=y(x)} = \\ &= n_m(x^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}x^{(2)}|\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11|2}) n_{N-m}(x^{(2)}|\mu^{(2)}, \Sigma_{22}) = \\ &= n_m(x^{(1)}|\mu_{1|2}, \Sigma_{11|2}) n_{N-m}(x^{(2)}|\mu^{(2)}, \Sigma_{22}), \end{aligned}$$

где m -вектор $\mu_{1|2}$ определен в (15.9).

С другой стороны, случайный вектор $x = \left((x^{(1)})^T : (x^{(2)})^T \right)^T$ имеет N -мерное распределение вероятностей с плотностью $p(x) = n_N(x|\mu, \Sigma)$, а его подвектор $x^{(2)}$ – маргинальное нормальное распределение с плотностью $n_{N-m}(x^{(2)}|\mu^{(2)}, \Sigma_{22})$. По известной формуле произведения плотностей [30] имеем

$$p(x) = n_{N-m}(x^{(2)}|\mu^{(2)}, \Sigma_{22}) p(x^{(1)}|x^{(2)}).$$

Приравнивая правые части последних двух соотношений, находим условную плотность:

$$p(x^{(1)}|x^{(2)}) = n_m(x^{(1)}|\mu_{1|2}, \Sigma_{11|2}),$$

что и завершает доказательство. \square

Следствие 15.3. В условиях теоремы 15.2 ковариационная $(m \times m)$ -матрица $\Sigma_{11|2} = \text{cov}\{x^{(1)}, x^{(1)}|x^{(2)}\}$ из (15.9) может быть записана поэлементно: $\Sigma_{11|2} = (\sigma_{kl|m+1, \dots, N})_{k,l=1}^m$, где

$$\sigma_{kl|m+1, \dots, N} = \text{cov}\{\tilde{x}_k, \tilde{x}_l|x^{(2)}\} = \sigma_{kl} - \sigma_{(k)}^T \Sigma_{22}^{-1} \sigma_{(l)} \quad (15.10)$$

условная ковариация k -й и l -й компонент $(k, l = 1, \dots, m)$ при фиксированных значениях остальных компонент $x^{(2)} = (\tilde{x}_{m+1}, \dots, \tilde{x}_N)^T$, а $\sigma_{(k)}^T = \text{cov}\{\tilde{x}_k, x_{(2)}\}$ – k -я строка матрицы Σ_{12} ($k = 1, \dots, m$).

Полученные соотношения (15.9) и (15.10) позволяют определить характеристику парной взаимосвязи на множестве выбранных компонент при фиксированных значениях остальных компонент, которая к тому же не зависит от фиксируемых значений компонент.

Определение 15.2. Частным коэффициентом корреляции между k -й и l -й компонентами \tilde{x}_k и \tilde{x}_l ($k, l = 1, \dots, m$) при фиксированных значениях компонент $x^{(2)} = (\tilde{x}_{m+1}, \dots, \tilde{x}_N)^T$ называется величина

$$\rho_{kl|m+1, \dots, N} = \frac{\sigma_{kl|m+1, \dots, N}}{\sqrt{\sigma_{kk|m+1, \dots, N} \sigma_{ll|m+1, \dots, N}}}, \quad (15.11)$$

где условная ковариация $\sigma_{kl|m+1, \dots, N}$ из (15.10) называется частной ковариацией, а $\sigma_{kk|m+1, \dots, N}$ – частной дисперсией.

Отметим, что частные корреляция, ковариация и дисперсия обладают всеми свойствами своих обычных (безусловных) аналогов.

15.3. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ: ПРОГНОЗИРОВАНИЕ, ФУНКЦИЯ РЕГРЕССИИ И МНОЖЕСТВЕННЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Пусть случайный N -вектор $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$ разбит на два подвектора: $x = \left((x^{(1)})^T : (x^{(2)})^T \right)^T$ и необходимо оценить неизвестные значения компонент из $x^{(1)} \in \mathbb{R}^m$ по наблюдаемым компонентам $x^{(2)} \in \mathbb{R}^{N-m}$ ($m < N$). Эта задача имеет большое прикладное значение при прогнозировании и оценивании («заполнении») пропущенных значений компонент в выборке.

Воспользуемся результатами предыдущего параграфа, предположив, что N -вектор $x \in \mathbb{R}^N$ по-прежнему имеет невырожденное нормальное распределение $N_N(\mu, \Sigma)$.

Теорема 15.3. В условиях теоремы 15.2 среди всех прогнозов $\hat{x}^{(1)} = (\hat{x}_j^*)_{j=1}^m$ для $x^{(1)}$ по $x^{(2)}$ прогноз $\hat{x}_*^{(1)} = (\hat{x}_j^*)_{j=1}^m = \mu_{1|2}(x^{(2)})$, где $\mu_{1|2} = \mu_{1|2}(x^{(2)}) = E\{x^{(1)}|x^{(2)}\}$ – так называемая функция регрессии, определенная в (15.9), обладает оптимальными свойствами ($j = 1, \dots, m$):

$$\hat{x}_j^* = \arg \min_{\hat{x}_j} E \left\{ \left(\hat{x}_j - \tilde{x}_j \right)^2 \right\} = \arg \max_{\hat{x}_j} \text{corr} \left\{ \hat{x}_j, \tilde{x}_j \right\},$$

причем

$$E \left\{ \left(\hat{x}_j^* - \tilde{x}_j \right)^2 \right\} = \sigma_{jj|m+1, \dots, N}; \quad (15.12)$$

$$\text{corr} \left\{ \hat{x}_j^*, \tilde{x}_j \right\} = \sqrt{\frac{\sigma_{(j)}^T \Sigma_{22}^{-1} \sigma_{(j)}}{\sigma_{jj}}}, \quad (15.13)$$

где $\Sigma_{22} = \text{cov} \{x^{(2)}, x^{(2)}\}$ – ковариационная матрица вектора $x^{(2)}$; $\sigma_{jj} = D\{\tilde{x}_j\}$ – дисперсия j -й компоненты; $(N-m)$ -вектор $\sigma_{(j)} = \text{cov} \{x^{(2)}, \tilde{x}_j\}$ определен в (15.10).

Доказательство. Оптимальность функции регрессии $\mu_{1|2} = \mu_{1|2}(x^{(2)})$ из (15.9) как прогноза $\hat{x}_*^{(1)} = \mu_{1|2}(x^{(2)})$ для $x^{(1)}$ по $x^{(2)}$ следует из общей теории стохастического прогнозирования, изложенной в п. 19.5.1.

Осталось доказать формулы (15.12) и (15.13). Воспользуемся (15.9) и запишем оптимальный прогноз $\hat{x}_*^{(1)} = (\hat{x}_j^*)_{j=1}^m$ покомпонентно ($j = 1, \dots, m$):

$$\hat{x}_j^* = \tilde{\mu}_j + \sigma_{(j)}^T \Sigma_{22}^{-1} (x^{(2)} - \mu^{(2)}), \quad (15.14)$$

где $\tilde{\mu}_j = E\{\tilde{x}_j\}$; $\mu^{(2)} = E\{x^{(2)}\}$. Учтем, что $E\{\hat{x}_j^*\} = E\{E\{\tilde{x}_j|x^{(2)}\}\} = E\{\tilde{x}_j\} = \tilde{\mu}_j$, и вычислим ковариацию:

$$\text{cov} \left\{ \hat{x}_j^*, \tilde{x}_j \right\} = E \left\{ \left(\hat{x}_j^* - \tilde{\mu}_j \right) \left(\tilde{x}_j - \tilde{\mu}_j \right) \right\} = E \left\{ \sigma_{(j)}^T \Sigma_{22}^{-1} (x^{(2)} - \mu^{(2)}) (\tilde{x}_j - \tilde{\mu}_j) \right\} = \sigma_{(j)}^T \Sigma_{22}^{-1} \sigma_{(j)}.$$

Отметим, что $D\{\tilde{x}_j\} = \sigma_{jj}$, а для $D\{\hat{x}_j^*\}$ имеем

$$\begin{aligned} D\{\hat{x}_j^*\} &= E \left\{ \left(\hat{x}_j^* - \tilde{\mu}_j \right)^2 \right\} = E \left\{ \left(\sigma_{(j)}^T \Sigma_{22}^{-1} (x^{(2)} - \mu^{(2)}) \right)^2 \right\} = \\ &= \sigma_{(j)}^T \Sigma_{22}^{-1} \text{cov} \{x^{(2)}, x^{(2)}\} \Sigma_{22}^{-1} \sigma_{(j)} = \sigma_{(j)}^T \Sigma_{22}^{-1} \sigma_{(j)}. \end{aligned}$$

Полученные соотношения и позволяют вычислить корреляцию в (15.13), а также доказать (15.12):

$$\begin{aligned} \mathbf{E} \left\{ \left(\hat{x}_j^* - \tilde{x}_j \right)^2 \right\} &= D \left\{ \hat{x}_j^* - \tilde{x}_j \right\} = D \left\{ \hat{x}_j^* \right\} - 2 \text{cov} \left\{ \hat{x}_j^*, \tilde{x}_j \right\} + D \left\{ \tilde{x}_j \right\} = \\ &= \sigma_{jj} - \sigma_{(j)}^T \Sigma_{22}^{-1} \sigma_{(j)} = \sigma_{jj|m+1, \dots, N}, \end{aligned}$$

где учтено (15.10). \square

Следствие 15.4. *Результаты теоремы 15.3 остаются верны, если не предполагать нормальность случайного вектора $x = \left((x^{(1)})^T : (x^{(2)})^T \right)^T$, считая лишь, что он имеет вторые моменты и его ковариационная матрица Σ невырождена, а класс допустимых прогнозов для $x^{(1)}$ по $x^{(2)}$ ограничить линейными по $x^{(2)}$ несмещенными прогнозами:*

$$\begin{aligned} \hat{x}^{(1)} &= \left(\hat{x}_j \right)_{j=1}^m, \quad \hat{x}_j = b_j^T x^{(2)} + d_j; \\ \mathbf{E} \left\{ \hat{x}^{(1)} - x^{(1)} \right\} &= \mathbf{0}_m. \end{aligned} \quad (15.15)$$

Доказательство. Воспользуемся условием несмещенности прогноза и найдем константы $\{d_j\}_{j=1}^m$ в (15.4):

$$d_j = \tilde{\mu}_j - b_j^T \mu^{(2)}, \quad j = 1, \dots, m.$$

Вычислим среднеквадратическую ошибку прогноза ($j = 1, \dots, m$):

$$\mathbf{E} \left\{ \left(\hat{x}_j - \tilde{x}_j \right)^2 \right\} = \mathbf{E} \left\{ \left(b_j^T (x^{(2)} - \mu^{(2)}) - (\tilde{x}_j - \tilde{\mu}_j) \right)^2 \right\} = b_j^T \Sigma_{22} b_j - 2b_j^T \sigma_{(j)} + \sigma_{jj}.$$

С учетом последнего соотношения задача определения оптимального в смысле минимума среднеквадратической ошибки прогноза сводится к задаче

$$b_j^T \Sigma_{22} b_j - 2b_j^T \sigma_{(j)} + \sigma_{jj} \rightarrow \min_{b_j \in \mathbb{R}^m},$$

необходимое условие минимума для которой

$$\nabla_{b_j} (b_j^T \Sigma_{22} b_j - 2b_j^T \sigma_{(j)} + \sigma_{jj}) = 2\Sigma_{22} b_j - 2\sigma_{(j)} = \mathbf{0}_m$$

позволяет найти оптимальное значение вектора b_j :

$$b_j = \Sigma_{22}^{-1} \sigma_{(j)}.$$

Достаточное условие максимума также выполняется: матрица вторых производных $2\Sigma_{22}$ положительно определена. Подставляя найденные значения d_j и b_j в (15.4), в качестве прогноза получаем (15.14), что является покомпонентной записью функции регрессии из (15.9), а для достигнутой среднеквадратической ошибки прогноза имеем (15.12). Для корреляции доказательство проводится аналогично. \square

Определение 15.3. *Множественным коэффициентом корреляции между компонентой \tilde{x}_j ($j = 1, \dots, m$) и компонентами из множества $\{\tilde{x}_{m+1}, \dots, \tilde{x}_N\}$ называется величина*

$$\bar{R}_{j,m+1, \dots, N} = \sqrt{\frac{\sigma_{(j)}^T \Sigma_{22}^{-1} \sigma_{(j)}}{\sigma_{jj}}}. \quad (15.16)$$

Свойства множественного коэффициента корреляции

C1. $0 \leq \bar{R}_{j,m+1,\dots,N} \leq 1$, $j = 1, \dots, m$.

Доказательство. Следует из определения (15.16) и соотношения (15.10):

$$D\{\tilde{x}_j|x^{(2)}\} = \sigma_{jj|m+1,\dots,N} = \sigma_{jj} - \sigma_{(j)}^T \Sigma_{22}^{-1} \sigma_{(j)} \geq 0, \quad j = 1, \dots, m.$$

□

C2. $\bar{R}_{j,m+1,\dots,N} = 0$ тогда и только тогда, когда j -я компонента \tilde{x}_j ($j = 1, \dots, m$) некоррелирована с компонентами из множества $\{\tilde{x}_{m+1}, \dots, \tilde{x}_N\}$.

Доказательство. Из (15.16) видно, что $\bar{R}_{j,m+1,\dots,N} = 0$ тогда и только тогда, когда $\sigma_{(j)} = \text{cov}\{x^{(2)}, \tilde{x}_j\} = \mathbf{0}_{N-m}$ ($j = 1, \dots, m$). □

C3. $\bar{R}_{j,m+1,\dots,N} = 1$ тогда и только тогда, когда j -я компонента \tilde{x}_j представима в виде $\tilde{x}_j \stackrel{\text{п.н.}}{=} b_j^T x^{(2)} + d_j$, где $b_j \in \mathbb{R}^{N-m}$, $d_j \in \mathbb{R}^1$ – детерминированы ($j = 1, \dots, m$).

Доказательство. Из (15.16), (15.13) и свойств обычного коэффициента корреляции следует, что $\bar{R}_{j,m+1,\dots,N} = \text{corr}\{\hat{\tilde{x}}_j^*, \tilde{x}_j\} = 1$ тогда и только тогда, когда $\tilde{x}_j \stackrel{\text{п.н.}}{=} \tilde{b}_j^T \hat{\tilde{x}}_j^* + \tilde{d}_j$, где $\tilde{b}_j \in \mathbb{R}^{N-m}$, $\tilde{d}_j \in \mathbb{R}^1$ – детерминированы ($j = 1, \dots, m$). Подставим вместо $\hat{\tilde{x}}_j^*$ его выражение через $x^{(2)}$ из (15.14) и получим доказываемое. □

C4. Справедливо соотношение $\sigma_{jj|m+1,\dots,N} = \left(1 - \bar{R}_{j,m+1,\dots,N}^2\right) \sigma_{jj}$, связывающее частную дисперсию $\sigma_{jj|m+1,\dots,N}$ из (15.10) и множественный коэффициент корреляции $\bar{R}_{j,m+1,\dots,N}$ ($j = 1, \dots, m$); частная дисперсия $\sigma_{jj|m+1,\dots,N}$ никогда не превосходит соответствующей безусловной дисперсии σ_{jj} : $\sigma_{jj|m+1,\dots,N} \leq \sigma_{jj}$.

Доказательство следует из соотношения (15.10) и свойства C1. □

C5. Если $N = 2$, $m = 1$, то множественный коэффициент корреляции (15.16) совпадает с точностью до знака с парным коэффициентом корреляции из (15.1): $\bar{R}_{1,2} = |\rho_{12}|$.

Замечание 15.4. В теории множественной регрессии множественный коэффициент корреляции служит мерой оптимальности прогноза: чем он ближе к единице, тем точнее прогноз.

15.4. ВЫБОРОЧНЫЕ ЧАСТНЫЙ И МНОЖЕСТВЕННЫЙ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ. ПРОВЕРКА ГИПОТЕЗ НЕЗАВИСИМОСТИ

Пусть случайный N -вектор признаков $x = (\tilde{x}_j)_{j=1}^N \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $N_N(\mu, \Sigma)$ с неизвестным математическим ожиданием μ и неизвестной ковариационной матрицей Σ ($|\Sigma| \neq 0$) и имеется выборка $X = \{x_1, \dots, x_n\}$, содержащая $n > N$ наблюдений над ним.

Оценим по выборке X частный коэффициент корреляции (15.11) и множественный коэффициент корреляции (15.16). Согласно разбиению вектора $x = \left((x^{(1)})^T : (x^{(2)})^T \right)^T$ на подвекторы $x^{(1)} \in \mathbb{R}^m$ и $x^{(2)} \in \mathbb{R}^{N-m}$ ($m < N$), наряду с разбиением (15.8) N -вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы Σ построим соответствующее разбиение их статистических оценок – арифметического среднего $\bar{x} = (\bar{x}_k)_{k=1}^N$ из (14.3), (14.14) и выборочной ковариационной матрицы $\hat{\Sigma} = A/n$

($A = (a_{kl})_{k,l=1}^N$) из (14.4), (14.14):

$$\bar{x} = \begin{pmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A_{21} = A_{12}^T. \quad (15.17)$$

Обозначим $a_{kl|m+1,\dots,N} = a_{kl} - a_{(k)}^T A_{22}^{-1} a_{(l)}$ как (k, l) -й элемент матрицы $A_{11|2} = (a_{kl|m+1,\dots,N})_{k,l=1}^m = A_{11} - A_{12} A_{22}^{-1} A_{12}^T$; $a_{(k)}^T$ — k -я строка матрицы A_{12} ($k = 1, \dots, m$).

Теорема 15.4. Пусть наблюдения $x_1, \dots, x_n \in \mathbb{R}^N$ из выборки $X = \{x_t\}_{t=1}^n$ объема $n > N$ независимы в совокупности и имеют невырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), тогда оценки максимального правдоподобия для частного коэффициента корреляции (15.11) и множественного коэффициента корреляции (15.16) являются статистиками:

$$r_{kl|m+1,\dots,N} = \frac{a_{kl|m+1,\dots,N}}{\sqrt{a_{kk|m+1,\dots,N} a_{ll|m+1,\dots,N}}}, \quad k, l = 1, \dots, m, \quad - \quad (15.18)$$

выборочный частный коэффициент корреляции,

$$R_{j,m+1,\dots,N} = \sqrt{\frac{a_{(j)}^T A_{22}^{-1} a_{(j)}}{a_{jj}}}, \quad j = 1, \dots, m, \quad - \quad (15.19)$$

выборочный множественный коэффициент корреляции.

Доказательство очевидно и основано на подстановке в (15.11), (15.16) с учетом обозначений (15.17) вместо неизвестной ковариационной матрицы Σ ее МП-оценки $\hat{\Sigma} = A/n$ по выборке X , полученной в теореме 14.3 главы 14, условия которой здесь выполняются. \square

Замечание 15.5. При вычислении выборочных частного (15.18) и множественного (15.19) коэффициентов корреляции вместо матрицы $A = (a_{kl})_{k,l=1}^N$ можно использовать выборочную ковариационную матрицу $\hat{\Sigma} = (\hat{\sigma}_{kl})_{k,l=1}^N$ с элементами $\hat{\sigma}_{kl} = a_{kl}/n$, $k, l = 1, \dots, N$, или несмещенную выборочную ковариационную матрицу $S = (s_{kl})_{k,l=1}^N$: $s_{kl} = a_{kl}/(n-1)$, $k, l = 1, \dots, N$.

Найдем распределения вероятностей случайных матриц $A_{11|2} = A_{11} - A_{12} A_{22}^{-1} A_{12}^T$ и $A_{12} A_{22}^{-1} A_{12}^T$, элементы которых служат для определения выборочных соответственно частного (15.18) и множественного (15.19) коэффициентов корреляции.

Докажем сначала известную лемму [4].

Лемма 15.2 (Андерсона). Предположим, что случайные векторы $Y_1, \dots, Y_p \in \mathbb{R}^q$ независимы в совокупности и имеют нормальные распределения: $\mathcal{L}\{Y_\alpha\} = N_q(\Gamma w_\alpha, \Psi)$, $\alpha = 1, \dots, p$, где Γ — $(q \times r)$ -матрица, а $\{w_\alpha \in \mathbb{R}^r\}_{\alpha=1}^p$ — r -векторы такие, что $(r \times r)$ -матрица $H = \sum_{\alpha=1}^p w_\alpha w_\alpha^T$ невырождена ($|H| \neq 0$).

Пусть $G = \sum_{\alpha=1}^p Y_\alpha w_\alpha^T H^{-1}$, тогда случайная матрица $\sum_{\alpha=1}^p Y_\alpha Y_\alpha^T - G H G^T$ распределена как матрица $\sum_{\alpha=1}^{p-r} U_\alpha U_\alpha^T$, где случайные q -векторы $\{U_\alpha\}_{\alpha=1}^{p-r}$ независимы в совокупности, одинаково распределены по закону $N_q(\mathbf{0}_q, \Psi)$ и не зависят от случайной матрицы $G H G^T$. При $\Gamma = \mathbf{0}_{q \times r}$ матрица $G H G^T$ распределена как $\sum_{\alpha=p-r+1}^p U_\alpha U_\alpha^T$, где все случайные векторы $\{U_\alpha\}_{\alpha=1}^p$ независимы в совокупности и одинаково распределены по закону $N_q(\mathbf{0}_q, \Psi)$.

Доказательство. По построению $(r \times r)$ -матрица $H = \sum_{\alpha=1}^p w_{\alpha} w_{\alpha}^T$ симметрична:

$H^T = H$, и по условию – невырождена: $|H| \neq 0$, поэтому она положительно определена: $H \succ 0$, и из матричной алгебры [34, 11] следует, что существует такая невырожденная квадратная $(r \times r)$ -матрица F ($|F| \neq 0$), что $FHF^T = I_r$.

Обозначим: $W = (w_1 : \dots : w_p)$ – $(r \times p)$ -матрица, столбцы которой – r -векторы $\{w_{\alpha}\}_{\alpha=1}^p$.

Положим $E_2 = FW$, тогда $E_2 E_2^T = FWW^T F^T = F \sum_{\alpha=1}^p w_{\alpha} w_{\alpha}^T F^T = FHF^T = I_r$, строки $(r \times p)$ -матрицы E_2 ортонормированы и их можно дополнить до ортонормированного базиса в \mathbb{R}^p $((p-r) \times p)$ -матрицей E_1 ($E_1 E_1^T = I_{p-r}$):

$$E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix}, \quad EE^T = E^T E = I_p.$$

Воспользуемся построенной выше ортогональной матрицей $E = (e_{\alpha\beta})_{\alpha,\beta=1}^p$ и по аналогии с доказательством теоремы 14.4 осуществим ортогональное преобразование случайных векторов $Y_1, \dots, Y_p \in \mathbb{R}^q$ в систему случайных векторов:

$$U_{\alpha} = \sum_{\beta=1}^p e_{\alpha\beta} Y_{\beta}, \quad \alpha = 1, \dots, p,$$

которые некоррелированы:

$$\begin{aligned} \text{cov}\{U_{\alpha}, U_{\gamma}\} &= E\{(U_{\alpha} - E\{U_{\alpha}\})(U_{\gamma} - E\{U_{\gamma}\})^T\} = \\ &= E\left\{\sum_{i=1}^p e_{\alpha i}(Y_i - E\{Y_i\}) \sum_{j=1}^p e_{\gamma j}(Y_j - E\{Y_j\})^T\right\} = \delta_{ij} \Psi, \end{aligned}$$

и поэтому как линейные преобразования нормальных случайных векторов независимы.

Построим матрицу $U = (U_1 : \dots : U_p)$, тогда $U = Y E^T$, где $Y = (Y_1 : \dots : Y_p)$. Воспользуемся тем, что $W = F^{-1} E_2$, и найдем математическое ожидание для U :

$$E\{U\} = E\{Y E^T\} = \Gamma W E^T = \Gamma F^{-1} E_2 (E_1^T : E_2^T) = (\mathbf{0}_{q \times (p-r)} : \Gamma F^{-1}).$$

Из последних двух соотношений заключаем, что все случайные векторы $\{U_{\alpha}\}_{\alpha=1}^p$ имеют одинаковые ковариационные матрицы Ψ , а первые $p-r$ из них $\{U_{\alpha}\}_{\alpha=1}^{p-r}$ – нулевые математические ожидания.

Воспользовавшись соотношениями $Y = UE$ и $(F^{-1})^T H^{-1} F^{-1} = I_r$, получим

$$\sum_{\alpha=1}^p Y_{\alpha} Y_{\alpha}^T = Y Y^T = U E (U E)^T = U E E^T U^T = U I_p U^T = U U^T = \sum_{\alpha=1}^p U_{\alpha} U_{\alpha}^T$$

и

$$\begin{aligned} GHG^T &= (Y W^T H^{-1}) H (Y W^T H^{-1})^T = Y W^T H^{-1} W Y^T = \\ &= U E (F^{-1} E_2)^T H^{-1} (F^{-1} E_2) (U E)^T = U E E_2^T (F^{-1})^T H^{-1} F^{-1} E_2 E^T U^T = \\ &= U E E_2^T E_2 E^T U^T = U \begin{pmatrix} \mathbf{0}_{(p-r) \times q} \\ I_r \end{pmatrix} (\mathbf{0}_{q \times (p-r)} : I_r) U^T = \sum_{\alpha=p-r+1}^p U_{\alpha} U_{\alpha}^T, \end{aligned}$$

откуда следует независимость от GHG^T матрицы

$$\sum_{\alpha=1}^p Y_{\alpha} Y_{\alpha}^T - GHG^T = \sum_{\alpha=1}^{p-r} U_{\alpha} U_{\alpha}^T,$$

а также тот факт, что при $\Gamma = \mathbf{0}_{q \times r}$ (влечет за собой $\mathbf{E}\{U\} = \mathbf{0}_{q \times p}$) случайная матрица GHG^T распределена как $\sum_{\alpha=p-r+1}^p U_{\alpha} U_{\alpha}^T$, где все $\{U_{\alpha}\}_{\alpha=1}^p$ независимы в совокупности и одинаково распределены по закону $N_q(\mathbf{0}_q, \Psi)$. \square

Теорема 15.5. В условиях теоремы 15.4 справедливы следующие утверждения.

1. Случайная матрица $A_{11|2} = A_{11} - A_{12}A_{22}^{-1}A_{12}^T$ распределена как матрица

$\sum_{\alpha=1}^{n-(N-m)-1} U_{\alpha} U_{\alpha}^T$ и не зависит от матрицы $A_{12}A_{22}^{-1}A_{12}^T$, а при $\Sigma_{12} = \mathbf{0}_{m \times (N-m)}$ матрица $A_{12}A_{22}^{-1}A_{12}^T$ распределена как $\sum_{\alpha=n-(N-m)}^{n-1} U_{\alpha} U_{\alpha}^T$, где случайные m -векторы $\{U_{\alpha}\}_{\alpha=1}^{n-1}$

независимы в совокупности и одинаково распределены по закону $N_m(\mathbf{0}_m, \Sigma_{11|2})$, $\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$.

Другими словами, случайная матрица $A_{11|2}$ имеет распределение Уишарта $W_m(\Sigma_{11|2}, n-(N-m)-1)$ и не зависит от матрицы $A_{12}A_{22}^{-1}A_{12}^T$, которая при $\Sigma_{12} = \mathbf{0}_{m \times (N-m)}$ также имеет распределение Уишарта: $\mathcal{L}\{A_{12}A_{22}^{-1}A_{12}^T | \Sigma_{12} = \mathbf{0}_{m \times (N-m)}\} = W_m(\Sigma_{11|2}, N-m)$.

2. Выборочный частный коэффициент корреляции $r_{kl|m+1, \dots, N}$ ($k, l = 1, \dots, m$), вычисленный по выборке объема n , распределен так же, как соответствующий ему обычный выборочный коэффициент корреляции r_{kl} , подсчитанный по выборке объема $n - (N - m)$ с истинным значением коэффициента корреляции $\rho_{kl} := \rho_{kl|m+1, \dots, N}$, $k, l = 1, \dots, m$.

3. Для выборочного коэффициента корреляции r_{kl} ($k, l = 1, \dots, N$), определенного по выборке объема n , при нулевом истинном значении $\rho_{kl} = 0$, статистика

$$\sqrt{n-2} \frac{|r_{kl}|}{\sqrt{1-r_{kl}^2}} \quad (15.20)$$

распределена как модуль случайной величины, имеющей t -распределение Стьюдента с $n-2$ степенями свободы.

4. Статистика $\frac{n-(N-m)-1}{N-m} \frac{R_{j,m+1, \dots, N}^2}{1-R_{j,m+1, \dots, N}^2}$ при нулевом истинном значении соответствующего множественного коэффициента корреляции ($\bar{R}_{j,m+1, \dots, N} = 0$) имеет F -распределение Фишера с $N-m$ и $n-(N-m)-1$ степенями свободы ($j = 1, \dots, m$):

$$\mathcal{L} \left\{ \frac{n-(N-m)-1}{N-m} \frac{R_{j,m+1, \dots, N}^2}{1-R_{j,m+1, \dots, N}^2} \middle| \bar{R}_{j,m+1, \dots, N} = 0 \right\} = F_{N-m, n-(N-m)-1}. \quad (15.21)$$

Доказательство основано на использовании леммы 15.2 и полученном при доказательстве теоремы 14.4 из главы 14 представлении для матрицы A : $A = \sum_{t=1}^{n-1} z_t z_t^T$, где случайные N -векторы $\{z_t\}_{t=1}^{n-1}$ независимы в совокупности и одинаково распределены по закону $N_N(\mathbf{0}_N, \Sigma)$. Докажем по очереди все утверждения теоремы.

1. В соответствии с (15.17) разобьем вектор z_t на два подвектора:

$$z_t = \begin{pmatrix} z_t^{(1)} \\ z_t^{(2)} \end{pmatrix}, \quad t = 1, \dots, n-1, \quad \text{тогда} \quad A_{ij} = \sum_{t=1}^{n-1} z_t^{(i)} \left(z_t^{(j)} \right)^T, \quad i, j = 1, 2.$$

Из теоремы 15.2 заключаем, что при фиксированных $\left\{ z_t^{(2)} \right\}_{t=1}^{n-1}$ случайные векторы $\left\{ z_t^{(1)} \right\}_{t=1}^{n-1}$ независимы в совокупности и имеют условные нормальные распределения $\left\{ N_m \left(\Sigma_{12} \Sigma_{22}^{-1} z_t^{(2)}, \Sigma_{11|2} \right) \right\}_{t=1}^{n-1}$.

Чтобы применить лемму, в ней полагаем: $p := n-1$, $Y_\alpha := z_\alpha^{(1)}$; $r := N-m$, $w_\alpha := z_\alpha^{(2)}$, $\Gamma := \Sigma_{12} \Sigma_{22}^{-1}$, $\Psi := \Sigma_{11|2}$, тогда $H = A_{22}$, $G = A_{12} A_{22}^{-1}$, и далее непосредственно используем результаты леммы, получая условные распределения, не зависящие от $\left\{ z_t^{(2)} \right\}_{t=1}^{n-1}$ и совпадающие с соответствующими безусловными распределениями.

2. Сравним формулу (15.2) для вычисления выборочного коэффициента корреляции по выборке объема n ($k, l = 1, \dots, N$):

$$r_{kl} = \frac{a_{kl}}{\sqrt{a_{kk} a_{ll}}},$$

с формулой (15.18) для выборочного частного коэффициента корреляции $r_{kl|m+1, \dots, N}$ ($k, l = 1, \dots, m$). Отметим, что обе эти формулы различаются лишь тем, что используют элементы матриц $A = (a_{kl})_{k,l=1}^N$ и $A_{11|2} = (a_{kl|m+1, \dots, N})_{k,l=1}^m$. Но согласно теореме 14.4 и доказанному выше матрицы A_{11} и $A_{11|2}$ имеют соответственно распределения Уишарта $W_m(\Sigma_{11}, n-1)$ и $W_m(\Sigma_{11|2}, n-(N-m)-1)$, что и доказывает утверждение.

3. С учетом формулы (15.2) преобразуем статистику из (15.20):

$$\sqrt{n-2} \frac{|r_{kl}|}{\sqrt{1-r_{kl}^2}} = \frac{|a_{kl}|}{\sqrt{\frac{1}{n-2}(a_{kk}a_{ll}-a_{kl}^2)}} = \frac{|a_{kl}|/\sqrt{a_{ll}}}{\sqrt{\frac{1}{n-2}(a_{kk}-a_{kl}a_{ll}^{-1}a_{kl})}},$$

где $a_{kl} = \sum_{t=1}^{n-1} z_{tk} z_{tl} - (k, l)$ -й элемент матрицы $A = (a_{kl})_{k,l=1}^N = \sum_{t=1}^{n-1} z_t z_t^T$, $z_t = (z_{t1}, \dots, z_{tN})^T$.

Зафиксируем $\{z_{tl}\}_{t=1}^{n-1}$, тогда случайные величины $\{z_{tk}\}_{t=1}^{n-1}$ независимы в совокупности и по теореме 15.2 имеют условные нормальные распределения $\{N_1(\sigma_{kl}\sigma_{ll}^{-1}z_{tl}, \sigma_{kk} - \sigma_{kl}\sigma_{ll}^{-1}\sigma_{kl})\}_{t=1}^{n-1}$. Применим лемму, положив: $p := n-1$, $Y_\alpha := z_{\alpha k}$; $r := 1$, $w_\alpha := z_{\alpha l}$, $\Gamma := \sigma_{kl}\sigma_{ll}^{-1}$, $\Psi := \sigma_{kk} - \sigma_{kl}\sigma_{ll}^{-1}\sigma_{kl}$, тогда $H = a_{ll}$, $G = a_{kl}a_{ll}^{-1}$. По аналогии с доказательством первого утверждения с учетом того, что по условию $\rho_{kl} = 0$ и $\sigma_{kl} = 0$, устанавливаем, что случайные величины $(a_{kl}/\sqrt{a_{ll}})^2$ и $a_{kk} - a_{kl}a_{ll}^{-1}a_{kl}$ при фиксированных $\{z_{tl}\}_{t=1}^{n-1}$ независимы и распределены соответственно как случайные величины $(U_{n-1})^2$ и $\sum_{\alpha=1}^{n-2} U_\alpha^2$,

где $\{U_\alpha\}_{\alpha=1}^{n-1}$ независимы в совокупности и одинаково распределены по закону $N_1(0, \sigma_{kk})$.

Разделим числитель и знаменатель в правой части полученного выше соотношения для статистики из (15.20) на $\sqrt{\sigma_{kk}}$, в результате получим, что статистика из (15.20) распределена как

$$\left| \frac{U_{n-1}^*}{\sqrt{\frac{1}{n-2} \sum_{\alpha=1}^{n-2} (U_\alpha^*)^2}} \right|, \quad U_\alpha^* = \frac{U_\alpha}{\sqrt{\sigma_{kk}}}, \quad \alpha = 1, \dots, n-1,$$

где случайная величина под модулем имеет условное t -распределение Стьюдента с $n - 2$ степенями свободы, которое не зависит от $\{z_{tl}\}_{t=1}^{n-1}$ и совпадает с ее безусловным распределением.

4. Воспользуемся формулой (15.19) для вычисления выборочного множественного коэффициента корреляции и обозначениями из (15.17) и получим для статистики из (15.21) следующее представление ($j = 1, \dots, m$):

$$\begin{aligned} & \frac{n - (N - m) - 1}{N - m} \frac{R_{j,m+1,\dots,N}^2}{1 - R_{j,m+1,\dots,N}^2} = \\ & = \frac{n - (N - m) - 1}{N - m} \frac{a_{(j)}^T A_{22}^{-1} a_{(j)}}{a_{jj} - a_{(j)}^T A_{22}^{-1} a_{(j)}} = \frac{\frac{1}{N - m} a_{(j)}^T A_{22}^{-1} a_{(j)}}{\frac{1}{n - (N - m) - 1} a_{jj|m+1,\dots,N}}. \end{aligned}$$

Но по условию теоремы $\bar{R}_{j,m+1,\dots,N} = 0$, и $\sigma_{(j)} = \mathbf{0}_{N-m}$. Далее по доказанному выше первому утверждению (положив в нем $N := N - m + 1$, $m := 1$) при фиксированных $\{z_t^{(2)}\}_{t=1}^{n-1}$ получаем, что $a_{jj|m+1,\dots,N}$ и $a_{(j)}^T A_{22}^{-1} a_{(j)}$ независимы и распределены соот-

ветственно как $\sum_{\alpha=1}^{n-(N-m)-1} U_{\alpha}^2$ и $\sum_{\alpha=n-(N-m)}^{n-1} U_{\alpha}^2$, где $\{U_{\alpha}\}_{\alpha=1}^{n-1}$ независимые в совокупности и одинаково распределенные случайные величины с законом распределения $N_1(0, \sigma_{jj|m+1,\dots,N})$. Разделив числитель и знаменатель последнего соотношения на $\sigma_{jj|m+1,\dots,N}$, получим при условии фиксированных $\{z_t^{(2)}\}_{t=1}^{n-1}$ распределение из (15.21), которое не зависит от $\{z_t^{(2)}\}_{t=1}^{n-1}$ и поэтому является также и безусловным распределением.

□

Второе утверждение теоремы позволяет использовать для проверки гипотез о значении частного коэффициента корреляции ($k, l = 1, \dots, m$): $\rho_{kl|m+1,\dots,N} = \rho_{kl|m+1,\dots,N}^o$, все критерии из п. 15.1, заменив в них n на $n - (N - m)$. Гипотеза независимости в данном случае: $\rho_{kl|m+1,\dots,N}^o = 0$, называется *гипотезой условной независимости*.

Воспользуемся выборочным множественным коэффициентом корреляции (15.19) для проверки гипотез о некоррелированности j -го признака с группой из $N - m$ выбранных признаков ($j = 1, \dots, m$):

$$\begin{aligned} H_0 & : \bar{R}_{j,m+1,\dots,N} = 0; \\ H_1 = \bar{H}_0 & : \bar{R}_{j,m+1,\dots,N} \neq 0. \end{aligned}$$

Гипотеза H_0 в предположении нормальности также называется *гипотезой независимости*.

Учтем, что статистика в (15.21) монотонно убывает с уменьшением $R_{j,m+1,\dots,N}^2$ (приближением его значения к нулю), и построим следующий критерий для проверки гипотез H_0, H_1 :

$$\begin{cases} H_0 & : \frac{n - (N - m) - 1}{N - m} \frac{R_{j,m+1,\dots,N}^2}{1 - R_{j,m+1,\dots,N}^2} \leq \Delta; \\ H_1 = \bar{H}_0 & : \frac{n - (N - m) - 1}{N - m} \frac{R_{j,m+1,\dots,N}^2}{1 - R_{j,m+1,\dots,N}^2} > \Delta, \end{cases} \quad (15.22)$$

где порог критерия Δ определяется с учетом (15.21) по наперед заданному малому значению уровня значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$:

$$\Delta = F_{N-m, n-(N-m)-1}^{-1}(1 - \alpha) \quad - \quad (15.23)$$

квантиль уровня $1 - \alpha$ от F -распределения с $N - m$ и $n - (N - m) - 1$ степенями свободы.

Замечание 15.6. Критерий (15.22), (15.23) в множественном регрессионном анализе называется проверкой регрессии на значимость. Если гипотеза H_0 отвергается, то регрессия j -го признака ($j = 1, \dots, m$) на выбранные $N - m$ признаков считается значимой и возможно прогнозирование значения \tilde{x}_j этого признака по $x^{(2)} \in \mathbb{R}^{N-m}$ при помощи подстановочной статистической оценки функции регрессии (15.9), (15.14):

$$\hat{\tilde{x}}_j = \bar{x}_j + a_{(j)}^T A_{22}^{-1} (x^{(2)} - \bar{x}^{(2)}). \quad (15.24)$$

В качестве примера статистического анализа зависимостей рассмотрим ставшие уже классическими известные данные Хукера [4] по урожайности сена в центнерах на акр (первый признак, \tilde{x}_1), весенним осадкам в дюймах (второй признак, \tilde{x}_2) и по количеству весенних дней с температурой воздуха выше 42 град. по Фаренгейту (третий признак, \tilde{x}_3 , характеризующий температуру воздуха весной) в Англии за 20 лет ($n = 20$). По этим данным получены следующие значения выборочных среднего и ковариационной матрицы ($N = 3$):

$$\bar{x} = \begin{pmatrix} 28,02 \\ 4,91 \\ 594 \end{pmatrix}, \quad \hat{\Sigma} = \frac{1}{n} A = \begin{pmatrix} 19,54 & 3,89 & -150,28 \\ 3,89 & 1,21 & -52,36 \\ -150,28 & -52,36 & 7225,00 \end{pmatrix}.$$

Для выборочной корреляционной матрицы имеем

$$\begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1,00 & 0,80 & -0,40 \\ 0,80 & 1,00 & -0,56 \\ -0,40 & -0,56 & 1,00 \end{pmatrix}.$$

Из полученных выборочных коэффициентов корреляции следует, что урожайность и количество осадков положительно коррелированы и находятся в прямой зависимости. Однако урожайность и температура воздуха, а также количество осадков и температура воздуха отрицательно коррелированы. Чтобы прояснить причины этого, вычислим выборочный частный коэффициент корреляции между урожайностью и температурой воздуха при фиксированном количестве осадков: $r_{13|2} = 0,097$. Таким образом, если действие осадков не учитывать, то урожайность и температура воздуха положительно коррелированы и находятся в прямой зависимости, хотя значение частной корреляции очень мало. Проверим с уровнем значимости $\alpha = 0,05$ гипотезу об условной некоррелированности (независимости) урожайности и температуры воздуха при фиксированном количестве осадков: $\rho_{13|2} = 0$. Используем критерий, основанный на Z -статистике Фишера:

$$\sqrt{20 - (3 - 2) - 1} Z(0,097) = 0,040 < \Phi^{-1}(1 - 0,05/2) = 1,96,$$

и гипотеза о некоррелированности не отвергается. Следовательно, на урожайность при фиксированном количестве осадков температура воздуха практически не оказывает влияния, тогда как на урожайность влияние количества осадков существенно, а отрицательная безусловная корреляция между урожайностью и температурой воздуха объясняется, скорее всего, тем, что большее количество осадков наблюдается при низкой температуре.

Оценим теперь регрессионную зависимость урожайности от количества осадков и температуры воздуха. Вычислим выборочный множественный коэффициент корреляции между ними: $R_{1,2,3} = 0,802$. Ясно, что корреляция значима, но для строгости проверим гипотезу независимости с уровнем значимости, например, $\alpha = 0,01$.

Из (15.22), (15.23) будем иметь

$$\frac{20 - (3 - 1) - 1}{3 - 1} \frac{R_{1,2,3}^2}{1 - R_{1,2,3}^2} = 15,3 > F_{3-1, 20-(3-1)-1}^{-1}(1 - 0,01) = 6,11,$$

и гипотеза независимости отвергается. Из (15.24) получаем соотношение

$$\hat{x}_1 = 28,02 + 3,37(\tilde{x}_2 - 4,91) + 0,00364(\tilde{x}_3 - 594),$$

позволяющее прогнозировать урожайность \tilde{x}_1 по количеству осадков \tilde{x}_2 и числу весенних дней \tilde{x}_3 с температурой выше 42 град. (по Фаренгейту).

15.5. ПРОВЕРКА ОБЩИХ ГИПОТЕЗ О НЕЗАВИСИМОСТИ

Пусть случайный N -вектор $x \in \mathbb{R}^N$, имеющий невырожденное нормальное распределение $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), разбит на $2 \leq K \leq N$ подвекторов-блоков $\{x^{(k)} \in \mathbb{R}^{N^{(k)}}\}_{k=1}^K$ ($N^{(1)} + \dots + N^{(K)} = N$). Соответствующим образом разобьем также N -вектор математического ожидания $\mu \in \mathbb{R}^N$ и ковариационную $(N \times N)$ -матрицу Σ :

$$x = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(K)} \end{pmatrix}; \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \vdots \\ \mu^{(K)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1K} \\ \vdots & \vdots & & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \dots & \Sigma_{KK} \end{pmatrix}.$$

По случайной выборке $X = \{x_1, \dots, x_n\}$ объема $n > N$, образованной независимыми в совокупности наблюдениями над случайным вектором $x \in \mathbb{R}^N$, необходимо проверить гипотезу о том, что подвекторы (группы признаков) $\{x^{(k)}\}_{k=1}^K$ независимы в совокупности между собой. В силу того что подвекторы $\{x^{(k)}\}_{k=1}^K$ имеют совместное нормальное распределение, их независимость:

$$n_N(x|\mu, \Sigma) = \prod_{k=1}^K n_{N^{(k)}}(x^{(k)}|\mu^{(k)}, \Sigma_{kk})$$

по следствию 14.3 эквивалентна их некоррелированности:

$$\text{cov}\{x^{(i)}, x^{(j)}\} = \Sigma_{ij} = \mathbf{0}_{N^{(i)} \times N^{(j)}}, \quad i \neq j \quad (i, j = 1, \dots, K),$$

и гипотеза независимости сводится к гипотезе

$$H_0 : \Sigma_{ij} = \delta_{ij}\Sigma_{ii}, \quad i, j = 1, \dots, K, \quad (15.25)$$

при альтернативе общего вида: $H_1 = \overline{H_0}$.

Воспользуемся для проверки гипотез H_0, H_1 , являющихся сложными гипотезами, критерием отношения правдоподобия (п. 12.5). Определим обобщенную статистику отношения правдоподобия:

$$\lambda = \frac{\max_{(\mu, \Sigma) \in \Theta_0} L_N(\mu, \Sigma)}{\max_{(\mu, \Sigma) \in \Theta} L_N(\mu, \Sigma)} \in [0, 1], \quad (15.26)$$

где

$$L_N(\mu, \Sigma) = \prod_{t=1}^n n_N(x_t|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}nN} |\Sigma|^{\frac{1}{2}n}} \exp\left(-\frac{1}{2} \sum_{t=1}^n (x_t - \mu)^T \Sigma^{-1} (x_t - \mu)\right) - \quad (15.27)$$

функция правдоподобия, вычисленная по выборке X объема n ;

$$\Theta = \{(\mu, \Sigma) : \mu \in \mathbb{R}^N, \Sigma = \Sigma^T \succ 0\} \quad -$$

множество всех допустимых значений параметров, а

$$\Theta_0 = \{(\mu, \Sigma) : \mu \in \mathbb{R}^N, \Sigma = \text{diag}\{\Sigma_{11}, \dots, \Sigma_{KK}\}, \Sigma_{ii} = \Sigma_{ii}^T \succ 0, i = 1, \dots, K\} \quad -$$

множество значений параметров, соответствующее гипотезе H_0 .

Критерий отношения правдоподобия, основанный на статистике (15.26), имеет вид

$$\begin{cases} H_0 & : \lambda \geq \lambda_o(\alpha); \\ H_1 = \overline{H_0} & : \lambda < \lambda_o(\alpha), \end{cases} \quad (15.28)$$

где порог критерия $\lambda_o = \lambda_o(\alpha)$ определяется по наперед заданному малому значению уровня значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$.

Разобьем вычисленные по выборке X объема $n > N$ выборочное среднее (14.3):

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

выборочную ковариационную матрицу (14.4):

$$\hat{\Sigma} = \frac{1}{n} A, \quad A = \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T = (a_{kl})_{k,l=1}^N,$$

а также выборочную корреляционную матрицу:

$$\mathcal{R} = (r_{kl})_{k,l=1}^N, \quad r_{kl} = \frac{a_{kl}}{\sqrt{a_{kk}a_{ll}}}, \quad k, l = 1, \dots, N,$$

на соответствующие блоки:

$$\bar{x} = \begin{pmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \\ \vdots \\ \bar{x}^{(K)} \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} & \dots & \hat{\Sigma}_{1K} \\ \vdots & \vdots & & \vdots \\ \hat{\Sigma}_{K1} & \hat{\Sigma}_{K2} & \dots & \hat{\Sigma}_{KK} \end{pmatrix}; \quad \Sigma_{ij} = \frac{1}{n} A_{ij}, \quad i, j = 1, \dots, K,$$

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1K} \\ \vdots & \vdots & & \vdots \\ A_{K1} & A_{K2} & \dots & A_{KK} \end{pmatrix}; \quad \mathcal{R} = \begin{pmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} & \dots & \mathcal{R}_{1K} \\ \vdots & \vdots & & \vdots \\ \mathcal{R}_{K1} & \mathcal{R}_{K2} & \dots & \mathcal{R}_{KK} \end{pmatrix}.$$

Теорема 15.6. Пусть выборка $X = \{x_t\}_{t=1}^n$ объема $n > N$ образована независимыми в совокупности, одинаково распределенными нормальными случайными N -векторами с невырожденным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), тогда критерий отношения правдоподобия для проверки гипотезы независимости H_0 из (15.25) против альтернативы $H_1 = \overline{H_0}$ может быть записан в виде

$$\begin{cases} H_0 & : V \geq V_o(\alpha); \\ H_1 = \overline{H_0} & : V < V_o(\alpha), \end{cases} \quad (15.29)$$

где статистика критерия является V -статистикой [4]:

$$V = \frac{|\hat{\Sigma}|}{\prod_{k=1}^K |\hat{\Sigma}_{kk}|} = \frac{|A|}{\prod_{k=1}^K |A_{kk}|} = \frac{|\mathcal{R}|}{\prod_{k=1}^K |\mathcal{R}_{kk}|} \in [0, 1], \quad (15.30)$$

а порог $V_o = V_o(\alpha)$ при $n \rightarrow +\infty$ определяется по уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$ из асимптотического соотношения:

$$V_o(\alpha) = \exp\left(-\frac{F_{\chi_f^2}^{-1}(1-\alpha)}{g}\right), \quad f = \frac{1}{2}\left(N^2 - \sum_{k=1}^K (N^{(k)})^2\right),$$

$$g = g(n) = n - \frac{3}{2} - \frac{1}{3} \frac{N^3 - \sum_{k=1}^K (N^{(k)})^3}{N^2 - \sum_{k=1}^K (N^{(k)})^2}, \quad (15.31)$$

$F_{\chi_f^2}^{-1}(\cdot)$ – квантиль χ^2 -распределения с f степенями свободы.

Доказательство. Воспользуемся теоремой 14.3, согласно доказательству которой выборочные среднее \bar{x} и ковариационная матрица $\hat{\Sigma}$ являются оценками максимального правдоподобия и максимизируют функцию правдоподобия $L_N(\mu, \Sigma)$ (логарифмическую функцию правдоподобия $l(\mu, \Sigma) = \ln L_N(\mu, \Sigma)$ – в доказательстве теоремы 14.3) по $(\mu, \Sigma) \in \Theta$:

$$\max_{(\mu, \Sigma) \in \Theta} L_N(\mu, \Sigma) = L_N(\bar{x}, \hat{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}nN} |\hat{\Sigma}|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN}.$$

При истинной гипотезе H_0 получим

$$\begin{aligned} \max_{(\mu, \Sigma) \in \Theta_0} L_N(\mu, \Sigma) &= \prod_{k=1}^K \max_{\mu^{(k)}, \Sigma_{kk}} L_{N^{(k)}}(\mu^{(k)}, \Sigma_{kk}) = \\ &= \prod_{k=1}^K L_{N^{(k)}}(\bar{x}^{(k)}, \hat{\Sigma}_{kk}) = \prod_{k=1}^K \frac{1}{(2\pi)^{\frac{1}{2}nN^{(k)}} |\hat{\Sigma}_{kk}|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN^{(k)}} = \\ &= \frac{1}{(2\pi)^{\frac{1}{2}nN} \left(\prod_{k=1}^K |\hat{\Sigma}_{kk}|\right)^{\frac{1}{2}n}} e^{-\frac{1}{2}nN}, \end{aligned}$$

где учтено, что $\bar{x}^{(k)}$ и $\hat{\Sigma}_{kk}$ являются оценками максимального правдоподобия для математического ожидания $\mu^{(k)}$ и ковариационной матрицы Σ_{kk} маргинального нормального распределения $N_{N^{(k)}}(\mu^{(k)}, \Sigma_{kk})$ и максимизируют соответствующую функцию правдоподобия $L_{N^{(k)}}(\mu^{(k)}, \Sigma_{kk})$ ($k = 1, \dots, K$).

Из последних двух соотношений для статистики отношения правдоподобия имеем

$$\lambda = \frac{\max_{(\mu, \Sigma) \in \Theta_0} L_N(\mu, \Sigma)}{\max_{(\mu, \Sigma) \in \Theta} L_N(\mu, \Sigma)} = \left(\frac{|\hat{\Sigma}|}{\prod_{k=1}^K |\hat{\Sigma}_{kk}|} \right)^{\frac{n}{2}} = V^{\frac{n}{2}} -$$

монотонно возрастающая функция от V -статистики $V = \frac{|\hat{\Sigma}|}{\prod_{k=1}^K |\hat{\Sigma}_{kk}|}$, что приводит к эк-

вивалентной записи критерия отношения правдоподобия (15.28), (15.26) в виде (15.29), (15.30).

Покажем, что для V также справедливы представления $V = \frac{|A|}{\prod_{k=1}^K |A_{kk}|} = \frac{|\mathcal{R}|}{\prod_{k=1}^K |\mathcal{R}_{kk}|}$.

Первое из них очевидно и следует из того, что $\hat{\Sigma} = A/n$. Докажем второе представление. Воспользуемся соотношением, связывающим корреляционную матрицу \mathcal{R} с матрицей $A = (a_{kl})_{k,l=1}^N$:

$$\mathcal{R} = CAC, \quad C = \text{diag} \left\{ \frac{1}{\sqrt{a_{11}}}, \dots, \frac{1}{\sqrt{a_{NN}}} \right\},$$

и разобьем диагональную матрицу C на соответствующие блоки:

$$C = \text{diag}\{C_{11}, \dots, C_{KK}\}, \quad |C| = \prod_{k=1}^K |C_{kk}|,$$

тогда

$$\frac{|\mathcal{R}|}{\prod_{k=1}^K |\mathcal{R}_{kk}|} = \frac{|CAC|}{\prod_{k=1}^K |C_{kk} A_{kk} C_{kk}|} = \frac{|C|^2 |A|}{\prod_{k=1}^K |C_{kk}|^2 |A_{kk}|} = \frac{|A|}{\prod_{k=1}^K |A_{kk}|} = V.$$

Осталось определить порог критерия $V_o = V_o(\alpha)$. В [4] установлено, что при $n \rightarrow +\infty$ в условиях гипотезы H_0 для распределения вероятностей V -статистики справедливо асимптотическое разложение ($g = g(n) \rightarrow +\infty$):

$$P_{H_0}\{-g \ln(V) \leq z\} = F_{\chi^2_g}(z) + O(g^{-2}), \quad z \in \mathbb{R},$$

из которого по заданному уровню значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$ и определяется порог (15.31). \square

Следствие 15.5. В условиях теоремы 15.6 критерий отношения правдоподобия для проверки гипотезы о независимости всех N компонент между собой ($K = N$; $N^{(k)} = 1$, $k = 1, \dots, N$) против общей альтернативы имеет вид

$$\begin{cases} H_0 & : |\mathcal{R}| \geq V_o(\alpha); \\ H_1 = \overline{H_0} & : |\mathcal{R}| < V_o(\alpha), \end{cases}$$

где $|\mathcal{R}|$ – определитель выборочной корреляционной матрицы, а для порога критерия $V_o = V_o(\alpha)$, $\alpha = P\{H_1|H_0\} \in (0, 1)$, при $n \rightarrow +\infty$ справедливо соотношение

$$V_o(\alpha) = \exp \left(- \left(n - \frac{3}{2} - \frac{N+1}{3} \right)^{-1} F_{\chi^2_{\frac{N(N-1)}{2}}}^{-1} (1 - \alpha) \right).$$

В качестве иллюстрации вернемся к примеру Хукера из предыдущего параграфа. По выборке объема $n = 20$ с уровнем значимости $\alpha = 0,01$ проверим гипотезу о независимости урожайности $x^{(1)} \in \mathbb{R}^1$ от количества осадков и температуры воздуха $x^{(2)} \in \mathbb{R}^2$ ($N = 3$, $K = 2$, $N^{(1)} = 1$, $N^{(2)} = 2$). В данном случае, конечно, лучше использовать для проверки гипотезы независимости критерий, основанный на выборочном множественном коэффициенте корреляции, поскольку он не требует асимптотики $n \rightarrow +\infty$. Однако в случае, когда гипотеза независимости проверяется между более чем двумя группами признаков либо когда все группы содержат более одного признака, применение этого критерия невозможно и приходится использовать асимптотический критерий,

описанный выше. Для него получаем

$$V = \frac{|\mathcal{R}|}{|r_{11}||\mathcal{R}_{22}|} = \frac{\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1,00 & 0,80 & -0,40 \\ 0,80 & 1,00 & -0,56 \\ -0,40 & -0,56 & 1,00 \end{vmatrix}}{\begin{vmatrix} 1 & -0,56 \\ -0,56 & 1 \end{vmatrix}} = 0,357,$$

что не превосходит значения порога

$$V_o(0,01) = \exp\left(-\frac{F_{x_f^2}^{-1}(1-0,01)}{g}\right) = \exp\left(-\frac{F_{x_2^2}^{-1}(0,99)}{17}\right) = 0,582,$$

где

$$f = \frac{1}{2}(3^2 - (1^2 + 2^2)) = 2, \quad g = 20 - \frac{3}{2} - \frac{1}{3} \frac{3^3 - (1^3 + 2^3)}{3^2 - (1^2 + 2^2)} = 17,$$

и гипотеза независимости, как и критерием на основе выборочного множественного коэффициента корреляции, отвергается.

15.6. УПРАЖНЕНИЯ

1. Предложить альтернативный вариант доказательства п. 3 теоремы 15.5, воспользовавшись геометрическим подходом для определения распределения вероятностей выборочного коэффициента корреляции из [4] (§ 4.2 в [4]).
2. Доказать следствие 15.5, воспользовавшись результатом теоремы 15.6.
3. Пусть случайные величины x_1, x_2, \dots, x_N имеют одно и то же математическое ожидание μ , общую дисперсию σ^2 и коэффициент корреляции любой пары величин равен ρ . Показать, что $-1/(N-1) \leq \rho \leq 1$ (для этого найти $D\{\bar{x}\}$, $\bar{x} = \sum_{i=1}^N x_i/N$).
4. Доказать, что $\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 \leq 1 + 2\rho_{12}\rho_{13}\rho_{23}$.
5. В условиях упражнения 6 к гл. 14 найти условное распределение \tilde{x}_1 и \tilde{x}_3 при \tilde{x}_2 .
6. Пусть $\mathcal{L}\{x\} = N_4(\mu, \Sigma)$,

$$\mu = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 15 & 3 & 1 & 0 \\ 3 & 16 & 6 & -2 \\ 1 & 6 & 4 & 1 \\ 0 & -2 & 1 & 3 \end{pmatrix};$$

определены подвекторы: $x_{(1)} = (\tilde{x}_1, \tilde{x}_2)^T$, $x_{(2)} = (\tilde{x}_3, \tilde{x}_4)^T$. Найти функцию регрессии $x_{(1)}$ на $x_{(2)}$ и матрицу частных ковариаций $\Sigma_{11|2}$.

7. Определить $\bar{R}_{1,2,3}$ в упражнении 6 из гл. 14.
8. Доказать, что $\bar{R}_{1,2}^2 \leq \bar{R}_{1,2,3}^2 \leq \dots \leq \bar{R}_{1,2,\dots,N}^2$.

Глава 16

ПРОВЕРКА ГИПОТЕЗ И СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ

16.1. ПРОВЕРКА ГИПОТЕЗ О ЗНАЧЕНИИ ВЕКТОРА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ. T^2 -СТАТИСТИКА ХОТЕЛЛИНГА

Пусть в пространстве \mathbb{R}^N наблюдается случайная выборка $X = \{x_1, \dots, x_n\}$ объема n из невырожденного N -мерного нормального распределения $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$). Истинные значения N -вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы Σ неизвестны. Необходимо проверить гипотезы о значении вектора математического ожидания:

$$\begin{aligned} H_0 &: \mu = \mu^o; \\ H_1 = \overline{H_0} &: \mu \neq \mu^o, \end{aligned} \quad (16.1)$$

где $\mu^o \in \mathbb{R}^N$ – предполагаемое (гипотетическое) значение математического ожидания.

Определение 16.1. Пусть

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t, \quad S = \frac{1}{n-1} A, \quad A = \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T, \quad -$$

выборочное среднее и несмещенная выборочная ковариационная матрица, построенные по выборке $X = \{x_1, \dots, x_n\}$ объема $n > N$, образованной N -мерными наблюдениями ($x_t \in \mathbb{R}^N$, $t = 1, \dots, n$).

T^2 -статистикой Хотеллинга называется следующая статистика:

$$T^2 = T^2(\mu) = n(\bar{x} - \mu)^T S^{-1}(\bar{x} - \mu) \geq 0, \quad (16.2)$$

где $\mu \in \mathbb{R}^N$ – фиксированный N -вектор.

Очевидно, что $T^2(\mu) = 0$ тогда и только тогда, когда $\bar{x} = \mu$. Более того, нетрудно заметить, что при $\mu = \mu^o$ с точностью до множителя n статистика $T^2(\mu^o) = T^2(\mu)|_{\mu=\mu^o}$ является оценкой квадрата расстояния Махаланобиса $\rho^2(\mu, \mu^o) = (\mu - \mu^o)^T \Sigma^{-1}(\mu - \mu^o)$ между истинным значением вектора математического ожидания μ и его предполагаемым значением μ^o . Чем $T^2(\mu^o)$ меньше, тем «ближе» истинное значение математического ожидания μ (выборочное среднее \bar{x} является оценкой μ) к гипотетическому значению μ^o . Эти соображения позволяют предложить следующий критерий для проверки гипотез (16.1):

$$\begin{cases} H_0 &: T^2(\mu^o) \leq T_o^2(\alpha); \\ H_1 = \overline{H_0} &: T^2(\mu^o) > T_o^2(\alpha), \end{cases} \quad (16.3)$$

где $T_o^2(\alpha)$ – пороговое значение критерия, подлежащее определению по наперед заданному уровню значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$.

Найдем распределение вероятностей T^2 -статистики (16.2) в случае, когда верна гипотеза H_0 (при $\mu = \mu^o$).

Лемма 16.1. Пусть N -вектор $y \in \mathbb{R}^N$ имеет невырожденное нормальное распределение $N_N(\mathbf{0}_N, \Sigma)$ ($|\Sigma| \neq 0$), а $(N \times N)$ -матрица W – распределение Уишарта $W_N(\Sigma, m)$ ($m \geq N$), тогда случайная величина $\tilde{T}^2 = y^T W^{-1} y$ с учетом нормировки имеет F -распределение Фишера с N и $m - N + 1$ степенями свободы:

$$\mathcal{L} \left\{ \frac{m - N + 1}{N} \tilde{T}^2 \right\} = F_{N, m - N + 1}.$$

Доказательство. Согласно условию леммы и определению распределения Уишарта матрица W распределена как $\sum_{t=1}^m z_t z_t^T$, где случайные N -векторы $\{z_t\}_{t=1}^m$ независимы в совокупности и одинаково распределены по закону $N_N(\mathbf{0}_N, \Sigma)$.

В силу невырожденности для ковариационной матрицы Σ ($\Sigma \succ 0$) справедливы представления, которые уже использовались ранее в п. 14.1:

$$\Sigma = \Sigma^{1/2} (\Sigma^{1/2})^T, \quad (\Sigma^{1/2})^T \Sigma^{-1} \Sigma^{1/2} = I_N;$$

$$\Sigma^{-1} = (\Sigma^{-1/2})^T \Sigma^{-1/2}, \quad \Sigma^{-1/2} = (\Sigma^{1/2})^{-1},$$

где $\Sigma^{1/2}$ – невырожденная $(N \times N)$ -матрица.

Выполним преобразования:

$$y^* = \Sigma^{-1/2} y, \quad W^* = \Sigma^{-1/2} W (\Sigma^{-1/2})^T.$$

Очевидно, что W^* распределена как $\sum_{t=1}^m z_t^* (z_t^*)^T$, где случайные N -векторы $\{z_t^* = \Sigma^{-1/2} z_t\}_{t=1}^m$ независимы в совокупности и одинаково распределены по закону $N_N(\mathbf{0}_N, I_N)$, и имеет место соотношение $(y^*)^T (W^*)^{-1} y^* = y^T W^{-1} y = \tilde{T}^2$.

Выполним теперь ортогональное преобразование:

$$\tilde{y}^* = Q y^*, \quad \tilde{W}^* = Q W^* Q^T,$$

где в случайной ортогональной $(N \times N)$ -матрице $Q = (q_{ij})_{i,j=1}^N$: $Q Q^T = Q^T Q = I_N$ первая строка выбрана следующим образом:

$$q_{1j} = \frac{y_j^*}{\sqrt{(y^*)^T y^*}}, \quad j = 1, \dots, N; \quad y^* = (y_j^*)_{j=1}^N.$$

После этого преобразования также выполняется $(\tilde{y}^*)^T (\tilde{W}^*)^{-1} \tilde{y}^* = (y^*)^T (W^*)^{-1} y^* = \tilde{T}^2$, где матрица \tilde{W}^* распределена как $\sum_{t=1}^m \tilde{z}_t^* (\tilde{z}_t^*)^T$, а случайные N -векторы $\{\tilde{z}_t^* = Q z_t^*\}_{t=1}^m$ по-прежнему независимы в совокупности и одинаково распределены по закону $N_N(\mathbf{0}_N, I_N)$.

Для компонент вектора $\tilde{y}^* = (\tilde{y}_i^*)_{i=1}^N$ имеем

$$\tilde{y}_1^* = \sum_{j=1}^N q_{1j} y_j^* = \sqrt{(y^*)^T y^*};$$

$$\tilde{y}_i^* = \sum_{j=1}^N q_{ij} y_j^* = \sqrt{(y^*)^T y^*} \sum_{j=1}^N q_{ij} q_{1j} = 0, \quad i = 2, \dots, N,$$

и для \tilde{T}^2 получаем $\tilde{T}^2 = (\tilde{y}^*)^T (\tilde{W}^*)^{-1} \tilde{y}^* = (\tilde{y}_1^*)^2 \tilde{w}^{*11}$, где \tilde{w}^{*11} – первый элемент матрицы $(\tilde{W}^*)^{-1} = (\tilde{w}^{*ij})_{i,j=1}^N$, обратной по отношению к матрице $\tilde{W}^* = (\tilde{w}_{ij}^*)_{i,j=1}^N$.

Представим \tilde{W}^* в блочном виде:

$$\tilde{W}^* = \begin{pmatrix} \tilde{w}_{11}^* & (\tilde{w}_{(1)}^*)^T \\ \tilde{w}_{(1)}^* & \tilde{W}_{22}^* \end{pmatrix},$$

воспользуемся известным соотношением для обращения невырожденной блочной матрицы [34, 11]:

$$\begin{pmatrix} E & F \\ G & H \end{pmatrix}^{-1} = \begin{pmatrix} (E - FH^{-1}G)^{-1} & E^{-1}F(GE^{-1}F - H)^{-1} \\ (GE^{-1}F - H)^{-1}GE^{-1} & (H - GE^{-1}F)^{-1} \end{pmatrix},$$

и получим

$$\tilde{w}^{*11} = \left(\tilde{w}_{11}^* - (\tilde{w}_{(1)}^*)^T (\tilde{W}_{22}^*)^{-1} \tilde{w}_{(1)}^* \right)^{-1} = (\tilde{w}_{11|2}^*)^{-1}.$$

Для \tilde{T}^2 окончательно имеем

$$\tilde{T}^2 = \frac{(y^*)^T y^*}{\tilde{w}_{11|2}^*}.$$

При фиксированной матрице Q согласно первому утверждению теоремы 15.5 случайная величина $\tilde{w}_{11|2}^* = \tilde{w}_{11}^* - (\tilde{w}_{(1)}^*)^T (\tilde{W}_{22}^*)^{-1} \tilde{w}_{(1)}^*$ имеет условное распределение $\mathcal{L}\{\tilde{w}_{11|2}^*|Q\} = W_1(1, m - (N - 1)) = \chi_{m-N+1}^2$. Случайная величина в числителе – распределение $\mathcal{L}\{(y^*)^T y^*|Q\} = \chi_N^2$. Полученные условные распределения не зависят от Q и поэтому являются также и безусловными распределениями для случайных величин $\tilde{w}_{11|2}^*$ и $(y^*)^T y^*$. Кроме того, эти случайные величины независимы. \square

Следствие 16.1. Пусть выборка $X = \{x_i\}_{i=1}^n$ объема $n > N$ образована независимыми в совокупности, одинаково распределенными случайными N -векторами с невырожденным нормальным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), тогда T^2 -статистика $T^2 = T^2(\mu)$ из (16.2) с учетом нормировки имеет F -распределение с N и $n - N$ степенями свободы:

$$\mathcal{L}\left\{\frac{n - N}{N(n - 1)} T^2(\mu)\right\} = F_{N, n-N}. \quad (16.4)$$

Доказательство. Согласно теореме 14.4 выборочное среднее \bar{x} и выборочная ковариационная матрица S независимы и имеют распределения $\mathcal{L}\{\bar{x}\} = N_N(\mu, \Sigma/n)$ и $\mathcal{L}\{n\hat{\Sigma}\} = W_N(\Sigma, n - 1)$. Поэтому будут независимыми статистики $y := \sqrt{n}(\bar{x} - \mu)$ и $W := (n - 1)S$, имеющие соответственно распределения $N_N(\mathbf{0}_N, \Sigma)$ и $W_N(\Sigma, m)$, $m := n - 1$. Далее применяем лемму 16.1 к $\tilde{T}^2 := T^2(\mu)/(n - 1)$. \square

Соотношение (16.4) позволяет найти порог критерия (16.3), вычислив вероятность ошибки первого рода:

$$\begin{aligned} P\{H_1|H_0\} &= \mathbf{P}\{T^2(\mu^o) > T_o^2|\mu = \mu^o\} = 1 - \mathbf{P}\{T^2(\mu^o) \leq T_o^2|\mu = \mu^o\} = \\ &= 1 - F_{N, n-N}\left(\frac{n - N}{N(n - 1)} T_o^2\right) := \alpha, \end{aligned}$$

откуда

$$T_o^2 = T_o^2(\alpha) = \frac{N(n - 1)}{n - N} F_{N, n-N}^{-1}(1 - \alpha), \quad (16.5)$$

где $F_{N, n-N}^{-1}(1 - \alpha)$ – квантиль уровня $1 - \alpha$ от F -распределения Фишера с N и $n - N$ степенями свободы.

Замечание 16.1. Порог $T_o^2 = T_o^2(\alpha)$ позволяет также построить доверительную область (см. п. 11.11) уровня $1 - \alpha$ для вектора математического ожидания:

$$\mathcal{M}_\alpha = \left\{ \mu \in \mathbb{R}^N : (\bar{x} - \mu)^T S^{-1} (\bar{x} - \mu) \leq \frac{T_o^2(\alpha)}{n} \right\},$$

являющуюся случайным N -мерным эллипсоидом в \mathbb{R}^N с центром в точке \bar{x} , форма и «объем» (мера Лебега) которого определяются матрицей S^{-1} и вероятностью $1 - \alpha$, с которой в него «попадает» истинное значение вектора математического ожидания μ :

$$P\{\mu \in \mathcal{M}_\alpha\} = 1 - \alpha.$$

Исследуем вероятностные свойства критерия (16.3), (16.5), называемого еще T^2 -критерием Стьюдента.

Теорема 16.1. T^2 -критерий (16.3), (16.5), построенный по выборке $X = \{x_t\}_{t=1}^n$ объема $n > N$, образованной независимыми в совокупности, одинаково распределенными N -векторами-наблюдениями с невырожденным нормальным распределением $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$), является критерием отношения правдоподобия для проверки гипотез (16.1) о значении вектора математического ожидания $H_0 : \mu = \mu^o$, $H_1 = \bar{H}_0$, при неизвестной ковариационной матрице Σ , и при любом фиксированном уровне значимости $\alpha = P\{H_1 | H_0\} \in (0, 1)$ является равномерно наиболее мощным критерием среди всех критериев для проверки гипотез (16.1), мощность которых $w = P\{H_1 | H_1\}$ зависит лишь от величины $n(\mu - \mu^o)^T \Sigma^{-1} (\mu - \mu^o)$.

Доказательство. Аналогично доказательству теоремы 15.6. Построим критерий отношения правдоподобия для проверки гипотез (16.1):

$$\begin{cases} H_0 & : \lambda \geq \lambda_o(\alpha); \\ H_1 = \bar{H}_0 & : \lambda < \lambda_o(\alpha), \end{cases}$$

где статистика отношения правдоподобия

$$\lambda = \frac{\max_{(\mu, \Sigma) \in \Theta_0} L_N(\mu, \Sigma)}{\max_{(\mu, \Sigma) \in \Theta} L_N(\mu, \Sigma)}$$

определяется функцией правдоподобия $L_N(\mu, \Sigma)$ из (15.27), максимальное значение которой на множестве всех допустимых значений параметров $\Theta = \{(\mu, \Sigma) : \mu \in \mathbb{R}^N, \Sigma = \Sigma^T \succ 0\}$ уже было вычислено при доказательстве теоремы 15.6 из главы 15:

$$\max_{(\mu, \Sigma) \in \Theta} L_N(\mu, \Sigma) = L_N(\bar{x}, \hat{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}nN} |\hat{\Sigma}|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN}, \quad \hat{\Sigma} = \frac{1}{n} A.$$

На множестве значений параметров Θ_0 , которому соответствует гипотеза H_0 из (16.1): $\Theta_0 = \{(\mu^o, \Sigma) : \Sigma = \Sigma^T \succ 0\}$, воспользовавшись ходом доказательства теоремы 14.3 из главы 14, имеем

$$\max_{(\mu, \Sigma) \in \Theta_0} L_N(\mu, \Sigma) = \max_{\Sigma = \Sigma^T \succ 0} L_N(\mu^o, \Sigma) = L_N(\mu^o, \hat{\Sigma}_0) = \frac{1}{(2\pi)^{\frac{1}{2}nN} |\hat{\Sigma}_0|^{\frac{1}{2}n}} e^{-\frac{1}{2}nN},$$

где

$$\hat{\Sigma}_0 = \frac{1}{n} \sum_{t=1}^n (x_t - \mu^o)(x_t - \mu^o)^T = \frac{1}{n} A + (\bar{x} - \mu^o)(\bar{x} - \mu^o)^T.$$

С учетом полученного соотношения преобразуем статистику отношения правдоподобия:

$$\lambda = \frac{L(\mu^o, \hat{\Sigma}_0)}{L(\bar{x}, \hat{\Sigma})} = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{\frac{n}{2}} = \left(\frac{|A|}{|A + n(\bar{x} - \mu^o)(\bar{x} - \mu^o)^T|} \right)^{\frac{n}{2}}.$$

Воспользуемся известными формулами для вычисления определителя блочной матрицы [34, 11]:

$$\begin{vmatrix} E & F \\ G & H \end{vmatrix} = |E| |H - GE^{-1}F|, \quad |E| \neq 0;$$

$$\begin{vmatrix} E & F \\ G & H \end{vmatrix} = |H| |E - FH^{-1}G|, \quad |H| \neq 0.$$

Положим $E := 1$, $F := \sqrt{n}(\bar{x} - \mu^o)^T$, $G := -\sqrt{n}(\bar{x} - \mu^o)$, $H = A$ и получим соотношение

$$\begin{vmatrix} 1 & \sqrt{n}(\bar{x} - \mu^o)^T \\ -\sqrt{n}(\bar{x} - \mu^o) & A \end{vmatrix} =$$

$$|1| |A - (-\sqrt{n}(\bar{x} - \mu^o))1^{-1}\sqrt{n}(\bar{x} - \mu^o)^T| =$$

$$= |A + n(\bar{x} - \mu^o)(\bar{x} - \mu^o)^T| = |A| |1 + n(\bar{x} - \mu^o)^T A^{-1}(\bar{x} - \mu^o)|,$$

с учетом которого статистика отношения правдоподобия

$$\lambda = \left(\frac{1}{1 + \frac{T^2(\mu^o)}{n-1}} \right)^{\frac{n}{2}} -$$

строго убывающая функция от T^2 -статистики $T^2(\mu^o)$ из (16.2), поэтому критерий отношения правдоподобия на ее основе эквивалентен T^2 -критерию Стьюдента (16.3), (16.5).

Доказательство того, что T^2 -критерий (16.3), (16.5) является равномерно наиболее мощным [4], носит технический характер и основано на анализе распределения вероятностей T^2 -статистики в случае, когда гипотеза H_0 неверна. \square

На практике иногда по выборке $X = \{x_1, \dots, x_n\}$ объема n из невырожденного нормального распределения $N_N(\mu, \Sigma)$ ($|\Sigma| \neq 0$) необходимо проверить еще одну гипотезу о значении вектора математического ожидания – *гипотезу симметрии*, которая состоит в предположении о том, что значения компонент вектора математического ожидания $\mu = (\mu_j)_{j=1}^N$ равны между собой при альтернативе общего вида:

$$\begin{aligned} H_0 &: \mu_1 = \dots = \mu_N; \\ H_1 = \overline{H_0} &: \exists i \neq j \in \{1, \dots, N\}, \mu_i \neq \mu_j. \end{aligned} \quad (16.6)$$

Пример 16.1. Пусть одна и та же операция на производстве выполняется N различными работниками, относительно которых нельзя сказать, что они работают независимо друг от друга. Фиксируется значение некоторого показателя качества выполнения операции каждым работником. Необходимо проверить, что все они выполняют свою работу одинаково эффективно.

Очевидно, что гипотеза симметрии H_0 из (16.6) может быть записана в виде $\mu = \tilde{\mu} \mathbf{1}_N$, где $\tilde{\mu} \in \mathbb{R}$ – произвольное общее значение компонент, а $\mathbf{1}_N$ – N -вектор, составленный из единиц.

Сведем гипотезы из (16.6) к гипотезам о значении вектора математического ожидания (16.1). Для этого введем в рассмотрение произвольную $((N-1) \times N)$ -матрицу C полного ранга: $\text{rank}(C) = N-1$, строки которой линейно независимы между собой и ортогональны к единичному N -вектору $\mathbf{1}_N$: $C\mathbf{1}_N = \mathbf{0}_{N-1}$. Осуществим линейное преобразование выборки $X = \{x_1, \dots, x_n\}$ в выборку $Y = \{y_1, \dots, y_n\}$: $y_t = Cx_t$, $t = 1, \dots, n$. Наблюдения из выборки Y имеют невырожденное нормальное распределение: $\mathcal{L}\{y_t\} = N_{N-1}(C\mu, C\Sigma C^T)$ ($|C\Sigma C^T| \neq 0$), и при истинной гипотезе H_0 , по построению, их математическое ожидание – нулевой $(N-1)$ -вектор: $C\mu = \tilde{\mu}C\mathbf{1}_N = \tilde{\mu}\mathbf{0}_{N-1} = \mathbf{0}_{N-1}$. Поэтому проверка гипотез (16.6) эквивалентна проверке гипотезы о том, что математическое ожидание наблюдений из выборки $Y = \{y_1, \dots, y_n\}$ – нулевое, против альтернативы общего вида. Остается лишь воспользоваться T^2 -критерием (16.3), (16.2), (16.5), положив в нем $N := N-1$; $x_t := Cx_t$, $t = 1, \dots, n$; $\mu^o := \mathbf{0}_{N-1}$.

Получим следующий критерий для проверки гипотез (16.6):

$$\begin{cases} H_0 & : n\bar{x}^T C^T (CSC^T)^{-1} C\bar{x} \leq \frac{(N-1)(n-1)}{n-(N-1)} F_{N-1, n-(N-1)}^{-1} (1-\alpha); \\ H_1 = \overline{H_0} & : n\bar{x}^T C^T (CSC^T)^{-1} C\bar{x} > \frac{(N-1)(n-1)}{n-(N-1)} F_{N-1, n-(N-1)}^{-1} (1-\alpha), \end{cases}$$

где $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$ – наперед заданный уровень значимости. Мощность полученного критерия зависит от матрицы линейного преобразования C : $w = w(C) = \mathbf{P}\{H_1|H_1\}$, и его можно оптимизировать по мощности, выбирая соответствующим образом матрицу C .

16.2. СРАВНЕНИЕ ВЕКТОРОВ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ ПО ДВУМ ВЫБОРКАМ. МНОГОМЕРНАЯ ПРОБЛЕМА БЕРЕНСА – ФИШЕРА

16.2.1. Сравнение векторов математических ожиданий при неизвестной одинаковой ковариационной матрице

Пусть в пространстве наблюдений \mathbb{R}^N имеются две случайные выборки $X^{(1)} = \{x_1^{(1)}, \dots, x_{n_1}^{(1)}\}$ и $X^{(2)} = \{x_1^{(2)}, \dots, x_{n_2}^{(2)}\}$ соответственно объемов $n_1 > N$ и $n_2 > N$. Наблюдения $\{x_t^{(i)}\}_{t=1}^{n_i}$ из выборки $X^{(i)}$ ($i = 1, 2$) независимы в совокупности и одинаково распределены по нормальному закону $N_N(\mu_i, \Sigma)$ с математическим ожиданием $\mu_i \in \mathbb{R}^N$ и невырожденной ковариационной $(N \times N)$ -матрицей Σ ($|\Sigma| \neq 0$). Выборка $X^{(1)}$ не зависит от выборки $X^{(2)}$. Необходимо проверить гипотезу о совпадении векторов математических ожиданий μ_1 и μ_2 против альтернативы общего вида:

$$\begin{aligned} H_0 & : \mu_1 = \mu_2; \\ H_1 = \overline{H_0} & : \mu_1 \neq \mu_2, \end{aligned} \quad (16.7)$$

при неизвестном значении одинаковой для обеих выборок ковариационной матрицы Σ .

Для проверки гипотез (16.7) воспользуемся предложенной в п. 16.1 T^2 -статистикой, а точнее – леммой 16.1. По выборкам $X^{(1)}$ и $X^{(2)}$ определим статистику:

$$y = \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} (\bar{x}_{(1)} - \bar{x}_{(2)}), \quad (16.8)$$

а также построим оценку для общей ковариационной матрицы Σ , которая учитывает все $n_1 + n_2$ наблюдений из выборок $X^{(1)}$ и $X^{(2)}$:

$$S = \frac{1}{n_1 + n_2 - 2} ((n_1 - 1)S_{(1)} + (n_2 - 1)S_{(2)}), \quad (16.9)$$

где

$$\bar{x}_{(i)} = \frac{1}{n_i} \sum_{t=1}^{n_i} x_t^{(i)}, \quad S_{(i)} = \frac{1}{n_i - 1} \sum_{t=1}^{n_i} (x_t^{(i)} - \bar{x}_{(i)}) (x_t^{(i)} - \bar{x}_{(i)})^T -$$

выборочные среднее и ковариационная матрица для выборки $X^{(i)}$ ($i = 1, 2$).

Согласно теореме 14.4, $\bar{x}_{(i)}$ и $S_{(i)}$ независимы и имеют распределения:

$$\mathcal{L}\{\bar{x}_{(i)}\} = N_N\left(\mu_i, \frac{1}{n_i}\Sigma\right);$$

$$\mathcal{L}\{(n_i - 1)S_{(i)}\} = W_N(\Sigma, n_i - 1), \quad i = 1, 2,$$

поэтому в силу независимости выборок $X^{(1)}$ и $X^{(2)}$ статистики y и S также независимы, а по свойствам многомерного нормального распределения и распределения Уишарта

$$\mathcal{L}\{y\} = N_N\left(\frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}(\mu_1 - \mu_2), \Sigma\right);$$

$$\mathcal{L}\{(n_1 + n_2 - 2)S\} = W_N(\Sigma, n_1 + n_2 - 2).$$

Если через μ_y обозначить

$$\mu_y = E\{y\} = \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}(\mu_1 - \mu_2),$$

то гипотезы (16.7) примут вид

$$\begin{aligned} H_0 &: \mu_y = \mathbf{0}_N; \\ H_1 = \overline{H_0} &: \mu_y \neq \mathbf{0}_N. \end{aligned}$$

Введем в рассмотрение *обобщенную \overline{T}^2 -статистику*:

$$\overline{T}^2 = y^T S^{-1} y, \quad (16.10)$$

основанный на ней критерий для проверки гипотез (16.7) имеет вид

$$\begin{cases} H_0 &: \overline{T}^2 \leq \overline{T}_o^2; \\ H_1 = \overline{H_0} &: \overline{T}^2 > \overline{T}_o^2, \end{cases} \quad (16.11)$$

где порог $\overline{T}_o^2 = \overline{T}_o^2(\alpha)$ определяется по наперед заданному малому значению уровня значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$.

Найдем пороговое значение $\bar{T}_o^2 = \bar{T}_o^2(\alpha)$, применив лемму 16.1 к \bar{T}^2 -статистике, задаваемой соотношениями (16.10), (16.8), (16.9):

$$\begin{aligned} P\{H_1|H_0\} &= 1 - P\{H_0|H_0\} = 1 - P\{\bar{T}^2 \leq \bar{T}_o^2 | H_0\} = \\ &= 1 - P\left\{\frac{n_1 + n_2 - N - 1}{N} \frac{\bar{T}^2}{n_1 + n_2 - 2} \leq \frac{n_1 + n_2 - N - 1}{N} \frac{\bar{T}_o^2}{n_1 + n_2 - 2}\right\} = \\ &= 1 - F_{N, n_1 + n_2 - N - 1}\left(\frac{n_1 + n_2 - N - 1}{N} \frac{\bar{T}_o^2}{n_1 + n_2 - 2}\right) := \alpha, \end{aligned}$$

откуда получаем

$$\bar{T}_o^2 = \bar{T}_o^2(\alpha) = \frac{N(n_1 + n_2 - 2)}{n_1 + n_2 - N - 1} F_{N, n_1 + n_2 - N - 1}^{-1}(1 - \alpha). \quad (16.12)$$

В п. 16.4.1 будет также установлено, что критерий (16.11), (16.12) является критерием отношения правдоподобия.

16.2.2. Сравнение векторов математических ожиданий при различных ковариационных матрицах. Многомерная проблема Беренса – Фишера

Обобщим задачу из п. 16.2.1. Пусть наблюдения из выборки $X^{(i)} = \{x_t^{(i)}\}_{t=1}^{n_i}$ объема $n_i > N$ ($i = 1, 2$) независимы в совокупности и одинаково распределены по нормальному закону $N_N(\mu_i, \Sigma_i)$ с математическим ожиданием $\mu_i \in \mathbb{R}^N$ и невырожденной ковариационной $(N \times N)$ -матрицей Σ_i ($|\Sigma_i| \neq 0$). Сами выборки $X^{(1)}$ и $X^{(2)}$ также независимы.

Необходимо проверить гипотезы (16.7) о совпадении векторов математических ожиданий при условии, что ковариационные матрицы Σ_1 и Σ_2 , вообще говоря, различны, а их значения неизвестны.

Рассмотрим два случая.

Выборки равного объема ($n_1 = n_2 = n$). Образует новую выборку $X = \{x_t\}_{t=1}^n$:

$$x_t = x_t^{(1)} - x_t^{(2)}, \quad t = 1, \dots, n,$$

где $\{x_t\}_{t=1}^n$ независимы в совокупности и одинаково распределены ($t = 1, \dots, n$):

$$\mathcal{L}\{x_t\} = N_N(\mu_1 - \mu_2, \Sigma_1 + \Sigma_2).$$

Обозначим $\mu = E\{x_t\} = \mu_1 - \mu_2$, тогда гипотезы (16.7) примут эквивалентный вид

$$\begin{aligned} H_0 &: \mu = \mathbf{0}_N; \\ H_1 = \bar{H}_0 &: \mu \neq \mathbf{0}_N. \end{aligned}$$

Построим выборочную ковариационную матрицу по выборке X :

$$S = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T = S_{(1)} + S_{(2)},$$

где $\bar{x} = \bar{x}_{(1)} - \bar{x}_{(2)}$, а $\bar{x}_{(1)}$, $\bar{x}_{(2)}$ и $S_{(1)}$, $S_{(2)}$ – выборочные средние и выборочные ковариационные матрицы из (16.9), вычисленные по выборкам $X^{(1)}$, $X^{(2)}$.

Далее применяем к выборке $X = \{x_t\}_{t=1}^n$ T^2 -критерий Стьюдента (16.3), (16.5) из п. 16.1. Положив в нем $\mu^o := \mathbf{0}_N$, для T^2 -статистики из (16.3) получим соотношение

$$T^2 = n\bar{x}^T S^{-1} \bar{x} = n(\bar{x}_{(1)} - \bar{x}_{(2)})^T (S_{(1)} + S_{(2)})^{-1} (\bar{x}_{(1)} - \bar{x}_{(2)}),$$

а сам T^2 -критерий (16.3) имеет вид

$$\begin{cases} H_0 & : T^2 \leq T_o^2(\alpha); \\ H_1 = \bar{H}_0 & : T^2 > T_o^2(\alpha), \end{cases}$$

где пороговое значение $T_o^2(\alpha)$ вычисляется по уровню значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$ из соотношения (16.5):

$$T_o^2(\alpha) = \frac{N(n-1)}{n-N} F_{N, n-N}^{-1}(1-\alpha).$$

Выборки различного объема ($n_1 \neq n_2$; пусть $n_1 < n_2$). В этом случае можно из выборки большего объема отбросить наблюдения, чтобы ее объем стал равен объему меньшей выборки, и воспользоваться критерием, полученным выше для случая равных объемов.

Однако чтобы учесть все наблюдения, целесообразнее поступить следующим образом [4]. Ввести в рассмотрение новую выборку $X = \{x_t\}_{t=1}^{n_1}$:

$$x_t = x_t^{(1)} - \frac{1}{n_2} \sum_{j=1}^{n_2} x_j^{(2)} - \sqrt{\frac{n_1}{n_2}} \left(x_t^{(2)} - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(2)} \right), \quad t = 1, \dots, n_1,$$

где $\{x_t\}_{t=1}^{n_1}$ некоррелированы:

$$\begin{aligned} \text{cov}\{x_t, x_l\} &= \text{cov} \left\{ x_t^{(1)} - \frac{1}{n_2} \sum_{j=1}^{n_2} x_j^{(2)} - \sqrt{\frac{n_1}{n_2}} \left(x_t^{(2)} - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(2)} \right); \right. \\ &\quad \left. x_l^{(1)} - \frac{1}{n_2} \sum_{j=1}^{n_2} x_j^{(2)} - \sqrt{\frac{n_1}{n_2}} \left(x_l^{(2)} - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(2)} \right) \right\} = \\ &= \delta_{tl} \Sigma_1 + \text{cov} \left\{ \frac{1}{n_2} \sum_{j=1}^{n_2} x_j^{(2)} + \sqrt{\frac{n_1}{n_2}} \left(x_t^{(2)} - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(2)} \right); \right. \\ &\quad \left. \frac{1}{n_2} \sum_{j=1}^{n_2} x_j^{(2)} + \sqrt{\frac{n_1}{n_2}} \left(x_l^{(2)} - \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(2)} \right) \right\} = \\ &= \delta_{tl} \Sigma_1 + \text{cov} \left\{ \sqrt{\frac{n_1}{n_2}} x_t^{(2)} + \left(\frac{1}{n_2} - \frac{1}{\sqrt{n_1 n_2}} \right) \sum_{j=1}^{n_1} x_j^{(2)} + \frac{1}{n_2} \sum_{j=n_1+1}^{n_2} x_j^{(2)}; \right. \\ &\quad \left. \sqrt{\frac{n_1}{n_2}} x_l^{(2)} + \left(\frac{1}{n_2} - \frac{1}{\sqrt{n_1 n_2}} \right) \sum_{j=1}^{n_1} x_j^{(2)} + \frac{1}{n_2} \sum_{j=n_1+1}^{n_2} x_j^{(2)} \right\} = \\ &= \delta_{tl} \Sigma_1 + \delta_{tl} \frac{n_1}{n_2} \Sigma_2 + 2 \left(\frac{1}{n_2} - \frac{1}{\sqrt{n_1 n_2}} \right) \sqrt{\frac{n_1}{n_2}} \Sigma_2 + \left(\frac{1}{n_2} - \frac{1}{\sqrt{n_1 n_2}} \right)^2 n_1 \Sigma_2 + \\ &\quad + \left(\frac{1}{n_2} \right)^2 (n_2 - n_1) \Sigma_2 = \delta_{tl} \left(\Sigma_1 + \frac{n_1}{n_2} \Sigma_2 \right), \quad t, l = 1, \dots, n_1, \end{aligned}$$

имеют нормальное распределение:

$$\mathcal{L}\{x_t\} = N_N \left(\mu_1 - \mu_2, \Sigma_1 + \frac{n_1}{n_2} \Sigma_2 \right), \quad t = 1, \dots, n_1,$$

и поэтому независимы в совокупности.

По построенной выборке $X = \{x_t\}_{t=1}^{n_1}$ объема n_1 вычислим T^2 -статистику из (16.3) ($\mu^o := \mathbf{0}_N$, $n := n_1$):

$$T^2 = n_1 \bar{x}^T S^{-1} \bar{x},$$

где

$$\bar{x} = \frac{1}{n_1} \sum_{t=1}^{n_1} x_t, \quad S = \frac{1}{n_1 - 1} \sum_{t=1}^{n_1} (x_t - \bar{x})(x_t - \bar{x})^T \quad -$$

выборочные среднее и ковариационная матрица по $X = \{x_t\}_{t=1}^{n_1}$.

Далее для проверки гипотез H_0, H_1 применяем к выборке $X = \{x_t\}_{t=1}^{n_1}$ T^2 -критерий Стьюдента (16.3), (16.5):

$$\begin{cases} H_0 & : T^2 \leq T_o^2(\alpha); \\ H_1 = \overline{H_0} & : T^2 > T_o^2(\alpha), \end{cases}$$

$$T_o^2(\alpha) = \frac{N(n_1 - 1)}{n_1 - N} F_{N, n_1 - N}^{-1}(1 - \alpha), \quad \alpha = P\{H_1|H_0\} \in (0, 1).$$

Замечание 16.2. Для построенных в п. 16.2.2 критериев не показано, что они являются оптимальными по мощности. В этом и состоит многомерная проблема Беренса – Фишера. Поэтому если заранее известно, что ковариационные матрицы в выборках совпадают, то лучше использовать критерий из п. 16.2.1, являющийся критерием отношения правдоподобия.

16.3. ПРОВЕРКА ГИПОТЕЗ ОТНОСИТЕЛЬНО ПАРАМЕТРОВ МНОГОМЕРНОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

В пространстве \mathbb{R}^N наблюдается случайная выборка $X = \{x_1, \dots, x_n\}$ объема $n > N$ из невырожденного нормального распределения $N_N(\mu, \Sigma)$ с неизвестными значениями параметров: вектора математического ожидания $\mu \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицы Σ ($|\Sigma| \neq 0$). Рассмотрим всевозможные гипотезы относительно μ и Σ .

16.3.1. Проверка гипотез о значении вектора математического ожидания

Пусть ковариационная матрица Σ ($|\Sigma| \neq 0$) неизвестна. Необходимо проверить гипотезы о том, что математическое ожидание μ совпадает с наперед заданным вектором $\mu^o \in \mathbb{R}^N$:

$$\begin{cases} H_0 & : \mu = \mu^o; \\ H_1 = \overline{H_0} & : \mu \neq \mu^o. \end{cases}$$

Данная задача была решена в п. 16.1.

16.3.2. Проверка гипотез о значении ковариационной матрицы

Пусть вектор математического ожидания $\mu \in \mathbb{R}^N$ неизвестен. Необходимо проверить гипотезы:

$$\begin{aligned} H_0 &: \Sigma = \Sigma^o; \\ H_1 = \overline{H_0} &: \Sigma \neq \Sigma^o, \end{aligned}$$

где Σ^o – наперед заданная невырожденная ковариационная $(N \times N)$ -матрица ($\Sigma^o = (\Sigma^o)^T \succ 0$).

Воспользуемся критерием отношения правдоподобия

$$\begin{cases} H_0 &: \lambda \geq \lambda_o(\alpha); \\ H_1 = \overline{H_0} &: \lambda < \lambda_o(\alpha), \end{cases}$$

где статистика отношения правдоподобия вычисляется по функции правдоподобия $L_N(\mu, \Sigma)$ из (15.27) аналогично доказательству теоремы 16.1:

$$\begin{aligned} \lambda &= \frac{\max_{\mu} L_N(\mu, \Sigma^o)}{\max_{\mu, \Sigma} L_N(\mu, \Sigma)} = \frac{L_N(\bar{x}, \Sigma^o)}{L_N(\bar{x}, \hat{\Sigma})} = \\ &= \frac{(2\pi)^{-\frac{nN}{2}} |\Sigma^o|^{-n/2} \exp\left(-\frac{1}{2} \sum_{t=1}^n (x_t - \bar{x})^T (\Sigma^o)^{-1} (x_t - \bar{x})\right)}{(2\pi)^{-\frac{nN}{2}} |\hat{\Sigma}|^{-n/2} e^{-\frac{nN}{2}}} = \\ &= \left(\frac{e}{n}\right)^{\frac{nN}{2}} |\Sigma^o|^{-n/2} |A|^{n/2} \exp\left(-\frac{1}{2} \text{tr}((\Sigma^o)^{-1} A)\right), \end{aligned}$$

$\bar{x} = \sum_{t=1}^n x_t/n$ – выборочное среднее, $\hat{\Sigma} = A/n$ – выборочная ковариационная матрица:

$$A = \sum_{t=1}^n (x_t - \bar{x})(x_t - \bar{x})^T.$$

Построим критерий, эквивалентный критерию отношения правдоподобия:

$$\begin{cases} H_0 &: -2 \ln \lambda \leq \Delta; \\ H_1 = \overline{H_0} &: -2 \ln \lambda > \Delta, \end{cases}$$

где относительно статистики $-2 \ln \lambda$ установлено [4]:

$$\mathcal{L}_{H_0}\{-2 \ln \lambda\} \rightarrow \chi_f^2, \quad n \rightarrow +\infty, \quad -$$

χ^2 -распределение с $f = \frac{N(N+1)}{2}$ степенями свободы. Этот факт позволяет при «большом» объеме выборки n ($n \rightarrow +\infty$) по наперед заданному уровню значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$ найти пороговое значение $\Delta = \Delta(\alpha)$:

$$\Delta(\alpha) = F_{\chi_f^2}^{-1}(1 - \alpha), \quad f = \frac{N(N+1)}{2}.$$

16.3.3. Проверка гипотез о совпадении многомерного нормального распределения с наперед заданным многомерным нормальным распределением

Необходимо проверить гипотезы:

$$\begin{aligned} H_0 &: \mu = \mu^o, \Sigma = \Sigma^o; \\ H_1 = \overline{H_0} &: \mu \neq \mu^o \text{ и/или } \Sigma \neq \Sigma^o, \end{aligned}$$

где μ^o и Σ^o – наперед заданные гипотетические значения N -вектора математического ожидания ($\mu^o \in \mathbb{R}^N$) и ковариационной $(N \times N)$ -матрицы ($\Sigma^o = (\Sigma^o)^T \succ 0$).

Критерий отношения правдоподобия имеет вид

$$\begin{cases} H_0 &: \lambda \geq \lambda_o(\alpha); \\ H_1 = \overline{H_0} &: \lambda < \lambda_o(\alpha), \end{cases}$$

где

$$\begin{aligned} \lambda &= \frac{L_N(\mu^o, \Sigma^o)}{\max_{\mu, \Sigma} L_N(\mu, \Sigma)} = \frac{L_N(\mu^o, \Sigma^o)}{L_N(\bar{x}, \hat{\Sigma})} = \\ &= \frac{(2\pi)^{-\frac{nN}{2}} |\Sigma^o|^{-n/2} \exp\left(-\frac{1}{2} \sum_{t=1}^n (x_t - \mu^o)^T (\Sigma^o)^{-1} (x_t - \mu^o)\right)}{(2\pi)^{-\frac{nN}{2}} |A/n|^{-n/2} e^{-\frac{nN}{2}}} = \\ &= \left(\frac{e}{n}\right)^{\frac{nN}{2}} |\Sigma^o|^{-n/2} |A|^{n/2} \exp\left(-\frac{1}{2} (x_t - \mu^o)^T (\Sigma^o)^{-1} (x_t - \mu^o)\right) = \\ &= \left(\frac{e}{n}\right)^{\frac{nN}{2}} |\Sigma^o|^{-n/2} |A|^{n/2} \exp\left(-\frac{1}{2} \text{tr}((\Sigma^o)^{-1} A) - \frac{n}{2} (\bar{x} - \mu^o)^T (\Sigma^o)^{-1} (\bar{x} - \mu^o)\right). \end{aligned}$$

Используем эквивалентный критерий

$$\begin{cases} H_0 &: -2 \ln \lambda \leq \Delta; \\ H_1 = \overline{H_0} &: -2 \ln \lambda > \Delta, \end{cases}$$

где для статистики $-2 \ln \lambda$ также установлено предельное распределение [4]:

$$\mathcal{L}_{H_0}\{-2 \ln \lambda\} \rightarrow \chi_f^2, \quad f = \frac{N(N+1)}{2} + N, \quad n \rightarrow +\infty,$$

что позволяет определить при $n \rightarrow +\infty$ пороговое значение $\Delta = \Delta(\alpha)$, $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$:

$$\Delta(\alpha) = F_{\chi_f^2}^{-1}(1 - \alpha), \quad f = \frac{N(N+1)}{2} + N.$$

16.4. ПРОВЕРКА ГИПОТЕЗ ОТНОСИТЕЛЬНО НЕСКОЛЬКИХ ВЫБОРОК ИЗ МНОГОМЕРНЫХ НОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ

Пусть в пространстве \mathbb{R}^N наблюдается $L \geq 2$ независимых в совокупности случайных выборок $X^{(i)} = \{x_1^{(i)}, \dots, x_{n_i}^{(i)}\}$, $i = 1, \dots, L$. Выборка $X^{(i)}$ объема $n_i > N$ ($i = 1, \dots, L$) образована независимыми в совокупности, одинаково распределенными N -векторами-наблюдениями, имеющими невырожденное нормальное распределение $N_N(\mu_i, \Sigma_i)$ с математическим ожиданием $\mu_i \in \mathbb{R}^N$ и ковариационной $(N \times N)$ -матрицей Σ_i ($|\Sigma_i| \neq 0$).

16.4.1. Гипотеза о совпадении векторов математических ожиданий при неизвестной одинаковой ковариационной матрице

Проверим гипотезу о равенстве между собой векторов математических ожиданий μ_1, \dots, μ_L против альтернативы общего вида

$$\begin{aligned} H_0 &: \mu_1 = \dots = \mu_L; \\ H_1 &= \overline{H_0}, \end{aligned}$$

при неизвестном значении одинаковой для всех выборок ковариационной матрицы

$$\Sigma_1 = \dots = \Sigma_L = \Sigma \quad (|\Sigma| \neq 0).$$

Запишем функцию правдоподобия для объединенной выборки $X = \bigcup_{i=1}^L X^{(i)}$ объема

$$n = \sum_{i=1}^L n_i:$$

$$\begin{aligned} L(\{\mu_i, \Sigma_i\}_{i=1}^L) &= \prod_{i=1}^L L_N^{(i)}(\mu_i, \Sigma_i) = \\ &= (2\pi)^{-\frac{nN}{2}} \prod_{j=1}^L |\Sigma_j|^{-n_j/2} \exp \left(-\frac{1}{2} \sum_{i=1}^L \sum_{t=1}^{n_i} \left(x_t^{(i)} - \mu_i \right)^T \Sigma_i^{-1} \left(x_t^{(i)} - \mu_i \right) \right), \end{aligned}$$

где учтено, что выборки $\{X^{(i)}\}_{i=1}^L$ независимы между собой, а функция правдоподобия для i -й выборки имеет вид

$$\begin{aligned} L_N^{(i)}(\mu_i, \Sigma_i) &= \prod_{t=1}^{n_i} n_N \left(x_t^{(i)} | \mu_i, \Sigma_i \right) = \\ &= \frac{1}{(2\pi)^{\frac{1}{2}n_i N} |\Sigma_i|^{\frac{1}{2}n_i}} \exp \left(-\frac{1}{2} \sum_{t=1}^{n_i} \left(x_t^{(i)} - \mu_i \right)^T \Sigma_i^{-1} \left(x_t^{(i)} - \mu_i \right) \right). \end{aligned}$$

Воспользуемся критерием отношения правдоподобия и определим статистику отношения правдоподобия:

$$\begin{aligned} \lambda &= \frac{\max_{\mu, \Sigma} L(\{\mu_i = \mu, \Sigma_i = \Sigma\}_{i=1}^L)}{\max_{\mu_1, \dots, \mu_L, \Sigma} L(\{\mu_i, \Sigma_i = \Sigma\}_{i=1}^L)} = \\ &= \frac{(2\pi)^{-\frac{nN}{2}} \left| \frac{1}{n} A \right|^{-n/2} e^{-\frac{nN}{2}}}{(2\pi)^{-\frac{nN}{2}} \left| \frac{1}{n} \sum_{i=1}^L A_i \right|^{-n/2} e^{-\frac{nN}{2}}} = \left(\left| \sum_{i=1}^L A_i \right| / |A| \right)^{\frac{n}{2}}, \end{aligned}$$

где

$$\bar{x} = \frac{1}{n} \sum_{i=1}^L \sum_{t=1}^{n_i} x_t^{(i)} -$$

выборочное среднее,

$$\hat{\Sigma}_0 = \frac{1}{n} A, \quad A = \sum_{i=1}^L \sum_{t=1}^{n_i} \left(x_t^{(i)} - \bar{x} \right) \left(x_t^{(i)} - \bar{x} \right)^T, \quad -$$

выборочная ковариационная матрица, вычисленные по объединенной выборке X объема n и являющиеся при истинной гипотезе H_0 оценками максимального правдоподобия для общих вектора математического ожидания μ и ковариационной матрицы Σ ;

$$\bar{x}_{(i)} = \frac{1}{n_i} \sum_{t=1}^{n_i} x_t^{(i)}, \quad i = 1, \dots, L, \quad -$$

выборочные средние,

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n A_i, \quad A_i = \sum_{t=1}^{n_i} \left(x_t^{(i)} - \bar{x}_{(i)} \right) \left(x_t^{(i)} - \bar{x}_{(i)} \right)^T, \quad i = 1, \dots, L, \quad -$$

выборочная ковариационная матрица, являющиеся оценками максимального правдоподобия для математических ожиданий $\{\mu_i\}_{i=1}^L$ и общей для всех выборок $\{X^{(i)}\}_{i=1}^L$ ковариационной матрицы Σ в общем случае (когда гипотеза H_0 , вообще говоря, неверна).

Вместо статистики λ воспользуемся так называемой U -статистикой:

$$U = \lambda^{\frac{2}{n}} = \left| \sum_{i=1}^L A_i \right| / |A|,$$

для которой известно [4]:

$$\mathcal{L}_{H_0} \{-g \ln U\} \rightarrow \chi_f^2, \quad n_i \rightarrow \infty, \quad i = 1, \dots, L \quad (n \rightarrow +\infty),$$

где $g = n - L - \frac{N - L}{2} - 1$, $f = N(L - 1)$. Тогда критерий отношения правдоподобия

$$\begin{cases} H_0 & : \lambda \geq \lambda_o(\alpha); \\ H_1 = \overline{H_0} & : \lambda < \lambda_o(\alpha), \end{cases}$$

запишется в эквивалентном виде:

$$\begin{cases} H_0 & : -g \ln U \leq \Delta(\alpha); \\ H_1 = \overline{H_0} & : -g \ln U > \Delta(\alpha), \end{cases}$$

где порог критерия $\Delta = \Delta(\alpha)$ определяется по наперед заданному уровню значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$:

$$\Delta(\alpha) = F_{\chi_f^2}^{-1}(1 - \alpha), \quad f = N(L - 1).$$

В случае двух выборок ($L = 2$), воспользовавшись соотношениями

$$A_1 + A_2 = A + n_1(\bar{x} - \bar{x}_{(1)})(\bar{x} - \bar{x}_{(1)})^T + n_2(\bar{x} - \bar{x}_{(2)})(\bar{x} - \bar{x}_{(2)})^T;$$

$$n\bar{x} = n_1\bar{x}_{(1)} + n_2\bar{x}_{(2)}$$

и формулой блочного определителя, по аналогии с доказательством теоремы 16.1 для U -статистики получим

$$\begin{aligned} U &= \frac{|A_1 + A_2|}{|A|} = \\ &= \frac{|A_1 + A_2|}{|A_1 + A_2 - n_1(\bar{x} - \bar{x}_{(1)})(\bar{x} - \bar{x}_{(1)})^T - n_2(\bar{x} - \bar{x}_{(2)})(\bar{x} - \bar{x}_{(2)})^T|} = \end{aligned}$$

$$\begin{aligned}
&= \frac{|A_1 + A_2|}{\left| A_1 + A_2 - \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_{(1)} - \bar{x}_{(2)}) (\bar{x}_{(1)} - \bar{x}_{(2)})^T \right|} = \\
&= \frac{|A_1 + A_2|}{\left| A_1 + A_2 \left| 1 + \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_{(1)} - \bar{x}_{(2)})^T (A_1 + A_2)^{-1} (\bar{x}_{(1)} - \bar{x}_{(2)}) \right| \right|} = \\
&= \frac{1}{1 + \frac{\bar{T}^2}{n_1 + n_2 - 2}}
\end{aligned}$$

выражается через обобщенную \bar{T}^2 -статистику из (16.10), и построенный ранее в п. 16.2.1 на ее основе критерий (16.11), (16.12) также является критерием отношения правдоподобия. Более того, он обладает очевидным преимуществом перед критерием, использующим U -статистику: пороговое значение (16.12) в нем не требует асимптотики по объемам выборок. Поэтому для проверки гипотезы о совпадении векторов математических ожиданий в случае двух выборок целесообразнее использовать критерий из п. 16.2.1.

16.4.2. Гипотеза о равенстве ковариационных матриц

Проверим гипотезы о равенстве ковариационных матриц

$$\begin{aligned}
H_0 &: \Sigma_1 = \dots = \Sigma_L; \\
H_1 &= \overline{H_0}.
\end{aligned}$$

Статистика отношения правдоподобия в этом случае имеет вид

$$\begin{aligned}
\lambda &= \frac{\max_{\mu_1, \dots, \mu_L, \Sigma} L(\{\mu_i, \Sigma_i = \Sigma\}_{i=1}^L)}{\max_{\{\mu_i, \Sigma_i\}_{i=1}^L} L(\{\mu_i, \Sigma_i\}_{i=1}^L)} = \\
&= \frac{\max_{\mu_1, \dots, \mu_L, \Sigma} L(\{\mu_i, \Sigma_i = \Sigma\}_{i=1}^L)}{\prod_{i=1}^L \max_{\mu_i, \Sigma_i} L_N^{(i)}(\mu_i, \Sigma_i)} = \frac{(2\pi)^{-\frac{nN}{2}} \left| \sum_{i=1}^L A_i / n \right|^{-n/2} e^{-\frac{nN}{2}}}{\prod_{i=1}^L (2\pi)^{-\frac{n_i N}{2}} |A_i / n_i|^{-n_i/2} e^{-\frac{n_i N}{2}}} = \\
&= \frac{n^{\frac{nN}{2}} \prod_{i=1}^L |A_i|^{n_i/2}}{\prod_{i=1}^L n_i^{\frac{n_i N}{2}} \left| \sum_{i=1}^L A_i \right|^{n/2}}.
\end{aligned}$$

Введем в рассмотрение «подправленную» статистику, предложенную Бартлеттом [4]:

$$\lambda^* = \frac{m^{\frac{mN}{2}} \prod_{i=1}^L |A_i|^{m_i/2}}{\prod_{i=1}^L m_i^{\frac{m_i N}{2}} \left| \sum_{i=1}^L A_i \right|^{m/2}},$$

где $m_i = n_i - 1$ ($i = 1, \dots, L$), $m = \sum_{i=1}^L m_i = n - L$, и построим критерий, асимптотически эквивалентный критерию отношения правдоподобия:

$$\begin{cases} H_0 & : -2\rho \ln \lambda^* \leq \Delta(\alpha); \\ H_1 = \overline{H_0} & : -2\rho \ln \lambda^* > \Delta(\alpha), \end{cases}$$

где

$$\rho = 1 - \left(\sum_{i=1}^L \frac{1}{m_i} - \frac{1}{m} \right) \frac{2N^2 + 3N - 1}{6(N+1)(L-1)}$$

и установлено [4]:

$$\mathcal{L}_{H_0}\{-2\rho \ln \lambda^*\} \rightarrow \chi_f^2, \quad f = \frac{L-1}{2}N(N+1), \quad n_i \rightarrow \infty, \quad i = 1, \dots, L,$$

что позволяет по уровню значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$ определить порог:

$$\Delta(\alpha) = F_{\chi_f^2}^{-1}(1 - \alpha), \quad f = \frac{L-1}{2}N(N+1).$$

16.4.3. Гипотеза об эквивалентности нормальных распределений (гипотеза однородности)

Проверим гипотезы о том, что наблюдения из всех выборок имеют одно и то же нормальное распределение:

$$\begin{aligned} H_0 & : \mu_1 = \dots = \mu_L, \quad \Sigma_1 = \dots = \Sigma_L; \\ H_1 = \overline{H_0}. \end{aligned}$$

Запишем статистику отношения правдоподобия:

$$\begin{aligned} \lambda &= \frac{\max_{\mu, \Sigma} L(\{\mu_i = \mu, \Sigma_i = \Sigma\}_{i=1}^L)}{\max_{\{\mu_i, \Sigma_i\}_{i=1}^L} L(\{\mu_i, \Sigma_i\}_{i=1}^L)} = \frac{(2\pi)^{-\frac{nN}{2}} |A/n|^{-n/2} e^{-\frac{nN}{2}}}{\prod_{i=1}^L (2\pi)^{-\frac{n_i N}{2}} |A_i/n_i|^{-n_i/2} e^{-\frac{n_i N}{2}}} = \\ &= \frac{n^{\frac{nN}{2}} \prod_{i=1}^L |A_i|^{n_i/2}}{\prod_{i=1}^L n_i^{\frac{n_i N}{2}} |A|^{n/2}}. \end{aligned}$$

Определим соответствующую «подправленную» статистику:

$$\bar{\lambda}^* = \frac{m^{\frac{mN}{2}} \prod_{i=1}^L |A_i|^{m_i/2}}{\prod_{i=1}^L m_i^{\frac{m_i N}{2}} |A|^{m/2}}$$

и построим критерий, асимптотически эквивалентный критерию отношения правдоподобия и имеющий вид

$$\begin{cases} H_0 & : -2\rho^* \ln \bar{\lambda}^* \leq \Delta(\alpha); \\ H_1 = \overline{H_0} & : -2\rho^* \ln \bar{\lambda}^* > \Delta(\alpha), \end{cases}$$

где

$$\rho^* = 1 - \left(\sum_{i=1}^L \frac{1}{m_i} - \frac{1}{m} \right) \frac{2N^2 + 3N - 1}{6(N+3)(L-1)} - \frac{1}{m} \frac{N-L+2}{N+3},$$

и также установлено [4]:

$$\mathcal{L}_{H_0}\{-2\rho^* \ln \bar{\lambda}^*\} \rightarrow \chi_f^2, \quad f = \frac{L-1}{2}N(N+3), \quad n_i \rightarrow \infty, \quad i = 1, \dots, L,$$

что позволяет по $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$ определить порог:

$$\Delta(\alpha) = F_{\chi_f^2}^{-1}(1 - \alpha), \quad f = \frac{L-1}{2}N(N+3).$$

16.5. ЗАДАЧИ СТАТИСТИЧЕСКОЙ КЛАССИФИКАЦИИ. ОПТИМАЛЬНОЕ (БАЙЕСОВСКОЕ) РЕШАЮЩЕЕ ПРАВИЛО

Пусть в пространстве \mathbb{R}^N регистрируются случайные наблюдения $x = x(\omega) \in \mathbb{R}^N$ над объектами $\omega \in \Omega$, принадлежащими к $L \geq 2$ классам $\{\Omega_1, \dots, \Omega_L\}$:

$$\Omega_i \cap \Omega_j = \emptyset, \quad i \neq j, \quad i, j \in \mathcal{S}, \quad \mathcal{S} = \{1, \dots, L\}; \quad \bigcup_{i \in \mathcal{S}} \Omega_i = \Omega.$$

Истинный номер класса $d^o \in \mathcal{S}$, к которому принадлежит наблюдение $x = x(\omega)$, является дискретной случайной величиной, принимающей значения из множества номеров классов \mathcal{S} :

$$d^o = d^o(\omega) = \sum_{i \in \mathcal{S}} i \mathbb{1}_{\Omega_i}(\omega), \quad \omega \in \Omega,$$

и имеющей распределение вероятностей

$$\mathbf{P}\{d^o(\omega) = i\} = \pi_i > 0, \quad i \in \mathcal{S}; \quad \pi_1 + \dots + \pi_L = 1, \quad (16.13)$$

где $\{\pi_i\}_{i \in \mathcal{S}}$ – так называемые *априорные вероятности классов*.

При фиксированном номере класса $d^o = i$, $i \in \mathcal{S}$, наблюдение $x \in \mathbb{R}^N$ описывается условными распределениями вероятностей, которые будем предполагать абсолютно непрерывными с условными плотностями

$$p_i(x) = p(x|d^o = i), \quad i \in \mathcal{S}, \quad (16.14)$$

называемыми *плотностями распределения классов*.

Исходя из (16.13), (16.14), наблюдение $x \in \mathbb{R}^N$ имеет безусловную плотность

$$p(x) = \sum_{i \in \mathcal{S}} \pi_i p_i(x),$$

называемую *плотностью-смесью*.

Таким образом, классы $\{\Omega_i\}_{i \in \mathcal{S}}$ однозначно определяются своими вероятностными характеристиками $\{\pi_i, p_i(\cdot)\}_{i \in \mathcal{S}}$ из (16.13), (16.14).

Задача статистической классификации заключается в отнесении объекта $\omega \in \Omega$ к одному из классов $\{\Omega_i\}_{i \in \mathcal{S}}$ по наблюдению $x = x(\omega)$ над ним, т. е. в оценивании неизвестного (ненаблюдаемого) номера класса $d^o = d^o(\omega) \in \mathcal{S}$ для ω по $x = x(\omega) \in \mathbb{R}^N$.

Рассмотрим *класс рандомизированных решающих правил* (п. 12.1): $\mathcal{D} = \{d(x), x \in \mathbb{R}^N\}$, являющихся случайными функциями со значениями из множества \mathcal{S} . При фиксированном x решающее правило (РП) $d(x)$ есть дискретная случайная величина с распределением

$$\mathbf{P}\{d(x) = i\} = \Phi_i(x) \geq 0, \quad x \in \mathbb{R}^N, \quad i \in \mathcal{S},$$

где критические функции $\{\phi_i(\cdot)\}_{i \in \mathcal{S}}$ удовлетворяют условию нормировки

$$\phi_1(x) + \dots + \phi_L(x) = 1, \quad x \in \mathbb{R}^N.$$

Таким образом, решение о принадлежности наблюдения x к классу с номером $d(x)$ выносится с помощью «жребия», определяемого набором критических функций $\{\phi_i(x)\}_{i \in \mathcal{S}}$. Если $\phi_i(x) \in \{0, 1\}$, $\forall x \in \mathbb{R}^N$, $i \in \mathcal{S}$, то РП $d(x)$ называется нерандомизированным и принадлежит к классу *нерандомизированных решающих правил* \mathcal{D}^* (очевидно, что $\mathcal{D}^* \subset \mathcal{D}$). Оптимальность (точность) РП $d(\cdot) \in \mathcal{D}$ будем характеризовать *функцией риска* (п. 12.3), имеющим смысл вероятности ошибочной классификации:

$$\begin{aligned} r &= r(d(\cdot)) = P\{d(x) \neq d^o\} = 1 - P\{d(x) = d^o\} = \\ &= 1 - \sum_{i \in \mathcal{S}} P\{d^o = i\} P\{d(x) = i | d^o = i\} = \\ &= 1 - \sum_{i \in \mathcal{S}} \pi_i \int_{\mathbb{R}^N} \phi_i(x) p_i(x) dx = 1 - \int_{\mathbb{R}^N} \sum_{i \in \mathcal{S}} \pi_i \phi_i(x) p_i(x) dx. \end{aligned}$$

Из правой части последнего соотношения очевидно, что минимум риска на классе всех рандомизированных решающих правил достигается на РП

$$d_o(x) = \arg \max_{i \in \mathcal{S}} (\pi_i p_i(x)), \quad x \in \mathbb{R}^N, \quad (16.15)$$

которое является нерандомизированным и имеет минимальную вероятность ошибки:

$$r_o = P\{d_o(x) \neq d^o\} = \min_{d(\cdot) \in \mathcal{D}} r(d(\cdot)) = 1 - \int_{\mathbb{R}^N} \max_{i \in \mathcal{S}} (\pi_i p_i(x)) dx. \quad (16.16)$$

Причем, если выполняется: $\mu(\{x : \pi_i p_i(x) = \pi_j p_j(x)\}) = 0$, $i \neq j$, $i, j \in \mathcal{S}$, то это РП является единственным оптимальным с точностью до множества меры Лебега $\mu = \mu(\cdot)$ нуль. РП (16.15) известно как *байесовское решающее правило* (БРП), а достигнутое на нем минимально возможное значение риска (16.16) называется *байесовским риском*.

Проясним смысл БРП (16.15). Введем в рассмотрение так называемые *апостериорные вероятности классов*:

$$P_i(x) = P\{d^o = i | x\} = \frac{\pi_i p_i(x)}{p(x)} = \frac{\pi_i p_i(x)}{\sum_{j \in \mathcal{S}} \pi_j p_j(x)}, \quad i \in \mathcal{S};$$

$$P_1(x) + \dots + P_L(x) = 1, \quad x \in \mathbb{R}^N,$$

$P_i(x) \in [0, 1]$ – вероятность того, что фиксированное наблюдение $x \in \mathbb{R}^N$ принадлежит классу Ω_i ($i \in \mathcal{S}$). Тогда БРП (16.15) может быть записано через $\{P_i(\cdot)\}_{i \in \mathcal{S}}$:

$$d_o(x) = \arg \max_{i \in \mathcal{S}} P_i(x), \quad x \in \mathbb{R}^N, \quad -$$

относит наблюдение x к тому классу, апостериорная вероятность которого больше остальных. На практике анализ значений апостериорных вероятностей позволяет сказать, насколько «уверенно» выносится решение: идеальным является случай, когда апостериорная вероятность класса, в пользу которого выносится решение для данного наблюдения, «сильно доминирует» по значению над остальными (близка к 1).

Пусть теперь характеристики $\{\pi_i, p_i(\cdot)\}_{i \in \mathcal{S}}$ классов $\{\Omega_i\}_{i \in \mathcal{S}}$, определенные в (16.13), (16.14), частично или полностью неизвестны. И в \mathbb{R}^N регистрируется случайная выборка $X = \{x_1, \dots, x_n\}$ объема n , образованная из n независимых в совокупности N -векторов-наблюдений x_1, \dots, x_n , принадлежащих $L \geq 2$ классам $\{\Omega_1, \dots, \Omega_L\}$. Наблюдение $x_t = x(\omega_t)$ принадлежит классу со случайным номером $d_t^o = d^o(\omega_t)$, определяемым дискретным распределением вероятностей (16.13), и при фиксированном $d_t^o \in \mathcal{S}$ имеет условную плотность $p_{d_t^o}(\cdot)$ ($t = 1, \dots, n$). Вектор истинной классификации $D^o = (d_1^o, \dots, d_n^o)^T \in \mathcal{S}^n$ выборки X может быть как известен, так и неизвестен (не наблюдается). В зависимости от этого выделяют две задачи.

Задача *кластерного анализа* (кластер-анализа) состоит в классификации выборки X , т. е. в построении статистической оценки $\hat{D} = \hat{D}(X) = (\hat{d}_1, \dots, \hat{d}_n)^T \in \mathcal{S}^n$ для неизвестного n -вектора истинной классификации $D^o \in \mathcal{S}^n$ по X объема n . Кроме того, может также ставиться задача классификации вновь поступающих наблюдений x_{n+1}, x_{n+2}, \dots , не зависящих от выборки X , по которой восполнялась неизвестная информация о классах $\{\Omega_i\}_{i \in \mathcal{S}}$ (оценивались их неизвестные характеристики $\{\pi_i, p_i(\cdot)\}_{i \in \mathcal{S}}$).

В *дискриминантном анализе* решается задача классификации вновь поступающих наблюдений x_{n+1}, x_{n+2}, \dots при наличии классифицированной обучающей выборки X объема n , для которой вектор истинной классификации $D^o \in \mathcal{S}^n$ также наблюдается. Широко используется подход, основанный на построении так называемых *подстановочных решающих правил* (plug-in-rules). Обычно строится подстановочное БРП, которое получается в результате подстановки в БРП $d_o(\cdot)$ вместо неизвестных характеристик $\{\pi_i, p_i(\cdot)\}_{i \in \mathcal{S}}$ классов $\{\Omega_i\}_{i \in \mathcal{S}}$ их статистических оценок $\{\hat{\pi}_i, \hat{p}_i(\cdot)\}_{i \in \mathcal{S}}$, вычисленных по X объема n и $D^o \in \mathcal{S}^n$:

$$\hat{d} = \hat{d}(x) = \arg \max_{i \in \mathcal{S}} (\hat{\pi}_i \hat{p}_i(x)), \quad x \in \mathbb{R}^N. \quad (16.17)$$

В кластерном анализе подстановочные РП типа (16.17) также используются, однако специфика их построения заключается в том, что оценивание неизвестных характеристик классов и принятие решений (классификация выборки) здесь производятся «одновременно», преимущественно с использованием итерационных вычислительных процедур.

Отметим также, что задача кластер-анализа, вообще говоря, может быть решена лишь с точностью до переобозначения номеров классов: находя оценку $\hat{D} = \hat{D}(X) \in \mathcal{S}^n$ для вектора истинной классификации $D^o \in \mathcal{S}^n$, мы на самом деле определяем последовательность из $L!$ возможных решений: $\hat{D}_{(1)} := \hat{D}$, $\hat{D}_{(2)}, \dots, \hat{D}_{(L!)}$, образованных всевозможными $L!$ перестановками (переобозначениями) в $\hat{D} \in \mathcal{S}^n$ номеров $1, \dots, L$ классов $\{\Omega_1, \dots, \Omega_L\}$. Чтобы избежать неоднозначности, далее везде будем предполагать, что классы $\{\Omega_i\}_{i \in \mathcal{S}}$ упорядочены: $\Omega_1 \prec \Omega_2 \prec \dots \prec \Omega_L$, где « \prec » – символ лексикографического сравнения, и решена так называемая проблема соотнесения. Данное предположение говорит о том, что среди возможных $L!$ решений $\hat{D}_{(j)} \in \mathcal{S}^n$, $j = 1, \dots, L!$, выбрано то, в котором нумерация классов $1, \dots, L$ совпадает с истинной $\{\Omega_1, \dots, \Omega_L\}$.

При данном предположении эффективность дискриминантного и кластерного анализа можно оценить при помощи следующей эмпирической характеристики:

$$\gamma_n = \frac{1}{n} \sum_{t=1}^n (1 - \delta_{\hat{d}_t, d_t^o}) \quad - \quad (16.18)$$

доли ошибочных решений, где \hat{d}_t – решение, выносимое для наблюдения x_t ($t = 1, \dots, n$).

Эта характеристика может рассматриваться как оценка риска, и для нее справедливо соотношение

$$E\{\gamma_n\} = \frac{1}{n} \sum_{t=1}^n P\{\hat{d}_t \neq d_t^o\} \geq r_o.$$

Разумеется, чтобы использовать характеристику (16.18) для оценивания эффективности в кластер-анализе, для выборки X необходимо знать истинное значение вектора классификации D^o .

16.6. СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ МНОГОМЕРНЫХ НОРМАЛЬНЫХ НАБЛЮДЕНИЙ

16.6.1. Байесовское решающее правило и его риск

Пусть теперь условные плотности $\{p_i(\cdot)\}_{i \in \mathcal{S}}$ из (16.14), описывающие классы $\{\Omega_i\}_{i \in \mathcal{S}}$, – многомерные нормальные:

$$p_i(x) = n_N(x|\mu_i, \Sigma_i), \quad x \in \mathbb{R}^N, \quad i \in \mathcal{S}, \quad (16.19)$$

где наблюдения $x \in \mathbb{R}^N$ из класса Ω_i ($d^o = i$) описываются условными: математическим ожиданием $\mu_i = E\{x | d^o = i\}$ (так называемый центр i -го класса) и невырожденной ковариационной $(N \times N)$ -матрицей $\Sigma_i = E\{(x - \mu_i)(x - \mu_i)^T | d^o = i\}$ ($|\Sigma_i| \neq 0$).

В приложениях при решении реальных задач часто наблюдения, подлежащие классификации, адекватно определяются частным случаем модели (16.13), (16.19) – *моделью Фишера*:

$$p_i(x) = n_N(x|\mu_i, \Sigma), \quad x \in \mathbb{R}^N, \quad i \in \mathcal{S}, \quad (16.20)$$

с общей для всех классов невырожденной ковариационной $(N \times N)$ -матрицей $\Sigma = E\{(x - \mu_i)(x - \mu_i)^T | d^o = i\}$ ($i \in \mathcal{S}$, $|\Sigma| \neq 0$), описывающей статистический характер ошибок наблюдения: $x = \mu_{d^o} + \xi$, где распределение вероятностей N -вектора ошибок ξ не зависит от номера класса d^o , к которому принадлежит наблюдение x , и является N -мерным нормальным вектором с нулевым математическим ожиданием и ковариационной матрицей Σ ($\mathcal{L}\{\xi\} = N_N(\mathbf{0}_N, \Sigma)$).

Построим байесовское решающее правило.

Теорема 16.2. Пусть классы $\{\Omega_i\}_{i \in \mathcal{S}}$ определяются моделью (16.13), (16.19) с априорными вероятностями $\{\pi_i\}_{i \in \mathcal{S}}$ и невырожденными нормальными распределениями $\{N_N(\mu_i, \Sigma_i), |\Sigma_i| \neq 0\}_{i \in \mathcal{S}}$, тогда БРП (16.15) допускает представление ($x \in \mathbb{R}^N$):

$$d_o(x) = \arg \min_{i \in \mathcal{S}} ((x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) + \ln |\Sigma_i| - 2 \ln \pi_i), \quad (16.21)$$

и для модели Фишера (16.13), (16.20):

$$d_o(x) = \arg \min_{i \in \mathcal{S}} ((x - \mu_i)^T \Sigma^{-1} (x - \mu_i) - 2 \ln \pi_i). \quad (16.22)$$

Доказательство. Очевидно и следует из соотношения

$$d_o(x) = \arg \max_{i \in \mathcal{S}} (\pi_i p_i(x)) = \arg \max_{i \in \mathcal{S}} \ln(\pi_i p_i(x)) = \arg \min_{i \in \mathcal{S}} (-2 \ln(\pi_i p_i(x))),$$

и вида плотности многомерного нормального распределения (14.1). \square

Следствие 16.2. В условиях модели Фишера (16.13), (16.20) при равновероятных классах: $\pi_i = 1/L$, $i \in \mathcal{S}$, БРП имеет вид

$$d_o(x) = \arg \min_{i \in \mathcal{S}} \rho(x, \mu_i), \quad x \in \mathbb{R}^N, \quad (16.23)$$

где

$$\rho(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}, \quad x, y \in \mathbb{R}^N, \quad - \quad (16.24)$$

метрика Махаланобиса.

Следствие 16.3. Для модели Фишера (16.13), (16.20) в случае двух классов ($L = 2$) байесовский риск r_o из (16.16) может быть вычислен из соотношения

$$r_o = \pi_1 \Phi\left(-\frac{\Delta}{2} - \frac{h}{\Delta}\right) + \pi_2 \Phi\left(-\frac{\Delta}{2} + \frac{h}{\Delta}\right),$$

$$h = \ln \frac{\pi_1}{\pi_2} = \ln \frac{\pi_1}{1 - \pi_1},$$

и при равновероятных классах ($\pi_1 = \pi_2 = 1/2$)

$$r_o = \Phi\left(-\frac{\Delta}{2}\right),$$

где $\Phi(\cdot)$ – функция распределения вероятностей стандартного нормального закона $N_1(0, 1)$; $\Delta = \rho(\mu_1, \mu_2) = \sqrt{(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)}$ – межклассовое расстояние Махаланобиса (расстояние Махаланобиса между «центрами» классов).

Доказательство. В случае двух классов ($L = 2$) БРП (16.22) может быть записано в виде

$$d_o(x) = \begin{cases} 1, & B(x) < 0, \\ 2, & B(x) \geq 0, \end{cases}$$

где $B(x) = 1/2 ((x - \mu_1)^T \Sigma^{-1} (x - \mu_1) - 2 \ln \pi_1 - ((x - \mu_2)^T \Sigma^{-1} (x - \mu_2) - 2 \ln \pi_2)) = b^T x - H$ – линейная по $x \in \mathbb{R}^N$ функция с коэффициентами

$$b = \Sigma^{-1}(\mu_2 - \mu_1), \quad H = \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_2 - \mu_1) + h.$$

Для байесовского риска r_o из (16.16) имеем

$$\begin{aligned} r_o &= \pi_1 P\{d_o(x) = 2 | d^o = 1\} + \pi_2 P\{d_o(x) = 1 | d^o = 2\} = \\ &= \pi_1 P\{B(x) \geq 0 | d^o = 1\} + \pi_2 P\{B(x) < 0 | d^o = 2\}. \end{aligned}$$

Найдем условные распределения вероятностей случайной величины $B(x)$ при фиксированном номере класса $d^o = i$ для наблюдения x . Учтем, что при $d^o = i$ согласно модели Фишера случайный N -вектор-наблюдение $x \in \mathbb{R}^N$ имеет многомерное нормальное распределение $N_N(\mu_i, \Sigma)$, и по теореме 14.2 $B(x) = b^T x - H \in \mathbb{R}^1$ также имеет условное нормальное распределение: $\mathcal{L}\{B(x) | d^o = i\} = N_1(m_i, \sigma_i^2)$, $i \in \mathcal{S}$. Найдем математическое ожидание m_i и дисперсию σ_i^2 ($i \in \mathcal{S}$, $\mathcal{S} = \{1, 2\}$):

$$\begin{aligned} m_i &= \mathbf{E}\{B(x) | d^o = i\} = b^T \mu_i - H = \\ &= (\mu_2 - \mu_1)^T \Sigma^{-1} \mu_i - \frac{1}{2}(\mu_1 + \mu_2)^T \Sigma^{-1}(\mu_2 - \mu_1) - h = \\ &= -\frac{1}{2}(\mu_1 + \mu_2 - 2\mu_i)^T \Sigma^{-1}(\mu_2 - \mu_1) - h = (-1)^i \frac{\Delta^2}{2} - h; \\ \sigma_i^2 &= \mathbf{D}\{B(x) | d^o = i\} = \text{cov}\{b^T x - H, b^T x - H | d^o = i\} = \\ &= b^T \text{cov}\{x, x | d^o = i\} b = b^T \Sigma b = (\mu_2 - \mu_1)^T \Sigma^{-1} \Sigma \Sigma^{-1} (\mu_2 - \mu_1) = \Delta^2. \end{aligned}$$

Продолжим преобразование риска r_o и получим

$$\begin{aligned} r_o &= \pi_1 P \left\{ \frac{B(x) - m_1}{\Delta} \geq \frac{-m_1}{\Delta} \mid d^o = 1 \right\} + \pi_2 P \left\{ \frac{B(x) - m_2}{\Delta} < \frac{-m_2}{\Delta} \mid d^o = 2 \right\} = \\ &= \pi_1 \left(1 - \Phi \left(\frac{-m_1}{\Delta} \right) \right) + \pi_2 \Phi \left(\frac{-m_2}{\Delta} \right) = \pi_1 \Phi \left(\frac{m_1}{\Delta} \right) + \pi_2 \Phi \left(\frac{-m_2}{\Delta} \right) = \\ &= \pi_1 \Phi \left(-\frac{\Delta}{2} - \frac{h}{\Delta} \right) + \pi_2 \Phi \left(-\frac{\Delta}{2} + \frac{h}{\Delta} \right), \end{aligned}$$

где учтено свойство функции распределения стандартного нормального закона: $\Phi(z) = 1 - \Phi(-z)$, $z \in \mathbb{R}$. \square

Замечание 16.3. Как видно из доказательства, в условиях модели Фишера БРП может быть выражено через функции, линейные по классифицируемому наблюдению, чего не скажешь о случае, когда ковариационные матрицы условных нормальных распределений, описывающих классы, различны. В связи с этим БРП для модели Фишера называется линейным, а при различных ковариационных матрицах – квадратичным.

16.6.2. Квадратичный и линейный дискриминантный анализ

Решим задачу дискриминантного анализа для модели (16.13), (16.19), когда классы $\{\Omega_i\}_{i \in \mathcal{S}}$ описываются неизвестными априорными вероятностями $\{\pi_i\}_{i \in \mathcal{S}}$ и невырожденными нормальными распределениями $\{N_N(\mu_i, \Sigma_i), |\Sigma_i| \neq 0\}_{i \in \mathcal{S}}$ с неизвестными значениями параметров $\{\mu_i, \Sigma_i\}_{i \in \mathcal{S}}$.

Наличие классифицированной обучающей выборки $X = \{x_1, \dots, x_n\}$ объема n , для которой известен вектор истинной классификации $D^o = (d_1^o, \dots, d_n^o)^T \in \mathcal{S}^n$, позволяет построить несмещенные оценки неизвестных характеристик $\{\pi_i, \mu_i, \Sigma_i\}_{i \in \mathcal{S}}$ классов $\{\Omega_i\}_{i \in \mathcal{S}}$ ($i \in \mathcal{S}$):

$$\hat{\pi}_i = \frac{n_i}{n}, \quad n_i = \sum_{t=1}^n \delta_{d_t^o, i}, \quad - \quad (16.25)$$

доля наблюдений, попавших, согласно вектору истинной классификации $D^o \in \mathcal{S}^n$, в i -й класс;

$$\hat{\mu}_i = \bar{x}_{(i)} = \frac{1}{n_i} \sum_{t=1}^n \delta_{d_t^o, i} x_t \quad - \quad (16.26)$$

арифметическое среднее наблюдений из выборки X , попавших в i -й класс (оценка «центра» i -го класса);

$$\hat{\Sigma}_i = S_{(i)} = \frac{1}{n_i - 1} \sum_{t=1}^n \delta_{d_t^o, i} (x_t - \hat{\mu}_i)(x_t - \hat{\mu}_i)^T \quad - \quad (16.27)$$

выборочная ковариационная матрица для i -го класса.

Несмещенность оценок априорных вероятностей (16.25) очевидна:

$$\mathbf{E}\{\hat{\pi}_i\} = \mathbf{E}\left\{\frac{1}{n} \sum_{t=1}^n \delta_{d_t^o, i}\right\} = \frac{1}{n} n P\{d_t^o = i\} = \pi_i, \quad i \in \mathcal{S}.$$

Оценки $\hat{\mu}_i$ и $\hat{\Sigma}_i$ из (16.26) и (16.27) при фиксированном $D^o \in \mathcal{S}^n$ согласно теореме 14.5 являются условно несмещенными оценками параметров многомерного нормального распределения $N_N(\mu_i, \Sigma_i)$, построенными по подвыборке $X^{(i)} = \{x_t \in X : d_t^o = i\}$ объема n_i : $\mathbf{E}\{\hat{\mu}_i | D^o\} = \mu_i$, $\mathbf{E}\{\hat{\Sigma}_i | D^o\} = \Sigma_i$, где условные математические ожидания не зависят от D^o , что означает безусловную несмещенность оценок (16.26) (16.27).

Подстановочное БРП, соответствующее (16.21) и основанное на оценках (16.25) – (16.27) имеет вид ($x \in \mathbb{R}^N$)

$$\hat{d}_o(x) = \arg \min_{i \in \mathcal{S}} \left((x - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1} (x - \hat{\mu}_i) + \ln |\hat{\Sigma}_i| - 2 \ln \hat{\pi}_i \right), \quad (16.28)$$

и определяет процедуру *квадратичного дискриминантного анализа* (применяется, если $|\hat{\Sigma}_i| \neq 0$, $i \in \mathcal{S}$).

В случае модели Фишера (16.13), (16.20) ($\Sigma_i = \Sigma$, $i \in \mathcal{S}$) можно продолжать пользоваться квадратичным РП (16.28), но с точки зрения точности оценивания лучше проводить *линейный дискриминантный анализ*, основанный на БРП (16.22) ($x \in \mathbb{R}^N$):

$$\hat{d}_o(x) = \arg \min_{i \in \mathcal{S}} \left((x - \hat{\mu}_i)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_i) - 2 \ln \hat{\pi}_i \right), \quad (16.29)$$

где в (16.29) вместо оценок ковариационных матриц $\{\hat{\Sigma}_i\}_{i \in \mathcal{S}}$ используется несмещенная оценка общей для всех классов ковариационной матрицы Σ :

$$\hat{\Sigma} = \frac{1}{n-L} \sum_{t=1}^n (x_t - \hat{\mu}_{d_t^o})(x_t - \hat{\mu}_{d_t^o})^T, \quad (16.30)$$

вычисляемая по всем n наблюдениям из выборки X (применяется, если $|\hat{\Sigma}| \neq 0$). Несмещенность оценки (16.30) следует из того, что при фиксированном векторе истинной классификации D^o матрица $(n-L)\hat{\Sigma}$ является суммой L независимых случайных матриц Уишарта:

$$(n-L)\hat{\Sigma} = \sum_{i \in \mathcal{S}} A_i, \quad A_i = \sum_{x_t \in X^{(i)}} (x_t - \hat{\mu}_i)(x_t - \hat{\mu}_i)^T;$$

$$\mathcal{L}\{A_i | D^o\} = W_N(\Sigma, n_i - 1), \quad i \in \mathcal{S},$$

а значит, по свойствам распределения Уишарта

$$\mathcal{L}\{(n-L)\hat{\Sigma} | D^o\} = W_N\left(\Sigma, \sum_{i \in \mathcal{S}} (n_i - 1)\right) = W_N(\Sigma, n-L)$$

и $E\{\hat{\Sigma} | D^o\} = \Sigma$ не зависит от D^o , поэтому $E\{\hat{\Sigma}\} = \Sigma$.

Замечание 16.4. Перед проведением дискриминантного анализа целесообразно проверить гипотезу однородности (см. пп. 16.4.3):

$$\mu_1 = \dots = \mu_L, \quad \Sigma_1 = \dots = \Sigma_L.$$

Если она принимается, то выборка считается однородной, и дискриминантный анализ не проводится. В противном случае проверяется гипотеза о совпадении ковариационных матриц (см. п. 16.4.2):

$$\Sigma_1 = \dots = \Sigma_L.$$

Ее принятие или отклонение позволяет соответственно сделать выбор в пользу линейного (модель Фишера) или квадратичного дискриминантного анализа.

16.6.3. Кластерный анализ. Метод L -средних

Пусть теперь вектор истинной классификации $D^o = (d_1^o, \dots, d_n^o)^T \in \mathcal{S}^n$ для выборки $X = \{x_1, \dots, x_n\}$ неизвестен. Для его оценивания по X в случае моделей (16.13), (16.19) и (16.13), (16.20) можно предложить итерационную процедуру, основанную на многократном повторении соответствующей процедуры дискриминантного анализа. Изложим ее в виде шагов.

Шаг 0. Из выборки $X = \{x_1, \dots, x_n\}$ выбираем какие-либо $L \geq 2$ наблюдений в качестве начальных приближений $\{\hat{\mu}_i^{(0)}\}_{i \in \mathcal{S}}$ для «центров» классов $\{\mu_i\}_{i \in \mathcal{S}}$, а начальные приближения априорных вероятностей классов полагаем равными между собой: $\hat{\pi}_i^{(0)} = 1/L$, $i \in \mathcal{S}$. В качестве начального приближения для ковариационных матриц $\{\Sigma_i\}_{i \in \mathcal{S}}$ (общей ковариационной матрицы Σ в модели Фишера) выбираем единичную матрицу: $\hat{\Sigma}_i^{(0)} = I_N$, $i \in \mathcal{S}$ ($\hat{\Sigma}^{(0)} = I_N$).

Шаг k . Для классификации выборки на k -м шаге ($k = 1, 2, \dots$) используем подстановочное БРП ($t = 1, \dots, n$):

$$\hat{d}_t^{(k)} = \arg \min_{i \in \mathcal{S}} \left((x_t - \hat{\mu}_i^{(k-1)})^T (\hat{\Sigma}_i^{(k-1)})^{-1} (x_t - \hat{\mu}_i^{(k-1)}) + \ln |\hat{\Sigma}_i^{(k-1)}| - 2 \ln \hat{\pi}_i^{(k-1)} \right);$$

$$\left(\hat{d}_t^{(k)} = \arg \min_{i \in \mathcal{S}} \left((x_t - \hat{\mu}_i^{(k-1)})^T (\hat{\Sigma}^{(k-1)})^{-1} (x_t - \hat{\mu}_i^{(k-1)}) - 2 \ln \hat{\pi}_i^{(k-1)} \right) \right)$$

и получаем оценку $\hat{D}^{(k)} = (\hat{d}_1^{(k)}, \dots, \hat{d}_n^{(k)})^T \in \mathcal{S}^n$ вектора истинной классификации $D^o \in \mathcal{S}^n$.

Уточняем оценки для $\{\pi_i, \mu_i\}_{i \in \mathcal{S}}$:

$$\hat{\pi}_i^{(k)} = \frac{n_i^{(k)}}{n}, \quad \hat{\mu}_i^{(k)} = \frac{1}{n_i^{(k)}} \sum_{t=1}^n \delta_{\hat{d}_t^{(k)}, i} x_t, \quad n_i^{(k)} = \sum_{t=1}^n \delta_{\hat{d}_t^{(k)}, i}, \quad i \in \mathcal{S},$$

и ковариационных матриц:

$$\hat{\Sigma}_i^{(k)} = \frac{1}{n_i^{(k)} - 1} \sum_{t=1}^n \delta_{\hat{d}_t^{(k)}, i} (x_t - \hat{\mu}_i^{(k)}) (x_t - \hat{\mu}_i^{(k)})^T, \quad i \in \mathcal{S};$$

$$\left(\hat{\Sigma}^{(k)} = \frac{1}{n - L} \sum_{t=1}^n (x_t - \hat{\mu}_{\hat{d}_t^{(k)}}^{(k)}) (x_t - \hat{\mu}_{\hat{d}_t^{(k)}}^{(k)})^T \right).$$

Шаг-остановка. При $\hat{D}^{(k)} = \hat{D}^{(k-1)}$ ($k \geq 2$) останавливаем итерационный процесс и полагаем: $\hat{\mu}_i := \hat{\mu}_i^{(k)}$, $i \in \mathcal{S}$, – оценки центров классов $\{\mu_i\}_{i \in \mathcal{S}}$; $\hat{\pi}_i := \hat{\pi}_i^{(k)}$, $i \in \mathcal{S}$, – оценки априорных вероятностей $\{\pi_i\}_{i \in \mathcal{S}}$; $\hat{\Sigma}_i := \hat{\Sigma}_i^{(k)}$, $i \in \mathcal{S}$, – оценки ковариационных матриц $\{\Sigma_i\}_{i \in \mathcal{S}}$ ($\hat{\Sigma} := \hat{\Sigma}^{(k)}$ – оценка общей ковариационной матрицы Σ); $\hat{D} := \hat{D}^{(k)}$ – оценка вектора истинной классификации $D^o \in \mathcal{S}^n$.

Однако на практике чаще используют упрощенный вариант приведенной выше процедуры кластер-анализа, основанный на модели Фишера и использовании решающего правила (16.23), в котором в качестве метрики наряду с метрикой Махаланобиса (16.24) может использоваться и любая другая метрика (например, евклидова). Данный подход известен как *метод L -средних*. Опишем его для метрик Махаланобиса и Евклида.

Шаг 0. Из выборки $X = \{x_1, \dots, x_n\}$ выбираем какие-либо $L \geq 2$ наблюдений в качестве начальных приближений $\{\hat{\mu}_i^{(0)}\}_{i \in \mathcal{S}}$ для «центров» классов $\{\mu_i\}_{i \in \mathcal{S}}$. При использовании метрики Махаланобиса в качестве начального приближения для ковариационной матрицы Σ выбираем единичную матрицу: $\hat{\Sigma}^{(0)} = I_N$.

Шаг k ($k = 1, 2, \dots$). Классифицируем наблюдения из выборки по «близости» к «центрам» классов:

$$\hat{d}_t^{(k)} = \arg \min_{i \in \mathcal{S}} \rho(x_t, \hat{\mu}_i^{(k-1)}), \quad t = 1, \dots, n,$$

где либо

$$\rho(x_t, \hat{\mu}_i^{(k-1)}) = \sqrt{(x_t - \hat{\mu}_i^{(k-1)})^T (\hat{\Sigma}^{(k-1)})^{-1} (x_t - \hat{\mu}_i^{(k-1)})} -$$

метрика Махаланобиса, либо

$$\rho(x_t, \hat{\mu}_i^{(k-1)}) = |x_t - \hat{\mu}_i^{(k-1)}| = \sqrt{(x_t - \hat{\mu}_i^{(k-1)})^T (x_t - \hat{\mu}_i^{(k-1)})} -$$

метрика Евклида.

Получаем $\hat{D}^{(k)} = (\hat{d}_1^{(k)}, \dots, \hat{d}_n^{(k)})^T \in \mathbb{S}^n$ – оценка вектора истинной классификации $D^o \in \mathbb{S}^n$ на k -м шаге.

Уточняем оценки «центров» классов:

$$\hat{\mu}_i^{(k)} = \frac{1}{n_i^{(k)}} \sum_{t=1}^n \delta_{\hat{d}_t^{(k)}, i} x_t, \quad n_i^{(k)} = \sum_{t=1}^n \delta_{\hat{d}_t^{(k)}, i}, \quad i \in \mathbb{S},$$

и ковариационной матрицы, если используется метрика Махаланобиса:

$$\hat{\Sigma}^{(k)} = \frac{1}{n-L} \sum_{t=1}^n \left(x_t - \hat{\mu}_{\hat{d}_t^{(k)}}^{(k)} \right) \left(x_t - \hat{\mu}_{\hat{d}_t^{(k)}}^{(k)} \right)^T.$$

Шаг-остановка. При $\hat{D}^{(k)} = \hat{D}^{(k-1)}$ ($k \geq 2$) итерационный процесс останавливаем и полагаем: $\hat{\mu}_i := \hat{\mu}_i^{(k)}$, $i \in \mathbb{S}$, – оценки «центров» классов $\{\mu_i\}_{i \in \mathbb{S}}$; $\hat{\Sigma} := \hat{\Sigma}^{(k)}$ – оценка ковариационной матрицы Σ при использовании метрики Махаланобиса; $\hat{D} := \hat{D}^{(k)}$ – оценка вектора истинной классификации $D^o \in \mathbb{S}^n$.

Замечание 16.5. Эффективность всех методов кластер-анализа существенно зависит от межклассовых расстояний $\rho(\mu_i, \mu_j)$, $i \neq j \in \mathbb{S}$: чем они больше, тем меньше доля ошибочных решений (16.18) и ниже «чувствительность» метода к выбору начальных приближений для «центров» классов и метрики (в методе L -средних).

С методом L -средних связана еще одна характеристика – *псевдо- F -статистика Фишера*:

$$\text{PFS}(L) = \frac{\frac{1}{L-1} \sum_{i \in \mathbb{S}} n_i |\hat{\mu}_i - \bar{x}|^2}{\frac{1}{n-L} \sum_{t=1}^n |x_t - \hat{\mu}_{\hat{d}_t}|^2};$$

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{t=1}^n \delta_{\hat{d}_t, i} x_t, \quad n_i = \sum_{t=1}^n \delta_{\hat{d}_t, i}, \quad i \in \mathbb{S}; \quad \bar{x} = \sum_{i \in \mathbb{S}} \frac{n_i}{n} \hat{\mu}_i = \frac{1}{n} \sum_{t=1}^n x_t,$$

которая при истинном числе классов L и $\hat{D} := D^o$ имеет F -распределение Фишера с $L-1$ и $n-L$ степенями свободы. Но более важно другое ее свойство: она имеет глобальный максимум на истинном числе классов. Это позволяет использовать ее для оценивания неизвестного числа классов L , задавая верхнюю границу L_+ и проводя кластер-анализ при каждом значении $L = 2, \dots, L_+$:

$$\hat{L} = \arg \max_{2 \leq L \leq L_+} \text{PFS}(L).$$

16.6.4. Статистический анализ данных Фишера

Рассмотрим типичные для задач статистической классификации известные данные Фишера по ирисам [4] (Fisher Iris Data, 1936), представляющие собой наблюдения $x = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4)^T$ над четырьмя признаками ($N = 4$): \tilde{x}_1, \tilde{x}_2 – длина и ширина чашелистика, \tilde{x}_3, \tilde{x}_4 – длина и ширина лепестка цветка ириса. Всего 150 наблюдений ($n = 150$), принадлежащих к $L = 3$ классам ($n_1 = n_2 = n_3 = 50$): Ω_1 – ирис цветной (*Iris versicolor*), Ω_2 – ирис махровый (*Iris setosa*), Ω_3 – ирис чистый (*Iris virginica*).

Согласно замечанию 16.4 сначала решаем вопрос о целесообразности проведения классификации, проверяя гипотезу однородности (она отклоняется с уровнем значимости $\alpha = 0,05$). Затем устанавливаем, что ковариационные матрицы по классам совпадают (с уровнем значимости $\alpha = 0,05$), что приводит нас к модели Фишера и линейному дискриминантному анализу, проведя который, вычисляем долю ошибочных решений: $\gamma_n = 3/150 = 0,02$. На рис. 16.1 данные Фишера отображены в пространстве первых

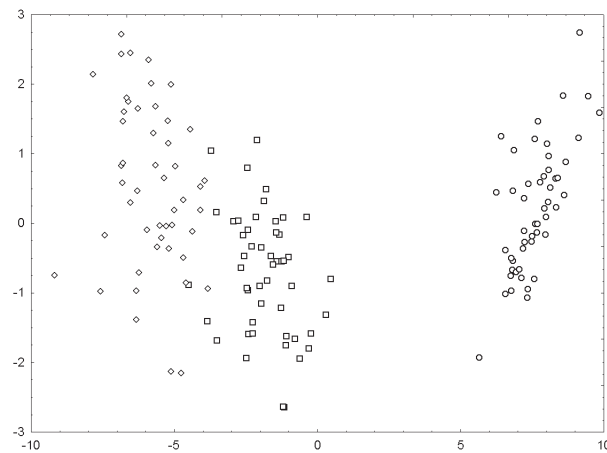


Рис. 16.1. Диаграмма рассеяния данных Фишера в пространстве двух главных компонент: \square – *Iris versicolor*; \circ – *Iris setosa*; \diamond – *Iris virginica*

двух главных компонент, вычисленных на основе построенной при дискриминантном анализе оценки ковариационной матрицы, а в табл. 16.1 приведен фрагмент полученной классификации с вычисленными оценками апостериорных вероятностей классов. Видно, что на ошибочных решениях (помечены «*») доминирующее значение апостериорной вероятности далеко от единицы.

Проведя кластер-анализ методом L -средних в метрике Евклида, имеем долю ошибочных решений: $\gamma_n = 16/150 = 0,11$, которая больше, чем в дискриминантном анализе, но по-прежнему приемлема. Полученные в кластер-анализе оценки «центров» классов показаны на рис. 16.2. Видно, что по каждому признаку они «достаточно» различаются. Отметим также, что для этих данных в предположении неизвестного числа классов псевдо F -статистика Фишера, подсчитанная по результатам метода L -средних при различных предполагаемых значениях числа классов, имеет глобальный максимум на истинном числе классов $L = 3$.

Таблица 16.1

Классификация данных Фишера

Номер наблюдения	Истинная классификация	Апостериорные вероятности		
		Setosa	Versicolor	Virginica
1	Setosa	1,000000	0,000000	0,000000
2	Virginica	0,000000	0,000003	0,999997
3	Versicolor	0,000000	0,995590	0,004410
4	Virginica	0,000000	0,000001	0,999999
* 5	Virginica	0,000000	0,729388	0,270612
6	Setosa	1,000000	0,000000	0,000000
7	Virginica	0,000000	0,000428	0,999572
8	Versicolor	0,000000	0,959573	0,040427
* 9	Versicolor	0,000000	0,253228	0,746772
10	Setosa	1,000000	0,000000	0,000000
11	Versicolor	0,000000	0,998093	0,001907
* 12	Versicolor	0,000000	0,143392	0,856608
13	Virginica	0,000000	0,003146	0,996854
14	Versicolor	0,000000	0,999997	0,000003
15	Virginica	0,000000	0,006084	0,993916
16	Virginica	0,000000	0,001078	0,998922
17	Virginica	0,000000	0,000001	0,999999
18	Setosa	1,000000	0,000000	0,000000
19	Versicolor	0,000000	0,998502	0,001498
20	Virginica	0,000000	0,000013	0,999987
21	Virginica	0,000000	0,000002	0,999998

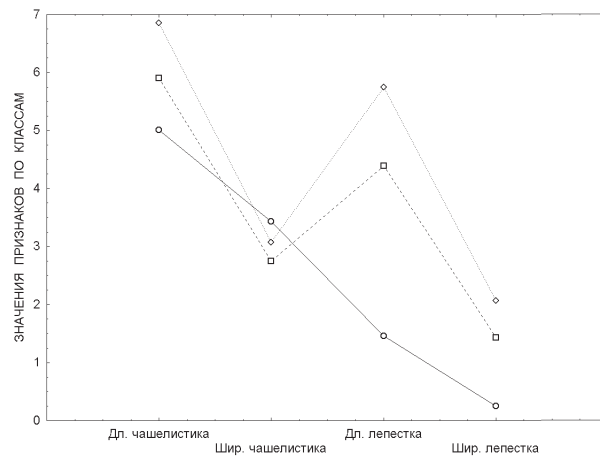


Рис. 16.2. Графики средних для данных Фишера по результатам кластер-анализа: □ — *Iris versicolor*; ○ — *Iris setosa*; ◇ — *Iris virginica*

16.7. ДИСПЕРСИОННЫЙ АНАЛИЗ

К задачам статистической проверки гипотез и классификации легко сводятся задачи так называемого *дисперсионного анализа*. Пусть имеется некоторый скрытый фактор, который может изменяться при переходе от одной серии экспериментов к другой и который неким неизвестным образом влияет на значения наблюдаемого показателя. Проводится $L \geq 2$ серий экспериментов: результат i -й серии – выборка $X^{(i)} = \{x_1^{(i)}, \dots, x_{n_i}^{(i)}\}$ объема n_i ($x_t^{(i)} \in \mathbb{R}$, $t = 1, \dots, n_i$; $i = 1, \dots, L$).

Необходимо решить следующие задачи статистического анализа выборок $X^{(1)}, \dots, X^{(L)}$:

- 1) проверить гипотезы об однородности всех L выборок;
- 2) оценить средние значения $\{\mu_i = E\{x_t^{(i)}\}\}_{i=1}^L$ - показателя в каждой выборке и построить для них доверительные интервалы;
- 3) построить доверительные интервалы для попарных разностей средних значений $\{\mu_i - \mu_j\}_{i \neq j=1}^L$ наблюдаемого показателя.

Задачи такого типа относятся к области *однофакторного дисперсионного анализа*.

Пример 16.2. Пусть имеется возможность выбора одного из L поставщиков некоторого изделия. Тогда измеряемая величина, характеризующая качество изделия, может быть различной у разных поставщиков. И по результатам выборочного обследования: $X^{(i)} - n_i$ изделий поставщика номер « i », $i = 1, \dots, L$, необходимо выяснить, различается ли качество изделий у этих поставщиков, и выбрать оптимального по качеству поставщика.

Сформулируем математическую модель однофакторного дисперсионного анализа. Результат t -го эксперимента в i -й выборке – наблюдение $x_t^{(i)} = \mu_i + \epsilon_{it}$, $t = 1, \dots, n_i$, $i = 1, \dots, L$, где μ_i – неизвестное математическое ожидание (среднее значение) наблюдений в i -й серии, а ϵ_{it} – случайная ошибка наблюдений: $\mathcal{L}\{\epsilon_{it}\} = N_1(0, \sigma^2)$. Эквивалентная запись:

$$x_t^{(i)} = \mu + \alpha_i + \epsilon_{it},$$

где $\mu = \frac{1}{n} \sum_{i=1}^L n_i \mu_i = \sum_{i=1}^L \frac{n_i}{n} \mu_i$ – взвешенное среднее для всех L выборок; $\frac{n_i}{n}$ – доля наблюдений из i -й выборки $\left(n = \sum_{i=1}^L n_i\right)$; $\alpha_i = \mu_i - \mu$ – влияние скрытого фактора в i -й выборке (i -й эффект).

Заметим, что по построению выполняется тождество $\sum_{i=1}^L \frac{n_i}{n} \alpha_i \equiv 0$.

Сформулируем математические задачи однофакторного дисперсионного анализа:

- 1) проверить гипотезу однородности:

$$H_0: \mu_1 = \dots = \mu_L \quad \text{или} \quad H_0: \alpha_1 = \dots = \alpha_L = 0,$$

означающую, что скрытый фактор или одинаков, или не влияет на наблюдаемый показатель, против альтернативы общего вида $H_1 = \overline{H_0}$;

- 2) построить точечные статистические оценки для μ , $\{\alpha_i\}_{i=1}^L$, σ^2 ;
- 3) построить доверительные интервалы для $\{\mu_i\}_{i=1}^L$, $\{\mu_i - \mu_j\}_{i \neq j=1}^L$.

Определим основные статистики:

$$\bar{x}_{(i)} = \frac{1}{n_i} \sum_{t=1}^{n_i} x_t^{(i)} = \hat{\mu}_i - \text{выборочное среднее для } i\text{-й выборки};$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^L \sum_{t=1}^{n_i} x_t^{(i)} = \sum_{i=1}^L \frac{n_i}{n} \bar{x}_{(i)} = \hat{\mu} - \text{общее выборочное среднее};$$

$$\hat{\alpha}_i = \bar{x}_{(i)} - \bar{x} - \text{оценка } i\text{-го эффекта};$$

$$s_{\text{вн. гр}}^2 = \frac{1}{n-L} \sum_{i=1}^L \sum_{t=1}^{n_i} (x_t^{(i)} - \bar{x}_{(i)})^2 = \hat{\sigma}^2 - \text{внутригрупповая дисперсия};$$

$$s_{\text{м. гр}}^2 = \frac{1}{L-1} \sum_{i=1}^L n_i (\bar{x}_{(i)} - \bar{x})^2 - \text{межгрупповая дисперсия};$$

$$F = \frac{s_{\text{м. гр}}^2}{s_{\text{вн. гр}}^2} - F\text{-статистика.}$$

Статистический критерий для проверки гипотез H_0, H_1 имеет вид

$$\begin{cases} H_0 & : F \leq \delta; \\ H_1 = \overline{H_0} & : F > \delta, \end{cases}$$

где порог критерия $\delta = F_{L-1, n-L}^{-1}(1-\epsilon)$ – квантиль уровня $1-\epsilon$ от F -распределения Фишера с $L-1$ и $n-L$ степенями свободы; ϵ – уровень значимости.

Пусть задан доверительный уровень ϵ . Тогда границы доверительного интервала для μ_i имеют вид ($i = 1, \dots, L$):

$$\bar{x}_{(i)} \pm F_{t_{n-L}}^{-1} \left(1 - \frac{\epsilon}{2}\right) \sqrt{\frac{s_{\text{вн. гр}}^2}{n_i}},$$

где $F_{t_{n-L}}^{-1}(1-\epsilon/2)$ – квантиль уровня $1-\epsilon/2$ от t -распределения Стьюдента с $n-L$ степенями свободы.

Границы доверительного интервала для $\mu_i - \mu_j$ при сравнении двух средних μ_i и μ_j имеют вид ($i \neq j = 1, \dots, L$):

$$(\bar{x}_{(i)} - \bar{x}_{(j)}) \pm F_{t_{n-L}}^{-1} \left(1 - \frac{\epsilon}{2}\right) \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right) s_{\text{вн. гр}}^2}.$$

Рассмотренная выше модель однофакторного дисперсионного анализа допускает обобщение по числу факторов. Например, *двухфакторный дисперсионный анализ*, в котором предполагается влияние двух скрытых факторов: r – число эффектов по первому фактору, а c – по второму. Тогда

$$x_t^{(ij)} = \mu_{ij} + e_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijt} -$$

наблюдение номер « t » в (ij) -й серии экспериментов ($i = 1, \dots, r, j = 1, \dots, c$), где e_{ijt} – случайная ошибка: $\mathcal{L}\{e_{ijt}\} = N_1(0, \sigma^2)$;

$$\mu = \frac{1}{r \cdot c} \sum_{i=1}^r \sum_{j=1}^c \mu_{ij}, \quad \alpha_i = \frac{1}{c} \sum_{j=1}^c \mu_{ij} - \mu, \quad \beta_j = \frac{1}{r} \sum_{i=1}^r \mu_{ij} - \mu,$$

а

$$(\alpha\beta)_{ij} = \mu_{ij} + \mu - \left(\frac{1}{r} \sum_{i=1}^r \mu_{ij} + \frac{1}{c} \sum_{j=1}^c \mu_{ij} \right) -$$

величина, описывающая эффект взаимодействия факторов.

16.8. УПРАЖНЕНИЯ

1. Проанализировать поведение многомерных критериев из пп. 16.1–16.4 в случае одномерного пространства наблюдений ($N = 1$) и сравнить их с известными одномерными критериями.
2. Минимизацией риска на классе всех рандомизированных решающих правил получить байесовское решающее правило (16.15) и соотношение для его риска (16.16).
3. По аналогии с однофакторным дисперсионным анализом рассмотреть подробно модель двухфакторного дисперсионного анализа из п. 16.7. Сформулировать и решить основные, связанные с ней статистические задачи [32].
4. Используя распределение \bar{x} и свойства S , доказать, что если $\mu = \mu_0$, то величина T^2 асимптотически распределена как χ^2 с N степенями свободы (при $n \rightarrow +\infty$).
5. Пусть $T^2 = n\bar{x}^T S^{-1} \bar{x}$, где \bar{x} и S определены по выборке объема n из распределения $N_N(\mu, \Sigma)$. Показать, что распределение T^2 не изменится, если заменить μ на $\lambda = (\tau, 0, \dots, 0)^T$, где $\tau^2 = \mu^T \Sigma^{-1} \mu$, и Σ заменить на I_N .
6. Используя данные Фишера при уровне значимости 0,05:
 - а) проверить гипотезу $H_0: \mu_1 = \mu_2 = \mu_3$, предполагая, что $\Sigma_1 = \Sigma_2 = \Sigma_3$;
 - б) проверить гипотезу $H_0: \Sigma_1 = \Sigma_2 = \Sigma_3$;
 - в) проверить гипотезу $H_0: \mu_1 = \mu_2 = \mu_3, \Sigma_1 = \Sigma_2 = \Sigma_3$.
 Данные Фишера. Наблюдается случайный вектор $x \in \mathbb{R}^4$. Каждая выборка состоит из 50 наблюдений и вычислены:

$$\bar{x}_{(1)} = \begin{pmatrix} 5,94 \\ 2,77 \\ 4,26 \\ 1,33 \end{pmatrix}, \bar{x}_{(2)} = \begin{pmatrix} 5,01 \\ 3,43 \\ 1,46 \\ 0,25 \end{pmatrix}, \bar{x}_{(3)} = \begin{pmatrix} 6,65 \\ 2,83 \\ 5,68 \\ 2,07 \end{pmatrix};$$

$$A_1 = \begin{pmatrix} 13,06 & 4,17 & 8,96 & 2,73 \\ 4,17 & 4,83 & 4,05 & 2,02 \\ 8,96 & 4,05 & 10,82 & 3,58 \\ 2,73 & 2,02 & 3,58 & 1,92 \end{pmatrix};$$

$$A_2 = \begin{pmatrix} 6,08 & 4,86 & 0,80 & 0,51 \\ 4,86 & 7,04 & 0,57 & 0,46 \\ 0,80 & 0,57 & 1,48 & 0,30 \\ 0,51 & 0,46 & 0,30 & 0,54 \end{pmatrix}; A_3 = \begin{pmatrix} 19,81 & 4,59 & 14,86 & 2,41 \\ 4,59 & 5,10 & 3,50 & 2,33 \\ 14,86 & 3,50 & 14,92 & 2,39 \\ 2,41 & 2,33 & 2,39 & 3,70 \end{pmatrix}.$$

ЧАСТЬ 5

**СТАТИСТИЧЕСКИЙ АНАЛИЗ
ВРЕМЕННЫХ РЯДОВ**

Глава 17

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВРЕМЕННЫХ РЯДОВ

17.1. СЛУЧАЙНАЯ ФУНКЦИЯ КАК МОДЕЛЬ СТАТИСТИЧЕСКИХ НАБЛЮДЕНИЙ В ДИНАМИКЕ. КЛАССИФИКАЦИЯ СЛУЧАЙНЫХ ФУНКЦИЙ

В математической статистике используются две основные модели для описания наблюдений.

1. Случайная выборка $X = \{x_1, \dots, x_n\}$ объема $n \geq 1$, где $x_t \in \mathbb{R}^N$, $t = 1, \dots, n$, – независимые в совокупности, одинаково распределенные случайные N -векторы-наблюдения с функцией распределения $F(x)$, $x \in \mathbb{R}^N$ (модель, рассмотренная в предыдущих главах).

2. Наблюдения в динамике над одним и тем же объектом, свойства которого изменяются во времени. Адекватной моделью такой ситуации является случайная функция, статистическому анализу которой посвящены последующие главы.

Определение 17.1. Случайной функцией называется параметрическое семейство случайных векторов:

$$x = x(t) = x(\omega, t) \in \mathbb{R}^N,$$

определенных на одном и том же вероятностном пространстве (Ω, \mathcal{F}, P) , $\omega \in \Omega$, где $t \in \mathcal{T} \subseteq \mathbb{R}^m$ – параметр, изменяющийся на множестве \mathcal{T} .

В зависимости от значений N и m размерностей пространства наблюдений и пространства значений параметра принята приведенная в табл. 17.1 классификация случайных функций (в скобках приводится пример).

Таблица 17.1

Классификация случайных функций

Размерность пространства наблюдений, N	Размерность пространства значений параметра, m	
	$m = 1$	$m > 1$
$N = 1$	Случайный процесс (цена акций [6], рис. 17.1; показатель солнечной активности [5], рис. 17.2)	Случайное поле (экран монитора в черно-белом режиме, $m = 2$)
$N > 1$	Векторный случайный процесс (несколько отведений электрокардиограммы, обычно $N = 8$)	Векторное случайное поле (радиационная карта местности с несколькими радиоизотопами, $m = 2$)

Если множество \mathcal{T} значений параметра t дискретно (конечно или счетно), то такой случайный процесс обычно называют *временным рядом* или *случайной последовательностью*, а одномерный параметр понимают как *время*. Векторный случайный процесс при этом называют *векторным временным рядом*.

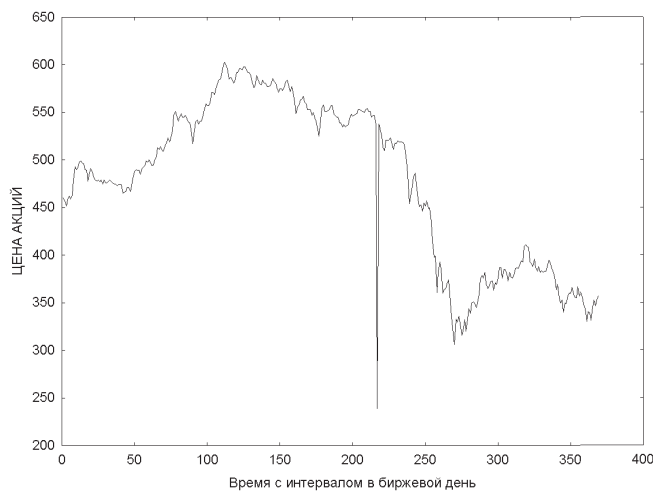


Рис. 17.1. Цена на акции IBM к закрытию биржи, ежедневно с 17 мая 1961 по 2 ноября 1962 г.

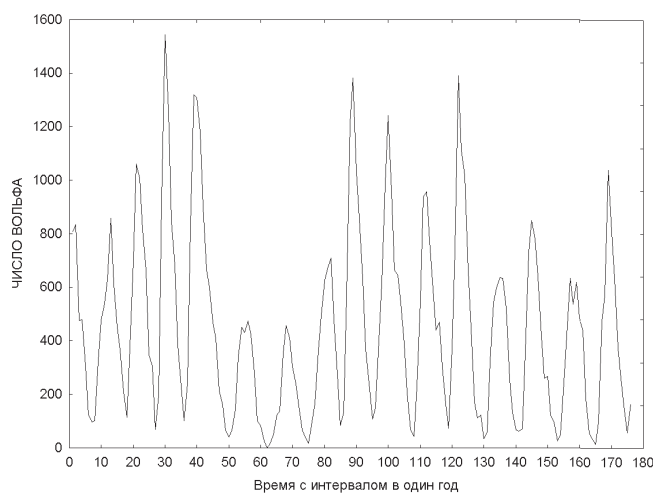


Рис. 17.2. Данные солнечной активности (число Вольфа) по ежегодным измерениям с 1749 по 1924 г.

17.2. ВРЕМЕННЫЕ РЯДЫ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

Дадим определение временного ряда и связанных с ним характеристик.

Определение 17.2. Под *временным рядом (ВР)* понимаются упорядоченные во времени статистические наблюдения над одним и тем же объектом в динамике. Другими словами, ВР – это случайная функция

$$x = x(t) = x(\omega, t) \in \mathbb{R},$$

где $t \in \mathcal{T} \subseteq \mathbb{R}$ – называется *временем*.

Определение 17.3. Отсчетом ВР называется случайное значение $x(t)$ этого ВР в некоторый фиксированный момент времени $t \in \mathcal{T}$.

Определение 17.4. Реализация ВР – это совокупность всех отсчетов при фиксированном эксперименте $\omega \in \Omega$:

$$x = x(t) = x(\omega, t), \quad t \in \mathcal{T}, \quad \omega = \text{const.}$$

Определение 17.5. Совокупность всех реализаций ВР называется ансамблем реализаций:

$$\{x = x(t) = x(\omega, t), \quad t \in \mathcal{T}\}_{\omega \in \Omega}.$$

Основные характеристики временного ряда

1. Совместное распределение вероятностей отсчетов.

Пусть выбраны $n \geq 1$ отсчетов ВР $x(t_1), \dots, x(t_n)$ в моменты времени $t_1 \leq t_2 \leq \dots \leq t_n$, тогда под совместным распределением вероятностей ВР понимается совместное распределение вероятностей отсчетов как случайных величин, однозначно определяемое совместной функцией распределения $(y_1, \dots, y_n \in \mathbb{R})$:

$$F_n(y_1, \dots, y_n; t_1, \dots, t_n) = P\{x(t_1) < y_1, \dots, x(t_n) < y_n\}. \quad (17.1)$$

Для того чтобы описать ВР, достаточно определить все возможные совместные распределения вероятностей отсчетов.

Если совместное распределение вероятностей отсчетов ВР абсолютно непрерывно, то вместо (17.1) можно использовать совместную плотность распределения: $p_n(y_1, \dots, y_n; t_1, \dots, t_n)$.

2. Математическое ожидание ВР:

$$m(t) = E\{x(t)\}, \quad t \in \mathcal{T}. \quad (17.2)$$

3. Дисперсия ВР:

$$D(t) = D\{x(t)\} = E\{(x(t) - m(t))^2\}, \quad t \in \mathcal{T}. \quad (17.3)$$

4. *Ковариационная функция ВР*, под которой понимается ковариация двух любых его отсчетов:

$$\begin{aligned} \sigma(t_1, t_2) &= \text{cov}\{x(t_1), x(t_2)\} = \\ &= E\{(x(t_1) - m(t_1))(x(t_2) - m(t_2))\}, \quad t_1, t_2 \in \mathcal{T}. \end{aligned} \quad (17.4)$$

5. Корреляционная функция ВР (корреляция отсчетов):

$$\rho(t_1, t_2) = \text{corr}\{x(t_1), x(t_2)\} = \frac{\sigma(t_1, t_2)}{\sqrt{D(t_1)D(t_2)}}.$$

Свойства ковариационной и корреляционной функций

C1. $|\sigma(t_1, t_2)| \leq \sqrt{D(t_1)D(t_2)}.$

C2. $|\rho(t_1, t_2)| \leq 1.$

C3. Ковариационная функция является неотрицательно определенной функцией, т. е. для любого $n \geq 1$, для любых $y_1, \dots, y_n \in \mathbb{R}$ и для любых моментов времени

$$t_1, \dots, t_n \in \mathcal{T} \text{ выполняется } \sum_{i,j=1}^n \sigma(t_i, t_j) y_i y_j \geq 0.$$

Доказательство. Рассмотрим ковариационную $(n \times n)$ -матрицу $\Sigma = (\sigma(t_i, t_j))_{i,j=1}^n$, образованную ковариациями (17.4). Поскольку любая ковариационная матрица неотрицательно определена, то $\sum_{i,j=1}^n \sigma(t_i, t_j) y_i y_j = y^T \Sigma y \geq 0$, где $y = (y_1, \dots, y_n)^T$ – произвольная точка в \mathbb{R}^n . \square

17.3. СТАЦИОНАРНЫЕ ВРЕМЕННЫЕ РЯДЫ. СПЕКТРАЛЬНАЯ ПЛОТНОСТЬ СТАЦИОНАРНОГО В ШИРОКОМ СМЫСЛЕ ВРЕМЕННОГО РЯДА

В теории временных рядов важным является понятие стационарности.

Определение 17.6. ВР $x = x(t)$, $t \in \mathcal{T}$, называется стационарным в узком смысле, если для любых $n \geq 1$ его отсчетов совместная функция распределения вероятностей (17.1) этих отсчетов не зависит от сдвига во времени, т. е. для любого $n \geq 1$, для любых $t_1 \leq t_2 \leq \dots \leq t_n \in \mathcal{T}$ и для любого $\tau \in \mathcal{T}$:

$$F_n(y_1, \dots, y_n; t_1, \dots, t_n) = F_n(y_1, \dots, y_n; t_1 + \tau, \dots, t_n + \tau). \quad (17.5)$$

Определение 17.7. ВР $x = x(t)$, $t \in T$, называется стационарным в широком смысле, если выполняются следующие условия:

1) отсчеты ВР как случайные величины имеют первый и второй моменты:

$$|E\{x(t)\}| < +\infty, \quad E\{(x(t))^2\} < +\infty;$$

2) математическое ожидание ВР не зависит от времени:

$$m(t) = \mu, \quad \mu \in \mathbb{R}, \quad \forall t \in \mathcal{T};$$

3) для ковариационной функции выполняется

$$\sigma(t_1, t_2) = \sigma(t_1 + \tau, t_2 + \tau), \quad \forall t_1, t_2 \in T, \quad \forall \tau \in \mathcal{T}.$$

Установим связь между стационарностью в узком и широком смысле.

Теорема 17.1. Пусть ВР $x = x(t)$, $t \in \mathcal{T}$, является стационарным в узком смысле и пусть существуют первый и второй моменты отсчетов этого ВР, тогда данный ВР является стационарным в широком смысле.

Доказательство. Воспользуемся свойством (17.5) стационарности в узком смысле и запишем математическое ожидание ВР (17.2):

$$m(t) = \int_{\mathbb{R}} y dF_1(y; t) = \int_{\mathbb{R}} y dF_1(y; 0) = \mu \in \mathbb{R},$$

что совпадает со вторым условием стационарности в широком смысле. Аналогично рас-

сматривается двухмерное распределение вероятностей для получения третьего условия стационарностей в широком смысле. \square

Отметим, что, как видно из теоремы 17.1, при определенных условиях из стационарности в узком смысле следует и стационарность в широком смысле. Однако из стационарности в широком смысле, вообще говоря, не следует стационарность в узком смысле.

На практике известен следующий важный класс ВР – гауссовские (нормальные) ВР, для которых понятия стационарности в узком и широком смысле эквивалентны.

Определение 17.8. ВР $x = x(t)$, $t \in \mathcal{T}$, называется гауссовским, если любые $n \geq 1$ его отсчетов имеют совместное нормальное распределение, т. е. $\forall n \geq 1, \forall t_1 \leq t_2 \leq \dots \leq t_n \in \mathcal{T}$:

$$p_n(y_1, \dots, y_n; t_1, \dots, t_n) = n_n(y_1, \dots, y_n | M, \Sigma),$$

где $n_n(y_1, \dots, y_n | M, \Sigma)$ – плотность n -мерного нормального распределения с математическим ожиданием $M = (m(t_1), \dots, m(t_n))^T \in \mathbb{R}^n$ и ковариационной $(n \times n)$ -матрицей $\Sigma = (\sigma(t_i, t_j))_{i,j=1}^n$, вычисленная в точке $y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$.

Замечание 17.1. Третье условие стационарности в широком смысле можно записать в другом виде, а именно: $\sigma(t, t + \tau) = \sigma(\tau)$, $\forall t, \tau \in \mathcal{T}$, либо $\sigma(t_1, t_2) = \sigma(t_2 - t_1)$, $\forall t_1, t_2 \in \mathcal{T}$.

Доказательство. Положив сдвиг $\tau = -t_1$, получаем $\sigma(t_1, t_2) = \sigma(t_1 + \tau, t_2 + \tau) = \sigma(0, t_2 - t_1)$. \square

Таким образом, ковариационная функция для стационарного в широком смысле ВР зависит лишь от одного аргумента – расстояния между отсчетами по времени, называемого *лагом*, и для нее легко доказать следующие свойства.

C1. Ковариационная функция стационарного в широком смысле ВР является четной (симметричной) функцией: $\sigma(\tau) = \sigma(-\tau)$, $\tau \in \mathcal{T}$.

C2. $|\sigma(\tau)| \leq \sigma(0)$, $\tau \in \mathcal{T}$, где $\sigma(0) = D(0)$ – дисперсия ВР, определенная в (17.3).

Со стационарностью в широком смысле связана еще одна характеристика – спектральная плотность ВР.

Пусть имеется стационарный в широком смысле ВР $x = x(t)$, $t \in \mathcal{T}$. Зафиксируем моменты времени, в которые производятся наблюдения: $\dots, t_{-1}, t_0, t_1, t_2, \dots \in \mathcal{T}$. Будем предполагать, что наблюдения проводятся по времени с одинаковым шагом $\delta = t_{j+1} - t_j$, $j \in \mathbb{Z}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, тогда промежуток времени δ для нас не важен, и время можно заменить последовательностью целых чисел $\dots, -1, 0, 1, 2, \dots$ ($t \in \mathbb{Z}$). Исходный ВР $x = x(t)$, $t \in \mathcal{T}$, можно в этом случае записать как $\{x_t = x(t)\}_{t=-\infty}^{+\infty}$.

Определение 17.9. Спектральной плотностью $f(\lambda)$, $\lambda \in [-\pi, \pi]$, стационарного в широком смысле ВР $\{x_t\}_{t=-\infty}^{+\infty}$, имеющего ковариационную функцию $\sigma(\tau)$, $\tau \in \mathbb{Z}$, называется прямое преобразование Фурье от ковариационной функции ($\lambda \in [-\pi, \pi]$):

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} \exp(i\lambda\tau) \sigma(\tau) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} \cos(\lambda\tau) \sigma(\tau), \quad (17.6)$$

при условии, что оно существует (i – мнимая единица).

Отметим, что спектральная плотность в виде (17.6) является явным выражением спектральной плотности, определенной в п. 7.7.

Для существования спектральной плотности необходимо потребовать выполнения определенных дополнительных условий. Получим эти условия. Оценим спектральную

плотность (17.6) сверху: $f(\lambda) = |f(\lambda)| \leq \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} |\sigma(\tau)|$, $\lambda \in [-\pi, \pi]$.

Следовательно, для существования спектральной плотности достаточно выполнения условия

$$\sum_{\tau=-\infty}^{+\infty} |\sigma(\tau)| < +\infty. \quad (17.7)$$

Если сходится ряд (17.7), то спектральная плотность $f(\lambda)$ из (17.6) существует, так как в этом случае она является равномерно сходящимся рядом по $\lambda \in [-\pi, \pi]$.

Заметим, что поскольку ковариационная функция является симметрической функцией, то при $\sigma(0) = D(0) < +\infty$ условие (17.7) эквивалентно условию $\sum_{\tau=1}^{+\infty} |\sigma(\tau)| < +\infty$.

Используя аппарат математического анализа, можно доказать следующие свойства спектральной плотности.

С1. $f(\lambda) \geq 0$, $\lambda \in [-\pi, \pi]$.

С2. $f(\lambda) = f(-\lambda)$, $\lambda \in [-\pi, \pi]$.

Установим связь между значениями ковариационной функции и спектральной плотности. Очевидно, что ковариационная функция может быть однозначно восстановлена по спектральной плотности обратным преобразованием Фурье.

Теорема 17.2. Пусть стационарный в широком смысле ВР имеет спектральную плотность $f(\lambda)$, $\lambda \in [-\pi, \pi]$, тогда его ковариационная функция $\sigma(\tau)$, $\tau \in \mathbb{Z}$, однозначно восстанавливается по $f(\cdot)$ обратным преобразованием Фурье:

$$\sigma(\tau) = \int_{-\pi}^{\pi} \exp(-i\lambda\tau) f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(\lambda\tau) f(\lambda) d\lambda, \quad \tau \in \mathbb{Z}.$$

Доказательство. Очевидно и следует из свойств обратного преобразования Фурье и симметричности спектральной плотности. \square

Замечание 17.2. Как следует из теоремы 17.2, если спектральная плотность существует, то она находится во взаимно-однозначном соответствии с ковариационной функцией, что говорит об избыточности спектральной плотности в теоретическом плане. Более того, как будет показано в следующей главе, ее статистические оценки проигрывают по свойствам и простоте построения оценкам ковариационной функции.

На практике спектральная плотность используется в основном для определения периода: $T^* = \frac{2\pi}{\lambda^*}$, $\lambda^* = \arg \max_{\lambda \in [0, \pi]} f(\lambda)$.

Величина T^* еще называется главным периодом и в некоторых случаях достаточно адекватно характеризует промежуток времени, через который свойства временного ряда «статистически повторяются».

Замечание 17.3. Случай одного ВР легко обобщается на случай нескольких ВР. Подходящей математической моделью такой ситуации является векторный ВР $x = x(t) = (\tilde{x}_1(t), \dots, \tilde{x}_N(t))^T \in \mathbb{R}^N$, $t \in \mathcal{T}$, где $N \geq 1$ – число компонент векторного ВР: $\tilde{x}_j = \tilde{x}_j(t)$, $j = 1, \dots, N$, – j -я компонента (одномерный ВР).

Для векторного ВР по аналогии с одномерным ВР можно определить математическое ожидание: $m(t) = E\{x(t)\} \in \mathbb{R}^N$, $t \in \mathcal{T}$, и матричную ковариационную функцию: $\Sigma = \Sigma(t_1, t_2) = \text{cov}\{x(t_1), x(t_2)\} = E\{(x(t_1) - m(t_1))(x(t_2) - m(t_2))^T\}$, $t_1, t_2 \in \mathcal{T}$, являющиеся соответственно N -вектором и $(N \times N)$ -матрицей, а также ввести понятие стационарности в широком смысле (аналогично определению 17.7) и определить матричную спектральную плотность (по аналогии с (17.6)).

Глава 18

НЕПАРАМЕТРИЧЕСКОЕ СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ХАРАКТЕРИСТИК СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

18.1. ВЫБОРОЧНОЕ СРЕДНЕЕ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА И ЕГО СВОЙСТВА

Пусть для наблюдений в динамике в качестве математической модели используется стационарный в широком смысле ВР $\{x_t\}_{t=-\infty}^{+\infty}$. Для данного ВР определены характеристики:

а) математическое ожидание

$$\mu = E\{x_t\}, \quad t \in \mathbb{Z};$$

б) ковариационная функция

$$\sigma(\tau) = E\{(x_t - \mu)(x_{t+\tau} - \mu)\}, \quad t, \tau \in \mathbb{Z};$$

в) спектральная плотность

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} \cos(\lambda\tau) \sigma(\tau), \quad \lambda \in [-\pi, \pi],$$

которая существует, если

$$\sum_{\tau=1}^{+\infty} |\sigma(\tau)| < +\infty. \quad (18.1)$$

На практике обычно математическое ожидание и ковариационная функция неизвестны, а для ВР $\{x_t\}_{t=-\infty}^{+\infty}$ наблюдается его реализация $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$, по которой необходимо оценить математическое ожидание, ковариационную функцию и спектральную плотность.

Оценим сначала математическое ожидание.

Определение 18.1. Выборочным средним для ВР $\{x_t\}_{t=-\infty}^{+\infty}$, вычисленным по реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$, называется статистика

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t. \quad (18.2)$$

Теорема 18.1. Пусть $\{x_t\}_{t=-\infty}^{+\infty}$ – стационарный ВР с математическим ожиданием $\mu = E\{x_t\}$, $t \in \mathbb{Z}$, тогда выборочное среднее (18.2) является несмещенной оценкой для μ :

$$E\{\bar{x}\} = \mu.$$

Если ВР $\{x_t\}_{t=-\infty}^{+\infty}$ является стационарным в широком смысле, то для дисперсии выборочного среднего (18.2) справедливо соотношение:

$$D\{\bar{x}\} = \frac{1}{T} \sum_{\tau=-(T-1)}^{T-1} \left(1 - \frac{|\tau|}{T}\right) \sigma(\tau), \quad (18.3)$$

где $\sigma(\tau)$ – ковариационная функция.

Если вдобавок к стационарности в широком смысле сходится ряд (18.1), то выборочное среднее (18.2) является состоятельной в среднеквадратическом оценке для μ :

$$\mathbf{E}\{(\bar{x} - \mu)^2\} \rightarrow 0, \quad T \rightarrow +\infty,$$

и для его дисперсии справедливо асимптотическое соотношение:

$$\lim_{T \rightarrow +\infty} (T \mathbf{D}\{\bar{x}\}) = \sum_{\tau=-\infty}^{+\infty} \sigma(\tau). \quad (18.4)$$

Доказательство. Докажем несмещенность оценки (18.2):

$$\mathbf{E}\{\bar{x}\} = \mathbf{E}\left\{\frac{1}{T} \sum_{t=1}^T x_t\right\} = \frac{1}{T} T \mu = \mu.$$

Для стационарного в широком смысле ВР существуют моменты первого и второго порядков. С учетом доказанной несмещенности вычислим дисперсию \bar{x} :

$$\begin{aligned} \mathbf{D}\{\bar{x}\} &= \mathbf{E}\{(\bar{x} - \mu)^2\} = \mathbf{E}\left\{\frac{1}{T} \sum_{t_1=1}^T (x_{t_1} - \mu) \frac{1}{T} \sum_{t_2=1}^T (x_{t_2} - \mu)\right\} = \\ &= \frac{1}{T^2} \sum_{t_1, t_2=1}^T \mathbf{E}\{(x_{t_1} - \mu)(x_{t_2} - \mu)\} = \frac{1}{T^2} \sum_{t_1, t_2=1}^T \sigma(t_1 - t_2) = \\ &= \frac{1}{T^2} \left(T \sigma(0) + 2 \sum_{\tau=1}^{T-1} \sigma(\tau)(T - \tau) \right). \end{aligned}$$

Учтено, что значение ковариационной функции $\sigma(t_1 - t_2) = \sigma(t_2 - t_1)$, $t_1 \neq t_2$, в ковариационной $(T \times T)$ -матрице $\Sigma = (\sigma(t_1 - t_2))_{t_1, t_2=1}^T$ встречается $2(T - |t_1 - t_2|)$ раз, а значение $\sigma(0)$ (при $t_1 = t_2$) – T раз. Далее, воспользовавшись симметричностью ковариационной функции, получаем соотношение (18.3).

Пусть теперь ряд (18.1) сходится, тогда в силу несмещенности \bar{x} для состоятельности в среднеквадратическом достаточно показать, что дисперсия $\mathbf{D}\{\bar{x}\} \rightarrow 0$, $T \rightarrow +\infty$. Оценим $\mathbf{D}\{\bar{x}\}$ сверху:

$$\begin{aligned} \mathbf{D}\{\bar{x}\} &= |\mathbf{D}\{\bar{x}\}| \leq \frac{1}{T} \sum_{\tau=-(T-1)}^{T-1} \left| 1 - \frac{|\tau|}{T} \right| |\sigma(\tau)| \leq \\ &\leq \frac{1}{T} \sum_{\tau=-(T-1)}^{T-1} |\sigma(\tau)| \rightarrow 0, \quad T \rightarrow +\infty. \end{aligned}$$

Для доказательства (18.4) оценим разность:

$$\begin{aligned} &\left| T \mathbf{D}\{\bar{x}\} - \sum_{\tau=-\infty}^{+\infty} \sigma(\tau) \right| = \\ &= \left| \sum_{\tau=-\infty}^{+\infty} \sigma(\tau) - \sum_{\tau=-(T-1)}^{T-1} \left(1 - \frac{|\tau|}{T} \right) \sigma(\tau) \right| = \end{aligned}$$

$$\begin{aligned}
&= \frac{2}{T} \left| T \sum_{\tau \geq T} \sigma(\tau) + \sum_{\tau=1}^{T-1} \tau \sigma(\tau) \right| = \frac{2}{T} \left| \sum_{\tau \geq 1} \sigma(\tau) + \sum_{s=1}^{T-1} \sum_{\tau=s+1}^{+\infty} \sigma(\tau) \right| \leq \\
&\leq \frac{2}{T} \left(\sum_{\tau \geq 1} |\sigma(\tau)| + \sum_{s=1}^{T-1} \sum_{\tau=s+1}^{+\infty} |\sigma(\tau)| \right) \rightarrow 0, T \rightarrow +\infty.
\end{aligned}$$

Справедливость последнего соотношения следует из сходимости ряда (18.1), в силу которой можно записать:

$$\exists M < +\infty : \sum_{\tau=l}^{+\infty} |\sigma(\tau)| < M, \quad \forall l \geq 1;$$

$$\forall \epsilon > 0, \exists T_\epsilon : \sum_{\tau=T}^{+\infty} |\sigma(\tau)| < \frac{\epsilon}{4}, \quad T > T_\epsilon,$$

откуда $\forall T > T_\epsilon$:

$$\begin{aligned}
&\frac{2}{T} \left(\sum_{\tau \geq 1} |\sigma(\tau)| + \sum_{s=1}^{T-1} \sum_{\tau=s+1}^{+\infty} |\sigma(\tau)| \right) = \\
&= \frac{2}{T} \left(\sum_{\tau \geq 1} |\sigma(\tau)| + \sum_{s=1}^{T_\epsilon-1} \sum_{\tau=s+1}^{+\infty} |\sigma(\tau)| + \sum_{s=T_\epsilon}^{T-1} \sum_{\tau=s+1}^{+\infty} |\sigma(\tau)| \right) < \\
&< 2 \left(\frac{T_\epsilon M}{T} + \frac{\epsilon}{4} \right),
\end{aligned}$$

и при $T > \tilde{T}_\epsilon$, $\tilde{T}_\epsilon = \max\{T_\epsilon, 4T_\epsilon M/\epsilon\}$:

$$\frac{2}{T} \left(\sum_{\tau \geq 1} |\sigma(\tau)| + \sum_{s=1}^{T-1} \sum_{\tau=s+1}^{+\infty} |\sigma(\tau)| \right) < \epsilon.$$

□

Следствие 18.1. В условиях теоремы 18.1, если выполняется (18.1), то выборочное среднее (18.2) является состоятельной оценкой для математического ожидания μ :

$$\bar{x} \xrightarrow{P} \mu, \quad T \rightarrow +\infty.$$

18.2. ВЫБОРОЧНАЯ КОВАРИАЦИОННАЯ ФУНКЦИЯ СТАЦИОНАРНОГО В ШИРОКОМ СМЫСЛЕ ВРЕМЕННОГО РЯДА И ЕЕ СВОЙСТВА

Рассмотрим стационарный в широком смысле временной ряд $\{x_t\}_{t=-\infty}^{+\infty}$. Для данного ВР существуют математическое ожидание $\mu = E\{x_t\}$, $t \in \mathbb{Z}$, и ковариационная функция $\sigma(\tau) = E\{(x_t - \mu)(x_{t+\tau} - \mu)\}$, $t, \tau \in \mathbb{Z}$.

Определение 18.2. Выборочной ковариационной функцией для ВР $\{x_t\}_{t=-\infty}^{+\infty}$, вычисленной по реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$, в случае известного математического ожидания μ называется следующая статистика ($\tau \in \{0, 1, \dots, T-1\}$):

$$c_\tau = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (x_t - \mu)(x_{t+\tau} - \mu), \quad c_{-\tau} = c_\tau. \quad (18.5)$$

Определение 18.3. Выборочной ковариационной функцией для ВР $\{x_t\}_{t=-\infty}^{+\infty}$ по реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$ при неизвестном математическом ожидании μ называется статистика ($\tau \in \{0, 1, \dots, T-1\}$)

$$c_\tau^* = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (x_t - \bar{x})(x_{t+\tau} - \bar{x}), \quad c_{-\tau}^* = c_\tau^*. \quad (18.6)$$

Исследуем свойства выборочной ковариационной функции (18.5). Это можно сделать и в общем случае для всех стационарных в широком смысле ВР, однако достаточно простыми аналитическими характеристиками она обладает только для гауссовских ВР.

Теорема 18.2. Пусть $\{x_t\}_{t=-\infty}^{+\infty}$ – стационарный гауссовский ВР с математическим ожиданием $\mu = \mathbf{E}\{x_t\}$ и ковариационной функцией $\sigma(\tau) = \mathbf{E}\{(x_t - \mu)(x_{t+\tau} - \mu)\}$ ($t, \tau \in \mathbb{Z}$), тогда выборочная ковариационная функция c_τ из (18.5) является несмещенной оценкой для $\sigma(\tau)$:

$$\mathbf{E}\{c_\tau\} = \sigma(\tau), \quad \tau \in \{0, 1, \dots, T-1\},$$

и справедливы следующие соотношения ($h, g \in \{0, 1, \dots, T-1\}$):

$$\begin{aligned} \text{cov}\{c_h, c_g\} &= \frac{1}{T-h} \frac{1}{T-g} \sum_{t_1=1}^{T-h} \sum_{t_2=1}^{T-g} (\sigma(t_1 - t_2) \sigma(t_1 - t_2 + h - g) + \\ &\quad + \sigma(t_1 - t_2 + h) \sigma(t_1 - t_2 - g)); \end{aligned} \quad (18.7)$$

$$\begin{aligned} D\{c_h\} &= \frac{1}{T-h} \sum_{\tau=-(T-h-1)}^{T-h-1} \left(1 - \frac{|\tau|}{T-h}\right) \times \\ &\quad \times (\sigma^2(\tau) + \sigma(\tau-h)\sigma(\tau+h)). \end{aligned} \quad (18.8)$$

Если сходится ряд (18.1), то выборочная ковариационная функция (18.5) является состоятельной в среднеквадратическом оценкой для $\sigma(\cdot)$, и для ее дисперсии справедливо асимптотическое соотношение:

$$\lim_{T \rightarrow +\infty} (T D\{c_h\}) = \sum_{\tau=-\infty}^{+\infty} (\sigma^2(\tau) + \sigma(\tau-h)\sigma(\tau+h)). \quad (18.9)$$

Доказательство. Докажем несмещенность ($\tau \in \{0, 1, \dots, T-1\}$):

$$\begin{aligned} \mathbf{E}\{c_\tau\} &= \mathbf{E} \left\{ \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (x_t - \mu)(x_{t+\tau} - \mu) \right\} = \\ &= \frac{1}{T-\tau} (T-\tau) \text{cov}\{x_t, x_{t+\tau}\} = \sigma(\tau). \end{aligned}$$

Вычислим ковариацию c_h и c_g ($h, g \in \{0, 1, \dots, T-1\}$):

$$\begin{aligned} \text{cov}\{c_h, c_g\} &= \mathbf{E}\{c_h c_g\} - \mathbf{E}\{c_h\} \mathbf{E}\{c_g\} = \\ &= \mathbf{E} \left\{ \frac{1}{T-h} \sum_{t_1=1}^{T-h} (x_{t_1} - \mu)(x_{t_1+h} - \mu) \frac{1}{T-g} \sum_{t_2=1}^{T-g} (x_{t_2} - \mu)(x_{t_2+g} - \mu) \right\} - \sigma(h)\sigma(g) = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T-h} \frac{1}{T-g} \sum_{t_1=1}^{T-h} \sum_{t_2=1}^{T-g} (E\{(x_{t_1} - \mu)(x_{t_1+h} - \mu)(x_{t_2} - \mu)(x_{t_2+g} - \mu)\} - \sigma(h)\sigma(g)) = \\
&= \frac{1}{T-h} \frac{1}{T-g} \sum_{t_1=1}^{T-h} \sum_{t_2=1}^{T-g} (\sigma(t_1-t_2)\sigma(t_1-t_2+h-g) + \sigma(t_1-t_2+h)\sigma(t_1-t_2-g)).
\end{aligned}$$

Выше использовано следующее известное свойство: пусть случайные величины ξ_1 , ξ_2 , ξ_3 и ξ_4 имеют совместное нормальное распределение, тогда

$$\begin{aligned}
E\{(\xi_1 - E\{\xi_1\})(\xi_2 - E\{\xi_2\})(\xi_3 - E\{\xi_3\})(\xi_4 - E\{\xi_4\})\} = \\
= \text{cov}\{\xi_1, \xi_2\}\text{cov}\{\xi_3, \xi_4\} + \text{cov}\{\xi_1, \xi_3\}\text{cov}\{\xi_2, \xi_4\} + \\
+ \text{cov}\{\xi_1, \xi_4\}\text{cov}\{\xi_2, \xi_3\}.
\end{aligned}$$

Положив в (18.7) $g = h$, получаем (18.8).

Пусть теперь ряд (18.1) сходится, тогда, поскольку оценка несмещенная, для состоятельности в среднеквадратическом достаточно показать, что дисперсия сходится к нулю.

Заметим, что сходимость ряда (18.1) влечет сходимость ряда

$$\sum_{\tau=1}^{+\infty} (\sigma(\tau))^2 < +\infty.$$

Используя неравенство Коши – Шварца – Буняковского $\left| \sum_{i=1}^N a_i b_i \right| \leq \left(\sum_{i=1}^N a_i^2 \sum_{j=1}^N b_j^2 \right)^{\frac{1}{2}}$, оценим дисперсию сверху:

$$D\{c_h\} \leq \frac{1}{T-h} b \sum_{\tau=1}^{+\infty} (\sigma(\tau))^2 \rightarrow 0, \quad T \rightarrow +\infty, \quad b = \text{const}.$$

Доказательство соотношения (18.9) аналогично доказательству соотношения (18.4) из теоремы 18.1. \square

Замечание 18.1. Из результатов теоремы (выражение (18.8)) видно, что точность оценивания ковариационной функции $\sigma(h)$ «падает» с увеличением лага h . Это связано с уменьшением числа слагаемых в (18.5). Хуже всего, если значение h приближается к $T-1$.

Замечание 18.2. Для оценки (18.6) можно получить похожие результаты. В этом случае c_h^* , $h \in \{0, 1, \dots, T-1\}$, будет асимптотически несмещенной и состоятельной.

Замечание 18.3. Для состоятельности в среднеквадратическом выборочных среднего и ковариационной функции требовалась сходимость ряда (18.6), необходимым (но не достаточным!) условием которой является:

$$\sigma(\tau) \rightarrow 0, \quad \tau \rightarrow +\infty.$$

Иными словами, ковариационная функция $\sigma(\tau)$ (корреляционная функция $\rho(\tau) = \sigma(\tau)/\sigma(0)$) «затухает» с ростом лага τ . Это условие используют на практике как эмпирическое правило для выявления временных рядов, подозрительных на стационарность. Для таких ВР выборочная ковариационная (корреляционная) функция должна «затухать».

На рис. 18.1 и 18.2 приведены выборочные корреляционные функции для ВР, изображенных соответственно на рис. 17.1 и 17.2. Видно, что данные солнечной активности могут оказаться стационарными, чего нельзя сказать о цене акций IBM.

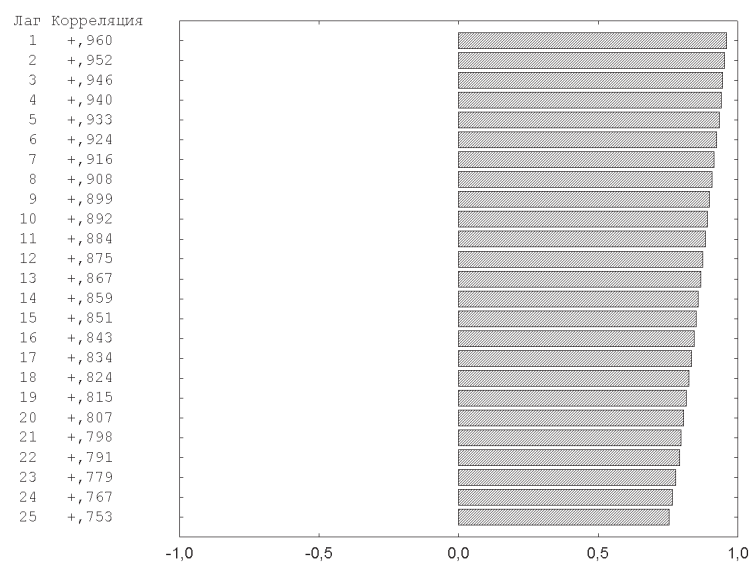


Рис. 18.1. Выборочная корреляционная функция для цены акций IBM

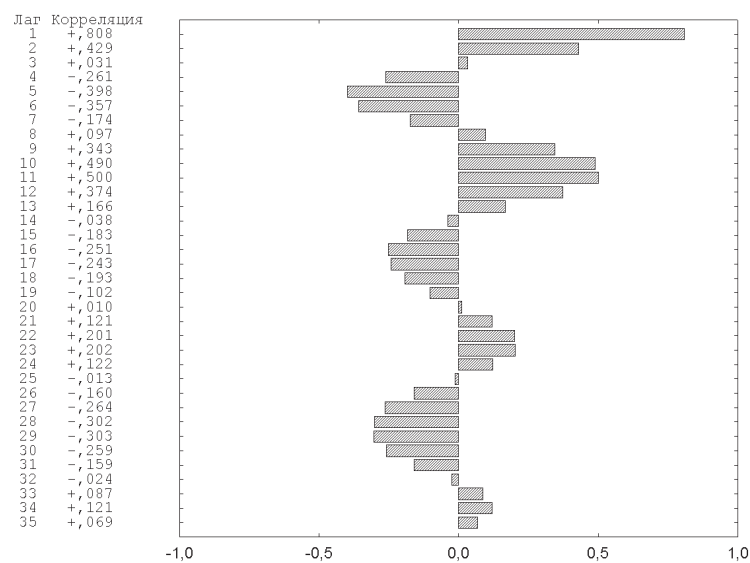


Рис. 18.2. Выборочная корреляционная функция для данных солнечной активности

18.3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ СПЕКТРАЛЬНОЙ ПЛОТНОСТИ

Пусть для стационарного в широком смысле ВР $\{x_t\}_{t=-\infty}^{+\infty}$ существует спектральная плотность $f(\lambda)$, $\lambda \in [-\pi, \pi]$. Рассмотрим задачу оценивания спектральной плотности $f(\cdot)$ по реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$.

Введем вспомогательные статистики ($\lambda \in [-\pi, \pi]$):

$$A(\lambda) = \frac{2}{T} \sum_{t=1}^T (x_t - \mu) \cos(\lambda t), \quad B(\lambda) = \frac{2}{T} \sum_{t=1}^T (x_t - \mu) \sin(\lambda t);$$

$$A^*(\lambda) = \frac{2}{T} \sum_{t=1}^T (x_t - \bar{x}) \cos(\lambda t), \quad B^*(\lambda) = \frac{2}{T} \sum_{t=1}^T (x_t - \bar{x}) \sin(\lambda t),$$

где $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ – выборочное среднее.

Определение 18.4. *Спектрограммой называется следующая статистика, определенная на реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$ ($\lambda \in [-\pi, \pi]$):*

$$R^2(\lambda) = A^2(\lambda) + B^2(\lambda) \quad -$$

в случае известного математического ожидания μ ;

$$R^{*2}(\lambda) = A^{*2}(\lambda) + B^{*2}(\lambda) \quad -$$

в случае неизвестного математического ожидания μ .

Определение 18.5. *Выборочной спектральной плотностью для ВР $\{x_t\}_{t=-\infty}^{+\infty}$, вычисленной по реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$, называется статистика ($\lambda \in [-\pi, \pi]$):*

$$I(\lambda) = \frac{T}{8\pi} R^2(\lambda) \quad -$$

в случае известного математического ожидания μ ;

$$I^*(\lambda) = \frac{T}{8\pi} R^{*2}(\lambda) \quad -$$

в случае неизвестного математического ожидания μ .

Замечание 18.4. На практике $\lambda \in [-\pi, \pi]$ как аргумент спектральной плотности $f(\lambda)$ не может меняться непрерывно. Обычно рассматривают сетку:

$$\lambda = \frac{2\pi j}{T}, \quad j = 0, \pm 1, \dots, \pm \frac{T}{2}.$$

В узлах этой сетки выполняется:

$$A(\lambda) = A^*(\lambda), \quad B(\lambda) = B^*(\lambda), \quad R^2(\lambda) = R^{*2}(\lambda).$$

Замечание 18.5. Коэффициенты $A(\lambda)$, $A^*(\lambda)$, спектрограмма и выборочная спектральная плотность – четные функции, а коэффициенты $B(\lambda)$, $B^*(\lambda)$ – нечетные.

Теорема 18.3. *Пусть через \tilde{c}_τ , $\tau \in \{0, 1, \dots, T-1\}$, обозначены специальным образом отнормированные значения выборочной ковариационной функции (18.5):*

$$\tilde{c}_\tau = \tilde{c}_{-\tau} = \left(1 - \frac{|\tau|}{T}\right) c_\tau,$$

тогда для выборочной спектральной плотности справедливо представление

$$I(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \tilde{c}_\tau \cos(\lambda\tau), \quad \lambda \in [-\pi, \pi], \quad (18.10)$$

и по ней можно однозначно восстановить выборочную ковариационную функцию ($\tau \in \{0, 1, \dots, T-1\}$):

$$\tilde{c}_\tau = \tilde{c}_{-\tau} = \int_{-\pi}^{\pi} \cos(\lambda\tau) I(\lambda) d\lambda. \quad (18.11)$$

Доказательство. Справедливость соотношения (18.11) очевидна и следует из взаимосвязи прямого и обратного преобразований Фурье.

Получим выражение (18.10) для выборочной спектральной плотности:

$$\begin{aligned} I(\lambda) &= \frac{T}{8\pi} R^2(\lambda) = \frac{T}{8\pi} (A^2(\lambda) + B^2(\lambda)) = \\ &= \frac{T}{8\pi} \frac{4}{T^2} \left(\sum_{t_1=1}^T \sum_{t_2=1}^T (x_{t_1} - \mu)(x_{t_2} - \mu) \cos(\lambda t_1) \cos(\lambda t_2) + \right. \\ &\quad \left. + \sum_{t_1=1}^T \sum_{t_2=1}^T (x_{t_1} - \mu)(x_{t_2} - \mu) \sin(\lambda t_1) \sin(\lambda t_2) \right) = \\ &= \frac{1}{2\pi T} \sum_{t_1=1}^T \sum_{t_2=1}^T (x_{t_1} - \mu)(x_{t_2} - \mu) \cos(\lambda(t_1 - t_2)) = \\ &= \frac{1}{2\pi T} \sum_{\tau=-(T-1)}^{T-1} \sum_{t \in S_\tau} (x_t - \mu)(x_{t+\tau} - \mu) \cos(\lambda\tau), \end{aligned}$$

где $S_\tau = \{1, \dots, T - \tau\}, \tau \geq 0; \{1 - \tau, \dots, T\}, \tau < 0\}$.

Учитывая, что мощность множества S_τ равна $T - |\tau|$, получаем требуемое. \square

Замечание 18.6. В случае неизвестного математического ожидания μ результаты аналогичны.

Исследуем свойства выборочной спектральной плотности.

Теорема 18.4. Пусть $\{x_t\}_{t=-\infty}^{+\infty}$ — стационарный гауссовский ВР, имеющий непрерывную спектральную плотность, т. е. $f(\lambda)$ непрерывна по $\lambda \in [-\pi, \pi]$, тогда выборочная спектральная плотность (18.10) является асимптотически несмещенной:

$$\mathbf{E}\{I(\lambda)\} \rightarrow f(\lambda), \quad \lambda \in [-\pi, \pi], \quad T \rightarrow +\infty,$$

но несостоятельной оценкой спектральной плотности:

$$\mathbf{E}\{(I(\lambda) - f(\lambda))^2\} \rightarrow \begin{cases} 2f^2(\lambda), & \lambda = 0, \pm\pi; \\ f^2(\lambda), & \lambda \neq 0, \pm\pi; \end{cases} \quad T \rightarrow +\infty, \quad \lambda \in [-\pi, \pi].$$

Доказательство можно найти в [5]. \square

На практике, чтобы избежать несостоятельности оценки спектральной плотности, используют так называемую *сглаженную спектральную плотность* [5]

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \tilde{\omega}_{\tau} \tilde{c}_{\tau} \cos(\lambda\tau), \quad \lambda \in [-\pi, \pi], \quad (18.12)$$

где $\{\tilde{\omega}_{\tau}\}$ – специальным образом подобранные веса.

Из определения спектральной плотности:

$$f(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} \cos(\lambda\tau) \sigma(\tau), \quad \lambda \in [-\pi, \pi],$$

видно, что можно было сразу записать сглаженную (взвешенную) оценку спектральной плотности как

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} \omega_{\tau} c_{\tau} \cos(\lambda\tau), \quad \lambda \in [-\pi, \pi],$$

где веса $\{\omega_{\tau}\}$ связаны с весами $\{\tilde{\omega}_{\tau}\}$ соотношением $\omega_{\tau} = \tilde{\omega}_{\tau} \left(1 - \frac{|\tau|}{T}\right)$.

Рассмотрим некоторые часто используемые на практике способы задания весов в (18.12).

1. $\tilde{\omega}_{\tau} = 1, \forall \tau$, тогда $\hat{f}(\lambda) = I(\lambda)$ – обычная выборочная спектральная плотность.
2. «Усеченная» выборочная спектральная плотность ($k < T$ – параметр «усечения»):

$$\tilde{\omega}_{\tau} = \begin{cases} 1, & |\tau| \leq k; \\ 0, & |\tau| > k. \end{cases}$$

3. Семейство оценок Парзена ($q = 1; 2; \dots$):

$$\tilde{\omega}_{\tau} = \begin{cases} 1 - \left(\frac{|\tau|}{k}\right)^q, & |\tau| \leq k; \\ 0, & |\tau| > k. \end{cases}$$

В частности, при $q = 1$ имеем оценку Бартлетта, а при $q = 2$ – собственно оценку Парзена.

Для асимптотической несмещенности и состоятельности взвешенной оценки спектральной плотности на практике требуется непрерывность спектральной плотности и выполнение асимптотики [5]

$$T \rightarrow +\infty, \quad k = k(T) \rightarrow +\infty, \quad \frac{k}{T} \rightarrow 0.$$

Замечание 18.7. Выше и в теореме 18.4 требуется, чтобы спектральная плотность $f(\lambda)$ была непрерывна по λ , но ее непрерывность следует, в частности, из сходимости ряда (18.1).

На рис. 18.3 построена оценка Бартлетта спектральной плотности по данным солнечной активности, приведенным на рис. 17.2. Для удобства на оси абсцисс помещены не значения аргумента λ спектральной плотности, а величины $\frac{2\pi}{\lambda}$, по которым согласно замечанию 17.2 к п. 17.3 легко оценить период. Видно, что оценка периода солнечной активности составляет приблизительно 11 лет.

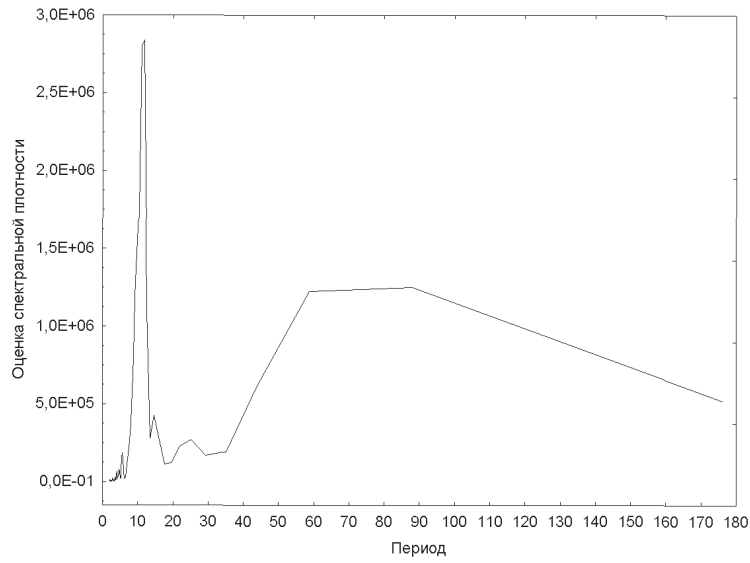


Рис. 18.3. Оценка спектральной плотности для данных солнечной активности

18.4. УПРАЖНЕНИЯ

1. Показать, что в условиях теоремы 18.1 для дисперсии выборочного среднего $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ выполняется

$$\lim_{T \rightarrow +\infty} TD\{\bar{x}\} = \sum_{\tau=-\infty}^{+\infty} \sigma(\tau) = 2\pi f(0),$$

где $\sigma(\cdot)$ и $f(\cdot)$ – ковариационная функция и спектральная плотность соответственно.

2. Строго обосновать утверждения, сделанные относительно спектрограммы и выборочной спектральной плотности в замечаниях 18.4 и 18.5.
3. Доказать теорему 18.4 о свойствах выборочной спектральной плотности, а также установить, указанные в п. 18.3, асимптотические свойства взвешенных оценок спектральной плотности.

Глава 19

ПАРАМЕТРИЧЕСКИЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

19.1. ВРЕМЕННЫЕ РЯДЫ АВТОРЕГРЕССИИ, СКОЛЬЗЯЩЕГО СРЕДНЕГО, АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО: УСЛОВИЯ СТАЦИОНАРНОСТИ И КОВАРИАЦИОННЫЕ ФУНКЦИИ, ЕДИНОЕ ЛИНЕЙНОЕ ПРЕДСТАВЛЕНИЕ

Прежде чем дать определения, сформулируем условие (Y), которое будет использоваться на протяжении всей главы.

У. Пусть $\{u_t\}_{t=-\infty}^{+\infty}$ – последовательность случайных величин, имеющих нулевое математическое ожидание и одинаковую ограниченную дисперсию:

$$E\{u_t\} = 0, \quad D\{u_t\} = \sigma^2 < +\infty, \quad t \in \mathbb{Z},$$

и пусть выполняется одно из следующих дополнительных условий.

У 1. Случайные величины $\{u_t\}_{t=-\infty}^{+\infty}$ некоррелированы:

$$E\{u_t u_l\} = \delta_{tl} \sigma^2, \quad t, l \in \mathbb{Z}.$$

У 2. Случайные величины $\{u_t\}_{t=-\infty}^{+\infty}$ независимы в совокупности и одинаково распределены.

У 3. Случайные величины $\{u_t\}_{t=-\infty}^{+\infty}$ некоррелированы и имеют нормальное распределение $N_1(0, \sigma^2)$.

Определение 19.1. ВР $\{x_t\}_{t=-\infty}^{+\infty}$ называется *временным рядом авторегрессии порядка p (АР(p))*, если

$$\beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} = u_t,$$

где $\beta_0 = 1, \beta_1, \dots, \beta_p$ – коэффициенты авторегрессии, $\beta_p \neq 0$.

Определение 19.2. ВР $\{x_t\}_{t=-\infty}^{+\infty}$ называется *временным рядом скользящего среднего порядка q (СС(q))*, если

$$x_t = \alpha_0 u_t + \alpha_1 u_{t-1} + \dots + \alpha_q u_{t-q},$$

где $\alpha_0 = 1, \alpha_1, \dots, \alpha_q$ – коэффициенты скользящего среднего, $\alpha_q \neq 0$.

Определение 19.3. ВР $\{x_t\}_{t=-\infty}^{+\infty}$ называется *временным рядом авторегрессии и скользящего среднего порядков p и q (АРСС(p, q))*, если $\sum_{j=0}^p \beta_j x_{t-j} = \sum_{i=0}^q \alpha_i u_{t-i}$,

где $\alpha_0 = \beta_0 = 1, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p$ – коэффициенты, $\alpha_q, \beta_p \neq 0$.

Очевидно, что модель АРСС(p, q) обобщает модели АР(p) (при $q = 0$) и СС(q) (при $p = 0$).

Убедимся, что условие $\alpha_0 = \beta_0 = 1$ не сужает модель. Пусть в модели АРСС(p, q) $\beta_0 \neq 1$, тогда введем в рассмотрение новые коэффициенты:

$$\tilde{\beta}_j = \frac{\beta_j}{\beta_0}, \quad j = 0, \dots, p; \quad \tilde{\alpha}_i = \frac{\alpha_i}{\beta_0}, \quad i = 0, \dots, q.$$

Если $\tilde{\alpha}_0 \neq 1$, то продолжаем:

$$\tilde{\tilde{\alpha}}_i = \frac{\tilde{\alpha}_i}{\tilde{\alpha}_0}, \quad i = 0, \dots, q; \quad \tilde{u}_t = \tilde{\alpha}_0 u_t.$$

В итоге получим следующую модель: $\sum_{j=0}^p \tilde{\beta}_j x_{t-j} = \sum_{i=0}^q \tilde{\alpha}_i \tilde{u}_{t-i}$, в которой $\tilde{\alpha}_0 = 1$, $\tilde{\beta}_0 = 1$, $\tilde{\alpha}_q \neq 0$, $\tilde{\beta}_p \neq 0$, $D\{\tilde{u}_t\} = (\tilde{\alpha}_0)^2 \sigma^2 < +\infty$, но это модель АРСС(p, q) по определению.

Приведем все три модели, АР(p), СС(q) и АРСС(p, q), к общему виду. Начнем с АР(p):

$$x_t = -\beta_1 x_{t-1} - \beta_2 x_{t-2} - \dots - \beta_p x_{t-p} + u_t.$$

Вместо x_{t-1} подставим его выражение через эту модель:

$$x_{t-1} = -\beta_1 x_{t-1-1} - \beta_2 x_{t-1-2} - \dots - \beta_p x_{t-1-p} + u_{t-1},$$

затем подставим x_{t-2} и т. д., в результате получим $x_t = \sum_{j=0}^{+\infty} \delta_j u_{t-j}$, где $\delta_0 = 1$,

а остальные $\{\delta_j\}_{j=1}^{+\infty}$ выражаются через $\{\beta_i\}_{i=1}^p$.

В случае модели СС(q) уже имеем данное представление, положив $\delta_j = \{\alpha_j, j = 0, \dots, q; 0, j > q\}$.

Для модели АРСС(p, q) получим то же представление, только коэффициенты $\{\delta_j\}_{j=1}^{+\infty}$ будут выражаться через коэффициенты $\{\beta_i\}_{i=1}^p$ и $\{\alpha_l\}_{l=1}^q$.

Обобщим данную модель следующим ВР:

$$x_t = \mu + \sum_{j=-\infty}^{+\infty} \gamma_j u_{t-j}, \quad (19.1)$$

где $|\mu| < +\infty$, $\{u_t\}_{t=-\infty}^{+\infty}$ удовлетворяют условию У, а $\{\gamma_j\}_{j=-\infty}^{+\infty}$ — коэффициенты.

Исследуем свойства ВР $\{x_t\}_{t=-\infty}^{+\infty}$, задаваемого соотношением (19.1).

Теорема 19.1. Пусть сходится ряд

$$\sum_{j=-\infty}^{+\infty} \gamma_j^2 < +\infty, \quad (19.2)$$

тогда если выполняется У1, то ряд из правой части (19.1) сходится в среднеквадратическом смысле, а сам ВР $\{x_t\}_{t=-\infty}^{+\infty}$ является стационарным в широком смысле с математическим ожиданием

$$E\{x_t\} = \mu, \quad t \in \mathbb{Z},$$

ковариационной функцией

$$\sigma(\tau) = E\{(x_t - \mu)(x_{t+\tau} - \mu)\} = \sigma^2 \sum_{j=-\infty}^{+\infty} \gamma_j \gamma_{j+\tau}, \quad \tau \in \mathbb{Z},$$

и дисперсией

$$D\{x_t\} = \sigma(0) = \sigma^2 \sum_{j=-\infty}^{+\infty} \gamma_j^2 < +\infty, \quad t \in \mathbb{Z}.$$

Доказательство. Докажем среднеквадратическую сходимость ряда (19.1). Не ограничивая общности, считаем $\mu = 0$:

$$\Delta_{m,n} = E\left\{\left(x_t - \sum_{j=-m}^n \gamma_j u_{t-j}\right)^2\right\} = E\left\{\left(\sum_{i=-\infty}^{-m-1} \gamma_i u_{t-i} + \sum_{j=n+1}^{+\infty} \gamma_j u_{t-j}\right)^2\right\} =$$

$$\begin{aligned}
&= E \left\{ \left(\sum_{i=-\infty}^{-m-1} \gamma_i u_{t-i} \right)^2 \right\} + E \left\{ \left(\sum_{j=n+1}^{+\infty} \gamma_j u_{t-j} \right)^2 \right\} = \\
&= \sigma^2 \sum_{i=-\infty}^{-m-1} \gamma_i^2 + \sigma^2 \sum_{j=n+1}^{+\infty} \gamma_j^2 \rightarrow 0, \quad n \rightarrow +\infty, \quad m \rightarrow +\infty.
\end{aligned}$$

Вычислим ковариационную функцию:

$$\sigma(\tau) = \text{cov}\{x_t, x_{t+\tau}\} = E \left\{ \sum_{j=-\infty}^{+\infty} \gamma_j u_{t-j} \sum_{i=-\infty}^{+\infty} \gamma_i u_{t+\tau-i} \right\} = \sigma^2 \sum_{j=-\infty}^{+\infty} \gamma_j \gamma_{j+\tau}.$$

Полученное выражение для ковариационной функции не зависит от t , а только от лага τ .

Выражение для дисперсии получается из последнего соотношения при $\tau = 0$:

$$D\{x_t\} = \text{cov}\{x_t, x_t\} = \sigma(0) = \sigma^2 \sum_{j=-\infty}^{+\infty} \gamma_j^2 < +\infty.$$

□

Следствие 19.1. Если в условиях теоремы 19.1 условие $Y1$ заменить на $Y2$ или $Y3$, то помимо стационарности в широком смысле для $BP \{x_t\}_{-\infty}^{+\infty}$ из (19.1) будет иметь место также и стационарность в узком смысле [5].

Следствие 19.2. В условиях теоремы 19.1 вместо сходимости ряда (19.2) можно потребовать сходимости ряда

$$\sum_{j=-\infty}^{+\infty} |\gamma_j| < +\infty. \quad (19.3)$$

Следствие 19.3. Для $CC(q)$ о стационарности в широком смысле можно говорить, если все $|\alpha_j| < +\infty$, $j = 0, \dots, q$, и выполнены остальные условия теоремы 19.1, ряд (19.2) при этом представляет собой конечную сумму и заведомо сходится.

Следствие 19.4. Для того чтобы временные ряды $AP(p)$ или $APCC(p, q)$ были стационарными в широком либо в узком смысле в условиях теоремы 19.1 либо следствия 19.1 к ней вместо сходимости ряда (19.2) достаточно потребовать выполнения следующего условия [5, 6]. Корни характеристического уравнения

$$\sum_{j=0}^p \beta_j z^{p-j} = 0 \quad (19.4)$$

должны лежать внутри единичного круга: $|z| < 1$.

Найдем ковариационные функции для всех трех моделей.

1. $AP(p)$:

$$\sum_{j=0}^p \beta_j x_{t-j} = u_t.$$

Домножим левую и правую части на $x_{t-\tau}$, $\tau \geq 0$, и вычислим математическое ожидание. При умножении правой части воспользуемся тем, что $x_{t-\tau} = \sum_{j=0}^{+\infty} \delta_j u_{t-\tau-j}$ и случайные величины $\{u_t\}_{t=-\infty}^{+\infty}$ некоррелированы, получим:

$$\sum_{j=0}^p \beta_j \sigma(\tau - j) = 0, \quad \tau \geq 1;$$

$$\sum_{j=0}^p \beta_j \sigma(j) = \sigma^2, \quad \tau = 0. \quad (19.5)$$

Соотношения (19.5) – система линейных алгебраических уравнений (СЛАУ), которую можно решать как относительно $\{\beta_j\}$, так и относительно $\{\sigma(j)\}$. Эта система называется *системой уравнений Юла – Уокера* (Yule – Walker).

2. $CC(q)$:

$$\sigma(\tau) = \text{cov}\{x_t, x_{t+\tau}\} = \sigma^2 \sum_{j=-\infty}^{+\infty} \gamma_j \gamma_{j+\tau} = \sigma^2 \sum_{j=0}^{q-\tau} \alpha_j \alpha_{j+\tau}.$$

3. $ARCC(p, q)$. Ковариационная функция $\sigma(\tau)$ определяется уравнениями Юла – Уокера (19.5), если значение лага $\tau \geq q + 1 - p$, и при этом не надо знать коэффициенты скользящего среднего. Если $\tau < q + 1 - p$, то в этом случае для ее вычисления получены рекуррентные формулы [5].

19.2. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНОГО СРЕДНЕГО И ВЫБОРОЧНОЙ КОВАРИАЦИОННОЙ ФУНКЦИИ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА

Будем рассматривать общую линейную модель (19.1):

$$x_t = \mu + \sum_{j=-\infty}^{+\infty} \gamma_j u_{t-j},$$

где на случайные величины $\{u_t\}_{t=-\infty}^{+\infty}$ и коэффициенты $\{\gamma_j\}_{j=-\infty}^{+\infty}$ необходимо наложить дополнительные условия, чтобы ВР (19.1) был стационарным в широком смысле.

ВР (19.1) характеризуется математическим ожиданием

$$E\{x_t\} = \mu, \quad t \in \mathbb{Z},$$

и ковариационной функцией

$$\sigma(\tau) = E\{(x_t - \mu)(x_{t+\tau} - \mu)\}, \quad \tau \in \mathbb{Z}.$$

На реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$ в качестве их статистических оценок в предыдущей главе были соответственно определены выборочное среднее (18.2):

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

и выборочная ковариационная функция (18.5):

$$c_\tau = c_{-\tau} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (x_t - \mu)(x_{t+\tau} - \mu), \quad \tau \in \{0, 1, \dots, T-1\}.$$

Если μ неизвестно, то оно заменяется на \bar{x} , и используется

$$c_\tau^* = c_\tau|_{\mu=\bar{x}}.$$

Для выборочных среднего и ковариационной функции в условиях модели (19.1) получены следующие асимптотические результаты [5].

Теорема 19.2. Пусть ВР (19.1) удовлетворяет условию У2 и сходится ряд (19.3):

$$\sum_{j=-\infty}^{+\infty} |\gamma_j| < +\infty,$$

тогда при $T \rightarrow \infty$ выборочное среднее \bar{x} асимптотически нормально распределено:

$$\mathcal{L}\left\{\sqrt{T}(\bar{x} - \mu)\right\} \rightarrow N_1(0, D^2),$$

где

$$D^2 = \sum_{\tau=-\infty}^{+\infty} \sigma(\tau) = \sigma^2 \left(\sum_{j=-\infty}^{+\infty} \gamma_j \right)^2.$$

Доказательство непосредственно следует из представления (19.1), центральной предельной теоремы и результатов теорем 18.1, 19.1. \square

Теорема 19.3. Пусть в условиях теоремы 19.2 для последовательности $\{u_t\}_{t=-\infty}^{+\infty}$ существует 4-й момент:

$$E\{u_t^4\} = 3\sigma^4 + \kappa_4 < +\infty,$$

тогда случайные величины в последовательности

$$\sqrt{T}(c_0 - \sigma(0)), \sqrt{T}(c_1 - \sigma(1)), \dots, \sqrt{T}(c_h - \sigma(h)), \dots$$

имеют при $T \rightarrow \infty$ нормальное распределение со следующими ковариациями ($h, g \geq 0$):

$$\begin{aligned} T \text{cov}\{c_h, c_g\} &\rightarrow \sum_{\tau=-\infty}^{+\infty} (\sigma(\tau+h)\sigma(\tau+g) + \\ &+ \sigma(\tau-h)\sigma(\tau+g)) + \frac{\kappa_4}{\sigma^4} \sigma(h)\sigma(g). \end{aligned} \quad (19.6)$$

Доказательство приведено в [5]. \square

Замечание 19.1. Если вместо условия У2 потребовать выполнения условия У3, то в соотношении (19.6) $\kappa_4 = 0$, и результат совпадает с полученным ранее в п. 18.2.

Замечание 19.2. Если μ неизвестно, то для оценки ковариационной функции c_τ^* из (18.6) остается справедливым результат теоремы 19.3.

19.3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ ПРИ ИЗВЕСТНОМ ПОРЯДКЕ АВТОРЕГРЕССИИ

Пусть имеется временной ряд АР(p) $\{x_t\}_{t=-\infty}^{+\infty}$:

$$x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} = u_t, \quad (19.7)$$

где относительно $\{u_t\}_{t=-\infty}^{+\infty}$ и параметров β_1, \dots, β_p предполагается выполнение каких-либо условий, обеспечивающих стационарность ВР (19.7) в широком смысле. В частности, это могут быть условия У на случайные величины $\{u_t\}_{t=-\infty}^{+\infty}$ и требование относительно корней характеристического уравнения (19.4) (они должны лежать внутри единичного круга).

Пусть коэффициенты $AP(p)$ β_1, \dots, β_p неизвестны и неизвестна дисперсия $\sigma^2 = D\{u_t\}$ в условиях \mathcal{U} . Необходимо по реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$ временного ряда (19.7) построить статистические оценки $\hat{\beta}_1, \dots, \hat{\beta}_p$ и $\hat{\sigma}^2$. Рассмотрим несколько способов оценивания неизвестных параметров в модели $AP(p)$.

19.3.1. Оценки Юла – Уокера

Для временного ряда $AP(p)$ в п. 19.1 была получена система уравнений Юла – Уокера (19.5), которые связывают коэффициенты β_1, \dots, β_p , дисперсию σ^2 и ковариационную функцию $\sigma(\tau)$. Воспользуемся тем, что математическое ожидание данного ВР равно нулю ($E\{x_t\} = 0$), и в (19.5) заменим ковариационную функцию $\sigma(\tau)$ на ее статистическую оценку:

$$c_\tau = c_{-\tau} = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} x_t x_{t+\tau}, \quad \tau \in \{0, 1, \dots, T-1\}.$$

Предположим, что β_1, \dots, β_p уже оценены, т. е. найдены $\hat{\beta}_1, \dots, \hat{\beta}_p$ ($\hat{\beta}_0 := 1$). Тогда из (19.5) при $\tau = 0$ получим оценку дисперсии $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \sum_{j=0}^p \hat{\beta}_j c_j. \quad (19.8)$$

Осталось найти $\hat{\beta}_1, \dots, \hat{\beta}_p$. Для этого достаточно p уравнений Юла – Уокера, где $\tau \neq 0$:

$$\sum_{j=0}^p \hat{\beta}_j c_{\tau-j} = 0, \quad \tau = 1, \dots, p. \quad (19.9)$$

Запишем (19.9) в матричном виде с учетом того, что $c_h = c_{-h}$, $h \geq 0$. Введем в рассмотрение p -вектор:

$$c = (c_1, \dots, c_p)^T \in \mathbb{R}^p,$$

и оценку ковариационной матрицы:

$$C = (c_{|i-j|})_{i,j=1}^p = \begin{pmatrix} c_0 & c_1 & \dots & c_{p-1} \\ c_1 & c_0 & \dots & c_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p-1} & c_{p-2} & \dots & c_0 \end{pmatrix}.$$

Тогда (19.9) можно переписать в виде $c + C\hat{\beta} = \mathbf{0}_p$, $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$, откуда находим

$$\hat{\beta} = -C^{-1}c. \quad (19.10)$$

Используя (19.10), перепишем также оценку для дисперсии (19.8):

$$\hat{\sigma}^2 = c_0 + \hat{\beta}^T c = c_0 - c^T C^{-1}c. \quad (19.11)$$

Соотношения (19.10), (19.11) являются оценками Юла – Уокера в матричном виде, и если не накладываются дополнительные условия, то эти оценки обладают свойствами оценок по методу моментов (см. п. 11.6).

Замечание 19.3. Если в уравнениях из (19.9) разделить обе части на c_0 , то получим

$$\sum_{j=0}^p \hat{\beta}_j \hat{\rho}_{\tau-j} = 0, \quad \tau = 1, \dots, p,$$

где $\hat{\rho}_h = c_h/c_0$ – оценка корреляционной функции $\rho_h = \sigma(h)/\sigma(0)$.

Обозначив через

$$\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_p)^T \in \mathbb{R}^p;$$

$$\mathcal{R} = (\hat{\rho}_{|i-j|})_{i,j=1}^p,$$

имеем эквивалентную запись для (19.10)

$$\hat{\beta} = -\mathcal{R}^{-1}\hat{\rho},$$

и оценки дисперсии (19.11):

$$\hat{\sigma}^2 = c_0 + \hat{\beta}^T c = c_0 - \hat{\rho}^T \mathcal{R}^{-1} c = c_0(1 - \hat{\rho}^T \mathcal{R}^{-1} \hat{\rho}).$$

19.3.2. Оценки максимального правдоподобия и их связь с МНК-оценками

Пусть выполняется условие УЗ, и корни уравнения (19.4) лежат внутри единичного круга, тогда временной ряд $AP(p)$ из (19.7) является стационарным в широком смысле.

Рассмотрим реализацию $X = (x_1, \dots, x_T)^T \in \mathbb{R}^T$ длительности $T \geq 1$. Воспользуемся тем, что в условиях УЗ вектор X имеет T -мерное нормальное распределение вероятностей с плотностью $p(X) = n_T(X|\mathbf{0}_T, \Sigma_{T,T})$, где $\Sigma_{T,T} = (\sigma_{|i-j|})_{i,j=1}^T$ – ковариационная $(T \times T)$ -матрица, однозначно определяемая ковариационной функцией $\sigma(\tau)$, $\tau \geq 0$. В свою очередь, согласно уравнениям Юла – Уокера (19.5) $\sigma(\tau)$ зависит от $\beta = (\beta_1, \dots, \beta_p)^T$ и σ^2 , поэтому $\Sigma_{T,T} = \Sigma_{T,T}(\beta, \sigma^2)$.

Воспользуемся методом максимального правдоподобия (ММП) и определим статистические оценки $\hat{\beta}$, $\hat{\sigma}^2$ для β и σ^2 . Экстремальная задача по нахождению оценок максимального правдоподобия (ОМП, МП-оценок) в данном случае имеет вид

$$\{\hat{\beta}, \hat{\sigma}^2\} = \arg \max_{\beta, \sigma^2} n_T(X|\mathbf{0}_T, \Sigma_{T,T}(\beta, \sigma^2)). \quad (19.12)$$

Теорема 19.4. Пусть $AP(p)$ (19.7) удовлетворяет условию УЗ и корни уравнения (19.4) лежат внутри единичного круга, тогда при $T \rightarrow +\infty$ ОМП (19.12) асимптотически эквивалентны оценкам по методу наименьших квадратов (МНК-оценкам):

$$\begin{aligned} \hat{\beta} &= - \left(\sum_{t=p+1}^T X_t X_t^T \right)^{-1} \sum_{t=p+1}^T x_t X_t; \\ \hat{\sigma}^2 &= \frac{1}{T-p} \sum_{t=p+1}^T (x_t + \hat{\beta}^T X_t)^2, \\ X_t &= (x_{t-1}, \dots, x_{t-p})^T \in \mathbb{R}^p. \end{aligned} \quad (19.13)$$

Оценки Юла – Уокера (19.10), (19.11) также асимптотически эквивалентны оценкам (19.13), а значит, и асимптотически эквивалентны МП-оценкам (19.12).

Доказательство. Разобьем реализацию X длительности T на два подвектора:

$$X = (x_1, \dots, x_T)^T = ((X^p)^T, x_{p+1}, \dots, x_T)^T.$$

Подвектор $X^p = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ имеет нормальное распределение $N_p(\mathbf{0}_p, \Sigma_{p,p})$, где $\Sigma_{p,p} = (\sigma_{|i-j|})_{i,j=1}^p$.

Запишем (19.7) в следующем виде:

$$x_t + \beta^T X_t = u_t, \quad t = p+1, \dots, T.$$

Заметим, что x_t и u_t как функциональные преобразования случайных величин связаны единичным якобианом, и учтем, что согласно УЗ случайные величины $\{u_t\}_{t=1}^T$ независимы в совокупности и одинаково распределены по закону $N_1(0, \sigma^2)$, а следовательно, имеют совместное нормальное распределение $N_T(\mathbf{0}_T, \text{diag}\{\sigma^2, \dots, \sigma^2\})$.

С учетом этих фактов преобразуем плотность из (19.12):

$$\begin{aligned} n_T(X|\mathbf{0}_T, \Sigma_{T,T}(\beta, \sigma^2)) &= n_p(X^p|\mathbf{0}_p, \Sigma_{p,p})p(x_{p+1}, \dots, x_T|X^p) = \\ &= \frac{1}{(2\pi)^{T/2}|\Sigma_{p,p}|^{1/2}(\sigma^2)^{\frac{T-p}{2}}} \times \exp\left(-\frac{1}{2}(X^p)^T \Sigma_{p,p}^{-1} X^p - \frac{1}{2\sigma^2} \sum_{t=p+1}^T (x_t + \beta^T X_t)^2\right). \end{aligned}$$

Исходя из последнего соотношения, экстремальную задачу (19.12) при $T \rightarrow +\infty$ можно свести к следующей задаче:

$$\sum_{t=p+1}^T (x_t + \beta^T X_t)^2 \rightarrow \min_{\beta}, \quad (19.14)$$

являющейся ничем иным, как МНК.

Вычислив первую производную левой части (19.14) по β и приравняв к нулю, получим уравнение для нахождения оценки $\hat{\beta}$:

$$\sum_{t=p+1}^T 2(x_t + \beta^T X_t) X_t = 2 \sum_{t=p+1}^T X_t X_t^T \beta + 2 \sum_{t=p+1}^T x_t X_t = \mathbf{0}_p,$$

откуда следует (19.13).

Оценку для $\hat{\sigma}^2$ ищем как в регрессионном анализе – это остаточная сумма квадратов.

Эквивалентность оценок Юла – Уокера (19.10), (19.11) оценкам (19.13) следует из соотношения: $(T-h)/(T-\tau) \rightarrow 1, 0 \leq h, \tau < +\infty$. □

Следствие 19.5. В условиях теоремы 19.4 МП-оценки (19.12), либо, что то же самое, оценки (19.13), являются состоятельными оценками для соответствующих параметров, причем $\hat{\beta}$ имеет асимптотически нормальное распределение:

$$\mathcal{L}\{\sqrt{T}(\hat{\beta} - \beta)\} \rightarrow N_p(\mathbf{0}_p, \sigma^2 \Sigma_{p,p}^{-1}), \quad T \rightarrow +\infty.$$

Доказательство приведено в [5]. □

19.4. ОЦЕНИВАНИЕ ПОРЯДКА АВТОРЕГРЕССИИ

Пусть порядок p в модели $AR(p)$ неизвестен. Очевидно, что задачу определения порядка можно свести к задаче проверки гипотез:

$$\begin{aligned} H_0 &: \beta_p = 0; \\ H_1 = \overline{H_0} &: \beta_p \neq 0, \end{aligned}$$

при заданном уровне значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$.

Если принимается H_0 , то считаем, что $\beta_p = 0$, и порядок может быть равным $p-1$ или менее того.

Для построения критерия проверки гипотез H_0, H_1 воспользуемся МП-оценками параметров авторегрессии из п. 19.3.

Теорема 19.5. Пусть выполнены условия теоремы 19.4 и $\beta_p = 0$, тогда для ОМП $\hat{\beta}_p$ справедливо асимптотическое соотношение

$$\mathcal{L} \left\{ \sqrt{T} \hat{\beta}_p \right\} \rightarrow N_1(0, 1), \quad T \rightarrow +\infty.$$

Доказательство. Согласно следствию 19.5:

$$\mathcal{L} \left\{ \sqrt{T} \hat{\beta}_p \right\} \rightarrow N_1 \left(0, \sigma^2 (\Sigma_{p,p}^{-1})_{p,p} \right), \quad T \rightarrow +\infty.$$

Найдем нижний правый элемент $(\Sigma_{p,p}^{-1})_{p,p}$ матрицы $\Sigma_{p,p}^{-1}$ при $\beta_p = 0$. Разобьем матрицу $\Sigma_{p,p}$ на блоки:

$$\Sigma_{p,p} = \begin{pmatrix} \Sigma_{p-1,p-1} & \sigma_{p-1} \\ \sigma_{p-1}^T & \sigma_{p,p} \end{pmatrix};$$

$$\sigma_{p-1} = (\sigma(p-1), \sigma(p-2), \dots, \sigma(1))^T, \quad \sigma_{p,p} = \sigma(0),$$

и, воспользовавшись формулами для обращения блочной матрицы (см. доказательство леммы 16.1 и тем, что при $\beta_p = 0$ имеет место модель $AR(p-1)$), получим

$$(\Sigma_{p,p}^{-1})_{p,p} = (\sigma_{p,p} - \sigma_{p-1}^T \Sigma_{p-1,p-1}^{-1} \sigma_{p-1})^{-1} = (\sigma(0) + \sigma_{p-1}^T \tilde{\beta})^{-1} = (\sigma^2)^{-1},$$

где $\tilde{\beta} = (\beta_{p-1}, \dots, \beta_1)^T$. □

Результат теоремы 19.5 позволяет построить следующий критерий для определения порядка авторегрессии:

$$\begin{cases} H_0 & : \sqrt{T} |\hat{\beta}_p| \leq \Delta; \\ H_1 = \bar{H}_0 & : \sqrt{T} |\hat{\beta}_p| > \Delta, \end{cases} \quad (19.15)$$

где

$$\Delta = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) -$$

квантиль уровня $1 - \frac{\alpha}{2}$ стандартного нормального закона $N_1(0, 1)$, а $\alpha = P\{H_1|H_0\} \in (0, 1)$ – заданный уровень значимости.

На практике применяются три схемы для определения порядка авторегрессии.

1. Известен максимальный порядок, на который мы согласны. Пусть этот порядок есть p^+ . Сначала проверяем гипотезы H_0, H_1 согласно (19.15) для $p = p^+$. Если H_0 принимается, то уменьшаем порядок на единицу и снова проверяем. И так до тех пор, пока H_0 не будет отвергнута. Значение p , при котором H_0 первый раз отвергается, и есть искомая оценка порядка авторегрессии.

2. Известен минимальный порядок, на который мы согласны. Пусть этот порядок есть p^- . Сначала проверяем гипотезы H_0, H_1 при $p = p^-$. Если H_0 не принимается, то увеличиваем порядок на единицу и снова проверяем. Так до тех пор, пока H_0 не будет принята. Если H_0 принята на значении p , то оценка порядка равна $p - 1$.

3. Известен примерный порядок p^* , на который мы согласны. Проверяем H_0, H_1 при $p = p^*$. Если H_0 не принимается, то переходим к схеме 2: увеличиваем порядок на единицу и снова проверяем. Если H_0 принимается, то переходим к схеме 1: уменьшаем порядок на единицу и снова проверяем.

Замечание 19.4. При последовательной проверке гипотез относительно порядка авторегрессии мы вынуждены производить оценивание коэффициентов авторегрессии для каждого предполагаемого значения порядка авторегрессии. Для этих целей наряду с (19.13) и (19.10) получены рекуррентные соотношения [5, 6].

19.5. ПРОГНОЗИРОВАНИЕ СТАЦИОНАРНЫХ В ШИРОКОМ СМЫСЛЕ ВРЕМЕННЫХ РЯДОВ

19.5.1. Общие принципы стохастического прогнозирования

Пусть имеются две случайные величины x и y , и необходимо спрогнозировать значение случайной величины y по значению случайной величины (или вектора) x , т. е. необходимо найти функцию

$$\hat{y} = \hat{y}(x) = \phi(x), \quad \phi(\cdot) \in \Psi,$$

где Ψ – некоторый допустимый класс функций.

В качестве критерия эффективности обычно используется среднеквадратическая ошибка прогнозирования:

$$\Delta^2(\phi) = E\{(\phi(x) - y)^2\}, \quad (19.16)$$

а оптимальным прогнозом $\hat{y}^* = \hat{y}^*(x) = \phi^*(x)$ считается прогноз

$$\phi^* = \arg \inf_{\phi \in \Psi} \Delta^2(\phi), \quad (19.17)$$

доставляющий минимум функционалу (19.16).

Теорема 19.6. Пусть конечны моменты второго порядка в (19.16): $E\{y^2\} < +\infty$, $E\{(\phi(x))^2\} < +\infty$, тогда оптимальный в смысле (19.17), (19.16) прогноз существует и совпадает с функцией регрессии y на x :

$$\phi^*(x) = E\{y|x\}. \quad (19.18)$$

Доказательство. Возьмем любой допустимый прогноз $\hat{y} = \phi(x)$, где $\phi(\cdot) \in \Psi$, и запишем для него функцию (19.16):

$$\begin{aligned} \Delta^2(\phi) &= E\{(\phi(x) - y)^2\} = E\{(\phi(x) - \phi^*(x) + \phi^*(x) - y)^2\} = \\ &= E\{(\phi(x) - \phi^*(x))^2\} + \Delta^2(\phi^*) + 2E\{(\phi(x) - \phi^*(x))(\phi^*(x) - y)\} = \\ &= E\{(\phi(x) - \phi^*(x))^2\} + \Delta^2(\phi^*) + 2E\{(\phi(x) - \phi^*(x))E\{(\phi^*(x) - y)|x\}\} = \\ &= E\{(\phi(x) - \phi^*(x))^2\} + \Delta^2(\phi^*) \geq \Delta^2(\phi^*). \end{aligned}$$

Равенство будет в том случае, когда $\phi(\cdot) \equiv \phi^*(\cdot)$. \square

Теорема 19.7. Оптимальный в среднеквадратическом прогноз $\hat{y}^* = \phi^*(x)$ из (19.18) максимально коррелирован с предсказываемым значением y .

Доказательство. Преобразуем ковариацию произвольного прогноза $\hat{y} = \phi(x)$ с y , учитывая, что $E\{y\} = E\{E\{y|x\}\} = E\{\phi^*(x)\}$:

$$\begin{aligned} \text{cov}\{\phi(x), y\} &= E\{(\phi(x) - E\{\phi(x)\})(y - E\{y\})\} = \\ &= E\{(\phi(x) - E\{\phi(x)\})E\{(y - E\{y\})|x\}\} = \\ &= E\{(\phi(x) - E\{\phi(x)\})(\phi^*(x) - E\{\phi^*(x)\})\} = \text{cov}\{\phi(x), \phi^*(x)\}; \\ \text{corr}^2\{\phi(x), y\} &= \frac{\text{cov}^2\{\phi(x), y\}}{D\{\phi(x)\}D\{y\}} = \frac{\text{cov}^2\{\phi(x), \phi^*(x)\}}{D\{\phi(x)\}D\{\phi^*(x)\}} \frac{D\{\phi^*(x)\}}{D\{y\}} = \\ &= \text{corr}^2\{\phi(x), \phi^*(x)\} \text{corr}^2\{\phi^*(x), y\}. \end{aligned}$$

Поскольку $\text{corr}^2\{\phi(x), \phi^*(x)\} \leq 1$, то получаем $|\text{corr}\{\phi(x), y\}| \leq |\text{corr}\{\phi^*(x), y\}|$. \square

19.5.2. Прогнозирование временных рядов авторегрессии

Пусть наблюдения описываются временным рядом $AP(p)$ $\{x_t\}_{t=-\infty}^{+\infty}$ из (19.7):

$$x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} = u_t.$$

Спрогнозируем значение x_t по «предыстории» $x_{t-1}, x_{t-2}, \dots, x_0, x_{-1}, \dots$.

Теорема 19.8. Пусть значения коэффициентов β_1, \dots, β_p временного ряда $AP(p)$ из (19.7) известны, а также выполняются какие-либо условия, обеспечивающие его стационарность в широком смысле. Тогда оптимальным в среднеквадратическом прогнозом для x_t по предыдущим значениям $x_{t-1}, x_{t-2}, \dots, x_0, x_{-1}, \dots$ является следующий прогноз:

$$\hat{x}_t^* = -\beta_1 x_{t-1} - \dots - \beta_p x_{t-p}, \quad (19.19)$$

причем достигнутая среднеквадратическая ошибка прогнозирования в условиях \mathbf{Y} равна $\sigma^2 = D\{u_t\}$.

Доказательство. Согласно (19.18) оптимальный прогноз имеет вид

$$\hat{x}_t^* = \Phi^*(x_{t-1}, \dots, x_{t-p}, \dots) = E\{x_t | x_{t-1}, \dots, x_{t-p}, \dots\},$$

что с учетом (19.7) приводит к (19.19). Для ошибки прогнозирования при этом имеем

$$\Delta^2(\Phi^*) = E\{(\hat{x}_t^* - x_t)^2\} = E\{u_t^2\} = D\{u_t\} = \sigma^2.$$

□

Замечание 19.5. Для того чтобы спрогнозировать $AP(p)$, из «предыстории» достаточно знать значения только p последних отсчетов, поэтому начинать прогнозирование можно при $t > p$. Это ограничение несущественно, поскольку на практике почти всегда имеется реализация $X = \{x_1, \dots, x_p, \dots, x_T\}$ длительности $T \geq p$.

Замечание 19.6. Если коэффициенты авторегрессии $\{\beta_j\}_{j=1}^p$ неизвестны, то при прогнозировании в (19.19) используются их статистические оценки $\{\hat{\beta}_j\}_{j=1}^p$ по реализации X длительности $T \geq p$. Если неизвестен порядок авторегрессии p , то сначала оценивается порядок (строится его оценка \hat{p}). В этом случае «подстановочный» прогноз будет иметь вид

$$\tilde{x}_t^* = -\hat{\beta}_1 x_{t-1} - \dots - \hat{\beta}_{\hat{p}} x_{t-\hat{p}}.$$

19.5.3. Гильбертово пространство стационарных в широком смысле временных рядов. Общая линейная модель как обобщение авторегрессионной модели и решение задачи прогнозирования

Построим гильбертово пространство стационарных в широком смысле временных рядов [5]. Для этого выполним следующие действия:

1. Возьмем все ВР $\{x_t\}_{t=-\infty}^{+\infty}$, стационарные в широком смысле с $E\{x_t\} = 0$.

2. Пополним это множество всевозможными конечными линейными комбинациями $\sum_{t \in S} b_t x_t$, где $S \subset \mathbb{Z}$ – конечные множества целых чисел, а $b_t \in \mathbb{R}$ – неслучайные коэффициенты.

3. Полученное на предыдущем шаге множество пополним всеми пределами в среднеквадратическом последовательностей из этого множества.

Обозначим полученное пространство через $H^{c. \kappa. \beta.}$. В этом пространстве определено скалярное произведение двух элементов

$$(x, y) = x \cdot y = E\{xy\},$$

порождающее соответствующую норму $\|x\| = \sqrt{(x, x)} = \sqrt{E\{x^2\}}$, и метрику

$$\rho(x, y) = \|x - y\| = \sqrt{E\{(x - y)^2\}}.$$

Для пространства $H^{c. \kappa. \beta.}$ известны два дополняющих друг друга результата [5].

Теорема 19.9. *Любой стационарный в широком смысле ВР $\{x_t\}_{t=-\infty}^{+\infty} \in H^{c. \kappa. \beta.}$ при условии, что для него существует спектральная плотность $f(\lambda)$, $\lambda \in [-\pi, \pi]$, и выполняется*

$$\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty, \text{ можно представить в виде } x_t = \sum_{j=0}^{+\infty} \gamma_j u_{t-j}, \text{ где } \{u_t\}_{t=-\infty}^{+\infty}$$

удовлетворяют условию У1; $\gamma_0 = 1$, а остальные коэффициенты $\{\gamma_j\}_{j=1}^{+\infty}$ таковы, что

$$\sum_{j=1}^{+\infty} \gamma_j^2 < +\infty.$$

Теорема 19.10. *Пусть в условиях теоремы 19.9 спектральная плотность $f(\cdot)$ ограничена: $f(\lambda) < +\infty$, $\lambda \in [-\pi, \pi]$, тогда справедливо представление*

$$u_t = \sum_{j=0}^{+\infty} \beta_j x_{t-j}; \quad \beta_0 = 1; \quad \sum_{j=1}^{+\infty} \beta_j^2 < +\infty, \quad (19.20)$$

которое имеет место тогда и только тогда, когда функция $G(z) = \sum_{j=0}^{+\infty} \gamma_j z^j \neq 0$,

$$|z| < 1, \text{ и следующий интеграл } \int_0^{2\pi} \frac{1}{|G(\rho e^{i\lambda})|^2} d\lambda \text{ ограничен при } \rho \rightarrow 1.$$

Замечание 19.7. Соотношение (19.20) известно как разложение Вольда (Wold) и является АР(+∞) (авторегрессией с бесконечным порядком), а результат теоремы 19.9 «обобщает» линейное представление из п. 19.1.

Воспользуемся результатами теорем 19.9, 19.10 и получим аналог уравнений Юла – Уокера (19.5). Умножим левую и правую части соотношения (19.20) (учтено, что

$$\beta_0 = 1) \quad x_t + \sum_{j=1}^{+\infty} \beta_j x_{t-j} = u_t \text{ на } x_{t-\tau} = \sum_{i=0}^{+\infty} \gamma_i u_{t-\tau-i}, \quad \tau \geq 1, \text{ возьмем математическое}$$

ожидание и получим уравнения Юла – Уокера вида

$$\sigma(\tau) + \sum_{j=1}^{+\infty} \beta_j \sigma(\tau - j) = 0, \quad \tau \geq 1. \quad (19.21)$$

Опишем общую схему идентификации стационарного в широком смысле ВР, основанную на (19.20), (19.21). Отметим, что если $\{\beta_j\}_{j=1}^{+\infty}$ известны, то наилучшим в сред-

неквадратическом прогнозом для x_t по x_{t-1}, x_{t-2}, \dots будет прогноз $\hat{x}_t^* = - \sum_{j=1}^{+\infty} \beta_j x_{t-j}$.

Очевидно, что задача идентификации сводится к оцениванию коэффициентов авторегрессии $\{\beta_j\}_{j=1}^{+\infty}$. Но сначала мы должны убедиться, что ВР стационарен в широком смысле. Не ограничивая общности, предположим, что $E\{x_t\} = 0$.

Изложим процедуру идентификации в виде шагов.

Шаг 0. По реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 1$ оцениваем ковариационную функцию $\sigma(h)$: $c_h = \frac{1}{T-h} \sum_{t=1}^{T-h} x_t x_{t+h}$, $h = 0, \dots, T-1$. Если c_h приближается к нулю с ростом лага ($h \rightarrow +\infty$), т. е. оценка ковариационной функции «затухает», то имеется надежда, что идентифицируемый ВР стационарен в широком смысле.

Шаг p ($p = 1, 2, \dots$). Считаем, что вместо модели $AP(+\infty)$ имеет место модель $AP(p)$, а коэффициенты $\beta_{p+1}, \beta_{p+2}, \dots$ полагаем равными нулю. Для того чтобы найти оценки $\hat{\beta}_1, \dots, \hat{\beta}_p$, решаем p уравнений Юла – Уокера вида (19.21):

$$c_\tau + \sum_{j=1}^p \hat{\beta}_j c_{\tau-j} = 0, \quad \tau = 1, \dots, p. \quad (19.22)$$

Решив СЛАУ (19.22), проверяем предположение $\beta_p = 0$.

Существуют два способа для проверки данного предположения:

1. Задать малое $\epsilon \in (0, 1)$ и проверять: если $|\hat{\beta}_p| < \epsilon$, то порядок считается равным $p-1$ и процесс идентификации останавливается, а в качестве оценок коэффициентов используются либо текущие оценки, либо оценки с предыдущего шага. Иначе, увеличиваем порядок на единицу и повторяем шаг p .
2. Использовать подход из п. 19.4, основанный на ОМП, поскольку оценки Юла – Уокера и ОМП асимптотически эквивалентны.

Замечание 19.8. В гл. 19 были получены различные типы оценок для параметров $AP(p)$, $p < +\infty$ (ОМП, МНК, Юла – Уокера). Для временных рядов $СС(q)$ и $АРСС(p, q)$ также существует теория построения статистических оценок, но все они носят рекуррентно-итерационный характер. Для любого стационарного в широком смысле ВР можно осуществлять «подгонку» к $AP(p)$, где порядок авторегрессии p отвечает за точность «подгонки». Это избавляет от необходимости производить идентификацию моделью $АРСС(p, q)$ и решать достаточно сложную задачу оценивания порядков p и q .

Замечание 19.9. Оценку последнего коэффициента авторегрессии $\hat{\beta}_p$, полученную в предположении, что для ВР $\{x_t\}_{t=-\infty}^{+\infty}$ имеет место модель $AP(p)$ (порядок авторегрессии равен p), взятую с противоположным знаком, иногда рассматривают как функцию от p , считая p лагом, и называют выборочной частной корреляционной функцией. Это связано с тем, что истинное значение $-\beta_p$ в условиях модели $AP(p)$ совпадает с частной корреляцией между отстоящими по времени на лаг p отсчетами x_t и x_{t+p} при фиксированных значениях отсчетов $x_{t+1}, \dots, x_{t+p-1}$ между ними, и все частные корреляции с лагами, большими p , в условиях модели $AP(p)$ равны нулю.

Выборочная частная корреляционная функция используется (см. также замечание 19.4) вместе с выборочной ковариационной функцией c_τ , $\tau \geq 0$ для идентификации модели $АРСС(p, q)$. Известен ряд эвристических правил [6].

Так, если $|\hat{\beta}_p|$ «резко обрывается» (стремится к 0) при значении лагов, больших p , а $|c_\tau|$ «плавно спадает» с ростом τ , то мы имеем дело с $AP(p)$.

Если $|c_\tau|$ «резко обрывается» при значении лагов больших, чем q ($\tau \geq q$), а $|\hat{\beta}_p|$ «плавно спадает» с ростом p , то имеет место $СС(q)$.

Вместо выборочной ковариационной функции c_τ можно использовать соответствующую ей выборочную корреляционную функцию: $\hat{\rho}_\tau = c_\tau / c_0$.

Приведенные выше правила достаточно просты. Однако для идентификации «смешанной» модели $АРСС(p, q)$ ($p, q \neq 0$) даже при $p = q = 1$ нет простых рекомендаций. Поэтому во избежание ошибки лучше производить идентификацию в рамках общей для всех стационарных в широком смысле временных рядов модели $AP(+\infty)$.

В качестве примера идентификации модели рассмотрим данные солнечной активности рис. 17.2. Для них на рис. 18.2 была построена выборочная корреляционная функция. Она «затухает» и «плавно спадает» по абсолютному значению с ростом лага. На рис. 19.1 приведена выборочная частная корреляционная функция, которая «обрывается» после лага, равного 2, что позволяет, согласно замечанию 19.9, идентифицировать исследуемый ВР как $AP(2)$ и произвести дальнейший статистический анализ. Однако существуют и другие мнения относительно порядка авторегрессии для этих данных [6, 5]: от 2 до 30. Это еще раз говорит о целесообразности использования модели $AP(+\infty)$ (замечание 19.8), исключающей неоднозначность в толковании.

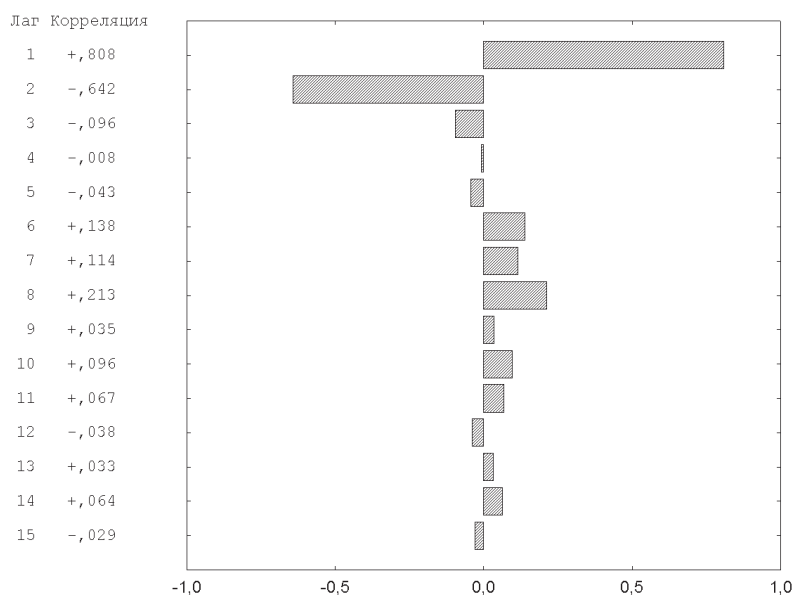


Рис. 19.1. Выборочная частная корреляционная функция для данных солнечной активности

19.6. УПРАЖНЕНИЯ

1. Показать, что для временного ряда $AP(1)$ с коэффициентом авторегрессии $\beta = \beta_1$ ($|\beta| < 1$) и дисперсией ошибок σ^2 ковариационная функция имеет вид $\sigma(\tau) = \frac{\sigma^2(-1)^\tau \beta^\tau}{1 - \beta^2}$, $\tau \geq 1$, а дисперсия отсчетов равна $\frac{\sigma^2}{1 - \beta^2}$.
2. Показать, что для временного ряда $APCC(1, 1)$ с коэффициентом авторегрессии $\beta = \beta_1$ ($|\beta| < 1$) и коэффициентом скользящего среднего $\alpha = \alpha_1$ ($|\alpha| < +\infty$), а также дисперсией ошибок $\sigma^2 < +\infty$, корреляционная функция $\rho(\tau) = \sigma(\tau)/\sigma(0) = (-1)^{\tau-1} \beta^{\tau-1} \rho(1)$, $\tau \geq 2$, где $\rho(1) = \frac{(1 - \alpha\beta)(\alpha - \beta)}{1 + \alpha^2 - 2\alpha\beta}$.
3. Доказать следствие 19.5 из п. 19.3.

Глава 20

СТАТИСТИЧЕСКИЙ АНАЛИЗ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

20.1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ С ТРЕНДОМ В ОРТОГОНАЛЬНОМ БАЗИСЕ

Пусть исследуемый ВР $\{x_t\}_{t=-\infty}^{+\infty}$, вообще говоря, нестационарен. Действительно, один и тот же интервал по времени нельзя прожить более чем один раз. Моделью нестационарного ВР, не противоречащей базовому предположению теории вероятностей и математической статистики о повторяемости исследуемых явлений, является *трендовая модель*:

$$x_t = f(t) + u_t, \quad (20.1)$$

где $f(t)$ – детерминированная (неслучайная) составляющая, зависящая от времени и называемая *трендом*. Последовательность случайных величин $\{u_t\}$ интерпретируется как *случайная ошибка измерений*, производимых над функцией $f(t)$ в различные моменты времени t . Считается, что эта ошибка порождена одним и тем же наблюдателем, и ее значения образуют стационарную в широком смысле последовательность во времени. Обычно случайные величины $\{u_t\}$ предполагаются некоррелированными с нулевыми математическими ожиданиями и конечной неизменной дисперсией:

$$E\{u_t\} = 0, \quad D\{u_t\} = \sigma^2, \quad (20.2)$$

т. е. $\{u_t\}$ удовлетворяют условию У1 из гл. 19.

На практике чаще всего применяются тренды двух видов: полиномиальные и тригонометрические. Причем тригонометрические тренды используются, если функция $f(t)$ периодическая, т. е. существует такое t^* , что $f(t) = f(t + t^*)$.

На практике модель (20.1), (20.2) уточняют и тренд $f(t)$ представляют в виде разложения в ортогональном базисе:

$$f(t) = b_1\phi_1(t) + \dots + b_m\phi_m(t), \quad (20.3)$$

где $\{\phi_j(\cdot)\}_{j=1}^{+\infty}$ – бесконечная система ортогональных базисных функций, удовлетворяющая следующим условиям:

$$\sum_{t=1}^T \phi_j(t)\phi_i(t) = \begin{cases} 0, & i \neq j; \\ \sum_{t=1}^T \phi_i^2(t), & i = j, \end{cases} \quad (20.4)$$

где T – длительность серии наблюдений $X = \{x_1, \dots, x_T\}$.

В модели (20.1)–(20.4) параметр m определяет необходимое для достижения требуемой точности «подгонки» число базисных функций и называется *порядком тренда*.

Если тренд полиномиальный, то $\phi_j(t)$, $j \geq 1$, – это полином порядка j по t ; если тригонометрический, то в качестве базисных функций используются тригонометрические функции.

Введем в рассмотрение следующие m -векторы:

$$b = (b_1, \dots, b_m)^T \in \mathbb{R}^m \quad -$$

вектор коэффициентов;

$$z_t = (\phi_1(t), \dots, \phi_m(t))^T \quad -$$

вектор значений базисных функций в момент времени t . Тогда модель (20.1)–(20.4) примет вид

$$x_t = b^T z_t + u_t,$$

где $\{u_t\}$ удовлетворяют условию У1.

Воспользуемся методом наименьших квадратов и запишем целевую функцию:

$$F(b) = \frac{1}{T} \sum_{t=1}^T (x_t - b^T z_t)^2,$$

минимизируя которую, найдем МНК-оценку $\hat{b} = \hat{b}(X) \in \mathbb{R}^m$ для неизвестного вектора коэффициентов $b \in \mathbb{R}^m$ по серии наблюдений $X = \{x_1, \dots, x_T\}$ длительности T . Необходимым условием минимума $F(b)$ по $b \in \mathbb{R}^m$ является

$$\nabla_b F(b) = \frac{1}{T} \sum_{t=1}^T 2(x_t - b^T z_t)(-z_t) = \mathbf{0}_m,$$

откуда

$$\hat{b} = \left(\sum_{t=1}^T z_t z_t^T \right)^{-1} \sum_{t=1}^T x_t z_t,$$

или через базисные функции ($\hat{b} = (\hat{b}_1, \dots, \hat{b}_m)^T \in \mathbb{R}^m$):

$$\hat{b}_i = \frac{\sum_{t=1}^T x_t \phi_i(t)}{\sum_{t=1}^T \phi_i^2(t)}, \quad i = 1, \dots, m. \quad (20.5)$$

Наконец, воспользовавшись известным результатом из регрессионного анализа, запишем оценку для дисперсии ошибок измерений σ^2 как остаточную сумму квадратов:

$$\hat{\sigma}^2 = \frac{1}{T-m} \sum_{t=1}^T \left(x_t - \sum_{i=1}^m \hat{b}_i \phi_i(t) \right)^2. \quad (20.6)$$

Исследуем статистические свойства оценок $\{\hat{b}_i\}_{i=1}^m$ из (20.5).

Теорема 20.1. Пусть в условиях модели (20.1)–(20.4) случайные ошибки измерений $\{u_t\}$ удовлетворяют условию У1, тогда оценки $\{\hat{b}_i\}_{i=1}^m$ несмещенные:

$$E\{\hat{b}_i\} = b_i, \quad i = 1, \dots, m,$$

и некоррелированные:

$$\text{cov}\{\hat{b}_i, \hat{b}_j\} = \begin{cases} 0, & i \neq j, \\ \frac{\sigma^2}{\sum_{t=1}^T \phi_i^2(t)}, & i = j, \end{cases} \quad i, j = 1, \dots, m.$$

Доказательство. Воспользуемся (20.5) и моделью (20.1) – (20.4):

$$\begin{aligned} E\{\hat{b}_i\} &= \frac{\sum_{t=1}^T E\{x_t\}\phi_i(t)}{\sum_{t=1}^T \phi_i^2(t)} = \frac{\sum_{t=1}^T f(t)\phi_i(t)}{\sum_{t=1}^T \phi_i^2(t)} = \\ &= \frac{\sum_{t=1}^T \sum_{j=1}^m b_j \phi_j(t) \phi_i(t)}{\sum_{t=1}^T \phi_i^2(t)} = \frac{b_i \sum_{t=1}^T \phi_i^2(t)}{\sum_{t=1}^T \phi_i^2(t)} = b_i, \quad i = 1, \dots, m. \end{aligned}$$

Найдем ковариацию ($i, j = 1, \dots, m$):

$$\begin{aligned} \text{cov}\{\hat{b}_i, \hat{b}_j\} &= E\left\{\left(\hat{b}_i - E\{\hat{b}_i\}\right)\left(\hat{b}_j - E\{\hat{b}_j\}\right)\right\} = E\left\{\left(\hat{b}_i - b_i\right)\left(\hat{b}_j - b_j\right)\right\} = \\ &= \frac{E\left\{\sum_{t=1}^T \phi_i(t)u_t \sum_{l=1}^T \phi_j(l)u_l\right\}}{\sum_{t=1}^T \phi_i^2(t) \sum_{l=1}^T \phi_j^2(l)} = \frac{\sigma^2 \sum_{t=1}^T \phi_i(t)\phi_j(t)}{\sum_{t=1}^T \phi_i^2(t) \sum_{l=1}^T \phi_j^2(l)} = \frac{\sigma^2}{\sum_{t=1}^T \phi_i^2(t)} \delta_{ij}. \end{aligned}$$

□

Следствие 20.1. Пусть в условиях модели (20.1)–(20.4) случайные величины $\{u_t\}$ удовлетворяют условию УЗ, тогда оценки $\{\hat{b}_i\}_{i=1}^m$ из (20.5) независимы в совокупности и распределены по нормальному закону:

$$\mathcal{L}\{\hat{b}_i\} = N_1\left(b_i, \frac{\sigma^2}{\sum_{t=1}^T \phi_i^2(t)}\right), \quad i = 1, \dots, m.$$

Доказательство. Используем результат теоремы 20.1, согласно которому $\{\hat{b}_i\}_{i=1}^m$ некоррелированы. По условию УЗ $\{u_t\}$ независимы в совокупности и распределены по нормальному закону $N_1(0, \sigma^2)$. Заметим, что оценки $\{\hat{b}_i\}_{i=1}^m$ из (20.5) являются линейными преобразованиями $\{u_t\}$. □

Как было отмечено выше, в модели (20.1)–(20.4) величина m (порядок тренда) отвечает за точность «подгонки» тренда под реальные данные. Теорема 20.1 и следствие 20.1 к ней доказаны в условиях известного m . На практике m неизвестно, и после того как выбрана система базисных функций, необходимо оценить порядок тренда. Оценивание m можно свести, как и при анализе порядка авторегрессии, к проверке гипотез ($m = 1, 2, \dots$):

$$\begin{aligned} H_0 &: b_m = 0; \\ H_1 = \bar{H}_0 &: b_m \neq 0. \end{aligned}$$

Если принята H_0 : $b_m = 0$, то полагаем порядок равным $m - 1$. Иначе, увеличивая порядок тренда на единицу, продолжаем до тех пор, пока не будет принята H_0 .

Построим критерий для проверки данных гипотез. Везде далее в этом пункте предполагаем выполнение условия УЗ, т. е. мы находимся в условиях следствия 20.1, из которого следует, что

$$\mathcal{L}_{H_0}\{\hat{b}_m\} = N_1\left(0, \frac{\sigma^2}{\sum_{t=1}^T \phi_m^2(t)}\right).$$

Значение дисперсии σ^2 считается неизвестным и для ее оценивания используется МНК-оценка $\hat{\sigma}^2$ из (20.6), для которой в регрессионном анализе установлено:

- 1) оценка $\hat{\sigma}^2$ независима от \hat{b}_m (в условиях УЗ);
- 2) $\mathcal{L}\left\{\frac{(T-m)\hat{\sigma}^2}{\sigma^2}\right\} = \chi_{T-m}^2$ – χ^2 -распределение с $T-m$ степенями свободы.

Введем в рассмотрение статистику

$$\gamma_T = \frac{\hat{b}_m \sqrt{\sum_{t=1}^T \phi_m^2(t)}}{\hat{\sigma}}, \quad \hat{\sigma} = \sqrt{\hat{\sigma}^2}, \quad (20.7)$$

которая при истинной гипотезе H_0 имеет t -распределение Стьюдента с $T-m$ степенями свободы $\mathcal{L}_{H_0}\{\gamma_T\} = t_{T-m}$.

Воспользуемся (20.7), свойствами распределения Стьюдента и получим критерий для проверки гипотез H_0, H_1 .

При истинной H_0 : $b_m = 0$ и оценка \hat{b}_m должна быть близка к нулю. Это влечет за собой требование, чтобы статистика γ_T также была близка к нулю.

Соответствующий критерий имеет вид

$$\begin{cases} H_0 & : |\gamma_T| \leq \Delta; \\ H_1 = \bar{H}_0 & : |\gamma_T| > \Delta. \end{cases}$$

Задав уровень значимости $\alpha = \mathbf{P}\{H_1|H_0\} \in (0, 1)$, находим порог критерия Δ :

$$\Delta = F_{t_{T-m}}^{-1}(1 - \alpha/2) \quad -$$

квантиль уровня $1 - \alpha/2$ от t -распределения Стьюдента с $T-m$ степенями свободы.

Замечание 20.1. В модели (20.1)–(20.4) на практике часто перед тем как проводить статистическое оценивание, производят так называемое сглаживание. Эта процедура не может быть теоретически обоснована, но иногда позволяет улучшить «подгонку». Суть ее состоит в том, что исходный ВР $\{x_t\}_{t=-\infty}^{+\infty}$ заменяется новым (сглаженным) ВР $\{y_t\}_{t=-\infty}^{+\infty}$:

$$y_t = \sum_{j=-m}^m a_j x_{t+j}, \quad (20.8)$$

где $a_j \in \mathbb{R}$, $j = -m, \dots, m$, – некоторые веса: $\sum_{j=-m}^m a_j = 1$.

Согласно (20.8) в условиях модели (20.1)–(20.4) можно записать

$$y_t = \sum_{j=-m}^m a_j f(t+j) + \sum_{j=-m}^m a_j u_{t+j} = f^*(t) + u_t^*.$$

Если в диапазоне значений времени $t + j$, $j = -m, \dots, m$: $f(t + j) \approx f(t)$, тогда значение нового тренда в любой момент времени t близко к $f(t)$: $f^*(t) \approx f(t)$. Однако новые случайные величины $\{u_t^*\}$ станут коррелированными, а их дисперсия

$$D\{u_t^*\} = \sum_{j=-m}^m a_j^2 \sigma^2.$$

Исходя из последнего соотношения, веса $\{a_j\}$ необходимо выбирать так, чтобы $D\{u_t^*\} < \sigma^2$; на практике часто используются веса $a_j = \frac{1}{2m+1}$, $j = -m, \dots, m$.

20.2. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ВРЕМЕННЫХ РЯДОВ АВТОРЕГРЕССИИ И ПРОИНТЕГРИРОВАННОГО СКОЛЬЗЯЩЕГО СРЕДНЕГО

Пусть имеется ВР $\{x_t\}_{t=-\infty}^{+\infty}$. Определим на нем два оператора:

а) B – оператор сдвига на один шаг назад:

$$Bx_t = x_{t-1};$$

б) ∇ – оператор взятия конечной разности на один шаг назад:

$$\nabla = 1 - B, \quad \nabla x_t = x_t - x_{t-1}.$$

Приведем основные свойства этих операторов.

С1. Оператор $S = \nabla^{-1}$, обратный оператору ∇ , имеет следующий вид:

$$S = (1 - B)^{-1} = \sum_{j=0}^{+\infty} B^j \quad (B^0 := 1), \quad Sx_t = \sum_{j=0}^{+\infty} x_{t-j}.$$

С2. Оператор $\nabla^d = \underbrace{\nabla \dots \nabla}_d$ конечной разности порядка d имеет вид

$$\nabla^d = (1 - B)^d = \sum_{j=0}^d (-1)^j C_d^j B^j, \quad \nabla^d x_t = \sum_{j=0}^d (-1)^j C_d^j x_{t-j},$$

где $\{C_d^j\}$ – биномиальные коэффициенты.

С3. Оператор S^d суммирования порядка d может быть записан следующим образом:

$$S^d = (1 - B)^{-d} = \sum_{j=0}^{+\infty} C_{d+j-1}^{d-1} B^j, \quad S^d x_t = \sum_{j=0}^{+\infty} C_{d+j-1}^{d-1} x_{t-j}.$$

С4. Оператор суммирования S^d порядка d является обратным оператором по отношению к ∇^d :

$$S^d \nabla^d = 1.$$

Запишем через оператор B временной ряд авторегрессии и скользящего среднего (АРСС(p, q)):

$$x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} = u_t + \alpha_1 u_{t-1} + \dots + \alpha_q u_{t-q}. \quad (20.9)$$

Введем в рассмотрение два полинома по B :

$$\Psi(B) = 1 + \beta_1 B + \dots + \beta_p B^p = \sum_{j=0}^p \beta_j B^j, \quad \beta_0 = 1;$$

$$\Theta(B) = 1 + \alpha_1 B + \dots + \alpha_q B^q = \sum_{j=0}^q \alpha_j B^j, \quad \alpha_0 = 1,$$

тогда модель АРСС(p, q) из (20.9) можно записать в следующем виде:

$$\Psi(B)x_t = \Theta(B)u_t. \quad (20.10)$$

Замечание 20.2. Для стационарности ВР (20.9) мы требовали, чтобы корни характеристического уравнения (19.4) лежали внутри единичного круга. Выполнение этого условия обеспечивает также и существование обратного оператора $\Psi^{-1}(B)$.

Определение 20.1. ВР $\{x_t\}_{t=-\infty}^{+\infty}$ называется *временным рядом авторегрессии и проинтегрированного скользящего среднего порядков p, d, q (АРПСС(p, d, q))*, если он удовлетворяет следующему соотношению:

$$\Psi(B) \nabla^d x_t = \Theta(B)u_t, \quad (20.11)$$

где $\Psi(B)$ – оператор АР(p); $\Theta(B)$ – оператор СС(q); $\nabla^d x_t$ – конечная разность порядка d ; а случайные величины $\{u_t\}$ удовлетворяют одному из условий U – чаще всего предполагается выполнение $U1$, т. е. некоррелированность $\{u_t\}$ и $\mathbf{E}\{u_t\} = 0$, $\mathbf{D}\{u_t\} = \sigma^2$.

Замечание 20.3. Если обозначить $w_t = \nabla^d x_t$, то в условиях (20.11) ВР $\{w_t\}$ является АРСС(p, q) и описывается моделью (20.9) или (20.10).

Получим три эквивалентных представления для модели (20.11). Введем в рассмотрение новый оператор:

$$\Phi(B) = \Psi(B) \nabla^d = \sum_{j=0}^p \beta_j B^j (1 - B)^d = \sum_{j=0}^{p+d} \phi_j B^j, \quad \phi_0 = 1.$$

Таким образом, модель АРПСС(p, d, q) из (20.11) формально может быть записана в виде АРСС($p + d, q$):

$$\Phi(B)x_t = \Theta(B)u_t. \quad (20.12)$$

Однако ряд (20.12) нестационарен, поскольку для него характеристическое уравнение

$$\sum_{j=0}^{p+d} \phi_j z^{p+d-j} = 0$$

имеет d корней z , по модулю равных единице: $|z| = 1$.

Выразим x_t через u_t и получим еще одно эквивалентное представление:

$$x_t = \Phi^{-1}(B)\Theta(B)u_t = \Psi^{-1}(B) \nabla^{-d} \Theta(B)u_t = \sum_{j=0}^{+\infty} \gamma_j u_{t-j}, \quad \gamma_0 = 1.$$

Возможно также представление, аналогичное АР($+\infty$):

$$u_t = \Theta^{-1}(B)\Psi(B) \nabla^d x_t = x_t + \sum_{j=1}^{+\infty} \pi_j x_{t-j}.$$

Приведем алгоритм идентификации для модели АРПСС(p, d, q).

Шаг d ($d = 0, 1, 2, \dots$). Вычисляем конечную разность порядка d :

$$w_t^{(d)} = \nabla^d x_t = \nabla w_t^{(d-1)}, \quad d \geq 1 \quad (w_t^{(0)} = x_t).$$

Проверяем ВР $\{w_t^{(d)}\}_{t=-\infty}^{+\infty}$ на стационарность. Например, строим выборочную ковариационную или корреляционную функцию. Если она «затухает», то мы считаем, что $\{w_t^{(d)}\}_{t=-\infty}^{+\infty}$ стационарен и переходим к *заключительному шагу*. Если нет, то увеличиваем d на единицу и повторяем этот шаг.

Заключительный шаг. Если при каком-то d ВР $\{w_t^{(d)}\}_{t=-\infty}^{+\infty}$ «оказался» стационарным, то идентифицируем его как АРСС(p, q), тем самым мы сможем построить прогноз $\hat{w}_t^{*(d)}$. Чтобы построить прогноз для x_t , необходимо осуществить обратное преобразование:

$$\hat{x}_t^* = \nabla^{-d} \hat{w}_t^{*(d)} = S^d \hat{w}_t^{*(d)}.$$

Замечание 20.4. Пусть на *шаге d* ($d \in \{0, 1, 2, \dots\}$) описанного выше алгоритма преобразованный ВР $\{w_t^{(d)}\}_{t=-\infty}^{+\infty}$ идентифицирован как АРСС(p, q). После того как найдены p и q , строятся оценки для коэффициентов $\{\beta_j\}_{j=1}^p$ и $\{\alpha_i\}_{i=1}^q$. Однако согласно соображениям, изложенным в замечаниях 19.8 и 19.9, целесообразнее этот преобразованный ряд идентифицировать общей для всех стационарных в широком смысле временных рядов моделью Вольда АР($+\infty$), добиваясь адекватности преобразованного ряда модели АР(p) за счет увеличения порядка p .

Для проверки адекватности на *заключительном шаге* алгоритма для уже имеющихся значений $\{x_t\}_{t=T_-}^T$ строятся прогнозы $\{\hat{x}_t^*\}_{t=T_-}^T$, где T_- – момент начала прогнозирования ($T_- < T$), и вычисляются остатки: $\hat{u}_t^* = x_t - \hat{x}_t^*$, $t = T_-, \dots, T$, которые должны быть близки к нулю и некоррелированы, если хорошо «подогнана» модель. Чтобы выяснить это, для ряда остатков $\{\hat{u}_t^*\}_{t=T_-}^T$ оценивается ковариационная (корреляционная) функция: ее значения должны быть близки к нулю при любом ненулевом лаге.

Замечание 20.5. Для приведения исходного ВР $\{x_t\}_{t=-\infty}^{+\infty}$ к стационарному на практике иногда вместо конечной разности на один шаг назад ∇ используется конечная разность на k шагов назад ($k \geq 1$): $\nabla(-k)x_t = x_t - x_{t-k}$, обобщающая оператор ∇ : $\nabla = \nabla(-1)$.

Такое преобразование целесообразно, если известно, что исходный ВР имеет так называемую «сезонную составляющую», т. е. его «поведение» повторяется через один и тот же промежуток времени k .

В качестве примера проведем идентификацию ВР цены на акции IBM (см. рис. 17.1). Ранее проведенный анализ выборочной корреляционной функции (см. рис. 18.1) показал его нестационарность. Возьмем конечную разность на один шаг назад, т. е. построим ВР $\{w_t^{(1)}\}$ (рис. 20.1).

Выборочные корреляционная и частная корреляционная функции для $\{w_t^{(1)}\}$ приведены соответственно на рис. 20.2 и рис. 20.3. Их анализ согласно замечанию 19.9 позволяет идентифицировать ВР $\{w_t^{(1)}\}$ как СС(1), а исходный ВР $\{x_t\}$ – как АРПСС(0,1,1) [6].

Однако руководствуясь замечанием 20.4, можно поступить иначе: проанализировав выборочную корреляционную функцию для ВР $\{w_t^{(1)}\}$ (рис. 20.2), заключить, что пре-

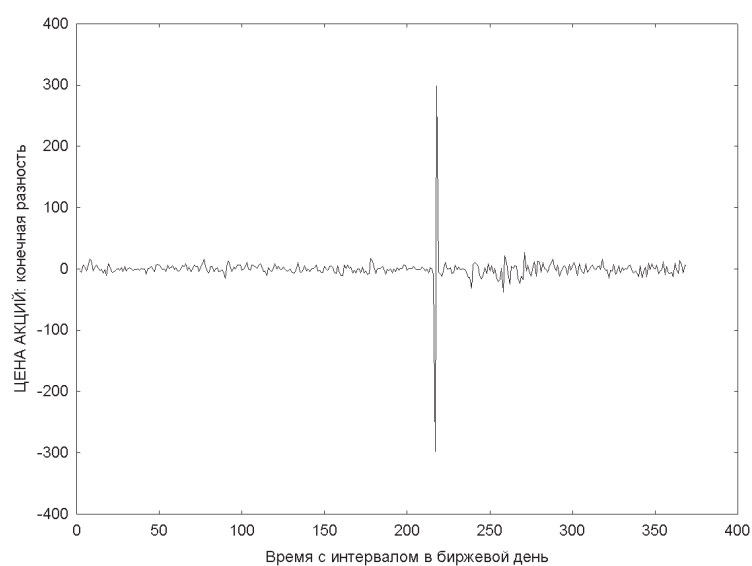


Рис. 20.1. Преобразованный временной ряд цены акций IBM:
конечная разность на один шаг назад

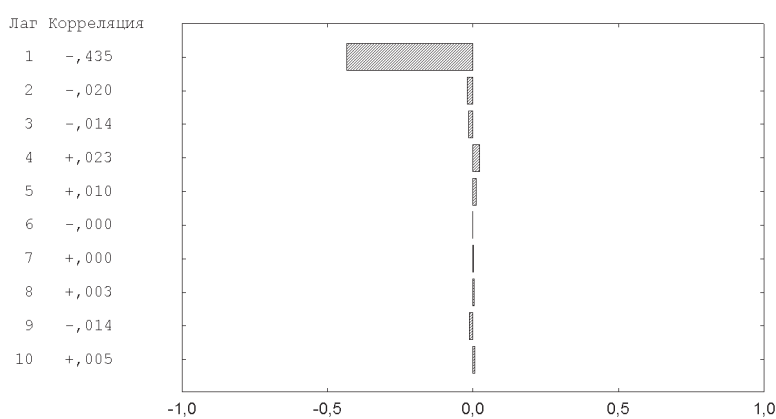


Рис. 20.2. Выборочная корреляционная функция
для преобразованного временного ряда

образованный ВР $\{w_t^{(1)}\}$ подозрителен на стационарность и идентифицировать его как $AP(+\infty)$, а исходный ВР $\{x_t\}$ – как $АРСС(p,1,0)$, добиваясь приемлемой точности «подгонки» за счет выбора порядка авторегрессии p .

При $p = 7$ были получены следующие МП-оценки коэффициентов авторегрессии:

$$\hat{\beta}_1 = -0,616, \hat{\beta}_2 = -0,404, \hat{\beta}_3 = -0,259, \hat{\beta}_4 = -0,127, \\ \hat{\beta}_5 = -0,047, \hat{\beta}_6 = -0,010, \hat{\beta}_7 = 0,003.$$

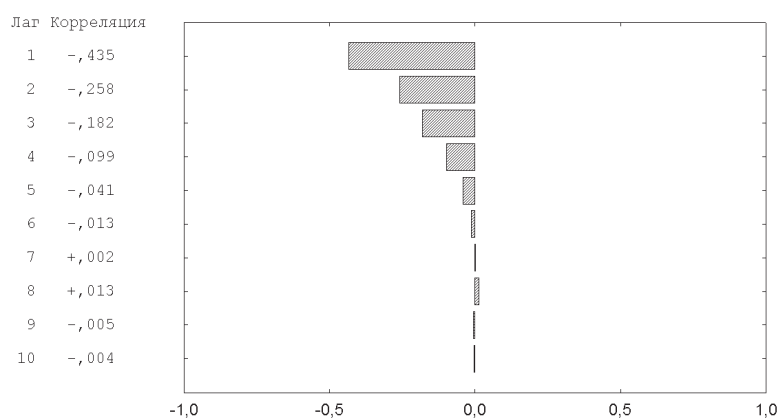


Рис. 20.3. Выборочная частная корреляционная функция для преобразованного временного ряда

Видно, что уже пятый коэффициент авторегрессии при выбранной модели АРПСС(7,1,0) близок к нулю, а остатки также близки к нулю (рис. 20.4) и некоррелированы (рис. 20.5).

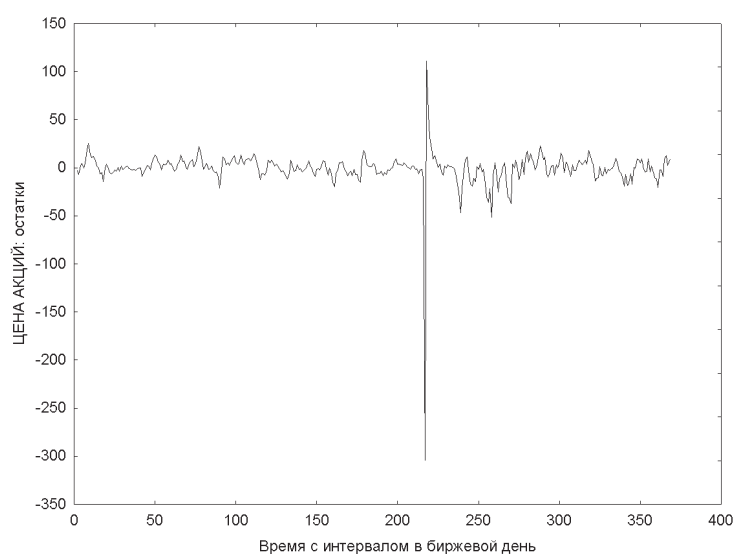


Рис. 20.4. График остатков

Наконец, на рис. 20.6 приведен результат прогнозирования (для иллюстрации проведенного, начиная с некоторого интервала времени в прошлом на такой же интервал времени в будущее).

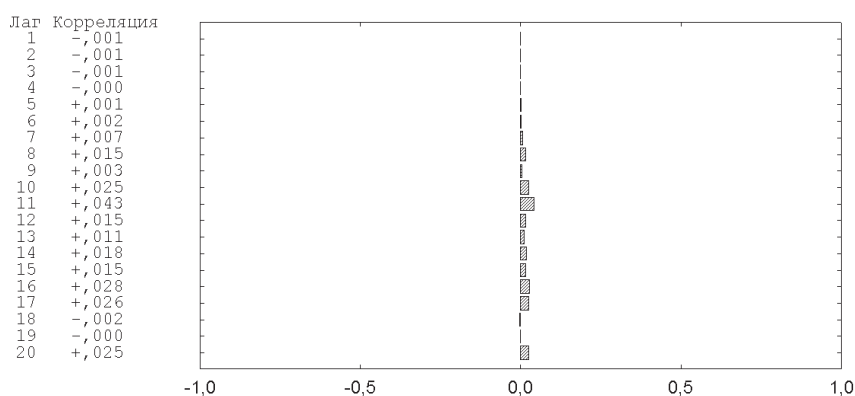


Рис. 20.5. Выборочная корреляционная функция остатков

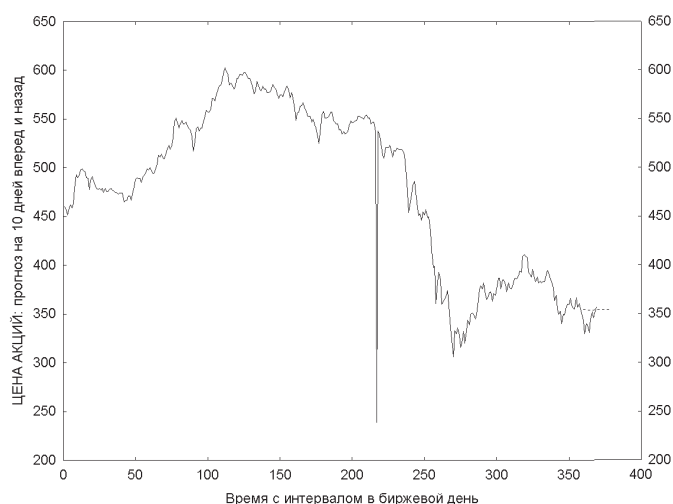


Рис. 20.6. Прогнозирование цены акций IBM

20.3. УПРАЖНЕНИЯ

1. Показать, что для временного ряда АРПСС(0, 1, 1) с коэффициентом скользящего среднего $\alpha = \alpha_1$ коэффициенты $\{\pi_j\}_{j=1}^{+\infty}$ в его представлении АР(+∞) (см. п. 20.2) имеют вид $\pi_j = (1 + \alpha)(-1)^j \alpha^{j-1}$, $j \geq 1$.
2. Показать, что для временного ряда АРПСС(1, 1, 1) с коэффициентом авторегрессии $\beta = \beta_1$ и коэффициентом скользящего среднего $\alpha = \alpha_1$ в представлении АР(+∞) коэффициенты могут быть вычислены из соотношений:

$$\pi_1 = \beta - (1 + \alpha), \quad \pi_2 = (\alpha - \beta)(1 + \alpha), \quad \pi_j = (\beta - \alpha)(1 + \alpha)(-1)^{j-1} \alpha^{j-2}, \quad j \geq 3.$$
3. Получить для временного ряда АРПСС(0, d, q) формальное представление в виде модели АРСС(d, q) со следующими коэффициентами авторегрессии: $\beta_j = (-1)^j C_d^j$, $j = 1, \dots, d$, где $\{C_d^j\}$ – биномиальные коэффициенты.

Глава 21

СТАТИСТИЧЕСКИЙ АНАЛИЗ ОДНОРОДНЫХ ЦЕПЕЙ МАРКОВА

21.1. ОДНОРОДНЫЕ ЦЕПИ МАРКОВА. ОСНОВНЫЕ ПОНЯТИЯ И СВОЙСТВА. УСЛОВИЯ СТАЦИОНАРНОСТИ

На практике для описания зависимости в последовательности событий (состояний) часто используется однородная цепь Маркова, являющаяся дискретным временным рядом.

Определение 21.1. ВР $x_1, \dots, x_T, x_{T+1}, \dots \in \mathcal{S}$, где $\mathcal{S} = \{1, \dots, L\}$ называется однородной цепью Маркова с пространством состояний \mathcal{S} , образованным из $L \geq 2$ состояний, если выполняется так называемое марковское свойство:

$$\begin{aligned} P\{x_{t+1} = d_{t+1} | x_t = d_t, x_{t-1} = d_{t-1}, \dots, x_1 = d_1\} = \\ = P\{x_{t+1} = d_{t+1} | x_t = d_t\} = p_{d_t, d_{t+1}}, \\ \forall d_t, d_{t+1}, \dots \in \mathcal{S}, \quad \forall t = 1, \dots, T, \dots \end{aligned} \quad (21.1)$$

Отметим, что данное определение является частным случаем понятия цепи Маркова из главы 9.

Для того чтобы однозначно определить однородную цепь Маркова, необходимо задать две вероятностные характеристики:

1) начальное распределение вероятностей:

$$\pi_i^{(1)} = P\{x_1 = i\}, \quad i \in \mathcal{S}; \quad \sum_{i \in \mathcal{S}} \pi_i^{(1)} = 1; \quad (21.2)$$

2) матрицу вероятностей одношаговых переходов:

$$\begin{aligned} P = (p_{ij})_{i,j \in \mathcal{S}}, \quad p_{ij} = P\{x_{t+1} = j | x_t = i\}; \\ \sum_{j \in \mathcal{S}} p_{ij} = 1, \quad i \in \mathcal{S}. \end{aligned} \quad (21.3)$$

Распределение вероятностей реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 2$ однородной цепи Маркова (21.1)–(21.3) однозначно определяется через $\pi^{(1)} = (\pi_i^{(1)})_{i \in \mathcal{S}}$ и $P = (p_{ij})_{i,j \in \mathcal{S}}$:

$$\begin{aligned} P_T(d_1, \dots, d_T) = P\{x_1 = d_1, \dots, x_T = d_T\} = \\ = P\{x_1 = d_1\} \prod_{t=1}^{T-1} P\{x_{t+1} = d_{t+1} | x_t = d_t, \dots, x_1 = d_1\} = \\ = \pi_{d_1}^{(1)} \prod_{t=1}^{T-1} p_{d_t, d_{t+1}}, \quad d_1, \dots, d_T \in \mathcal{S}. \end{aligned} \quad (21.4)$$

Цепь Маркова характеризуется также своим *текущим распределением вероятностей* (в момент времени t):

$$\pi_i^{(t)} = P\{x_t = i\}, \quad i \in \mathcal{S}, \quad t = 1, 2, \dots, T, \dots \quad (21.5)$$

Теорема 21.1. В условиях модели (21.1)–(21.3) *текущее распределение вероятностей* (21.5) может быть найдено из соотношения

$$\pi^{(t)} = (P^T)^{t-1} \pi^{(1)} = (P^{t-1})^T \pi^{(1)}, \quad (21.6)$$

где $\pi^{(t)} = (\pi_1^{(t)}, \dots, \pi_L^{(t)})^T$, $t = 1, 2, \dots, T, \dots$; а $\pi^{(1)} = \pi^{(t)}|_{t=1}$ – начальное распределение вероятностей.

Доказательство. Вычислим $\pi_i^{(t)}$, $i \in \mathcal{S}$, через предыдущее:

$$\pi_i^{(t)} = P\{x_t = i\} = \sum_{j \in \mathcal{S}} P\{x_{t-1} = j\} P\{x_t = i | x_{t-1} = j\} = \sum_{j \in \mathcal{S}} \pi_j^{(t-1)} p_{ji},$$

или в матричном виде:

$$\pi^{(t)} = P^T \pi^{(t-1)} = (P^T)^2 \pi^{(t-2)} = (P^T)^{t-1} \pi^{(1)}.$$

□

Замечание 21.1. Матрица $P^t = \underbrace{P \times \dots \times P}_t$ имеет самостоятельный смысл: ее (i, j) -й элемент – вероятность перехода однородной цепи Маркова из i -го состояния в j -е за t шагов.

Определение 21.2. Распределение вероятностей $\pi = (\pi_1, \dots, \pi_L)^T$ называется *стационарным распределением для однородной цепи Маркова* (21.1)–(21.3), если выполняется

$$\begin{cases} P^T \pi = \pi, \\ \sum_{i \in \mathcal{S}} \pi_i = 1. \end{cases} \quad (21.7)$$

Теорема 21.2. Пусть для однородной цепи Маркова (21.1)–(21.3) начальное распределение вероятностей $\pi^{(1)}$ совпадает со стационарным из (21.7): $\pi^{(1)} = \pi$, тогда данная цепь Маркова является стационарной, и текущее распределение вероятностей в любой момент времени совпадает со стационарным:

$$\pi^{(t)} = \pi, \quad t = 1, 2, \dots, T, \dots \quad (21.8)$$

Доказательство. Подставим в (21.6) $\pi^{(1)} = \pi$ и воспользуемся соотношением (21.7): $\pi^{(t)} = (P^T)^{t-1} \pi = (P^T)^{t-2} P^T \pi = (P^T)^{t-2} \pi = \pi$. □

Замечание 21.2. Если в условиях теоремы 21.2 начальное распределение $\pi^{(1)}$ не совпадает со стационарным распределением (21.7), то при некоторых ограничениях на матрицу вероятностей одношаговых переходов $P = (p_{ij})_{i,j \in \mathcal{S}}$ для текущего распределения (21.5) выполняется [36] (см. также п. 9.2):

$$\pi_i^{(t)} \rightarrow \pi_i, \quad t \rightarrow +\infty, \quad i \in \mathcal{S}.$$

21.2. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ДЛЯ ОЦЕНИВАНИЯ МАТРИЦЫ ВЕРОЯТНОСТЕЙ ОДНОШАГОВЫХ ПЕРЕХОДОВ И СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Пусть имеется реализация $X = \{x_1, \dots, x_T\}$ длительности $T \geq 2$ однородной цепи Маркова (21.1)–(21.3). Ставится задача оценить матрицу вероятностей одношаговых переходов P и стационарное распределение π . Распределение вероятностей $\pi^{(1)}$ начального состояния оценить невозможно, поскольку оно встречается один раз и для него невозможно «набрать статистику». Для построения оценок \hat{P} , $\hat{\pi}$ по реализации X воспользуемся методом максимального правдоподобия, основываясь на распределении вероятностей для X из (21.4).

Введем следующие обозначения:

$$n_{ij} = \sum_{t=1}^{T-1} \delta_{x_t, i} \delta_{x_{t+1}, j} \quad - \quad (21.9)$$

число одношаговых переходов из состояния i в состояние j ($i, j \in \mathcal{S}$), встретившихся в реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 2$;

$$n_i = \sum_{j \in \mathcal{S}} n_{ij}, \quad \tilde{n}_i = \sum_{t=1}^T \delta_{x_t, i} \quad (\tilde{n}_i = n_i + \delta_{x_T, i}), \quad i \in \mathcal{S}. \quad (21.10)$$

Теорема 21.3. Пусть для реализации $X = \{x_1, \dots, x_T\}$ длительности $T \geq 2$ однородной цепи Маркова (21.1)–(21.3) выполняется условие $n_i > 0$, $i \in \mathcal{S}$, тогда оценка максимального правдоподобия (МП-оценка) матрицы вероятностей одношаговых переходов задается соотношением

$$\hat{P} = (\hat{p}_{ij})_{i,j \in \mathcal{S}}, \quad \hat{p}_{ij} = \frac{n_{ij}}{n_i}. \quad (21.11)$$

Доказательство. С учетом (21.4) построим логарифмическую функцию правдоподобия по $X = \{x_1, \dots, x_T\}$ длительности $T \geq 2$:

$$l_T(X; P) = \ln \left(\pi_{x_1}^{(1)} \prod_{t=1}^{T-1} p_{x_t, x_{t+1}} \right) = \ln \pi_{x_1}^{(1)} + \sum_{t=1}^{T-1} \ln p_{x_t, x_{t+1}} = \ln \pi_{x_1}^{(1)} + \sum_{i,j \in \mathcal{S}} n_{ij} \ln p_{ij},$$

где статистики $n_{ij} = n_{ij}(X)$, $i, j \in \mathcal{S}$, определены в (21.9).

Согласно (21.3) элементы матрицы вероятностей одношаговых переходов должны удовлетворять условию нормировки, а сама экстремальная задача по построению МП-оценки будет иметь вид

$$\begin{cases} l_T(X; P) = \ln \pi_{x_1}^{(1)} + \sum_{i,j \in \mathcal{S}} n_{ij} \ln p_{ij} \rightarrow \max_{\{p_{ij}\}_{i,j \in \mathcal{S}}}, \\ \sum_{j \in \mathcal{S}} p_{ij} = 1, \quad i \in \mathcal{S}. \end{cases} \quad (21.12)$$

Задача (21.12) распадается на $L \geq 2$ отдельных задач ($i \in \mathcal{S}$):

$$\begin{cases} \sum_{j \in \mathcal{S}} n_{ij} \ln p_{ij} \rightarrow \max_{\{p_{ij}\}_{j \in \mathcal{S}}}, \\ \sum_{j \in \mathcal{S}} p_{ij} = 1. \end{cases}$$

Решим каждую задачу по отдельности. Для i -й задачи запишем функцию Лагранжа:

$$l_T^{(i)}(X; \{p_{ij}\}_{j \in \mathcal{S}}, \lambda_i) = \sum_{j \in \mathcal{S}} n_{ij} \ln p_{ij} + \lambda_i \left(\sum_{j \in \mathcal{S}} p_{ij} - 1 \right), \quad (21.13)$$

где λ_i – множитель Лагранжа.

Необходимым условием максимума функции Лагранжа (21.13) является равенство первых производных нулю, где производные вычисляются по переменным $\{p_{ij}\}_{j \in \mathcal{S}}$ и множителю λ_i :

$$\begin{aligned} \frac{\partial}{\partial p_{ij}} l_T^{(i)}(X; \{p_{ij}\}_{j \in \mathcal{S}}, \lambda_i) &= \frac{n_{ij}}{p_{ij}} + \lambda_i = 0, \quad j \in \mathcal{S}; \\ \frac{\partial}{\partial \lambda_i} l_T^{(i)}(X; \{p_{ij}\}_{j \in \mathcal{S}}, \lambda_i) &= \sum_{j \in \mathcal{S}} p_{ij} - 1 = 0. \end{aligned}$$

Из первых L уравнений найдем

$$p_{ij} = -\frac{n_{ij}}{\lambda_i}, \quad j \in \mathcal{S},$$

и подставим в последнее уравнение:

$$\sum_{j \in \mathcal{S}} \left(-\frac{n_{ij}}{\lambda_i} \right) - 1 = 0,$$

откуда с учетом обозначений (21.10)

$$\lambda_i = -\sum_{j \in \mathcal{S}} n_{ij} = -n_i,$$

и окончательно имеем $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$, что совпадает с (21.11). \square

Чтобы оценить стационарное распределение вероятностей $\pi = (\pi_1, \dots, \pi_L)^T$, согласно подстановочному принципу и определению (21.7) достаточно решить следующую СЛАУ:

$$\begin{cases} \hat{P}^T \hat{\pi} = \hat{\pi}, \\ \sum_{i \in \mathcal{S}} \hat{\pi}_i = 1, \quad \hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_L)^T, \end{cases} \quad (21.14)$$

где $\hat{P} = (\hat{p}_{ij})_{i,j \in \mathcal{S}}$ – МП-оценка матрицы вероятностей одношаговых переходов из (21.11).

Оценку $\hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_L)^T$, определяемую (21.14), также иногда называют МП-оценкой, однако это неверно.

Для оценивания стационарного распределения вероятностей интуитивно можно предложить и другой метод, построив оценку

$$\tilde{\pi} = (\tilde{\pi}_1, \dots, \tilde{\pi}_L)^T; \quad \tilde{\pi}_i = \frac{1}{T} \sum_{t=1}^T \delta_{x_t, i} = \frac{\tilde{n}_i}{T}, \quad i \in \mathcal{S}. \quad (21.15)$$

Если начальное распределение совпадает со стационарным, то из теоремы 21.2 (соотношение (21.8)) следует, что

$$\mathbf{E}\{\tilde{\pi}_i\} = \pi_i, \quad i \in \mathcal{S}.$$

Иначе, исходя из замечания 21.2 к теореме 21.2, можно считать, что

$$\mathbf{E}\{\tilde{\pi}_i\} \rightarrow \pi_i, \quad T \rightarrow +\infty, \quad i \in \mathcal{S}.$$

Однако оценка (21.15) также не является МП-оценкой в традиционном смысле, поскольку даже при $\pi^{(1)} = \pi$, когда цепь Маркова стационарна: $\pi^{(t)} = \pi$, ее отсчеты $\{x_t\}_{t=1}^T$, вообще говоря, зависимы между собой, а оценка (21.15) является МП-оценкой вероятностей состояний лишь для схемы независимых испытаний.

21.3. ПРОВЕРКА ГИПОТЕЗ О ЦЕПИ МАРКОВА

По реализации ВР $X = \{x_1, \dots, x_T\}$ длительности $T \geq 2$ на практике часто необходимо выяснить, зависимы или независимы отсчеты $\{x_t\}_{t=1}^T$.

В условиях модели (21.1)–(21.3) это приводит к задаче проверки следующих гипотез:

$$\begin{aligned} H_0 &: p_{ij} = \pi_j, \quad i, j \in \mathcal{S}; \\ H_1 = \overline{H_0} &: \exists i, j \in \mathcal{S} : p_{ij} \neq \pi_j. \end{aligned}$$

Гипотеза H_0 означает, что вероятность перехода в j -е состояние ($j \in \mathcal{S}$) не зависит от предыдущего состояния:

$$P\{x_{t+1} = j | x_t = i\} = \pi_j, \quad \forall i \in \mathcal{S},$$

т. е. исследуемая однородная цепь Маркова является схемой независимых испытаний.

Проверка гипотез H_0, H_1 основана на следующем критерии, который использует МП-оценку (21.9)–(21.11) для матрицы вероятностей одношаговых переходов, полученную в теореме 21.3:

$$\begin{cases} H_0 &: \gamma_T \leq \Delta; \\ H_1 = \overline{H_0} &: \gamma_T > \Delta, \end{cases}$$

где статистика критерия γ_T является обобщением χ^2 -статистики:

$$\gamma_T = \sum_{i,j \in \mathcal{S}} \frac{(n_{ij} - n_i n_j / T)^2}{n_i n_j / T}.$$

Определим порог критерия Δ . Используя теорему 21.3 и свойства МП-оценок, можно показать, что в условиях H_0 статистика γ_T имеет при $T \rightarrow +\infty$ χ^2 -распределение с $(L^2 - L) - (L - 1) = (L - 1)^2$ степенями свободы. Зададим уровень значимости $\alpha = P\{H_1 | H_0\} \in (0, 1)$, тогда

$$\Delta = F_{\chi_{(L-1)^2}^2}^{-1}(1 - \alpha) -$$

квантиль уровня $1 - \alpha$ от χ^2 -распределения с $(L - 1)^2$ степенью свободы.

Для однородной цепи Маркова решаются и другие задачи по проверке гипотез. Статистики критериев для решения таких задач, как и в задаче проверки на независимость, являются обобщениями χ^2 -статистики.

Например, пусть проверяются гипотезы о том, что матрица вероятностей одношаговых переходов $P = (p_{ij})_{i,j \in \mathcal{S}}$ совпадает с наперед заданной матрицей $P^o = (p_{ij}^o)_{i,j \in \mathcal{S}}$:

$$\begin{aligned} H_0 &: P = P^o; \\ H_1 = \overline{H_0} &: P \neq P^o \quad (\exists i, j \in \mathcal{S} : p_{ij} \neq p_{ij}^o). \end{aligned}$$

Тогда H_0 принимается, если

$$\gamma_T = \sum_{\substack{i,j \in \mathcal{S} \\ p_{ij}^o \neq 0}} \frac{(n_{ij} - n_i p_{ij}^o)^2}{n_i p_{ij}^o} \leq \Delta,$$

$\Delta = F_{\chi^2_{m-L}}^{-1}(1 - \alpha)$ – порог критерия, определяемый по уровню значимости $\alpha = P\{H_1|H_0\} \in (0, 1)$ через квантиль χ^2 -распределения с $m - L$ степенями свободы, где $m = L^2 - \sum_{i,j \in \mathcal{S}} \delta_{p_{ij}^o, 0}$ – число ненулевых вероятностей одношаговых переходов в гипотетической $(L \times L)$ -матрице $P^o = (p_{ij}^o)_{i,j \in \mathcal{S}}$.

21.4. УПРАЖНЕНИЯ

1. Пусть матрица вероятностей одношаговых переходов $P = (p_{ij})_{i,j \in \mathcal{S}}$ однородной цепи Маркова с пространством состояний $\mathcal{S} = \{1, \dots, L\}$ является бистохастической: $\sum_{j \in \mathcal{S}} p_{ij} = \sum_{j \in \mathcal{S}} p_{ji} = 1, i \in \mathcal{S}$. Показать, что стационарное распределение вероятностей $\pi = (\pi_i)_{i \in \mathcal{S}}$ имеет вид $\pi_i = 1/L, i \in \mathcal{S}$.
2. Установить, что однородная цепь Маркова с $L \geq 2$ состояниями превращается в схему независимых испытаний с равновероятными исходами: $\pi_i = 1/L, i \in \mathcal{S}$, в случае, когда ее матрица вероятностей одношаговых переходов имеет вид $P = (p_{ij})_{i,j \in \mathcal{S}}, p_{ij} = 1/L, i, j \in \mathcal{S}$.
3. Доказать, что статистика γ_T , используемая в п. 21.3 для проверки гипотез независимости, в случае отсутствия зависимости (верна гипотеза H_0) имеет при $T \rightarrow +\infty$ χ^2 -распределение с $(L-1)^2$ степенями свободы (T – длительность реализации, $L \geq 2$ – число состояний цепи Маркова или число исходов в схеме независимых испытаний).
4. Монета подбрасывается T раз: $X = \{x_1, \dots, x_T\}$, где $x_t = 1$, если выпала «решка»; $x_t = 2$, если – «орел» ($t = 1, \dots, T$). Предполагая возможное наличие марковской зависимости между бросаниями, а также возможную несимметричность монеты, проверить, что на самом деле монета симметрична, а зависимость отсутствует. Указание: воспользоваться критерием для проверки гипотез о значении матрицы вероятностей одношаговых переходов из п. 21.3, положив в нем $P^0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ (см. также упражнение 2).

ЛИТЕРАТУРА

1. Айвазян, С. А. Прикладная статистика: Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. М. : Финансы и статистика, 1985.
2. Айвазян, С. А. Прикладная статистика: Классификация и снижение размерности / С. А. Айвазян [и др.]. М. : Финансы и статистика, 1989.
3. Айвазян, С. А. Прикладная статистика: Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. М. : Финансы и статистика, 1983.
4. Андерсон, Т. Введение в многомерный статистический анализ / Т. Андерсон. М. : Физматгиз, 1963.
5. Андерсон, Т. Статистический анализ временных рядов / Т. Андерсон. М. : Мир, 1976.
6. Бокс, Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. М. : Мир, Вып. 1. 1974.
7. Большев, Л. Н. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. М. : Наука, 1983.
8. Боровков, А. А. Математическая статистика / А. А. Боровков. М. : Наука, 2002.
9. Боровков, А. А. Теория вероятностей / А. А. Боровков. М. : Наука, 1998.
10. Булинский, А. В. Теория случайных процессов / А. В. Булинский, А. Н. Ширяев. М. : ФИЗМАТЛИТ, 2003.
11. Гантмахер, Ф. Р. Теория матриц / Ф. Р. Гантмахер. М. : Наука, 1967.
12. Гихман, И. И. Теория вероятностей и математическая статистика / И. И. Гихман, А. В. Скороход, М. И. Ядренко. Киев : Наук. думка, 1983.
13. Гнеденко, Б. В. Курс теории вероятностей / Б. В. Гнеденко. М. : Наука, 1996.
14. Зуев, Н. М. Предельные теоремы для слабозависимых случайных величин / Н. М. Зуев. Минск : БГУ, 2000.
15. Зуєв, М. М., Сячко, У. У. Теорія імавернастей і математична статистика / М. М. Зуєв, У. У. Сячко. Мазыр : Видавець Дом «Белы Вецер», 2000.
16. Ивченко, Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. М. : Высш. шк., 1984.
17. Карлин, С. Основы теории случайных процессов / С. Карлин. М. : Мир, 1971.
18. Колмогоров, А. Н. Основные понятия теории вероятностей / А. Н. Колмогоров. М. : Наука, 1974.
19. Крамер, Г. Стационарные случайные процессы / Г. Крамер, М. Лидбеттер. М. : Мир, 1969.
20. Липцер, Р. А. Теория мартингалов / Р. А. Липцер, А. Н. Ширяев. М. : Наука, 1986.
21. Розанов, Ю. А. Введение в теорию случайных процессов / Ю. А. Розанов. М. : Наука, 1982.
22. Розанов, Ю. А. Стационарные случайные процессы / Ю. А. Розанов. 2-е изд. М. : Наука, 1990.
23. Севастьянов, Б. А. Курс теории вероятностей и математической статистики / Б. А. Севастьянов. М. : Наука, 1988.
24. Скороход, А. В. Лекції з теорії випадкових процесів: навч. посібник / А. В. Скороход. Київ : Либідь, 1990.

25. Справочник по теории вероятностей и математической статистике / под ред. В. С. Королука. М. : Наука, 1985.
26. *Феллер, В.* Введение в теорию вероятностей и ее приложения: в 2 т. / В. Феллер. М. : Мир, 1994.
27. *Феллер, В.* Теория вероятностей и ее приложения / В. Феллер. М. : Мир, 1984. Т. 1, 2.
28. *Хампель, Ф.* Робастность в статистике: Подход на основе функций влияния / Ф. Хампель [и др.]. М. : Мир, 1989.
29. *Харин, Ю. С.* Математическая и прикладная статистика / Ю. С. Харин, Е. Е. Жук. Минск : БГУ, 2005.
30. *Харин, Ю. С.* Теория вероятностей / Ю. С. Харин, Н. М. Зуев. Минск : БГУ, 2004.
31. *Харин, Ю. С.* Теория вероятностей, математическая статистика, случайные процессы в задачах, упражнениях и тестовых заданиях: уч. пособие с грифом Минобразования / Ю. С. Харин, [и др.]. Минск : БГУ, 2010.
32. *Харин, Ю. С.* Практикум на ЭВМ по математической статистике / Ю. С. Харин, М. Д. Степанова. Минск : Университетское, 1987.
33. *Харин, Ю. С.* Сборник задач по теории вероятностей, теории случайных процессов и математической статистике / Ю. С. Харин, Г. А. Хацкевич, В. И. Лобач. Минск : БГУ, 1995.
34. *Хорн, Р.* Матричный анализ / Р. Хорн, Ч. Джонсон. М. : Мир, 1989.
35. *Хьюбер, П.* Робастность в статистике / П. Хьюбер. М. : Мир, 1984.
36. *Ширяев, А. Н.* Вероятность / А. Н. Ширяев. М. : Наука, 1989.
37. *Grimmett, G.* One Thousand Exercises in Probability / G. Grimmett, D. Stirzaker. Oxford : Oxford Univ. Press, 2001.
38. *Grimmett, G.* Probability and Random Processes / G. Grimmett, D. Stirzaker. Oxford : Oxford Univ. Press, 2001.
39. *Nguyen, H. T.* Fundamentals of mathematical statistics / H. T. Nguyen, G. S. Rogers. N.Y., Vol. 1, 2. 1989.

ПРИЛОЖЕНИЯ

1. Электронный учебно-методический комплекс по теории вероятностей и математической статистике

В качестве электронного помощника для изучения теории вероятности и математической статистики может быть использован «Электронный учебно-методический комплекс по теории вероятностей и математической статистике», разработанный на кафедре математического моделирования и анализа данных БГУ.

«Электронный учебно-методический комплекс» (ЭУМК) – это система нормативной и учебно-методической документации, средств обучения и контроля, необходимых и достаточных для качественной организации основных и дополнительных образовательных программ, согласно учебному плану, представленная на электронных носителях. ЭУМК является одним из элементов организации образовательной деятельности по очной, заочной, очно-заочной и дистанционной формам обучения.

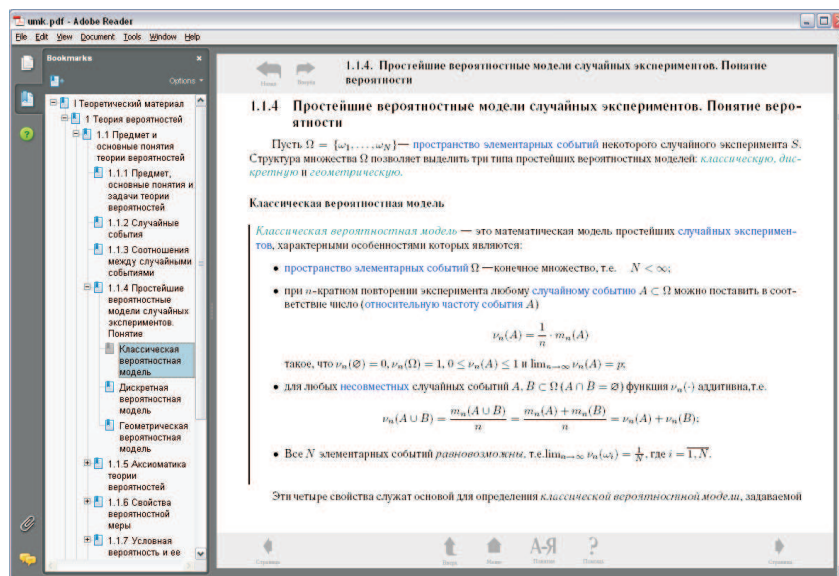
Основная цель создания ЭУМК – предоставить студенту полный комплект учебно-методических материалов для самостоятельного изучения дисциплины. ЭУМК по курсу «Теория вероятностей и математическая статистика» предназначен также для использования в учебном процессе высших учебных заведений Республики Беларусь по всем специальностям, учебные планы которых включают учебную дисциплину «Теория вероятностей и математическая статистика» и разделы с использованием современных образовательных технологий.

Структура ЭУМК

ЭУМК для достижения максимального эффекта составлено несколько иначе нежели традиционно изданные пособия: главы являются более короткими, что соответствует меньшему размеру компьютерных экранных страниц по сравнению с книжными. Каждый раздел, соответствующий рубрикам нижнего уровня, разбит на дискретные фрагменты, содержащие необходимый и достаточный материал по конкретному узкому вопросу. Как правило, такой фрагмент (статья) содержит один – три текстовых абзаца (абзацы также должны быть короче книжных), формулу, таблицу или рисунок и подпись к нему, включающую краткое пояснение смысла рисунка.

Статья – это минимальный информационный объект ЭУМК. Элементы знаний связаны между собой гиперссылками. Информационное содержание элемента знаний определяется как упорядоченная совокупность связанных с ним учебных фрагментов. Пример статьи ЭУМК по курсу «Теория вероятностей и математическая статистика» приведен на рисунке. Таким образом, все содержание информационных модулей структурируется в виде статей (учебных фрагментов) по разделам ЭУМК и представляется в виде гипертекста.

Пользователь ЭУМК просматривает не непрерывно излагаемый материал, а отдельные экранные статьи, дискретно следующие друг за другом. Изучив на данном экране представленные сведения, студент нажимает кнопку «Следующий», размещенную ниже текста, и получает следующий фрагмент материала. Если он видит, что не все понял или запомнил с предыдущего экрана, то нажимает кнопку «Предыдущий» и возвращается на один шаг назад.



Пример статьи ЭУМК

Дискретная последовательность экранов находится внутри (и в пределах) наименьшей структурной единицы, позволяющей прямую адресацию, т. е. внутри пункта или подпункта (того, что характеризуется заголовком третьего уровня) содержится одна или несколько статей, последовательно связанных друг с другом гипертекстовыми связями. На основе таких фрагментов проектируется структура учебного материала, которая в соответствии с главами 1–21 данного учебника содержит: учебные программы; список основных понятий (гlossарий); теоретический материал; комплекс задач и упражнений; сборник тестовых заданий; список рекомендуемой литературы (библиография).

Такая организация учебного материала обеспечивает дифференцированный подход к обучаемым в зависимости от уровня их подготовленности; результатом этого является более высокий уровень мотивации обучения, что приводит к лучшему и ускоренному усвоению материала.

ЭУМК содержит два основных фрейма: в левом фрейме организована навигация в пределах пособия в целом, которая оформлена в виде оглавления документа. Поскольку размеры подробного оглавления велики, то во фрейме обязательно содержится слайдер (окно прокрутки) и возможность раскрывать/скрывать детальное оглавление подуровней.

Для удобства в отдельный раздел выделен список определений, переход к которым организован по ссылкам от терминов, встречающихся в основном тексте и требующих пояснения. Предметный или алфавитный указатель пособия позволяет перейти от соответствующих терминов и понятий к основному тексту, в котором они упоминаются, с помощью гипертекстовых ссылок. Таким образом, содержимое этого фрейма обеспечивает дополнительные возможности навигации в пределах ЭУМК. Однако реализация этого способа требует от студента хотя бы предварительного знакомства с тематикой пособия. Иначе говоря, этим способом навигации целесообразно пользоваться лишь тем студентам, которые стремятся усовершенствовать или повысить уровень своих знаний по данному предмету.

Ориентация и перемещение внутри ЭУМК

Ориентация студента в ЭУМК достигается несколькими путями. Прежде всего, как и в печатном издании, с помощью заголовков. Рубрикация ЭУМК характеризуется большей глубиной (большим числом уровней), чем у печатного учебника. Каждый экран содержит заголовок раздела. Выше уже упоминалось, что разделы ЭУМК достаточно короткие и каждый раздел содержит исчерпывающую информацию по одному конкретному вопросу.

Другой вариант ориентации связан с использованием в ЭУМК колонтитулов (как и в печатном пособии). Колонтитул предусмотрен на каждой экранной странице и позволяет контролировать название изучаемой главы и параграфа, т. е. пользователь не теряет ориентации в учебнике.

Наконец, в левом фрейме «Содержание» выбранный раздел отмечается другим цветом по сравнению со всеми оставшимися. Здесь преимущество ЭУМК особенно наглядно, так как студент, наряду с чтением раздела, всегда видит, где он находится, какой раздел изучает и расположение этого раздела среди рубрикаций.

Перемещение внутри ЭУМК также может производиться с помощью гипертекстовых ссылок.

В ЭУМК применяется иерархическое структурирование оглавления. Скажем, исходя на экране в левом фрейме помещается список заголовков первого уровня. Если щелкнуть по одному из заголовков, то в этом же фрейме появляется совокупность всех заголовков второго уровня для соответствующей части учебника. Аналогично при щелчке указателем мыши по заголовку второго уровня фрейм заполняется заголовками третьего уровня и так далее.

Принципы построения ЭУМК

В основе построения ЭУМК лежат следующие принципы.

1. *Полнота.* Содержание ЭУМК охватывает все основные разделы теории вероятностей и математической статистики, включая теоретический материал, тестовые задания, сборники задач.

ЭУМК включает учебные материалы для разных специальностей вузов, объем учебных материалов значительно превышает потребности каждой отдельно взятой программы. Элементы знаний связаны между собой гипер-ссылками.

2. *Гибкие средства адаптации* к изменению учебных планов и программ. В состав ЭУМК в качестве самостоятельного раздела включаются типовые программы курсов. Привязка образовательного контента ЭУМК к действующим учебным программам по конкретным специальностям осуществляется с помощью установленных связей.

3. *Универсальность* технологического процесса создания и представления информации. ЭУМК является документом в кроссплатформенном формате электронных документов Adobe PDF, что позволяет пользоваться им практически на любых современных компьютерных платформах и операционных системах. Для подготовки исходных текстов ЭУМК создан специальный пакет для разработчиков УМК, включающий электронные шаблоны представления образовательного контента.

4. *Адаптация* к потребностям пользователя. Наличие «избыточного» учебного материала различных уровней сложности и детализации дает возможность обеспечить полноценную индивидуализацию обучения.

2. Основные вероятностные распределения

Распределение	Обозначение	Параметры	Плотность или дискретное распределение вероятностей	Характеристическая функция	Математическое ожидание	Мода	Дисперсия
Равномерное	$R[a, b]$	$a, b \in \mathbb{R}^1, a < b$	$\frac{1}{b-a}, a \leq x \leq b$	$\frac{e^{iat}(e^{i(b-a)t}-1)}{i(b-a)t}$	$\frac{a+b}{2}$	$M \in [a, b]$	$\frac{(b-a)^2}{12}$
Одномерное нормальное (Гаусса)	$N_1(\mu, \sigma^2)$	$\mu \in \mathbb{R}^1, \sigma > 0$	$\frac{1}{\sqrt{2\pi}\sigma} \times \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right)$	μ	μ	σ^2
Экспоненциальное	$E(\lambda)$	$\lambda > 0$	$\lambda \exp(-\lambda x), x \geq 0$	$(1 - it/\lambda)^{-1}$	λ^{-1}	0	λ^{-2}
Логистическое	$l(a, b)$	$a \in \mathbb{R}^1, b > 0, k = \sqrt{3}b/\pi$	$\frac{k^{-1} \exp((x-a)/k)}{(1 + \exp((x-a)/k))^2}$	$e^{ita} \Gamma(1 - ikt) \Gamma(1 + ikt)$	a	a	b^2
Гамма	$\gamma(\lambda, \alpha)$	$\lambda, \alpha > 0$	$\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, x \geq 0$	$(1 - it/\lambda)^{-\alpha}$	$\alpha \lambda^{-1}$	$(\alpha - 1)/\lambda, \lambda \geq 1$	$\alpha \lambda^{-2}$
Вейбулла – Гнеденко	$W(\lambda, \alpha)$	$\lambda, \alpha > 0$	$\alpha \lambda x^{\alpha-1} \times \exp(-\lambda x^\alpha), x \geq 0$	$\sum_{k=0}^{\infty} \frac{(it)^k \Gamma(1+k/\alpha)}{k! \lambda^{k/\alpha}}$	$\frac{\Gamma(1+\alpha^{-1})}{\lambda^{1/\alpha}}$	0, если $\alpha \leq 1$; $\left(\frac{1-\alpha^{-1}}{\lambda}\right)^{1/\alpha}$	$\lambda^{-2/\alpha} \times (2\alpha^{-1} \Gamma(2/\alpha) - \alpha^{-2} \Gamma^2(1/\alpha))$
Бета	$\beta(v, w)$	$v, w > 0$	$\frac{\Gamma(v+w)}{\Gamma(v)\Gamma(w)} x^{v-1} \times (1-x)^{w-1}, 0 \leq x \leq 1$	$\frac{\Gamma(v+w)}{\Gamma(v)} \times \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \frac{\Gamma(u+k)}{\Gamma(v+w+k)}$	$\frac{v}{v+w}$	$\frac{v-1}{v+w-2}, v, w > 1$	$\frac{vw}{(v+w)^2(v+w+1)}$
Хи-квадрат	χ_v^2	$v > 0$ – целое	$x^{\frac{v}{2}-1} e^{-\frac{x}{2}} \times \left(2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)\right)^{-1}, x \geq 0$	$(1 - 2it)^{-v/2}$	v	$v - 2, v \geq 2$	$2v$
F (Фишера)	$F_{v,w}$	$v, w > 0$ – целые	$\left(\frac{v}{w}\right)^{v/2} \Gamma\left(\frac{v+w}{2}\right) \times \left(\Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{w}{2}\right)\right)^{-1} \times x^{\frac{v}{2}-1} \times \left(1 + \frac{v}{w} x\right)^{-\frac{v+w}{2}}, x \geq 0$	$\left(\frac{w}{v}\right)^{it/2} \times \left(\Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{w}{2}\right)\right)^{-1} \times \Gamma\left(\frac{v+it}{2}\right) \Gamma\left(\frac{w-it}{2}\right)$	$\frac{w}{w-2}, w > 2$	$\frac{w(v-2)}{v(w-2)}, v > 2$	$\frac{2w^2(v+w-2)}{v(w-2)^2(w-4)}, w > 4$

t (Стюдента)	t_v	$v > 0 -$ целое	$\Gamma\left(\frac{v+1}{2}\right) \times$ $\times \left(\sqrt{v}\Gamma\left(\frac{v}{2}\right)\right)^{-1} \times$ $\times (1+x^2/v)^{-\frac{v+1}{2}}$	$\frac{\sqrt{\pi}\Gamma((v+1)/2)\exp(-\sqrt{v} t)}{\Gamma(v/2)2^{2(n-1)}(n-1)!} \times$ $\times \sum_{k=0}^{n-1} (2k)! C_{n-1+k}^{2k} \times$ $\times (2\sqrt{v} t)^{n-1-k},$ если $(v+1)/2 = n - \text{целое}$	0	0	$\frac{v}{v-2},$ $v > 2$
Коши	$C(a, b)$	$a \in \mathbb{R}^1,$ $b > 0$	$(\pi b(((x-a)/b)^2 +$ $+1))^{-1}$	$\exp(iat - b t)$	Не существует	a	Не существует
Логнормальное	$L(m, \sigma^2)$	$m, \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma x}} \times$ $\times \exp\left(-\frac{(\ln(x/m))^2}{2\sigma^2}\right),$ $x \geq 0$	$\sum_{k=0}^{\infty} \frac{(it)^k}{k!} \times$ $\times \exp\left(k \ln m + \frac{k^2 \sigma^2}{2}\right)$	$m\sqrt{w},$ $w =$ $= \exp(\sigma^2)$	m/w	$m^2 w(w-1)$
Лапласа	$EE(a, \lambda)$	$a \in \mathbb{R}^1,$ $\lambda > 0$	$(\lambda/2) \times$ $\times \exp(-\lambda x-a)$	$(1+(t/\lambda)^2)^{-1} \exp(ita)$	a	a	$2/\lambda^2$
Многомерное нормальное	$N_N(\mu, \Sigma)$	$\mu \in \mathbb{R}^N,$ $\Sigma > 0$	$\frac{(2\pi)^{-\frac{N}{2}}}{\sqrt{ \Sigma }} \times$ $\times \exp\left(-\frac{1}{2}(x-\mu)^T \times$ $\times \Sigma^{-1}(x-\mu)\right),$ $x \in \mathbb{R}^N$	$\exp(it^T \mu - \frac{1}{2} t^T \Sigma t)$	μ	μ	$\Sigma -$ ковариационная матрица
Бернулли	$Bi(1, p)$	$p \in [0, 1]$	$p^x(1-p)^{1-x},$ $x \in \{0, 1\}$	$1 + p(e^{it} - 1)$	p	1, если $p > 1/2;$ 0, если $p < 1/2;$ 0 и 1, если $p = 1/2$	$p(1-p)$
Биномиальное	$Bi(n, p)$	$n \geq 2,$ $p \in [0, 1]$	$C_n^x p^x (1-p)^{n-x},$ $x \in \{0, \dots, n\}$	$(1 + p(e^{it} - 1))^n$	np	$(n+1)p - 1/2 + \epsilon,$ $-1/2 \leq \epsilon \leq 1/2$	$np(1-p)$
Гипергеометрическое	$H(N, L, n)$	$L, n < N$	$\frac{C_N^x C_{N-L}^{n-x}}{C_N^n},$ $0 \leq x \leq \min\{L, n\}$	$\sum_{k=0}^L \frac{C_N^k C_{N-L}^{n-k}}{C_N^n} e^{itk}$	$\frac{nL}{N}$	—	$\frac{nL(N-L)(N-n)}{N^2(N-1)}$
Пуассона	$\Pi(\lambda)$	$\lambda > 0$	$\lambda^x (x!)e^{-\lambda},$ $x \in \{0, 1, \dots\}$	$\exp(\lambda(e^{it} - 1))$	λ	λ и $\lambda - 1$, если $\lambda - \text{целое};$ иначе целая часть λ	λ
Геометрическое	$G(p)$	$p \in [0, 1]$	$p(1-p)^x,$ $x \in \{0, 1, \dots\}$	$p(1 - (1-p)e^{it})^{-1}$	$\frac{1-p}{p}$	0	$\frac{1-p}{p^2}$
Отрицательное биномиальное	$\overline{Bi}(r, p)$	$r \geq 1,$ $p \in [0, 1]$	$C_{r+x-1}^x p^r (1-p)^x,$ $x \geq 0$	$\left(\frac{p}{1-(1-p)e^{it}}\right)^r$	$\frac{r(1-p)}{p}$	λ и $\lambda - 1$, если $\lambda \geq 1 - \text{целое};$ иначе целая часть $\lambda;$ $\lambda = \frac{(r-1)(1-p)}{p}$	$\frac{r(1-p)}{p^2}$

3. Алфавиты, используемые в математике

Латинский алфавит

<i>Aa</i>	а	<i>Jj</i>	йот	<i>Ss</i>	эс
<i>Bb</i>	бэ	<i>Kk</i>	ка	<i>Tt</i>	тэ
<i>Cc</i>	цэ	<i>Ll</i>	эль	<i>Uu</i>	у
<i>Dd</i>	дэ	<i>Mm</i>	эм	<i>Vv</i>	вэ
<i>Ee</i>	е	<i>Nn</i>	эн	<i>Ww</i>	дубль-вэ
<i>Ff</i>	эф	<i>Oo</i>	о	<i>Xx</i>	икс
<i>Gg</i>	гэ	<i>Pp</i>	пэ	<i>Yy</i>	игрек
<i>Hh</i>	аш	<i>Qq</i>	ку	<i>Zz</i>	зет
<i>Ii</i>	и	<i>Rr</i>	эр		

Готический алфавит

<i>Aa</i>	а	<i>Jj</i>	йот	<i>Ss</i>	эс
<i>Bb</i>	бэ	<i>Kk</i>	ка	<i>Tt</i>	тэ
<i>Cc</i>	цэ	<i>Ll</i>	эль	<i>Uu</i>	у
<i>Dd</i>	дэ	<i>Mm</i>	эм	<i>Vv</i>	фау
<i>Ee</i>	э	<i>Nn</i>	эн	<i>Ww</i>	вэ
<i>Ff</i>	эф	<i>Oo</i>	о	<i>Xx</i>	икс
<i>Gg</i>	гэ	<i>Pp</i>	пэ	<i>Yy</i>	ипсилон
<i>Hh</i>	ха	<i>Qq</i>	ку	<i>Zz</i>	цэт
<i>Ii</i>	и	<i>Rr</i>	эр		

Греческий алфавит

<i>Aα</i>	альфа	<i>Iι</i>	йота	<i>Pρ</i>	ро
<i>Bβ</i>	бета	<i>Kκ</i>	каппа	<i>Σσ</i>	сигма
<i>Γγ</i>	гамма	<i>Λλ</i>	ламбда	<i>Tτ</i>	тау
<i>Δδ</i>	дельта	<i>Μμ</i>	мю	<i>Υυ</i>	ипсилон
<i>Eε</i>	эпсилон	<i>Nν</i>	ню	<i>Φφ</i>	фи
<i>Zζ</i>	дзета	<i>Ξξ</i>	кси	<i>Χχ</i>	хи
<i>Ηη</i>	эта	<i>Οο</i>	омикрон	<i>Ψψ</i>	пси
<i>Θθϑ</i>	тэта	<i>Ππ</i>	пи	<i>Ωω</i>	омега

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- 0 — 1-функция потерь, 280
- L -оценки, 304
- M -оценки, 304
- T^2 -критерий Стьюдента, 372
- T^2 -статистика Хотеллинга, 369
- U -статистика, 382
- V -статистика, 365
- Z -преобразование Фишера, 351
- Z -статистика Фишера, 351
- \mathcal{F} -измеримая функция, 37
- χ^2 -критерий согласия Пирсона, 319
- χ^2 -статистика Пирсона, 318
- σ -алгебра, 24
- k -мерная маргинальная (частная) функция распределения, 50
- k -мерная маргинальная плотность распределения вероятностей, 52
- n -мерная случайная величина, 49
- n -мерная функция распределения, 49
- n -мерное нормальное распределение вероятностей, 52
- p -квантиль, 260
- I достаточный признак сходимости п. н., 118
- II достаточный признак сходимости п. н., 120
- МНК-оценка, 289, 290
- РП Неймана — Пирсона, 314
- УЗБЧ для н. о. р. СВ, 141
- ЦПТ Леви — Линдберга, 153
- ЦПТ Линдберга, 150
- ЦПТ Феллера, 152
- ЦПТ для слабозависимых СВ, 157
- ЦПТ для случайных векторов, 156
- абсолютно непрерывная функция распределения, 43
- аксиома конечной аддитивности, 20
- аксиома нормировки, 25
- аксиома равновероятности, 20
- аксиома счетной аддитивности, 22
- аксиомы Колмогорова, 25
- алгебра случайных событий, 23
- ансамбль реализаций ВР, 403
- апостериорные вероятности классов, 386
- априорные вероятности классов, 385
- асимптотически наикратчайший центральный доверительный интервал, 301
- асимптотически несмещенная оценка, 254
- асимптотически нормальная оценка, 261
- байесовская оценка параметра, 280
- байесовский принцип оптимальности, 280, 315
- байесовский риск, 386
- байесовское решающее правило, 315, 386
- безусловная вероятность, 29
- белый шум, 183
- биномиальное распределение вероятностей, 35
- биномиальный закон распределения вероятностей, 42
- борелевская σ -алгебра, 25
- борелевская функция, 56
- борелевское множество, 24
- вариация оценки, 268, 270
- вероятностное пространство, 25
- вероятность, 25
- вероятность ошибки I рода, 311
- вероятность ошибки II рода, 311
- взаимная спектральная плотность, 188
- винеровский процесс, 192
- временной ряд (ВР), 403
- временной ряд авторегрессии, 417
- временной ряд авторегрессии и проинтегрированного скользящего среднего, 436
- временной ряд авторегрессии и скользящего среднего, 417
- временной ряд скользящего среднего, 417
- выборочная p -квантиль, 260
- выборочная ковариационная матрица, 341
- выборочная медиана, 261
- выборочная спектральная плотность, 413
- выборочная функция распределения, 255
- выборочная характеристическая функция, 255
- выборочное пространство, 252
- выборочное среднее, 341
- выборочный коэффициент корреляции, 349
- выборочный множественный коэффициент корреляции, 358
- выборочный момент, 255
- выборочный частный коэффициент корреляции, 358
- гауссовский случайный вектор, 52
- гауссовский случайный процесс, 162
- геометрическая вероятностная модель, 23
- геометрическое распределение вероятностей, 43
- гипергеометрический закон распределения вероятностей, 42
- гипергеометрическое распределение вероятностей, 21
- гипотеза H_k , 309
- гипотеза независимости, 362
- гипотеза симметрии, 373

- гипотеза согласия, 317
 гипотеза условной независимости, 362
 гистограмма, 257
 главные компоненты, 335
- двухфакторный дисперсионный анализ, 397
 дискретная вероятностная модель, 22
 дискретная случайная величина, 41
 дискретная функция распределения, 41
 дискретное распределение вероятностей, 40
 дискретный равномерный на множестве X закон распределения вероятностей, 42
 дискриминантный анализ, 387
 дисперсионный анализ, 396
 дисперсия ВР, 403
 дисперсия случайной величины, 98
 доверительная вероятность, 295
 доверительный интервал, 295
 доля ошибочных решений, 387
 достаточная статистика, 283
 достаточное условие Колмогорова УЗБЧ, 140
 достаточное условие Маркова ЗБЧ, 136
 достаточное условие Чебышева ЗБЧ, 137
- закон $0 \vee 1$ Бореля, 120
 закон больших чисел (ЗБЧ), 135
 закон распределения вероятностей Бернулли, 41
- измеримое пространство, 24
 импульсная характеристика фильтра, 178
 интеграл Лебега по вероятностной мере, 65
 интеграл Лебега – Стильбеса, 82
 интеграл Римана – Стильбеса, 83
 интервальная оценка, 295
 интервальное оценивание, 295
 информационная матрица Фишера, 265–267
- квадратичная функция потерь, 280
 квадратичный дискриминантный анализ, 391
 классическая вероятностная модель, 20
 кластерный анализ, 387
 ковариационная матрица, 100
 ковариационная функция, 171
 ковариация случайных величин, 100
 количество информации по Фишеру, 270
 количество информации по Шеннону, 104
 корреляционная функция ВР, 403
 коэффициент корреляции, 101
 критерий L_2 -сходимости, 126
 критерий ЗБЧ, 135
 критерий Коши L_r -сходимости, 124
 критерий Коши сходимости п. н., 118
 критерий независимости случайных величин, 56
 критерий отношения правдоподобия, 365
 критерий отношения правдоподобия для проверки сложных гипотез, 321
 критерий равномерной интегрируемости, 128
 критерий согласия, 317
- критерий согласия Колмогорова, 321
 критерий сходимости ФР, 144
 критерий факторизации Неймана – Фишера, 283
 критерий эффективности оценки, 270
 критические функции, 310, 386
 кумулянтная функция СВ, 110
- левосторонний доверительный интервал, 295
 линейная регрессионная модель, 287
 линейное преобразование, 177
 линейный дискриминантный анализ, 391
 логарифмическая функция правдоподобия Фишера, 273
- марковский момент, 194
 марковский процесс, 213
 мартингал, 235
 математическое ожидание, 71
 математическое ожидание ВР, 403
 математическое ожидание функции от СВ, 78
 матрица вариаций, 254
 матрица вероятностей одношаговых переходов, 441
- медиана распределения вероятностей, 261
 мера, 23
 мера Лебега, 48
 метод L -средних, 392
 метод главных компонент, 334
 метод максимального правдоподобия, 272
 метод моментов, 271
 метод обратной функции, 296
 метод стьюдентизации, 298
 метрика Махаланобиса, 340
 минимальная достаточная статистика, 283
 многомерная проблема Беренса – Фишера, 378
 многомерная регрессионная модель, 287
 многомерное нормальное распределение, 337
 множественная регрессионная модель, 287
 множественный коэффициент корреляции, 356
 модель Тьюки – Хьюбера, 302
 модель Фишера, 388
 монотонная последовательность случайных событий, 26
 мощность решающего правила, 312
- начальный момент k -го порядка, 97
 независимость случайных событий, 32
 независимые в совокупности, 33
 независимые случайные величины, 56
 нелинейная регрессионная модель, 287
 непрерывная случайная величина, 43
 неприводимая цепь Маркова, 224
 неравенство Берри – Эссеена, 153
 неравенство Гаека – Реньи, 138
 неравенство Гельдера, 81
 неравенство Йенсена, 80
 неравенство Колмогорова, 139

- неравенство Коши – Буняковского (Шварца), 81
 неравенство Крамера – Рао, 270
 неравенство Ляпунова, 81
 неравенство Минковского, 82
 неравенство Чебышева, 79
 неравенство Чебышева относительно дисперсии, 99
 неравенство информации, 267
 нерадомизированное РП, 310
 несмещенность статистической оценки, 253
 нижняя и верхняя доверительные границы, 295
 нормальный (гауссовский) закон распределения вероятностей, 46
 обобщенная \bar{T}^2 -статистика, 375
 обобщенная статистика отношения правдоподобия, 364
 обобщенная формула сложения вероятностей, 28
 обобщенная формула умножения вероятностей, 31
 однородная цепь Маркова, 441
 однородный случайный процесс, 191
 однофакторный дисперсионный анализ, 396
 отсчетом ВР, 403
 оценка Бартлетта, 415
 оценка максимального правдоподобия, 274
 оценка минимального контраста (ОМК, МК-оценка), 304
 оценка по методу моментов, 272
 оценки Юла – Уокера, 422
 параметрическое пространство, 252
 парная корреляция, 349
 парные зависимости, 349
 план эксперимента, 290
 плотность распределения вероятностей, 43
 подстановочные решающие правила, 387
 подстановочный принцип («plug-in»), 256
 полиномиальное распределение вероятностей, 52
 попарно независимые случайные события, 33
 порядковая статистика, 258
 последовательность авторегрессии, 183
 последовательностью скользящего среднего, 183
 последовательный анализ Вальда, 324
 правило «3 σ », 100
 правосторонний доверительный интервал, 296
 предел последовательности случайных событий, 26
 предмет математической статистики, 251
 прикладная статистика, 331
 принцип максимума правдоподобия, 274
 принцип метода моментов, 272
 принцип метода наименьших квадратов, 289
 принцип оптимальности Вальда, 325
 принцип оптимальности Неймана – Пирсона, 312
 производная Радона – Никодима, 96
 производящая функция СВ, 111
 простая гипотеза H_k , 309
 простая регрессионная модель, 287
 пространство решений, 309
 пространство элементарных событий, 18
 псевдо- F -статистика Фишера, 393
 пуассоновский процесс, 200
 равномерно интегрируемая (СВ), 127
 равномерный (прямоугольный) закон распределения вероятностей, 45
 разведочный анализ, 332
 разложение Вольда, 428
 ранговые статистики, 262
 радомизированное РП, 310
 распределение Уипарта, 343
 распределение вероятностей, 38
 расстояние Колмогорова, 320
 расстояние Махаланобиса, 389
 реализация ВР, 403
 решающее правило Вальда, 325
 решающее правило Неймана – Пирсона, 312
 решающее правило, решающая функция, критерий, тест, 309
 робастное статистическое оценивание, 301
 свойства МП-оценки, 276
 свойства вероятностной меры (вероятности), 27
 свойства достаточной статистики, 285
 свойства интеграла Лебега, 74
 свойство строгой марковости, 194
 сглаженная спектральная плотность, 415
 семейство конечномерных распределений, 162
 семейство оценок Парзена, 415
 сепарабельный случайный процесс, 164
 сжатие данных, 334
 сильная состоятельность оценки, 253
 сингулярная функция распределения, 48
 система обратных дифференциальных уравнений Колмогорова, 219
 система прямых дифференциальных уравнений Колмогорова, 220
 система уравнений Юла – Уокера, 420
 слабая сходимость последовательности ФР, 144
 сложная гипотеза H_k , 309
 случайная величина, 37
 случайная выборка, 251
 случайная последовательность, 115
 случайная функция, 401
 случайное событие, 18
 случайный n -вектор, 49
 случайный процесс, 161
 случайный процесс с независимыми приращениями, 191
 случайный процесс с ортогональными приращениями, 173

- случайный процесс, дифференцируемый в среднем квадратичном, 165
 случайный процесс, непрерывный в среднем квадратичном, 164
 случайный эксперимент, 17
 смешанная модель авторегрессии и скользящего среднего, 184
 смещение, 253
 соотношение Вальда, 90
 соотношения между видами сходимости случайных последовательностей, 130
 состояние возвратное, 224
 состояние несущественное, 224
 состояние нулевое, 224
 состояние периодическое, 224
 состоятельность статистической оценки, 253
 спектральная плотность, 172
 спектральная функция, 172
 спектральное представление, 172, 174
 спектрограмма, 413
 среднеквадратическая ошибка прогнозирования, 426
 среднеквадратическое (стандартное) отклонение случайной величины, 98
 стандартная нормальная плотность распределения вероятностей, 47
 стандартная нормальная функция распределения, 47
 стандартный нормальный закон $N(0, 1)$, 150
 статистика, 251
 статистика отношения правдоподобия, 312, 322
 статистическая оценка, 252
 статистические выводы, 251
 статистический анализ данных, 331
 стационарное распределение для однородной цепи Маркова, 442
 стационарные вероятности, 216
 стохастическое дифференциальное уравнение (Ито), 207
 субмартигал, 235
 супермартигал, 235
 схема независимых испытаний, 34
 схема независимых испытаний Бернулли, 34
 схема серий, 155
 сходимость в среднем квадратическом, 116
 сходимость в среднем порядка r , 116
 сходимость по вероятности, 116
 сходимость по распределению, 115
 сходимость почти на верное, 116
 теорема Ляпунова, 153
 теорема Муавра – Лапласа, 154
 теорема единственности для характеристической функции, 109
 теорема непрерывности, 145
 теорема обращения для характеристической функции, 108
 теория статистической проверки гипотез, 309
 тест максимального правдоподобия, 317
 точечная оценка, 295
 траектория, 161
 тренд, 431
 уравнение Колмогорова – Чепмена, 214
 усиленный закон больших чисел (УЗБЧ), 135
 условие Линдеберга, 149
 условие Эйкера, 291
 условие нормировки, 22
 условие ординарности, 201
 условия регулярности, 265
 условная вероятность, 29
 условная функция распределения, 53
 условное математическое ожидание, 87
 условное математическое ожидание относительно σ -алгебры, 90
 фильтрация, 178
 формула Байеса, 31
 формула Байеса для плотностей распределения вероятностей, 55
 формула полной вероятности, 31
 формула свертки, 60
 формула сложения вероятностей, 27
 формула умножения вероятностей, 30
 функционал риска, 280, 315, 386
 функция Линдеберга, 148
 функция неупреждающая, 202
 функция потерь, 280
 функция правдоподобия Фишера, 272
 функция распределения СВ, 38
 функция регрессии, 426
 характеристическая функция СВ, 104
 характеристическое уравнение, 419
 центральная предельная теорема Линдеберга – Феллера, 152
 центральный доверительный интервал, 296
 центральный момент k -го порядка, 97
 центральный смешанный момент порядка k , 100
 центрированная случайная величина, 97
 цепь Маркова, 213
 частная дисперсия, 354
 частная ковариация, 354
 частный коэффициент корреляции, 354
 частотная характеристика, 177
 экспоненциальное семейство, 286
 элементарные события, 18
 эллипсоид рассеяния, 340
 энтропия случайного вектора, 103
 эргодическая цепь Маркова, 226
 эффективная оценка, 269
 эффективность статистической оценки, 269

ОГЛАВЛЕНИЕ

.....	1
.....	2
.....	2
.....	4
ПРЕДИСЛОВИЕ	5
ВВЕДЕНИЕ	9
ОСНОВНЫЕ ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	11
ЧАСТЬ 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ	15
.....	16
ГЛАВА 1. Случайные события и их вероятности	17
1.1. ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ	17
1.2. СЛУЧАЙНЫЕ СОБЫТИЯ И СООТНОШЕНИЯ МЕЖДУ НИМИ	18
1.3. ПОНЯТИЕ ВЕРОЯТНОСТИ. ПРОСТЕЙШИЕ ВЕРОЯТНОСТНЫЕ МОДЕЛИ .	20
1.4. АЛГЕБРА, σ -АЛГЕБРА И ИХ СВОЙСТВА. ИЗМЕРИМОЕ ПРОСТРАНСТВО .	23
1.5. АКСИОМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ. ВЕРОЯТНОСТНОЕ ПРОСТРАН- СТВО	25
1.6. СВОЙСТВА ВЕРОЯТНОСТНОЙ МЕРЫ (ВЕРОЯТНОСТИ)	27
1.7. УСЛОВНАЯ ВЕРОЯТНОСТЬ И ЕЕ СВОЙСТВА. ФОРМУЛА ПОЛНОЙ ВЕ- РОЯТНОСТИ. ФОРМУЛА БАЙЕСА	29
1.8. НЕЗАВИСИМЫЕ СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ СВОЙСТВА	32
1.9. СХЕМА НЕЗАВИСИМЫХ ИСПЫТАНИЙ БЕРНУЛЛИ. БИНОМИАЛЬ- НОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ	34
1.10. УПРАЖНЕНИЯ	35
ГЛАВА 2. Случайные величины и их распределения вероятностей	37
2.1. ПОНЯТИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И РАСПРЕДЕЛЕНИЯ ВЕРОЯТНО- СТЕЙ. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ И ЕЕ СВОЙСТВА	37
2.2. КЛАССИФИКАЦИЯ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ, РАСПРЕДЕЛЕНИЕ ВЕ- РОЯТНОСТЕЙ И СЛУЧАЙНЫХ ВЕЛИЧИН. ОСНОВНЫЕ ЗАКОНЫ РАС- ПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ	40
2.3. n -МЕРНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА И СЛУЧАЙНЫЙ n -ВЕКТОР. n - МЕРНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, n -МЕРНАЯ ПЛОТНОСТЬ РАС- ПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ И ИХ СВОЙСТВА	49
2.4. УСЛОВНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, УСЛОВНАЯ ПЛОТНОСТЬ РАС- ПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ И ИХ СВОЙСТВА	53
2.5. НЕЗАВИСИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН	56
2.6. ФУНКЦИОНАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН	58
2.7. УПРАЖНЕНИЯ	61
ГЛАВА 3. Числовые характеристики случайных величин	65
3.1. СХЕМА ПОСТРОЕНИЯ ИНТЕГРАЛА ЛЕБЕГА ПО ВЕРОЯТНОСТНОЙ МЕ- РЕ. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ПРОСТОЙ СЛУЧАЙНОЙ ВЕЛИЧИ- НЫ И ЕГО СВОЙСТВА	65

3.2. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ (ИНТЕГРАЛ ЛЕБЕГА) ДЛЯ ПРОИЗВОЛЬНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН	70
3.3. СВОЙСТВА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ (ИНТЕГРАЛА ЛЕБЕГА) ...	74
3.4. НЕРАВЕНСТВА ДЛЯ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ	79
3.5. ВЫЧИСЛЕНИЕ ИНТЕГРАЛА ЛЕБЕГА. ИНТЕГРАЛЫ ЛЕБЕГА – СТИЛТЬЕСА И РИМАНА – СТИЛТЬЕСА	82
3.6. УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ И ЕГО СВОЙСТВА	87
3.7. УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ОТНОСИТЕЛЬНО σ -АЛГЕБРЫ И ЕГО СВОЙСТВА	90
3.8. МОМЕНТЫ СКАЛЯРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И ИХ СВОЙСТВА	97
3.9. ДИСПЕРСИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ И ЕЕ СВОЙСТВА	98
3.10. МОМЕНТЫ МНОГОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ, КОВАРИАЦИЯ, КОВАРИАЦИОННАЯ МАТРИЦА, КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ И ИХ СВОЙСТВА	100
3.11. ЭНТРОПИЯ, КОЛИЧЕСТВО ИНФОРМАЦИИ ПО ШЕННОНУ И ДРУГИЕ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН	103
3.12. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ И ЕЕ СВОЙСТВА	104
3.13. УПРАЖНЕНИЯ	111
ГЛАВА 4. Случайные последовательности и их сходимость	115
4.1. ОПРЕДЕЛЕНИЕ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ И ВИДОВ ЕЕ СХОДИМОСТИ	115
4.2. СХОДИМОСТЬ ПОЧТИ НАВЕРНОЕ. ЗАКОН $0 \vee 1$ БОРЕЛЯ	116
4.3. СХОДИМОСТЬ ПО ВЕРОЯТНОСТИ	120
4.4. СХОДИМОСТЬ В СРЕДНЕМ	124
4.5. РАВНОМЕРНАЯ ИНТЕГРИРУЕМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН	127
4.6. СООТНОШЕНИЯ МЕЖДУ ВИДАМИ СХОДИМОСТИ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ	130
4.7. УПРАЖНЕНИЯ	133
ГЛАВА 5. Предельные теоремы	135
5.1. ЗАКОН БОЛЬШИХ ЧИСЕЛ И УСЛОВИЯ ЕГО ВЫПОЛНЕНИЯ	135
5.2. УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ И УСЛОВИЯ ЕГО ВЫПОЛНЕНИЯ	138
5.3. СХОДИМОСТЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ И ХАРАКТЕРИСТИЧЕСКИХ ФУНКЦИЙ	144
5.4. УСЛОВИЕ ЛИНДЕБЕРГА И ЕГО ВЕРОЯТНОСТНЫЙ СМЫСЛ	148
5.5. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ЛИНДЕБЕРГА – ФЕЛЛЕРА	150
5.6. ЧАСТНЫЕ СЛУЧАИ ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЫ	153
5.7. СХЕМА СЕРИЙ. СХОДИМОСТЬ К ЗАКОНАМ ГАУССА И ПУАССОНА	155
5.8. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ СЛУЧАЙНЫХ ВЕКТОРОВ И СЛАБОЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН	156
5.9. УПРАЖНЕНИЯ	157
ЧАСТЬ 2. ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ	159
.....	160
ГЛАВА 6. Основные определения и общие свойства случайных процессов ..	161
6.1. ОПРЕДЕЛЕНИЕ И ЗАДАНИЕ СЛУЧАЙНЫХ ПРОЦЕССОВ	161
6.2. ЭКВИВАЛЕНТНЫЕ, ТОЖДЕСТВЕННЫЕ И СЕПАРАБЕЛЬНЫЕ ПРОЦЕССЫ	163
6.3. НЕПРЕРЫВНОСТЬ И ДИФФЕРЕНЦИРУЕМОСТЬ СЛУЧАЙНОГО ПРОЦЕССА В СРЕДНЕМ КВАДРАТИЧНОМ	164
6.4. СТОХАСТИЧЕСКИЕ ИНТЕГРАЛЫ В СРЕДНЕМ КВАДРАТИЧНОМ	166
6.5. НЕПРЕРЫВНОСТЬ ТРАЕКТОРИЙ СЛУЧАЙНОГО ПРОЦЕССА	167
6.6. УПРАЖНЕНИЯ	170

ГЛАВА 7. Стационарные в широком смысле случайные процессы	171
7.1. ОСНОВНЫЕ ПОНЯТИЯ И СВОЙСТВА	171
7.2. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ КОВАРИАЦИОННОЙ ФУНКЦИИ СЛУЧАЙНОГО ПРОЦЕССА	172
7.3. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ СТАЦИОНАРНОГО В ШИРОКОМ СМЫСЛЕ СЛУЧАЙНОГО ПРОЦЕССА	173
7.4. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ДЕЙСТВИТЕЛЬНЫХ СЛУЧАЙНЫХ ПРОЦЕССОВ	176
7.5. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ СЛУЧАЙНОГО ПРОЦЕССА	176
7.6. ПРОГНОЗИРОВАНИЕ СЛУЧАЙНОГО ПРОЦЕССА	178
7.7. СПЕКТРАЛЬНОЕ ПРЕДСТАВЛЕНИЕ КОВАРИАЦИОННОЙ ФУНКЦИИ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ И САМОЙ СТАЦИОНАРНОЙ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ	180
7.8. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ	181
7.9. ПОСЛЕДОВАТЕЛЬНОСТИ АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО	183
7.10. ПРОГНОЗИРОВАНИЕ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ	185
7.11. ИНТЕРПОЛЯЦИЯ И ФИЛЬТРАЦИЯ СЛУЧАЙНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ	187
7.12. УПРАЖНЕНИЯ	189
ГЛАВА 8. Случайные процессы с независимыми приращениями	191
8.1. ВИД ХАРАКТЕРИСТИЧЕСКОЙ ФУНКЦИИ СЛУЧАЙНОГО ПРОЦЕССА С НЕЗАВИСИМЫМИ ПРИРАЩЕНИЯМИ	191
8.2. ВИНЕРОВСКИЙ ПРОЦЕСС	192
8.3. ПУАССОНОВСКИЙ ПРОЦЕСС	200
8.4. СТОХАСТИЧЕСКИЙ ИНТЕГРАЛ ИТО	202
8.5. СУЩЕСТВОВАНИЕ И ЕДИНСТВЕННОСТЬ РЕШЕНИЙ СТОХАСТИЧЕСКИХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ	207
8.6. УПРАЖНЕНИЯ	210
ГЛАВА 9. Марковские случайные процессы	213
9.1. ЦЕПИ МАРКОВА. ОПРЕДЕЛЕНИЯ И ОСНОВНЫЕ СВОЙСТВА	213
9.2. СХОДИМОСТЬ ВЕРОЯТНОСТЕЙ ЦЕПИ МАРКОВА. СТАЦИОНАРНЫЕ ВЕРОЯТНОСТИ	216
9.3. ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ ЦЕПЕЙ МАРКОВА С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ	218
9.4. ВЕТВЯЩИЕСЯ ПРОЦЕССЫ С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ	220
9.5. КЛАССИФИКАЦИЯ СОСТОЯНИЙ ЦЕПИ МАРКОВА С ДИСКРЕТНЫМ ВРЕМЕНЕМ	223
9.6. ЭРГОДИЧЕСКИЕ ЦЕПИ МАРКОВА	226
9.7. УПРАЖНЕНИЯ	227
ГЛАВА 10. Мартингалы, полумартингалы и марковские моменты	235
10.1. МАРТИНГАЛЫ, ПОЛУМАРТИНГАЛЫ И МАРКОВСКИЕ МОМЕНТЫ	235
10.2. НЕРАВЕНСТВА ДЛЯ СУБМАРТИНГАЛОВ	238
10.3. ТЕОРЕМЫ О СХОДИМОСТИ	241
10.4. СХОДИМОСТЬ РЯДОВ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН	245
10.5. УПРАЖНЕНИЯ	248
ЧАСТЬ 3. МАТЕМАТИЧЕСКИЕ МЕТОДЫ СТАТИСТИКИ	249
	250
ГЛАВА 11. Теория статистического оценивания параметров	251

11.1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ	252
11.2. ВЫБОРОЧНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, ВЫБОРОЧНАЯ ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ, ВЫБОРОЧНЫЕ МОМЕНТЫ, ГИСТОГРАММА И ИХ СВОЙСТВА	254
11.3. ПОРЯДКОВЫЕ СТАТИСТИКИ, ИХ СВОЙСТВА И ПРИМЕНЕНИЕ В ТЕОРИИ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ	258
11.4. РАНГОВЫЕ СТАТИСТИКИ И ИХ СВОЙСТВА	262
11.5. НЕРАВЕНСТВО ИНФОРМАЦИИ. НЕРАВЕНСТВО КРАМЕРА – РАО. ЭФФЕКТИВНЫЕ ОЦЕНКИ	264
11.6. МЕТОД МОМЕНТОВ. ОЦЕНКИ ПО МЕТОДУ МОМЕНТОВ И ИХ СВОЙСТВА	271
11.7. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ И ЕГО СВОЙСТВА	272
11.8. БАЙЕСОВСКИЙ МЕТОД СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ	279
11.9. ДОСТАТОЧНЫЕ СТАТИСТИКИ И ИХ СВОЙСТВА	282
11.10. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И ЕГО СВОЙСТВА	287
11.11. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ	294
11.12. РОБАСТНОЕ СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ	301
11.13. УПРАЖНЕНИЯ	306
ГЛАВА 12. Теория статистической проверки гипотез	309
12.1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ СТАТИСТИЧЕСКОЙ ПРОВЕРКИ ГИПОТЕЗ	309
12.2. РЕШАЮЩЕЕ ПРАВИЛО НЕЙМАНА – ПИРСОНА	311
12.3. БАЙЕСОВСКОЕ РЕШАЮЩЕЕ ПРАВИЛО	314
12.4. ПРОВЕРКА ГИПОТЕЗ О РАСПРЕДЕЛЕНИИ ВЕРОЯТНОСТЕЙ. КРИТЕРИИ СОГЛАСИЯ	317
12.5. КРИТЕРИЙ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ ДЛЯ ПРОВЕРКИ СЛОЖНЫХ ГИПОТЕЗ	321
12.6. ПОНЯТИЕ О ПОСЛЕДОВАТЕЛЬНОМ АНАЛИЗЕ ВАЛЬДА	324
12.7. УПРАЖНЕНИЯ	327
ЧАСТЬ 4. СТАТИСТИЧЕСКИЙ АНАЛИЗ МНОГОМЕРНЫХ ДАННЫХ	329
.....	330
ГЛАВА 13. Разведочный анализ и сжатие данных	331
13.1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ И ЕГО ОСНОВНЫЕ ЭТАПЫ	331
13.2. ВЫБОРКА ИЗ МНОГОМЕРНЫХ НАБЛЮДЕНИЙ КАК МОДЕЛЬ СТАТИСТИЧЕСКИХ ДАННЫХ	331
13.3. РАЗВЕДОЧНЫЙ АНАЛИЗ ДАННЫХ	332
13.4. ПРОБЛЕМА СЖАТИЯ ДАННЫХ. МЕТОД ГЛАВНЫХ КОМПОНЕНТ	334
ГЛАВА 14. Многомерное нормальное распределение и оценивание параметров	337
14.1. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И ЕГО ОСНОВНЫЕ СВОЙСТВА	337
14.2. ОЦЕНКИ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ПАРАМЕТРОВ МНОГОМЕРНОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ. ВЫБОРОЧНОЕ СРЕДНЕЕ И ВЫБОРОЧНАЯ КОВАРИАЦИОННАЯ МАТРИЦА	341
14.3. ВЕРОЯТНОСТНЫЕ СВОЙСТВА ВЫБОРОЧНОГО СРЕДНЕГО И ВЫБОРОЧНОЙ КОВАРИАЦИОННОЙ МАТРИЦЫ. НЕСМЕЩЕННАЯ ВЫБОРОЧНАЯ КОВАРИАЦИОННАЯ МАТРИЦА	343
14.4. УПРАЖНЕНИЯ	347
ГЛАВА 15. Исследование зависимостей	349

15.1. ИССЛЕДОВАНИЕ ПАРНОЙ ЗАВИСИМОСТИ ПРИЗНАКОВ. ВЫБОРОЧНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ. Z-СТАТИСТИКА ФИШЕРА И ПРОВЕРКА ГИПОТЕЗ О ЗНАЧЕНИИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ	349
15.2. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ: УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ И ЧАСТНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ	352
15.3. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ: ПРОГНОЗИРОВАНИЕ, ФУНКЦИЯ РЕГРЕССИИ И МНОЖЕСТВЕННЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ	355
15.4. ВЫБОРОЧНЫЕ ЧАСТНЫЙ И МНОЖЕСТВЕННЫЙ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ. ПРОВЕРКА ГИПОТЕЗ НЕЗАВИСИМОСТИ	357
15.5. ПРОВЕРКА ОБЩИХ ГИПОТЕЗ О НЕЗАВИСИМОСТИ	364
15.6. УПРАЖНЕНИЯ	368
ГЛАВА 16. Проверка гипотез и статистическая классификация	369
16.1. ПРОВЕРКА ГИПОТЕЗ О ЗНАЧЕНИИ ВЕКТОРА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ. T^2 -СТАТИСТИКА ХОТЕЛЛИНГА	369
16.2. СРАВНЕНИЕ ВЕКТОРОВ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ ПО ДВУМ ВЫБОРКАМ. МНОГОМЕРНАЯ ПРОБЛЕМА БЕРЕНСА – ФИШЕРА	374
16.3. ПРОВЕРКА ГИПОТЕЗ ОТНОСИТЕЛЬНО ПАРАМЕТРОВ МНОГОМЕРНОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ	378
16.4. ПРОВЕРКА ГИПОТЕЗ ОТНОСИТЕЛЬНО НЕСКОЛЬКИХ ВЫБОРОК ИЗ МНОГОМЕРНЫХ НОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ	380
16.5. ЗАДАЧИ СТАТИСТИЧЕСКОЙ КЛАССИФИКАЦИИ. ОПТИМАЛЬНОЕ (БАЙЕСОВСКОЕ) РЕШАЮЩЕЕ ПРАВИЛО	385
16.6. СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ МНОГОМЕРНЫХ НОРМАЛЬНЫХ НАБЛЮДЕНИЙ	388
16.7. ДИСПЕРСИОННЫЙ АНАЛИЗ	396
16.8. УПРАЖНЕНИЯ	398
ЧАСТЬ 5. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ	399
.....	400
ГЛАВА 17. Основные понятия теории временных рядов	401
17.1. СЛУЧАЙНАЯ ФУНКЦИЯ КАК МОДЕЛЬ СТАТИСТИЧЕСКИХ НАБЛЮДЕНИЙ В ДИНАМИКЕ. КЛАССИФИКАЦИЯ СЛУЧАЙНЫХ ФУНКЦИЙ	401
17.2. ВРЕМЕННЫЕ РЯДЫ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ	403
17.3. СТАЦИОНАРНЫЕ ВРЕМЕННЫЕ РЯДЫ. СПЕКТРАЛЬНАЯ ПЛОТНОСТЬ СТАЦИОНАРНОГО В ШИРОКОМ СМЫСЛЕ ВРЕМЕННОГО РЯДА	404
ГЛАВА 18. Непараметрическое статистическое оценивание	407
18.1. ВЫБОРОЧНОЕ СРЕДНЕЕ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА И ЕГО СВОЙСТВА	407
18.2. ВЫБОРОЧНАЯ КОВАРИАЦИОННАЯ ФУНКЦИЯ СТАЦИОНАРНОГО В ШИРОКОМ СМЫСЛЕ ВРЕМЕННОГО РЯДА И ЕЕ СВОЙСТВА	409
18.3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ СПЕКТРАЛЬНОЙ ПЛОТНОСТИ	413
18.4. УПРАЖНЕНИЯ	416
ГЛАВА 19. Параметрический анализ стационарных временных рядов	417
19.1. ВРЕМЕННЫЕ РЯДЫ АВТОРЕГРЕССИИ, СКОЛЬЗЯЩЕГО СРЕДНЕГО, АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО: УСЛОВИЯ СТАЦИОНАРНОСТИ И КОВАРИАЦИОННЫЕ ФУНКЦИИ, ЕДИНОЕ ЛИНЕЙНОЕ ПРЕДСТАВЛЕНИЕ	417
19.2. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНОГО СРЕДНЕГО И ВЫБОРОЧНОЙ КОВАРИАЦИОННОЙ ФУНКЦИИ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА	420

19.3. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ АВТОРЕГРЕССИОН- НЫХ ВРЕМЕННЫХ РЯДОВ ПРИ ИЗВЕСТНОМ ПОРЯДКЕ АВТОРЕГРЕССИИ	421
19.4. ОЦЕНИВАНИЕ ПОРЯДКА АВТОРЕГРЕССИИ	424
19.5. ПРОГНОЗИРОВАНИЕ СТАЦИОНАРНЫХ В ШИРОКОМ СМЫСЛЕ ВРЕ- МЕННЫХ РЯДОВ	426
19.6. УПРАЖНЕНИЯ	430
ГЛАВА 20. Статистический анализ нестационарных временных рядов	431
20.1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ С ТРЕНДОМ В ОР- ТОГОНАЛЬНОМ БАЗИСЕ	431
20.2. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ВРЕМЕННЫХ РЯДОВ АВТОРЕГРЕС- СИИ И ПРОИНТЕГРИРОВАННОГО СКОЛЬЗЯЩЕГО СРЕДНЕГО	435
20.3. УПРАЖНЕНИЯ	440
ГЛАВА 21. Статистический анализ однородных цепей Маркова	441
21.1. ОДНОРОДНЫЕ ЦЕПИ МАРКОВА. ОСНОВНЫЕ ПОНЯТИЯ И СВОЙСТВА. УСЛОВИЯ СТАЦИОНАРНОСТИ	441
21.2. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ДЛЯ ОЦЕНИВАНИЯ МАТ- РИЦЫ ВЕРОЯТНОСТЕЙ ОДНОШАГОВЫХ ПЕРЕХОДОВ И СТАЦИОНАР- НОГО РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ	443
21.3. ПРОВЕРКА ГИПОТЕЗ О ЦЕПИ МАРКОВА	445
21.4. УПРАЖНЕНИЯ	446
ЛИТЕРАТУРА	447
Приложения	449
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ	455

Учебное издание

Харин Юрий Семенович

Учебник

Редактор
Технический редактор
Корректоры

Подписано в печать 00.00.2011. Формат 60х84/16. Бумага офсетная.
Гарнитура Таймс. Печать офсетная. Усл. печ. л.
Уч.-изд. л. Тираж 500 экз. Зак. .

Белорусский государственный университет.
Лицензия на осуществление издательской деятельности
№ 02330/0056804 от 02.03.2004.
220030, Минск, проспект Независимости, 4.

Отпечатано с оригинала-макета заказчика.
Республиканское унитарное предприятие
«Издательский центр Белорусского государственного университета».
Лицензия на осуществление полиграфической деятельности
№ 02330/0056850 от 30.04.2004.
220030, Минск, ул. Красноармейская, 6.