

М. В. Козлов, А. В. Прохоров

ВВЕДЕНИЕ
В МАТЕМАТИЧЕСКУЮ
СТАТИСТИКУ

ИЗДАТЕЛЬСТВО
МОСКОВСКОГО
УНИВЕРСИТЕТА
1987

УДК 519

Козлов М. В., Прохоров А. В. Введение в математическую статистику.—
М.: Изд-во МГУ, 1987.— 264 с.

Книга предназначена для начального изучения математической статистики.
Основные понятия, задачи и методы математической статистики вводятся на
примере простых статистических моделей. Значительное внимание уделено, с од-
ной стороны, численным и графическим иллюстрациям, с другой — логическим
основам математической статистики.

Для студентов университетов, обучающихся по специальностям «Матема-
тика», «Прикладная математика», «Механика».

Библиогр. 26 назв. Ил. 13.

Р е ц е н з е н т ы:

академик *A. Н. Колмогоров*,
д-р физ.-мат. наук *Ю. Н. Тюрин*

Печатается по постановлению
Редакционно-издательского совета
Московского университета

Учебное издание

Михаил Васильевич Козлов,
Александр Владимирович Прохоров

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ

Зав. редакцией *С. И. Зеленский*
Редакторы *А. А. Локшин, Ф. И. Горобец*
Технический редактор *Г. Д. Колоскова*
Корректоры *Л. А. Мушникова, Т. С. Милякова*

ИБ № 2612

Сдано в набор 05.06.86. Подписано к печати 05.02.87.
Л-62050 Формат 60×90/16 Бумага тип. № 3
Гарнитура литературная. Высокая печать
Усл. печ. л. 16,5 Уч.-изд. л. 17,04
Тираж 11000 экз. Заказ № 405
Цена 75 коп. Изд. № 4492

Ордена «Знак Почета» издательство Московского университета.
103009, Москва, ул. Герцена, 5/7.
Типография ордена «Знак Почета» изд-ва МГУ.
119899, Москва, Ленинские горы

K 1702060000—076
077(02)—87 120—87

© Издательство Московского
университета, 1987 г.

ОГЛАВЛЕНИЕ

Предисловие	6
Глава I. СТАТИСТИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ: НАЧАЛЬНЫЕ ПОНЯТИЯ	8
§ 1. Вероятность и частота	8
1. Введение. 2. Таблица случайных чисел. 3. Парадокс де Мере. 4. Общая вероятностная модель. 5. Закон больших чисел	8
§ 2. Эмпирическое распределение вероятностей	15
1. Эмпирическое распределение дискретной случайной величины. 2. Эмпирическая функция распределения. 3. Статистика Колмогорова. 4. Выборочные среднее и дисперсия.	15
§ 3. Порядковые статистики в задачах оценивания	22
1. Порядковые статистики. 2. Оценивание содержимого урны. 3. Оценка параметра равномерного распределения. 4. Оценивание параметра сдвига экспоненциального распределения. 5. Доверительный интервал. 6. Выборочные квантили. 7. Равномерное распределение. 8. Доверительные интервалы для квантилей. 9. Совместное распределение пары порядковых статистик.	22
§ 4. Параметры сдвига и масштаба: графический анализ	31
1. Семейство сдвига-масштаба. 2. Вероятностная бумага. 3. Примеры графического анализа.	31
§ 5. Экспоненциальное распределение и пуассоновский процесс	36
1. Экспоненциальное распределение. 2. Пуассоновский процесс. 3. Условное распределение точек пуассоновского процесса. 4. Пример.	36
§ 6. Оценивание параметров экспоненциального распределения	44
1. Оценивание параметра сдвига при известном параметре масштаба. 2. Оценивание параметра масштаба при известном параметре сдвига. 3. Точечные оценки μ и σ . 4. Распределение интервалов между порядковыми статистиками. 5. Дисперсии несмещенных оценок μ^* , σ^* ; интервальное оценивание μ , σ . 6. Оценивание по первым r порядковым статистикам.	44
§ 7. Сведения о важнейших непрерывных распределениях в R^1	51
1. Гамма-распределение. 2. Экспоненциальное распределение. 3. Нормальное распределение. 4. Распределение χ^2 с l степенями свободы. 5. Бета-распределение. 6. Равномерное распределение. 7. Распределение Фишера, или F -распределение. 8. Распределение Стьюдента, или t -распределение.	51
§ 8. Нормальное распределение: оценивание параметров, сравнение двух выборок	58
1. Доверительные оценки для μ и σ^2 , когда один из параметров известен. 2. Несмещенная оценка дисперсии. 3. Доверительные интервалы для μ и σ^2 , когда оба параметра неизвестны. 4. Сравнение дисперсий в двух выборках. 5. Задача сравнения средних μ_1 и μ_2 . 6. Пример.	58
Глава II. ЛИНЕЙНАЯ СТАТИСТИЧЕСКАЯ МОДЕЛЬ	69
§ 9. Оценивание коэффициентов линейной модели	69
1. Введение. 2. Примеры линейных моделей. 3. Линейная статистическая модель. 4. Несмешенные оценки с минимальной дисперсией:	69

матрица полного ранга. 5. Наилучшие оценки в случае матрицы неполного ранга. 6. Пример: модель с матрицей неполного ранга. 7. Наилучшие оценки как оценки наименьших квадратов	
§ 10. Ковариации, каноническая форма линейной модели, обобщения	80
1. Матрицы из случайных элементов. 2. Матрица ковариаций. 3. Ковариации наилучших оценок. 4. Пример оптимального выбора матрицы модели. 5. Каноническая форма линейной модели. 6. Оценивание дисперсии. 7. Обобщенная линейная модель	
§ 11. Порядковые линейные оценки для параметров сдвига и масштаба	89
1. Обобщенная линейная модель для семейства сдвига-масштаба. 2. Наилучшие несмещенные оценки. 3. Симметричное распределение. 4. Пример: равномерное распределение на интервале (μ , $\mu + \sigma$). 5. Пример: экспоненциальное распределение. 6. Пример: нормальное распределение. 7. Цензурированная выборка. 8. Упрощенные линейные оценки.	
§ 12. Многомерное нормальное распределение	101
1. Невырожденное нормальное распределение. 2. Случайные векторы с вырожденной матрицей ковариаций. 3. Вырожденное нормальное распределение. 4. Распределение проекций стандартного нормального вектора.	
§ 13. Доверительное оценивание и проверка гипотез в линейной модели с нормальными наблюдениями	109
1. Распределение вектора оценок. 2. Доверительные области для параметров и параметрических функций. 3. Проверка гипотез с помощью доверительных эллипсоидов. 4. Пример: сравнение средних в нескольких нормальных выборках.	
Глава III. ДОСТАТОЧНЫЕ СТАТИСТИКИ	116
§ 14. Статистическая модель, подобные статистики	116
1. Статистическая модель. 2. Подобные статистики.	
§ 15. Достаточные статистики в дискретной модели	120
1. Введение. 2. Примеры. 3. Достаточные статистики: определение. 4. Критерий достаточности. 5. Достаточные и минимальные достаточные разбиения.	
§ 16. Достаточные статистики в непрерывной модели	133
1. Примеры достаточных статистик. 2. Определение достаточной статистики, теорема факторизации. 3. Экспоненциальные семейства, минимальная достаточность.	
§ 17. Достаточность и несмещение оценивание	152
1. Полные достаточные статистики. 2. Наилучшие несмещенные оценки в дискретной модели. 3. Наилучшие несмещенные оценки в непрерывной модели.	
§ 18. Информация в статистике	168
1. Байесовский подход в статистике. 2. Информация по Шеннону. 3. Информация по Кульбаку. 4. Информация по Фишеру.	
§ 19. Неравенство Фреше—Ро—Крамера	185
1. Скалярный параметр. 2. Векторный параметр. 3. Границы дисперсии при нарушении условий регулярности.	
Глава IV. ПРАВДОПОДОБИЕ	198
§ 20. Метод максимума правдоподобия	198
1. Функция правдоподобия. 2. Оценки максимального правдоподобия. 3. Метод Монте-Карло в модели сдвига-масштаба.	
§ 21. Критерий отношения правдоподобий	211
1. Проверка статистических гипотез. 2. Лемма Неймана—Пирсона. 3. Примеры. Равномерно наиболее мощные критерии. 4. Близкие гипотезы, 5. Сложные гипотезы.	

§ 22. Последовательный критерий отношения правдоподобий	226
1. Метод последовательного статистического анализа. 2. Последовательный критерий отношения правдоподобий. 3. Среднее число наблюдений в последовательном критерии.	
Глава V. БОЛЬШИЕ ВЫБОРКИ	235
§ 23. Асимптотические свойства оценок	235
1. Состоительность. 2. Состоительность оценок максимального правдоподобия. 3. Асимптотическая нормальность. 4. Преобразование статистик. 5. Асимптотическая нормальность выборочных квантилей. 6. Асимптотическая нормальность оценок максимума правдоподобия. 7. Асимптотическая эффективность оценок максимума правдоподобия. 8. Асимптотическая достаточность. 9. Векторный параметр. 10. Оценивание параметров сдвига и масштаба.	
§ 24. Асимптотические свойства критерия отношения правдоподобий	256
1. Скалярный параметр. 2. Векторный параметр. 3. Полиномиальное распределение.	
Литература	264

ПРЕДИСЛОВИЕ

По замыслу авторов книга должна служить учебным пособием для студентов-математиков, знакомых с элементами теории вероятностей в объеме семестрового курса. Она написана на основе курсов, читавшихся авторами для студентов механико-математического факультета Московского университета, специализирующихся в области теории вероятностей и математической статистики.

Среди книг, рассчитанных на первоначальное изучение предмета, особое место занимают классические труды Г. Крамера [20] и Б. Л. Ван дер Вардена [10]. Написанные более 30 лет назад и воспитавшие поколения вероятностников, они пользуются заслуженным авторитетом и сейчас. Элементы математической статистики являются составной частью общих курсов по теории вероятностей, среди которых следует назвать учебники Б. В. Гнеденко [17], Ю. А. Розанова [18], Б. А. Севастьянова [19], В. Н. Тутубалина [21]. Недавно изданное учебное пособие по математической статистике А. А. Боровкова [12] рассчитано на студентов старших курсов и аспирантов. Отметим также учебник для вузов Г. И. Ивченко, Ю. И. Медведева [22]. При работе над книгой авторы использовали методические достижения отечественной и зарубежной литературы по математической статистике, а также опыт и традиции преподавания вероятностных дисциплин на механико-математическом факультете и факультете вычислительной математики и кибернетики Московского университета. Особо отметим работу Д. Кокса, Д. Хинкли [14], а также уже упоминавшуюся книгу Б. Л. Ван дер Вардена, логическая ясность, простота изложения и отчетливое представление прикладной стороны предмета в которой служили нам образцом.

Книга начинается со знакомства с некоторыми основополагающими понятиями математической статистики на примере простейших статистических моделей (гл. I, II). Значительное внимание уделяется здесь числовым и графическим иллюстрациям, тогда как логика предмета отнесена на второй план. Достаточно детально изучается линейная статистическая модель. Эта модель, во-первых, служит хорошим источником материала для педагогической практики, во-вторых, открывает подход к теоретическим и прикладным аспектам регрессионного и дисперсионного анализа. В настоящей книге линейная модель должна еще и подготовить

читателя к восприятию теоретических основ статистики, которым уделяется серьезное внимание в последующих разделах.

Вопросам использования информации, содержащейся в статистических данных, для построения статистических выводов и формализации самого понятия информации в статистике посвящена гл. III, содержащая разделы: достаточность, несмещенное оценивание, информация в статистике, нижние границы дисперсии оценок. Гл. IV объединяет некоторые приемы построения статистических оценок и критериев, основанных на понятии правдоподобия. Гл. V посвящена асимптотическим свойствам оценок максимального правдоподобия и критерия отношения правдоподобий.

Ограниченный объем книги не позволил включить некоторые важные статистические процедуры и затронуть другие разделы современной математической статистики. Вместо этого авторы более детально, с большим количеством примеров и повышенным вниманием к логической стороне вопроса разобрали такие фундаментальные понятия статистики, как статистическая модель, достаточность, правдоподобие. Авторы надеются, что тщательная проработка материала книги поможет читателю самостоятельно разобрать темы, затронутые нами лишь частично, по другим источникам, среди которых мы выделим капитальный труд М. Дж. Кендалла, А. Стьюарта [24—26]. Библиографический список в конце книги включает в основном издания на русском языке и ни в какой мере не претендует на полноту.

Авторы пользуются возможностью выразить благодарность своим учителям и коллегам — сотрудникам кафедры теории вероятностей и кафедры математической статистики и случайных процессов механико-математического факультета МГУ. Авторы благодарны сотрудникам факультета Н. Н. Марчук, Т. В. Нистратовой, Т. В. Козьминой, О. Б. Фисуненко, а также студентам и аспирантам кафедры математической статистики и случайных процессов за помощь в оформлении рукописи. Рассматривая книгу как очередной шаг в разработке методики преподавания математической статистики, авторы отдают себе ясный отчет в несовершенстве своего труда и будут благодарны всем, кто выскажет свои замечания и предложения.

М. В. Козлов, А. В. Прохоров

ГЛАВА I

СТАТИСТИЧЕСКИЕ МОДЕЛИ И МЕТОДЫ: НАЧАЛЬНЫЕ ПОНЯТИЯ

§ 1. ВЕРОЯТНОСТЬ И ЧАСТОТА

1. Введение.

Процесс познания окружающего мира включает наблюдение и эксперимент. Результаты наблюдений во многих случаях можно представить последовательностью действительных чисел

$$(x_1, x_2, \dots, x_n) = x.$$

Для того чтобы из ряда наблюдений можно было извлечь полезную информацию, необходимо иметь модель явления. Вероятностные модели оказываются чрезвычайно плодотворными при анализе явлений, исходы $x \in \mathcal{X}$ которых обладают некоторой степенью неопределенности. Понятие неопределенности в теории вероятностей формализуется путем введения распределения вероятностей на множестве \mathcal{X} всех возможных наблюдений x . В простейшем случае, когда \mathcal{X} — конечное или счетное множество, задаются вероятности $p(x)$ всех его элементов x , так что $0 \leq p(x) \leq 1$ и

$$\sum_{x \in \mathcal{X}} p(x) = 1.$$

Любое подмножество $A \subseteq \mathcal{X}$ называют в этом случае *событием*, а его *вероятность* определяют формулой

$$P(A) = \sum_{x \in A} p(x).$$

Пример 1 (*случайный выбор с возвращением*). Представим себе урну с N шарами, занумерованными числами от 0 до $N - 1$, и рассмотрим опыт, который заключается в n -кратном извлечении шара «наугад» с последующим его возвращением в урну. Все возможные исходы опыта могут быть представлены множеством из N^n последовательностей:

$$\mathcal{X} = \{x = (x_1, \dots, x_n) : x_i = 0, 1, \dots, N - 1; i = 1, \dots, n\}.$$

На \mathcal{X} задается равномерное распределение вероятностей: $p(x) =$

$=N^n$ для всех $x \in \mathcal{X}$, что отражает равную степень неопределенности каждого возможного исхода.

Пример 2 (*случайный выбор без возвращения*). Если из урны с N шарами последовательно извлекаются «наугад» n шаров без возвращения, $n < N$, то возможные исходы описываются множеством из $N(N-1)\dots(N-n+1) = N!/(N-n)!$ последовательностей:

$$\mathcal{X} = \{x = (x_1, x_2, \dots, x_n) : x_i = 0, 1, \dots, N-1, x_i \neq x_j \text{ при } i \neq j; i, j = 1, \dots, n\},$$

а распределение вероятностей на \mathcal{X} снова равномерное: для любого x

$$p(x) = (N-n)!/N!$$
 ■

Взаимоотношение явления и его вероятностной модели имеет статистический характер, т. е. обнаруживается при повторных наблюдениях за явлением. Частоты исходов в длинном ряду испытаний стабилизируются, их колебания с ростом числа испытаний уменьшаются. Этот эмпирический факт, называемый законом устойчивости частот, наблюдается в самых различных ситуациях. Уже выходя за пределы реального опыта, полагают, что при неограниченном повторении частоты стремятся к пределам, которые и принимают за вероятности соответствующих исходов или событий.

2. Таблица случайных чисел.

Обратимся к табл. 7.1, а книги Л. Н. Большева и Н. В. Смирнова [1], содержащей 12500 «случайных цифр», которые можно представлять себе как реализацию опыта, описанного в примере 1 при $N=10$, $n=12500$. Проследим за изменчивостью частот появления цифры «0» с ростом числа наблюдений. Для облегчения анализа сгруппируем первые 100 цифр таблицы (читываемой по строкам) по 10 цифр, последующие 900 — по 100 цифр, следующие 9000 — по 1000. Наблюдаемые числа и частоты нулей по группам и суммарно собраны в табл. 1.

Нетрудно подметить, глядя на последний столбец таблицы 1, что колебания частоты m/n с ростом n затухают.

3. Парадокс де Мере.

Рассмотрим задачу, которая, как утверждают, возникла за игрорным столом и была предложена в 1654 г. Паскалю. Игрок де Мере полагал, что вероятность p_1 получить хотя бы одну единицу при бросании четырех игральных костей (событие A) меньше вероятности p_2 выпадения одновременно двух единиц хотя бы раз при 24 бросаниях двух костей (событие B), и ставил в игре на событие B .

Таблица 1

Колебание частоты появления нуля в 10 000 случайных цифр

Число наблюдений в группе	Число нулей в группе	Частота нулей в группе	Накопленное число наблюдений, n	Накопленное число нулей, m	Накопленная частота нулей, m/n
10	2	0,2	10	2	0,2
10	1	0,1	20	3	0,15
10	0	0,0	30	3	0,10
10	2	0,2	40	5	0,15
10	0	0,0	50	5	0,16
10	2	0,2	60	7	0,116
10	0	0,0	70	7	0,100
10	2	0,2	80	9	0,112
10	3	0,3	90	12	0,133
10	2	0,2	100	14	0,140
100	14	0,14	200	28	0,140
100	9	0,09	300	37	0,123
100	13	0,13	400	50	0,125
100	8	0,08	500	58	0,116
100	16	0,16	600	74	0,123
100	14	0,14	700	88	0,125
100	6	0,06	800	94	0,117
100	9	0,09	900	103	0,114
100	8	0,08	1 000	111	0,1110
1000	97	0,097	2 000	208	0,1040
1000	106	0,106	3 000	314	0,1046
1000	99	0,099	4 000	413	0,1032
1000	91	0,091	5 000	504	0,1008
1000	97	0,097	6 000	601	0,1001
1000	91	0,091	7 000	692	0,0988
1000	100	0,100	8 000	792	0,0990
1000	117	0,117	9 000	909	0,1010
1000	89	0,089	10 000	998	0,0998

Однако эти вероятности легко вычисляются:

$$\begin{aligned} P(A) = p_1 &= 1 - (5/6)^4 = 0,5177 \dots, P(B) = p_2 = 1 - (35/36)^{24} = \\ &= 0,4914 \dots \end{aligned}$$

Поскольку де Мере не были доступны эти вычисления, то свои заключения он мог бы основывать на опытных данных. Представим себе, что опыт с бросанием четырех игральных костей и опыт из 24 бросаний двух игральных костей повторяются n раз. Результат i -го испытания, $i=1, 2, \dots, n$, опишем парой (x_i, y_i) , полагая $x_i=1$ ($y_i=1$), если произошло событие A (B), и полагая $x_i=0$ ($y_i=0$), если событие A (B) не произошло. Обозначим через

$$m_n^{(1)}/n = \frac{1}{n} \sum_{t=1}^n x_t, \quad m_n^{(2)}/n = \frac{1}{n} \sum_{t=1}^n y_t.$$

Естественный вопрос заключается в том, насколько основательным является сравнение вероятностей p_1 и p_2 по их частотам

$m_n^{(1)}/n$ и $m_n^{(2)}/n$. Ниже приводятся результаты моделирования игры с помощью таблицы случайных цифр из [1].

n	25	50	100	250
$m_n^{(1)}/n$	0,68	0,54	0,53	0,488
$m_n^{(2)}/n$	0,52	0,50	0,48	0,508

Как видно, хотя $p_1 > p_2$, но при $n=250$ нам «не повезло» — $m_n^{(1)}/n < m_n^{(2)}/n$. Остановив «игру» на этом n , мы могли бы сделать ошибочное заключение, что $p_1 < p_2$.

В теории описанному опыту соответствуют две независимые последовательности испытаний Бернулли с вероятностями положительного исхода («успеха») в одном случае p_1 , в другом — p_2 . Введем случайные величины (сл. в.) v_1 и v_2 , равные числу успехов в каждой из этих последовательностей.

Вероятность события $\{v_1 < v_2\}$ для больших n рассчитаем приближенно, пользуясь теоремой Муавра—Лапласа. Сл. в.

$$v_i^* = (v_i - np_i)/\sqrt{np_i q_i}, \quad q_i = 1 - p_i, \quad i = 1, 2, \quad (1)$$

независимы и приближенно нормально распределены с нулевым математическим ожиданием и единичной дисперсией. Неравенство $v_1 < v_2$ после подстановки в него (1) запишем в виде

$$\frac{\sqrt{p_1 q_1} v_1^* - \sqrt{p_2 q_2} v_2^*}{\sqrt{p_1 q_1 + p_2 q_2}} < \frac{\sqrt{n} (p_2 - p_1)}{\sqrt{p_1 q_1 + p_2 q_2}}. \quad (2)$$

Так как разность независимых нормальных сл. в. распределена нормально, то левая часть (2) при больших n имеет распределение, близкое к нормальному, с математическим ожиданием и дисперсией, равными

$$M(\sqrt{p_1 q_1} v_1^* - \sqrt{p_2 q_2} v_2^*) = \sqrt{p_1 q_1} \cdot 0 - \sqrt{p_2 q_2} \cdot 0 = 0$$

и

$$D\left(\frac{\sqrt{p_1 q_1} v_1^* - \sqrt{p_2 q_2} v_2^*}{\sqrt{p_1 q_1 + p_2 q_2}}\right) = \frac{p_1 q_1 \cdot 1 + p_2 q_2 \cdot 1}{p_1 q_1 + p_2 q_2} = 1.$$

Таким образом, имеем приближенное равенство (обозначение \approx)

$$P\{v_1 < v_2\} \approx \Phi\left(\frac{\sqrt{n} (p_2 - p_1) + 0,5}{\sqrt{p_1 q_1 + p_2 q_2}}\right),$$

$$\text{где } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt -$$

функция распределения стандартного нормального закона, а слагаемое 0,5 — поправка на дискретность. Используя таблицу нормального распределения, для различных n получаем

n	25	50	100	250	1000
$P(v_1 < v_2)$	0,42	0,39	0,35	0,18	0,12

Итак, даже при $n=1000$ приблизительно 12 шансов из 100 за то, что v_1 окажется меньше v_2 и будет сделан неправильный вывод, что $p_1 < p_2$.

4. Общая вероятностная модель.

В современной теории вероятностей под вероятностной моделью понимают тройку объектов $(\Omega, \mathcal{A}, \mathcal{P})$, называемую *вероятностным пространством*. Здесь Ω — некоторое множество, элементы которого ω называются *элементарными событиями*, \mathcal{A} — система подмножества множества Ω , называемых *событиями*, \mathcal{P} — вероятностная функция, ставящая в соответствие каждому событию A число $\mathcal{P}(A)$, $0 \leq \mathcal{P}(A) \leq 1$, называемое его *вероятностью*. В соответствии с аксиоматикой Колмогорова требуется, чтобы система событий была замкнута относительно перехода к дополнению и конечному или счетному объединению:

$$A \in \mathcal{A} \rightarrow \bar{A} = \{\omega : \omega \notin A\} \in \mathcal{A}, \\ A_i \in \mathcal{A}, i = 1, 2, \dots \rightarrow \bigcup A_i \in \mathcal{A},$$

а также содержала все Ω : $\Omega \in \mathcal{A}$ и притом $\mathcal{P}(\Omega) = 1$ (легко видеть, что конечное или счетное пересечение событий также является событием); вероятностная мера \mathcal{P} должна обладать свойством счетной аддитивности: если $A_i \cap A_j = \emptyset$ при $i \neq j$, то

$$\mathcal{P}\left(\bigcup_i A_i\right) = \sum_i \mathcal{P}(A_i),$$

где индекс i пробегает конечное или счетное множество.

Вероятностная модель явления ставит в соответствие результатам наблюдений

$$x_1, x_2, \dots, x_n \tag{3}$$

последовательность случайных величин (сл. в.) — функций на Ω :

$$X_1(\omega), X_2(\omega), \dots, X_n(\omega). \tag{4}$$

Полагают, что наблюдения (3) являются значениями величин (4) при осуществлении ω . Несмотря на различие объектов (3) и (4), в математической статистике принято называть и то и другое

выборкой. Статистическим закономерностям числовых данных реального эксперимента (3) отвечают вероятностные утверждения для случайных величин (4).

5. Закон больших чисел.

Табл. 1 составлена по результатам испытаний, порождающих последовательность «случайных» цифр. Здесь $x_i \in \{0, 1, 2, \dots, 9\}$, $X_i(\omega)$ — независимые сл. в., принимающие значения $0, 1, 2, \dots, 9$ с равными вероятностями $p = 1/10$. Обозначим через m_n число появлений цифры 0 в первых n испытаниях. Анализ последнего столбца 1 привел нас к заключению, что частота нуля m_n/n с ростом n совершает затухающие колебания вблизи $1/10$. Это — выражение эмпирического закона устойчивости частот. Соответствующее математическое утверждение — закон больших чисел (З.Б.Ч.) для сл. в. X_i . Введем индикаторную сл. в.

$$I_{\{X_i(\omega)=0\}} = \begin{cases} 1, & \text{если } X_i(\omega)=0, \\ 0, & \text{если } X_i(\omega) \neq 0, \end{cases}$$

$$v_n = \sum_{i=1}^n I_{\{X_i(\omega)=0\}}.$$

Тогда З.Б.Ч. в усиленной форме выражается соотношением

$$\mathcal{P}(\lim_{n \rightarrow \infty} v_n/n = p) = 1.$$

Для практических целей полезно иметь оценки вероятностей

$$\mathcal{P}(|v_n/n - p| < \delta)$$

при конечных n . Это можно сделать, используя распределение вероятностей величины v_n :

$$\mathcal{P}(v_n = m) = C_n m p^m (1-p)^{n-m}, \quad m=0, 1, 2, \dots, n, \quad (5)$$

где $p = 1/10$. Для $n = 10$ расчет по формуле (5) дает следующие значения (см. табл. 5.1 биномиального распределения в [1]):

m	0	1	2	3	4	5	6
$\mathcal{P}(v_n = m)$	0,3486	0,3874	0,1937	0,574	0,0111	0,0014	0,0001
$\mathcal{P}(v_n < m)$	0,3486	0,7360	0,9297	0,9871	0,9982	0,9997	0,9998

При $n = 100$ непосредственные вычисления по формуле (5) нецелесообразны. Хорошее приближение к биномиальному распре-

делению (5) дает в этом случае распределение Пуассона:

$$\mathcal{P}(v_n = m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \quad m = 0, 1, 2, \dots,$$

где $\lambda = np = 10$. По табл. 5.3 пуассоновского распределения в [1] находим приближенные значения вероятностей

$$p_i = \mathcal{P}(|v_n - 10| \leq i) = \mathcal{P}(|v_n/n - 0.1| \leq i \cdot 0.01):$$

i	1	2	3	4	5	6	7	8
p_i	0,3639	0,5711	0,7340	0,8490	0,9215	0,9620	0,9822	0,9914

При $n=1000$ для расчета (5) надо воспользоваться нормальным приближением

$$\mathcal{P}\left(\frac{v_n - np}{\sqrt{npq}} < x\right) \approx \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Выпишем для нескольких значений x приближенные значения вероятностей

$$\mathcal{P}(|v_n/n - 0.1| \leq x \cdot 0.0095) = p_x$$

из таблицы значений функции нормального распределения:

x	1	2	2,6	3
p_x	0,6827	0,9545	0,9916	0,9973

Подведем итог приведенным расчетам в форме следующего соотношения:

$$\mathcal{P}(|v_n/n - 0.1| > \delta) = a \approx 0.01, \quad (6)$$

где при $n=10$ следует брать $\delta=0.3$; при $n=100$ — $\delta=0.08$; при $n=1000$ — $\delta=0.025$.

Статистический смысл неравенства (6) выясняется при многократном повторении опыта, порождающего n случайных цифр. Допустим, результаты большого числа N таких опытов представлены последовательностями

$$x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)}, \quad j = 1, 2, \dots, N, \quad (7)$$

и пусть $m_n^{(j)}$ — число нулей в (7). Обозначим через M_δ число тех последовательностей (7), для которых выполнено неравенство

$$|m_n^{(j)}/n - p| < \delta.$$

Тогда имеет место приближенное равенство

$$M_n/N \approx 1 - a,$$

выражающее закон устойчивости частот в применении к событию

$$\{|v_n/n - p| < \delta\}.$$

§ 2. ЭМПИРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

1. Эмпирическое распределение дискретной случайной величины.

Предположим сначала, что X дискретна и принимает значения a, b, c, \dots с вероятностями

$$p(a), p(b), p(c), \dots$$

По ряду наблюдений

$$x_1, x_2, \dots, x_n \quad (1)$$

вычисляем частоты $\rho(x) = m_n(x)/n$ исходов $x=a, b, c, \dots$. Набор частот

$$\{\rho(a), \rho(b), \dots\} \quad (2)$$

представляет собой дискретное распределение вероятностей:

$$0 \leq \rho(x) \leq 1, \quad \rho(a) + \rho(b) + \dots = 1;$$

его называют *эмпирическим распределением* (э. р.), отвечающим наблюдениям (1). Э. р. (2) является оценкой теоретического распределения сл. в. X :

$$\{p(a), p(b), \dots\}. \quad (3)$$

Пример 1. Пусть теоретическое распределение — биномиальное:

$$p(j) = C_k j p^j (1-p)^{k-j}, \quad j=0, 1, \dots, k. \quad (4)$$

Смоделируем выборку объема $n=40$ из распределения (4) с параметрами $k=250$, $p=0.1$, объединяя первые 10 000 случайных цифр из [1] в $n=40$ групп по 250 цифр и подсчитывая число x_i нулей в i -й группе, $i=1, 2, \dots, 40$. Результаты представлены ниже:

32	26	31	22	31	13	25	28	24	26
30	26	22	18	29	30	19	24	23	25
25	21	24	27	31	24	19	17	17	28
30	25	24	31	32	30	22	24	23	20

Эмпирическое и теоретическое распределения представлены в табл. 2.

Таблица 2

Вероятности $p(j) = C_k p^j (1-p)^{k-j}$;
 $k = 250$: $p = 0,1$ и их оценки $\hat{p}(j)$
 по выборке объема $n = 40$

j	$\hat{p}(j)$	$p(j)$
<13	0,0	—
13	0,025 (0,003)	0,0019
17	0,05	0,020
18	0,025	0,028
19	0,05	0,038
20	0,025	0,048
21	0,025	0,059
22	0,075	0,068
23	0,05	0,076
24	0,15	0,081
25	0,1	0,083
26	0,075	0,081
27	0,025	0,076
28	0,05	0,068
29	0,025	0,059
30	0,1	0,048
31	0,1	0,038
32	0,05	0,028
>32	0,0	—

Таблица 3

Эмпирическая и теоретическая функции распределения $\hat{F}(x)$, $F(x)$ для выборки из биномиального распределения;
 $\Delta(x) = |F(x) - \hat{F}(x)|$

x	$\hat{F}(x)$	$F(x)$	$\Delta(x)$
13	0,025	0,01	0,015
17	0,075	0,05	0,025
18	0,1	0,07	0,03
19	0,15	0,10	0,04
20	0,175	0,15	0,025
21	0,2	0,21	0,01
22	0,275	0,27	0,005
23	0,325	0,34	0,015
24	0,475	0,42	0,055
25	0,575	0,5	0,075
26	0,65	0,58	0,07
27	0,675	0,66	0,015
28	0,725	0,73	0,005
29	0,75	0,79	0,04
30	0,85	0,85	0,00
31	0,95	0,89	0,06
32	1	0,92	0,08

Теоретическое распределение рассчитано приближенно с помощью локальной предельной теоремы Муавра—Лапласа:

$$p(j) = \frac{1}{\sqrt{kp(1-p)}} \Phi \left(\frac{j - kp}{\sqrt{kp(1-p)}} \right), \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; \quad (5)$$

значение $p(13)$ вычислено непосредственно по формуле (4)
 (в скобках указано приближенное значение по формуле (5)).

Как видно из табл. 2, относительная погрешность оценки

$$[(\hat{p}(j) - p(j))/p(j)] 100\%$$

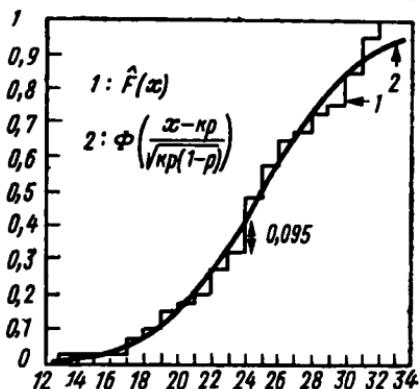
в интервале $17 \leq j \leq 32$ достигает 150%. Очевидно, для хорошей оценки распределения (3) необходимо, чтобы доля наблюдений, приходящихся на каждую из представляющих интерес вероятностей, была достаточно велика. Этого можно добиться группировкой, оценивая вероятности не отдельных значений, а группы значений.

С теоретической точки зрения намного более полезным является представление эмпирической информации в виде функции распределения (ф. р.)

$$\hat{F}(x) = \sum_{j \leq x} \hat{p}(j).$$

В табл. 3 и на рис. 1 проводится сравнение $F(x)$ с $F(x) = \Phi\left(\frac{x-kp}{\sqrt{kp(1-p)}}\right)$, рассчитанной с помощью интегральной теоремы Муавра — Лапласа.

Абсолютная погрешность $\Delta(x) = |F(x) - F(x)|$ достигает 0,08 при $13 < x < 32$, а относительная погрешность $[\Delta(x)/F(x)] \times 100\%$ при $17 < x < 32$ не превосходит 40%.



2. Эмпирическая функция распределения.

Распределение вероятностей произвольной случайной величины X характеризуют функцией распределения (ф.р.) $F(x)$. При фиксированном x $F(x)$ определяется как вероятность события $\{X \leq x\}$, и ее можно оценить частотой этого события в ряду наблюдений (1). Соответствующую частоту принято обозначать $F_n(x)$. Иногда желательно отметить в обозначениях зависимость $F_n(x)$ от $x = (x_1, x_2, \dots, x_n)$:

$$F_n(x) = F_n(x; x_1, x_2, \dots, x_n) = F_n(x; x).$$

Используя индикаторы, получаем следующую формулу:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{x_i \leq x\}}$$

($I_A = 1$, если утверждение A истинно, и $I_A = 0$ — в противном случае).

Функция $F_n(x)$, так, как она определена, есть функция распределения дискретной сл.в., принимающей значения x_1, x_2, \dots, x_n с вероятностью $1/n$ (если какие-то k из x_i совпадают между собой, то их общему значению приписывается вероятность k/n), и ее называют *эмпирической функцией распределения* (э. ф. р.).

Перейдем к вероятностной модели *повторных испытаний* — последовательности

$$X_1, X_2, \dots, X_n \quad (6)$$

независимых сл.в. с одинаковой ф.р. $F(x)$. В математической статистике принято употреблять одни и те же термины для характеристик последовательности сл.в. (6) и принятых ими в ре-

зультате опыта значений (1). В частности,

$$F_n(x; \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}, \quad \mathbf{X} = (X_1, \dots, X_n),$$

также называют э. ф. р. Подчеркнем, что $F_n(x; \mathbf{X})$ при каждом фиксированном x — случайная величина. Так как

$$MF_n(x; \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n M I_{\{X_i \leq x\}} = F(x),$$

$$DF_n(x; \mathbf{X}) = \frac{1}{n^2} \sum_{i=1}^n D I_{\{X_i \leq x\}} = \frac{F(x)(1 - F(x))}{n},$$

то из неравенства Чебышева получаем следующую оценку для отклонения $F_n(x)$ от $F(x)$:

$$\mathcal{P}(|\bar{n}|F_n(x; \mathbf{X}) - F(x)| > t) \leq F(x)(1 - F(x))t^{-2}. \quad (7)$$

Лучшую оценку для вероятности в левой части соотношения (7) при больших n можно получить применением интегральной теоремы Муавра—Лапласа:

$$\mathcal{P}(|\bar{n}|F_n(x; \mathbf{X}) - F(x)| > t) \approx 2(1 - \Phi(t/\sqrt{F(x)(1 - F(x))})). \quad (8)$$

Заметим, что правые части (7) и (8) зависят от величины $F(x)$, которую мы как раз и пытаемся оценить по результатам наблюдений. Заменяя $F(x)(1 - F(x))$ ее наибольшим возможным значением $1/4$, эту зависимость мы устраним, но ценой огрубления оценки.

3. Статистика Колмогорова.

Рассмотрим следующую меру расхождения э. ф. р. и теоретической ф. р.

$$D_n(\mathbf{X}) = \sup_x |F_n(x; \mathbf{X}) - F(x)|,$$

называемую статистикой Колмогорова.

Л е м м а. *Распределение вероятностей сл. в. $D_n(\mathbf{X})$ не зависит от ф. р. $F(x) = \mathcal{P}(X_i \leq x)$, если только $F(x)$ — непрерывна.*

Доказательство. Мы покажем, что распределение сл. в. $D_n(\mathbf{X})$ совпадает с распределением $D_n(\mathbf{Y})$, где $\mathbf{Y} = (Y_1, \dots, Y_n)$ — вектор из независимых равномерно распределенных на $[0, 1]$ сл. в. Предположим вначале, что $F(x)$ строго монотонная и $F^{-1}(y)$ — обратная к ней функция. Случайные величины $Y_i = F(X_i)$ независимы, а их общая ф. р. при $0 < y < 1$ равна

$$R(y) = \mathcal{P}(Y_1 \leq y) = \mathcal{P}(F(X_1) \leq y) = \mathcal{P}(X_1 \leq F^{-1}(y)) = F(F^{-1}(y)) = y,$$

т. е. Y_i равномерно распределена на $[0, 1]$. Для э. ф. р.

$$R_n(y; \mathbf{Y}), \quad \mathbf{Y} = (Y_1, \dots, Y_n),$$

имеем при $0 < y < 1$

$$\begin{aligned} R_n(y; \mathbf{Y}) &= \frac{1}{n} \sum_{i=1}^n I_{\{Y_i \leq y\}} = \frac{1}{n} \sum_{i=1}^n I_{\{F(X_i) \leq y\}} = \\ &= \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq F^{-1}(y)\}} = F_n(F^{-1}(y); \mathbf{X}). \end{aligned}$$

Отсюда получаем

$$\begin{aligned} \sup_{x: 0 < F(x) < 1} |F_n(x; \mathbf{X}) - F(x)| &= \sup_{0 < y < 1} |F_n(F^{-1}(y); \mathbf{X}) - \\ &- F(F^{-1}(y))| = \sup_{0 < y < 1} |R_n(y; \mathbf{Y}) - y|. \end{aligned} \quad (9)$$

Остается заметить, что в точках x , где $F(x)=0$ или $F(x)=1$, с вероятностью 1 $F_n(x; \mathbf{X})=F(x)$, а ввиду монотонности F_n и F по x также $\sup_{x: 0 < F(x) < 1} |F_n(x; \mathbf{X}) - F(x)|$ по множеству всех таких x равен 0 . Таким образом, правая и левая части соотношения (9) с вероятностью единица равны $D_n(\mathbf{X})$ и $D_n(\mathbf{Y})$ соответственно. Следовательно, распределение $D_n(\mathbf{X})$ для сл. в. \mathbf{X} , распределенных по закону $F(x)$, совпадает с распределением $D_n(\mathbf{Y})$ для сл. в. \mathbf{Y} , распределенных равномерно на отрезке $[0, 1]$.

Если $F(x)$ не обязательно строго монотонна, то положим при $0 < y < 1$

$$F^{-1}(y) = \sup \{x : F(x) = y\}.$$

Заметив, что супремум достигается ввиду непрерывности $F(x)$ и что поэтому при $0 < y < 1$

$$x \leq F^{-1}(y) \Leftrightarrow F(x) \leq y,$$

нетрудно обнаружить, что проведенные выше рассуждения для случая строго монотонной $F(x)$ сохраняются без изменения и для общего случая. ■

Практическое значение доказанной леммы заключается в том, что можно выбрать распределение вероятностей $F(x)$ из соображений наибольшей простоты численного расчета вероятностей $P(D_n > \delta) = a$ и табулировать набор значений (n, δ, a) . Например, если взять $F(x) = R(x) = x$, $0 \leq x \leq 1$, то расчеты сводятся к вычислению объема n -мерной области, высекаемой в единичном кубе условием

$$\sup_{0 \leq x \leq 1} |R_n(x; x_1, x_2, \dots, x_n) - x| > \delta.$$

Как бы трудоемки ни оказались вычисления, но, проделав их однажды, мы обеспечиваем возможность в случае любой непре-

рывной ф. р. $F(x)$ указать α и δ такие, что при данном n выполнено соотношение

$$\mathcal{P}(\sup |F_n(x; X) - F(x)| > \delta) = \alpha. \quad (10)$$

Иначе говоря, полоса $F_n(x; X) \pm \delta$, $-\infty < x < +\infty$, ширины 2δ с границами, строящимися по выборке, с вероятностью $1 - \alpha$ заключает внутри себя неизвестную ф. р. $F(x)$.

Таблицы для значений $n=1(1)100$ (от 1 до 100 с шагом 1) и $\alpha=0,01; 0,02; 0,05; 0,1; 0,2$ приведены в [1].

Пример 1 (продолжение). Для биномиальной сл. в. X с параметрами $k=250$, $p=0,1$, $n=40$ нормированную сл. в. $X^* = (X - kp)/\sqrt{kp(1-p)}$ можно в силу теоремы Муавра—Лапласа с известной степенью приближения считать распределенной по нормальному закону с параметрами 0 и 1. Учитывая сказанное, можно полагать, что для X^* соотношение (10) будет выполняться с некоторой небольшой погрешностью. Выпишем из табл. 6.2 [1] несколько пар значений α , δ , для которых $\mathcal{P}(D_{40} > \delta) = \alpha$:

α	0,2	0,1	0,05	0,02	0,01
δ	0,16547	0,18913	0,21012	0,23494	0,25205

(11)

Из рис. 1 и табл. 3 с учетом равенства

$$\sup_x |F_n(x; x^*) - \Phi(x)| = \sup_x |F_n(x; x) - \Phi((x - kp)/\sqrt{kp(1-p)})|,$$

$$x_i^* = (x_i - kp)/\sqrt{kp(1-p)},$$

находим $D_n = 0,095$. Полученное отклонение надо рассматривать как умеренное, если взглянуть на (11), откуда вытекает, что отклонение на большую величину 0,165 и то происходит в среднем в одном случае из пяти. ■

А. Н. Колмогоровым было найдено предельное распределение статистики D_n при $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \mathcal{P}(\sqrt{n} D_n \leq y) = K(y), \quad y > 0,$$

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2}.$$

Ряд для $K(y)$ быстро сходится, и три члена этого ряда уже могут обеспечить достаточную степень точности. Таблицу функции

$K(y)$ можно найти в [1] (табл. 6.1). Выпишем несколько значений $\alpha = 1 - K(y)$:

y	0,60	0,65	0,77	0,90	1,08	1,23	1,36	1,52	1,63
α	0,84	0,8	0,6	0,4	0,2	0,1	0,05	0,02	0,01

(12)

Пример 1 (продолжение). Имеем $\sqrt{n} \cdot D_n = \sqrt{40} \cdot 0,095 = 0,60$. Из первого столбца (12) получаем, что примерно в 84 случаях из 100 следовало бы ожидать отклонение 0,60 или более, т. е. мы имеем дело с довольно «удачными» результатами наблюдений.

Полезно сравнить точные значения для α , представленные в (11), со следующими приближенными, получающимися из (12):

α	0,2	0,1	0,05	0,02	0,01
$\delta = y/\sqrt{n}$	0,17077	0,19449	0,21505	0,24035	0,25774

Итак, уже при $n=40$ приближение ф. р. D_n с помощью асимптотической формулы очень хорошее. Абсолютная ошибка в определении δ , такого, что $\mathcal{P}(D_{40} > \delta) = \alpha$, для α от 0,01 до 0,2 составляет около 0,005; относительная ошибка в худшем случае равна примерно 2,2%.

4. Выборочные среднее и дисперсия.

Поскольку э. ф. р. $F_n(x)$ оценивает теоретическую ф. р. $F(x)$, то можно ожидать, что для функционалов $T(F)$ от F оценкой будет служить значение $T(F_n)$, вычисленное по э. ф. р. F_n . К числу важнейших функционалов относятся среднее и дисперсия:

$$T_1(F) = MX_1, \quad T_2(F) = DX_1.$$

Поскольку $F_n(x; x_1, x_2, \dots, x_n)$ отвечает дискретному распределению с вероятностями $1/n$ в точках x_i , то

$$T_1(F_n(x; x)) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x},$$

$$T_2(F_n(x; x)) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = s_0^2$$

или, переходя к случайным величинам,

$$T_1(F_n(x; \mathbf{X})) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X},$$

$$T_2(F_n(x; \bar{X})) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X^2 - \bar{X}^2 = S_0^2.$$

Выборочные среднее \bar{X} и дисперсия S_0^2 , как мы увидим в дальнейшем, оказываются во многих случаях хорошими и, в некоторых отношениях, оптимальными оценками теоретических значений.

§ 3. ПОРЯДКОВЫЕ СТАТИСТИКИ В ЗАДАЧАХ ОЦЕНИВАНИЯ

1. Порядковые статистики.

Расположим наблюдения

$$x_1, x_2, \dots, x_n \quad (1)$$

в порядке возрастания и перенумеруем их заново:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}. \quad (2)$$

Фактически $x_{(i)}$, $i = 1, 2, \dots, n$, — действительные функции переменных (1):

$$x_{(1)} = \min(x_1, x_2, \dots, x_n);$$

$$x_{(2)} = \max \{ \min (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n), i=1, \dots, n \};$$

.....

$$x_{(n)} = \max(x_1, \dots, x_n).$$

$$x_{(n)} = \max(x_1, \dots, x_n).$$

Функцию от результатов наблюдений принято называть *статистикой*. Статистики $x_{(k)}$, $k = 1, 2, \dots, n$, называются *порядковыми*, последовательность (2) часто называют *вариационным рядом*. При переходе от наблюдений (1) к случайным величинам

$$X_1, X_2, \dots, X_n \quad (3)$$

терминология сохраняется: $X_{(k)}$, определенное тем же способом, что и $x_{(k)}$, называется k -й порядковой статистикой, а

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} \quad (4)$$

— вариационным рядом. Отметим, что э. ф. р. зависит от наблюдений через порядковые статистики:

$$F_n(x; X_1, \dots, X_n) = F_n(x; X_{(1)}, \dots, X_{(n)}).$$

Иногда бывает желательно подчеркнуть зависимость порядковых статистик $X_{(i)}$ от n . В таком случае наряду с обозначениями (4) будем употреблять следующие:

$$X_{in} = X_{(i)}, \quad i = 1, 2, \dots, n. \quad (5)$$

Порядковые статистики широко используются в статистических задачах. Рассмотрим некоторые примеры.

2. Оценивание содержимого урны.

Предположим, что урна содержит неизвестное число шаров N , занумерованных числами от 1 до N . С целью определения N совершается повторный выбор с возвращением объема n . Пусть x_1, x_2, \dots, x_n — номера извлеченных шаров. Естественно в качестве оценки \hat{N} предложить максимальное из наблюдений:

$$\hat{N} = x_{(n)}. \quad (6)$$

Шуточную интерпретацию этой задачи дает В. Феллер [2], предлагая для оценки числа зарегистрированных в городе автомобилей встать на перекрестке и записывать номера проезжающих автомобилей. Приведенные ниже результаты моделирования этой задачи с помощью таблицы случайных чисел показывают, что предлагаемая статистическая процедура приводит к хорошим оценкам. Пусть $N=10\,000$, $n=10$; по табл. 7.1, а [1] получены следующие значения $x_{(n)}$ для 10 независимых выборок x_1, \dots, x_n :

$$9590, 8947, 9533, 9901, 9511, 9885, 7990, 9776, 9055, 9352. \quad (7)$$

Сл. в. $Y_i = X_i \cdot N^{-1}$ с хорошей степенью приближения могут рассматриваться как равномерно распределенные на $[0, 1]$, так что

$$\mathcal{P}(X_{(n)} \leqslant Ny) = \mathcal{P}(Y_{(n)} \leqslant y) = (\mathcal{P}(Y_1 \leqslant y))^n \approx y^n, \quad 0 \leqslant y \leqslant 1.$$

Для ряда значений y и $n=10$ имеем

y	0.79	0.74	0.63
$\mathcal{P}(Y_{(n)} \leqslant y)$	0.1	0.05	0.01

Отсюда можно сделать заключение, что в среднем в одном случае из 10 в качестве оценки $\hat{N}=10\,000$ будет появляться значение $\hat{N} \leqslant 7900$. Поэтому появление в (7) значения 7990 не следует считать особым.

Совершенно ясно, что оценка (6) всегда дает заниженный результат: $\hat{N} \leqslant N$. Вычисляя

$$M Y_{(n)} \approx \int_0^1 y^n y^{n-1} dy = \frac{n}{n+1},$$

найдем

$$M\hat{N} = NMY_{(n)} \approx N \frac{n}{n+1}.$$

Оценку \hat{N} можно поправить, устранив систематическое смещение. Пренебрегая погрешностью, запишем

$$M(1+1/n)\hat{N} = N. \quad (8)$$

Исправленные значения (7) выглядят так (с точностью до единицы):

$$10549, 9841, 10486, 10891, 10462, 10873, 8789, 10753, 9960, 10287. \quad (9)$$

Свойство оценки $(1+1/n)\hat{N}$, выражаемое формулой (8), называется *несмешенностью*. Вообще говоря, несмешенность — полезное свойство оценки. Заметим, однако, что устранение смещения в рассматриваемом случае приводит к увеличению в $(1+1/n)^2 = 1,2!$ раз дисперсии оценки $(1+1/n)\hat{N}$ по сравнению с \hat{N} (и в 20 раз увеличивается выборочная дисперсия ряда (9) по сравнению с (7)).

Возвращаясь к ситуации, когда N неизвестно, но достаточно велико, можно по результатам наблюдений $x_i, i=1, \dots, n$, подобрать такое число M , что последовательность x_i/M могла бы с достаточной степенью точности рассматриваться как выборка из равномерного распределения из отрезка $[0, \theta]$, где $\theta=N/M$ неизвестно. В следующем примере эта задача рассматривается более подробно.

3. Оценка параметра равномерного распределения.

Пусть (1) — выборка из равномерного распределения на $[0, \theta]$ с неизвестным θ , (3) — соответствующие независимые равномерно распределенные сл. в. Поскольку $MX_i=\theta/2$, то для оценки θ можно предложить статистику

$$T(X) = 2\bar{X} = \frac{2}{n} \sum_{i=1}^n X_i. \quad (10)$$

Оценка (10), очевидно, несмешенная: $MT(X)=\theta$. Учитывая $DX_1=\theta/12$, получаем

$$DT(X) = \frac{4}{n} DX_1 = \frac{\theta^2}{3n}.$$

Статистика $X_{(n)}$ имеет следующие ф.р., среднее и дисперсию:

$$\mathcal{P}(X_{(n)} \leq x) = \mathcal{P}(X_1 \leq x, \dots, X_n \leq x) = \left(\frac{x}{\theta}\right)^n, \quad 0 < x \leq \theta,$$

$$MX_{(n)} = \frac{n}{n+1} \theta, \quad DX_{(n)} = \frac{n}{(n+2)(n+1)^2} \theta^2. \quad (11)$$

Сравнивая статистики $(1+1/n)X_{(n)}$ и $T(\mathbf{X})$, замечаем, что обе они несмешенно оценивают θ , но дисперсия первой из них, равная $(n(n+2))^{-1}\theta^2$, в $(n+2)/3$ раза меньше дисперсии второй. Для измерения качества оценки часто используют ту или иную характеристику разброса ее распределения вероятностей, чаще всего дисперсию. С этой точки зрения оценка $X_{(n)}$ (или $(1+1/n) \times X_{(n)}$) существенно лучше, нежели $T(\mathbf{X})$. При больших n сравнение $X_{(n)}$ и $T(\mathbf{X})$ можно провести более детально. В соответствии с центральной предельной теоремой при $t \geq 0$

$$\mathcal{P}(|T(\mathbf{X}) - \theta| < t\theta/\sqrt{3n}) \approx 1 - 2\Phi(-t). \quad (12)$$

В свою очередь из (11) мы получаем при $t \geq 0$

$$\begin{aligned} \mathcal{P}(0 < \theta - X_{(n)} < t\theta/n) &= \mathcal{P}(X_{(n)} > \theta(1-t/n)) = \\ &= 1 - (1-t/n)^n \approx 1 - e^{-t}. \end{aligned} \quad (13)$$

Фиксируем t_1 и t_2 так, что

$$1 - a = 1 - 2\Phi(-t_1) = 1 - e^{-t_1}.$$

Из (12) и (13) вытекает, что с вероятностью, приблизительно равной $1 - a$, выполняется каждое из следующих неравенств:

$$T(\mathbf{X})(1+t_1/\sqrt{3n})^{-1} < \theta < T(\mathbf{X})(1-t_1/\sqrt{3n})^{-1}, \quad (14)$$

$$X_{(n)} < \theta < X_{(n)}(1-t_2/n)^{-1}. \quad (15)$$

Таким образом, при больших n ширина первого и второго интервалов, накрывающих неизвестное значение θ с вероятностью $1 - a$, равна соответственно

$$2T(\mathbf{X})t_1/\sqrt{3n} \text{ и } X_{(n)}t_2/n.$$

Из соотношений (12), (13) вытекает, что сл. в. $T(\mathbf{X})/X_{(n)}$ с высокой вероятностью принимает значение, близкое к 1. Поэтому интервал, основанный на статистике $T(\mathbf{X})$, шире интервала, построенного по $X_{(n)}$, примерно в $\sqrt{n} \cdot 2t_1/\sqrt{3}t_2$. При $a = 0,05$ имеем $t_1 = 1,96$, $t_2 = 2,99$, и множитель при \sqrt{n} равняется 0,76.

4. Оценивание параметра сдвига экспоненциального распределения.

Предположим, что сл. в. (3) независимы и имеют одну и ту же плотность $f(x-\theta)$, где

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

Распределение с плотностью $f(x)$ и ф. р. $F(x) = 1 - e^{-x}$, $x \geq 0$, называется *экспоненциальным*. Оно часто употребляется в теории надежности как распределение времени до выхода из строя (от-

каза) изделия. «Сдвинутое» экспоненциальное распределение $f(x-\theta)$, $\theta > 0$, интерпретируют как наличие «гарантийного» срока θ , в течение которого отказ произойти не может.

Порядковая статистика $X_{(1)}$ служит естественной оценкой θ , ее распределение дается формулой

$$\mathcal{P}(X_{(1)} > x) = \mathcal{P}(X_1 > x, \dots, X_n > x) = e^{-n(x-\theta)}, \quad x > 0,$$

откуда получаем

$$\mathcal{P}(0 \leq X_{(1)} - \theta \leq x/n) = 1 - \mathcal{P}(X_{(1)} > x/n + \theta) = 1 - e^{-x}, \quad (16)$$

т. е. интервал $X_{(1)} + \frac{\ln(1-\alpha)}{n} \leq \theta \leq X_{(1)}$ с вероятностью $1-\alpha$

накрывает неизвестное значение θ .

Среднее значение распределения $f(x-\theta)$ равно $\theta+1$, и потому для оценки θ можно также предложить статистику $T(\mathbf{X}) = \bar{X} - 1$. Легко подсчитать, что

$$MT(\mathbf{X}) = \theta, \quad DT(\mathbf{X}) = 1/n.$$

С другой стороны, из (16) находим

$$MX_{(1)} = \theta + 1/n, \quad DX_{(1)} = 1/n^2.$$

Сравнивая дисперсии оценки $T(\mathbf{X})$ и оценки $X_{(1)} - 1/n$ (с устраниенным смещением), приходим к выводу, что если измерять качество оценок значениями их первых двух моментов, то оценка $X_{(1)} - 1/n$ предпочтительнее оценки $T(\mathbf{X})$. Используя центральную предельную теорему, найдем, что распределение $\bar{X} - 1 - \theta$ при больших n приближенно нормально с нулевым средним и дисперсией $1/n$:

$$\mathcal{P}(|\sqrt{n}(\bar{X} - 1 - \theta)| \leq t) \approx 1 - 2\Phi(-t), \quad t \geq 0,$$

откуда интервал

$$\bar{X} - 1 - t/\sqrt{n} \leq \theta \leq \bar{X} - 1 + t/\sqrt{n}$$

ширины порядка $1/\sqrt{n}$ накрывает неизвестное θ с вероятностью примерно $1 - 2\Phi(-t)$, в то время как аналогичный интервал (16) имеет длину порядка $1/n$.

5. Доверительный интервал.

Для оценивания неизвестного параметра θ по результатам наблюдений в разобранных примерах сначала предлагались так называемые *точечные оценки* — статистики, значения которых считались приближением к θ . Погрешность оценивания характеризовалась двумя первыми моментами оценки — средним и дисперсией. С другой стороны, рассматривалась *интервальная оценка*

$$T_1(\mathbf{X}, \alpha) \leq \theta \leq T_2(\mathbf{X}, \alpha),$$

такая, что вне зависимости от того, каково истинное значение параметра θ , оно заключено в данном интервале с вероятностью $1 - \alpha$. В этом случае говорят, что построен *α -доверительный интервал для θ* , или, иначе, *доверительный интервал с коэффициентом доверия $1 - \alpha$* (если написанное выше неравенство выполняется с вероятностью, приближенно равной $1 - \alpha$, то говорят о *приближенном α -доверительном интервале*). Отметим, что в предыдущем параграфе была построена *α -доверительная полоса*, заключающая внутри себя неизвестную ф. р. $F(x)$.

6. Выборочные квантили.

В приведенных выше примерах крайние порядковые статистики выступают в качестве оценок неизвестного параметра, который определяет «крайнюю» точку носителя распределения вероятностей выборки. Другое важное применение порядковых статистик возникает в задачах оценивания функции, обратной к теоретической ф. р. Назовем *p -квантилью* непрерывной ф. р. $F(x)$ решение $x_p = x_p(F)$ уравнения

$$F(x_p) = p, \quad 0 < p < 1. \quad (17)$$

Для $p = 1/2$ x_p называется *медианой* распределения, для $p = 1/4$ и $p = 3/4$ употребляется название *квартиль*. Если F строго монотонна, то $x_p = F^{-1}(p)$ определяется соотношением (17) однозначно, в противном случае для некоторых p уравнение (17) имеет в качестве решения целый отрезок $[\underline{x}, \bar{x}]$ значений x_p . Так как при этом $F(\underline{x}) = F(\bar{x}) = p$, то

$$\mathcal{P}(X \in [\underline{x}, \bar{x}]) = F(\bar{x}) - F(\underline{x}) = 0.$$

С точки зрения теории вероятностей значения x из $[\underline{x}, \bar{x}]$ вообще можно не принимать во внимание. Таким образом, неоднозначность решения уравнения (17) несущественна. Чтобы устранить связанные с отмеченной неоднозначностью формальные неудобства, можно принять за x_p при $p \neq 1/2$ наименьший корень уравнения (17): $x_p = \underline{x}$. Для медианы $x_{1/2}$ в случае неоднозначности ее определения удобнее принять середину отрезка $[\underline{x}, \bar{x}]$.

Если $p = k/n$, $1 \leq k \leq n - 1$, то уравнение (17), записанное для э. ф. р., имеет своим наименьшим решением $x_{(k)}$:

$$F_n(x_{(k)}; x_1, \dots, x_n) = p = k/n,$$

и поэтому $x_{(k)}$ может рассматриваться как естественная оценка квантили x_p . В случае произвольного p выборочной квантилью обычно называют $x_{[np]+1}$, где $[a]$ обозначает целую часть числа a .

Нетрудно найти ф. р. порядковой статистики (в предположении независимости сл. в. (3)):

$$\begin{aligned} \mathcal{P}(X_{(k)} \leq x) &= \mathcal{P}(F_n(x, \mathbf{X}) \geq k/n) = \\ &= \mathcal{P}\left(\sum_{i=1}^n I_{\{X_i \leq x\}} \geq k\right) = \sum_{i=k}^n C_n^i F(x)^i (1-F(x))^{n-i}, \end{aligned} \quad (18)$$

где было использовано, что сл. в. $\sum_{i=1}^n I_{\{X_i \leq x\}}$ есть число успехов в n испытаниях Бернулли с вероятностью успеха $F(x)$.

7. Равномерное распределение.

В случае независимой выборки из равномерного распределения: $F(x) = x$, $0 \leq x \leq 1$, из (18) получаем

$$\mathcal{P}(X_{(k)} \leq x) = \sum_{i=k}^n C_n^i x^i (1-x)^{n-i}. \quad (19)$$

Дифференцируя (19) по x , имеем

$$\begin{aligned} \sum_{i=k}^n \frac{n!}{i!(n-i)!} i x^{i-1} (1-x)^{n-i} - \sum_{i=k}^{n-1} \frac{n!}{i!(n-i)!} (n-i) x^i (1-x)^{n-i-1} = \\ = \sum_{i=k}^n n C_{n-1}^{i-1} x^{i-1} (1-x)^{(n-1)-(i-1)} - \sum_{i=k}^{n-1} n C_{n-1}^i x^i (1-x)^{(n-1)-i} = \\ = n C_{n-1}^{k-1} x^{k-1} (1-x)^{n-k}, \end{aligned}$$

или

$$\frac{d}{dx} \mathcal{P}(X_{(k)} \leq x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k}, \quad 0 < x < 1. \quad (20)$$

Таким образом, для ф. р. (19) имеет место следующее представление

$$\mathcal{P}(X_{(k)} \leq x) = \frac{n!}{(k-1)!(n-k)!} \int_0^x t^{k-1} (1-t)^{n-k} dt, \quad 0 < x < 1. \quad (21)$$

Функция двух переменных

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt, \quad a > 0, b > 0, \quad (22)$$

носит название *бета-функции Эйлера*. Плотность распределения (21)

$$[B(a, b)]^{-1} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1, \quad (23)$$

называется *бета-плотностью* (с параметрами a, b), а соответствующая ф. р.

$$I_x(a, b) = [B(a, b)]^{-1} \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad 0 \leq x \leq 1, \quad (24)$$

называется *неполной бета-функцией*. Как вытекает из (19), (21),

$$I_x(k, n-k+1) = \sum_{i=k}^n C_n^i x^i (1-x)^{n-i}. \quad (25)$$

Для вычисления $I_x(a, b)$ составлены таблицы (см. [1]). Из (18) и (25) получаем, что ф. р. k -й порядковой статистики от независимой выборки объема n с непрерывной ф. р. $F(x)$ может быть выражена через неполную бета-функцию:

$$\mathcal{P}(X_{(k)} \leq x) = I_{F(x)}(k, n-k+1). \quad (26)$$

8. Доверительные интервалы для квантилей.

Формула (26) приводит к важному статистическому приложению — доверительным границам и интервалам для квантилей непрерывной теоретической ф. р. Именно, подставляя в (26) p -квантиль, получаем

$$\mathcal{P}(X_{(k)} \leq x_p) = I_p(k, n-k+1), \quad (27)$$

т. е. $X_{(k)}$ является *нижней a -доверительной границей* для p -квантили x_p с $a = 1 - I_p(k, n-k+1)$. Аналогично

$$\mathcal{P}(X_{(l)} \geq x_p) = 1 - I_p(l, n-l+1), \quad (28)$$

т. е. $X_{(l)}$ является *верхней a -доверительной границей* для квантили x_p .

При $k < l$ с вероятностью $1 - X_{(k)} < X_{(l)}$, поэтому

$$\mathcal{P}(X_{(k)} \leq x_p \leq X_{(l)}) = 1 - \mathcal{P}(x_p < X_{(k)}) - \mathcal{P}(X_{(l)} < x_p),$$

т. е. доверительный интервал

$$X_{(k)} \leq x_p \leq X_{(l)} \quad (29)$$

имеет коэффициент доверия

$$1 - a = I_p(k, n-k+1) - I_p(l, n-l+1). \quad (30)$$

Наиболее часто используют доверительный интервал (29) для медианы распределения $x_{1/2}$, так как для симметричных распределений медиана совпадает с математическим ожиданием (если оно существует). Правда, как мы увидим, в случае нормальной выборки, а также если выборка большая, интервальная оценка, основанная на порядковых статистиках, проигрывает по сравнению с другими методами интервального оценивания.

9. Совместное распределение пары порядковых статистик.

Аналогично (18) можно найти совместную ф.р. двух порядковых статистик $X_{(k)}$ и $X_{(l)}$, $k < l$. При $x < y$

$$\begin{aligned} \mathcal{P}(X_{(k)} \leq x, X_{(l)} \leq y) &= \mathcal{P}\left(\sum_{i=1}^n I_{\{X_i \leq x\}} \geq k,\right. \\ &\quad \left.\sum_{i=1}^n I_{\{X_i \leq y\}} \geq l\right) = \sum_{r=k}^n \sum_{s=l-r}^{n-r} \mathcal{P}\left(\sum_{i=1}^n I_{\{X_i \leq x\}} = r,\right. \\ &\quad \left.\sum_{i=1}^n I_{\{x < X_i \leq y\}} = s\right) = \sum_{r=k}^n \sum_{s=l-r}^{n-r} \frac{n!}{r! s! (n-r-s)!} \times \\ &\quad \times F(x)^r (F(y) - F(x))^s (1 - F(y))^{n-r-s}. \end{aligned}$$

Если ф.р. $F(x)$ имеет плотность $f(x)$, то пара $(X_{(k)}, X_{(l)})$ имеет совместную плотность, которую можно получить, дифференцируя по x и y их совместную ф.р. Полезнее, однако, сделать предельный переход при $\delta \rightarrow 0$, $\delta > 0$, в следующем выражении (предполагаем $x < y$):

$$\delta^{-2} \mathcal{P}(x < X_{(k)} \leq x + \delta, y < X_{(l)} \leq y + \delta). \quad (31)$$

Используя равновероятность всех $n!$ перестановок $X_1 < X_2 < \dots < X_n$, запишем (31) в виде

$$\delta^{-2} n! \mathcal{P}(x < X_k \leq x + \delta, y < X_l \leq y + \delta, X_1 < \dots < X_n). \quad (32)$$

Выражение (32) эквивалентно при $\delta \rightarrow 0$ следующему:

$$\begin{aligned} &\delta^{-2} n! \mathcal{P}(X_1 < \dots < X_{k-1} \leq x < X_k \leq x + \delta < \\ &< X_{k+1} < \dots < X_{l-1} < y < X_l \leq y + \delta < X_{l+1} < \dots < X_n). \end{aligned} \quad (33)$$

Действительно, если, скажем, в интервал $(x, x + \delta]$ попало значение еще одной сл.в., то вероятность в (32) оценивается сверху величиной

$$(F(x + \delta) - F(x))^2 (F(y + \delta) - F(y)) \sim f^2(x) f(y) \delta^3, \quad \delta \rightarrow 0,$$

в то время как вероятность в (33) имеет порядок δ^2 (при $f(x)f(y) \neq 0$). Используя независимость сл.в. X_i , $i = 1, \dots, n$, равновероятность $(k-1)!$ перестановок из X_1, \dots, X_{k-1} и $(l-k-1)!$ перестановок из X_{k+1}, \dots, X_n , получаем для (33) выражение

$$\begin{aligned} &\delta^{-2} n! \mathcal{P}(X_1 < \dots < X_{k-1} \leq x) \mathcal{P}(x < X_k \leq x + \delta) \times \\ &\times \mathcal{P}(x + \delta < X_{k+1} < \dots < X_{l-1} \leq y) \mathcal{P}(y < X_l \leq y + \delta) \times \\ &\times \mathcal{P}(y + \delta < X_{l+1} < \dots < X_n) = \delta^{-2} n! \frac{1}{(k-1)!} F(x)^{k-1} (F(x + \delta) - \end{aligned}$$

$$-F(x)) \frac{1}{(l-k-1)!} (F(y)-F(x+\delta))^{l-k-1} \times \\ \times (F(y+\delta)-F(y)) \frac{1}{(n-l)!} (1-F(y+\delta))^{n-l}. \quad (34)$$

Пределом выражения (34) является совместная плотность пары порядковых статистик $(X_{(k)}, X_{(l)})$, $k < l$, в точке (x, y) , $x < y$, и этот предел равен

$$\frac{n!}{(k-1)! (l-k-1)! (n-l)!} F(x)^{k-1} f(x) \times \\ \times (F(y)-F(x))^{l-k-1} f(y) (1-F(y))^{n-l}. \quad (35)$$

§ 4. ПАРАМЕТРЫ СДВИГА И МАСШТАБА: ГРАФИЧЕСКИЙ АНАЛИЗ

1. Семейство сдвига-масштаба.

Значительный теоретический и прикладной интерес представляет статистическая модель независимых наблюдений

$$X_1, X_2, \dots, X_n \quad (1)$$

с общей ф.р. вида $F((x-\mu)/\sigma)$, где $F(x)$ — известная непрерывная ф.р., а параметры сдвига μ и масштаба $\sigma > 0$, вообще говоря, неизвестны. Нормированные сл. в.

$$Y_i = (X_i - \mu)/\sigma, \quad i = 1, 2, \dots, n, \quad (2)$$

очевидно, также независимы и одинаково распределены, а их общая ф.р. равна

$$\mathcal{P}(Y_i \leq x) = \mathcal{P}((X_i - \mu)/\sigma \leq x) = \mathcal{P}(X_i \leq \sigma x + \mu) = F(x).$$

Семейство ф.р. $F((x-\mu)/\sigma)$ может быть образовано из любой ф.р. $F(x)$, однако в математической статистике и ее приложениях употребляют сравнительно небольшой набор распределений, которые благодаря особым теоретическим свойствам часто используются в реальных задачах. К числу таких распределений относится прежде всего нормальное

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \equiv \Phi(x).$$

Параметры μ и σ семейства $\Phi((x-\mu)/\sigma)$ имеют хорошо известный вероятностный смысл:

$$\mathbf{M}X_i = \mu + \sigma \mathbf{M}Y_i = \mu + \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} te^{-t^2/2} dt = \mu,$$

$$\mathbf{D}X_i = \sigma^2 \mathbf{M}Y_i^2 = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt = \sigma^2.$$

Отметим также семейство экспоненциальных распределений, задающееся плотностью

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

Здесь $MY_i=1$, $DY_i=1$ и $MX_i=\mu+\sigma$, $DX_i=\sigma^2$.

Приведем еще один интересный пример, когда семейство ф. р., не являющееся сдвигом-масштабным, становится таким после преобразования. Распределение с ф. р.

$$G(x)=1-\exp[-(x/\beta)^{\gamma}], \quad x>0, \quad (3)$$

возникает в теории экстремальных значений и часто (под названием *распределение Вейбулла*) употребляется в качестве распределения времени жизни изделия в задачах надежности. Пусть сл. в. X имеет ф. р. $G(x)$. Тогда сл. в. $Z=\ln X$ имеет ф. р.

$$\begin{aligned} \mathcal{P}(Z \leq x) &= \mathcal{P}(\ln X \leq x) = \mathcal{P}(X \leq e^x) = \\ &= 1 - \exp[-\beta^{-\gamma} e^{\gamma x}] = 1 - \exp[-\exp(\gamma(x - \ln \beta))]. \end{aligned}$$

Таким образом, если случайная выборка X_i , $i=1, \dots, n$, взята из семейства (3) с параметрами $\beta>0$, $\gamma>0$, то $Z_i=\ln X_i$, $i=1, \dots, n$, является случайной выборкой из семейства $F((x-\mu)/\sigma)$, где

$$F(x)=1-\exp[-\exp x], \quad (4)$$

а параметры μ , σ связаны с β , γ формулами

$$\mu=\ln \beta, \quad \sigma=\gamma^{-1}.$$

Отметим, что ф. р. (4) также возникает в теории экстремальных значений.

2. Вероятностная бумага.

Во многих случаях теоретические соображения или предшествующий опыт позволяют высказать предположение о том, что случайная выборка x_1, x_2, \dots, x_n взята из распределения $F((x-\mu)/\sigma)$ с известной непрерывной $F(x)$, но неизвестными μ , σ . Используя статистический материал x_1, x_2, \dots, x_n , желательно проверить это предположение. Соответствующая задача относится к области теории проверки статистических гипотез. Существуют различные принципы и разработаны различные методы решения подобных задач, однако предварительный графический анализ данных остается полезным инструментом исследования, особенно для небольшой выборки.

Эмпирическая ф. р. является естественным приближением теоретической ф. р., но, за исключением выборки из равномерного распределения (когда теоретическая ф. р.—прямая линия), графическая подгонка ф. р. из семейства $F((x-\mu)/\sigma)$ к э. ф. р.

представляет, очевидно, определенные трудности. Преодолеть их можно, «распрямив» график $F((x-\mu)/\sigma)$. С этой целью, предполагая, что $F(x)$ непрерывна и строго монотонна в точках x , в которых $0 < F(x) < 1$, введем отображение полосы $-\infty < x < +\infty$, $0 < y < 1$ в плоскость по формуле

$$(x, y) \rightarrow (x, F^{-1}(y)), \quad (5)$$

где $F^{-1}(y)$ — функция, обратная к $F(x)$. Отображение (5) по условиям, наложенным на $F(x)$, взаимно-однозначно и взаимно-непрерывно в области своего определения. При этом отображении

$$(x, F(x)) \rightarrow (x, F^{-1}(F(x))) = (x, x), \quad (6)$$

т. е. график $y = F(x)$, $0 < y < 1$, переходит в прямую линию $y = x$ (или ее интервал, если существуют такие значения x , что $F(x) = 1$ или $F(x) = 0$). Далее,

$$(x, F((x-\mu)/\sigma)) \rightarrow (x, F^{-1}(F((x-\mu)/\sigma))) = (x, (x-\mu)/\sigma),$$

т. е. график $F((x-\mu)/\sigma)$, ф. р. из рассматриваемого семейства, переходит в прямую линию

$$y = (x - \mu)/\sigma.$$

Опишем вытекающую отсюда методику. Пусть $F_n(x)$ — э. ф. р. случайной выборки из семейства $F((x-\mu)/\sigma)$. График $y = F_n(x)$ преобразуем в $y = F^{-1}(F_n(x))$ и подберем наиболее тесно прилегающую к нему прямую линию

$$y = (\hat{x} - \hat{\mu})/\hat{\sigma}. \quad (7)$$

При этом $\hat{\sigma}$ — котангенс угла наклона прямой (7) к оси x — будет служить оценкой параметра σ . Оценкой $\hat{\mu}$ будет абсцисса точки пересечения прямой (7) с осью x . Если график $y = F^{-1}(F_n(x))$ систематически уклоняется от линейной зависимости, то есть все основания сомневаться в том, что наблюдаемые данные взяты из распределения $F((x-\mu)/\sigma)$.

При фактической реализации указанного приема нет необходимости строить целиком график $y = F^{-1}(F_n(x))$, а следует отметить только точки $(x_{(i)}, F^{-1}(i/n))$, $i = 1, \dots, n$, отвечающие скачкам $F_n(x)$, и подгонять прямую к этим точкам. Ввиду неудобств, связанных с точкой $(x_{(n)}, F^{-1}(1))$, обычно строят точки $(x_{(i)}, F^{-1}(i/(n+1)))$, $i = 1, \dots, n$. Более удачным является выбор точек в виде $(x_{(i)}, F^{-1}((i-0,5)/n))$, $i = 1, 2, \dots, n$.

Для графической работы с семейством сдвига-масштаба $F((x-\mu)/\sigma)$ удобно изготовить так называемую вероятностную бумагу, выбрав по оси ординат неравномерный масштаб и прислав точке с декартовой координатой y новую координату $y' = F(y)$. В новой системе координат xy' непосредственно наносятся точки $(x_{(i)}, (i-0,5)/n)$. Оцифровка оси ординат для нормального закона показана на рис. 2.

3. Примеры графического анализа.

С помощью таблицы случайных чисел образована выборка объема $n=10$ из нормального закона с $\mu=1$, $\sigma=2$:

$-2,504; 0,342; -1,512; 1,636; 4,062; 1,698; -0,916; 0,882;$
 $1,830; -1,168.$

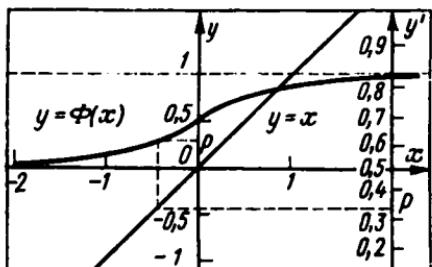


Рис. 2. Изготовление нормальной вероятностной бумаги

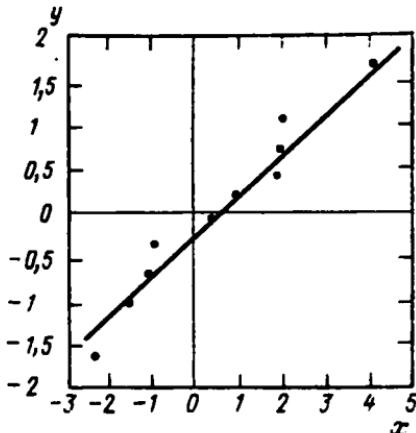


Рис. 3. Точки $(x_{(i)}, \Phi^{-1}((i-0.5)/10))$ и прилегающая прямая для искусственной нормальной выборки

По таблице функции, обратной к нормальной, из [1] находим

i	$6(5)$	$7(4)$	$8(3)$	$9(2)$	$10(1)$
$\Phi^{-1}\left(\frac{i-0,5}{10}\right)$	(-)0,125	(-)0,385	(-)0,675	(-)1,036	(-)1,645

Точки $(x_{(i)}, \Phi^{-1}((i-0.5)/10))$, $i=1, 2, \dots, 10$, отмечены на графике (рис. 3), и подобрана на глаз прямая

$$y = (x - \hat{\mu})/\hat{\sigma}, \text{ или } x = \hat{\sigma}y + \hat{\mu},$$

где $\hat{\mu}=0,44$; $\hat{\sigma}=1,87$. Выборочные среднее и дисперсия равны

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,435; s_0^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = 2,976; s_0 = 1,725.$$

Как видно, графическое приближение дает оценки $\hat{\mu}$, $\hat{\sigma}$, близкие к выборочным. ■

В следующей таблице приведены данные о ежегодном потреблении в фунтах на душу населения мясных продуктов, включая птицу и рыбу, в США с 1919 по 1941 г. (считывать по строкам) [3]:

171,5	167,0	164,5	169,3	179,4	179,2	172,6	170,5
168,6	164,7	163,0	162,1	160,2	161,2	165,8	163,5
146,7	160,2	156,8	156,8	165,4	174,7	178,7	

Эти данные предлагаются описать моделью

$$x_t = a + bt + ct^2 + dt^3 + \varepsilon_t, \quad (9)$$

где t обозначает год, x_t — потребление в этот год. Полином $a + bt + ct^2 + dt^3$ призван отразить неслучайные изменения переменной x_t и называется трендом, коэффициенты a, b, c, d подлежат

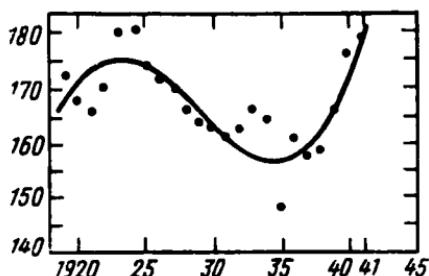


Рис. 4 Ежегодное потребление мяса в США в 1919—1941 гг. и полиномиальное приближение

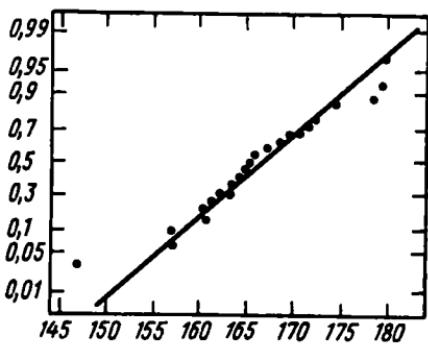


Рис. 5. Данные о потреблении мяса в США в 1919—1941 гг., нанесенные на нормальную вероятностную бумагу

подбору по наблюдениям (8); ε_t — случайные колебания в ежегодном потреблении, их полагают независимыми, одинаково нормально распределенными сл. в. с нулевым средним и неизвестной дисперсией. Значения (8) и аппроксимирующий их полином изображены на рис. 4.

Попробуем с помощью описанного выше графического приема проанализировать данные (8) на предмет того, нельзя ли их рассматривать как выборку из нормального распределения, т. е. считать, что коэффициенты b, c, d в (9) равны нулю, и полагать колебания в ежегодном потреблении чисто случайными, не содержащими тренда. Графическое построение на рис. 5 показывает удовлетворительное согласие с проведенной на глаз прямой

$$x = \hat{\sigma}y + \hat{\mu}, \quad \hat{\mu} = 166,5, \quad \hat{\sigma} = 7,33. \quad (10)$$

С другой стороны, приведенный в книге [3] статистический анализ показывает, что данные (8) следует описывать моделью (9),

как это и представлено на рис. 4. Противоречие в толковании статистических данных (8) позволяет обратить внимание на следующий важный вывод.

Любой статистический прием направлен на проверку одних сторон вероятностной модели и не улавливает отклонения в других направлениях. Поэтому при анализе данных надо ясно представлять себе, какие предположения о модели не ставятся под сомнение, а что подлежит проверке по результатам наблюдений. При графическом анализе (см. рис. 4) предполагалось, что наблюдения независимы и одинаково распределены, и проверялась гипотеза о нормальности их общего распределения. Если исходные предположения нарушены, то и выводы, которые делаются в отношении проверяемой гипотезы, несостоятельны.

Отметим, кстати, что э. ф. р. не меняется при перестановке наблюдений x_1, x_2, \dots, x_n ; иначе говоря, она является функцией порядковых статистик $x_{(1)}, x_{(2)}, \dots, x_{(n)}$. В частности, наблюдения $y_i = x_{(i)}, i=1, \dots, n$, имеют ту же э. ф. р., что и $x_i, i=1, 2, \dots, n$, и если не обратить на это внимание, то по результатам анализа э. ф. р. можно для значений $y_i, i=1, \dots, n$, ошибочно принять тот же вероятностный закон, что и для x_i .

§ 5. ЭКСПОНЕНЦИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И ПУАССОНОВСКИЙ ПРОЦЕСС

1. Экспоненциальное распределение.

Стандартное экспоненциальное распределение определяется плотностью

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (1)$$

и имеет ф. р. $F(x) = 1 - e^{-x}$, $x > 0$. Обычно в распределение (1) вводится масштабный параметр и рассматривается ф. р. $F(x/\sigma)$ или $F(\lambda x)$, $\lambda = \sigma^{-1}$. Параметр σ является средним значением распределения. Важность распределения (1) в теории и приложениях связана с его характеристическим свойством, называемым «отсутствием памяти»: если сл. в. X имеет плотность (1), то при $x, y > 0$

$$\begin{aligned} \mathcal{P}(X > x + y | X > x) &= \mathcal{P}(X > x + y, X > x) / \mathcal{P}(X > x) = \\ &= \mathcal{P}(X > x + y) / \mathcal{P}(X > x) = e^{-y} = \mathcal{P}(X > y). \end{aligned}$$

Допустим, что X описывает случайное время до наступления некоторого события. Тогда знание того, что это событие не произошло до момента x , никак не оказывается на его появлении после момента x : остающееся время ожидания имеет то же самое экспоненциальное распределение.

Отмеченное свойство распределения (1) становится более прозрачным, если обратиться к схеме Бернулли. Допустим, что в моменты времени, кратные некоторому шагу $h > 0$, совершаются испытания Бернулли с вероятностью успеха $p = \lambda h$. Пусть v_1 — номер испытания, при котором произошел первый успех, и $X_1^{(h)} = hv_1$ — случайное время до появления первого успеха. Так как

$$\mathcal{P}(v_1 = k) = q^{k-1}p, \quad \mathcal{P}(v_1 > k) = q^k, \quad q = 1 - p,$$

то

$$\mathcal{P}(X_1^{(h)} > x) = (1 - \lambda h)^{\lfloor x/h \rfloor}, \quad (2)$$

где $\lfloor a \rfloor$ означает целую часть a . Устремим теперь h к нулю. Очевидно, правая часть (2) будет иметь предел $e^{-\lambda x}$, т. е. в рассматриваемом предельном процессе распределение времени ожидания первого успеха сходится к экспоненциальному (с масштабным параметром λ^{-1}).

2. Пуассоновский процесс.

С экспоненциальным распределением тесно связан так называемый пуассоновский процесс. Обратимся к уже использовавшейся схеме испытаний Бернулли. Обозначим через $v_1 + v_2 + \dots + v_k$ номер k -го по порядку успеха, $k = 1, 2, \dots$. Нетрудно видеть, что

$$\mathcal{P}(v_1 = n_1, v_2 = n_2, \dots, v_m = n_m) = \prod_{i=1}^m (q^{n_i-1}p),$$

$$\mathcal{P}(v_1 > n_1, v_2 > n_2, \dots, v_m > n_m) = \prod_{i=1}^m q^{n_i}.$$

Для величин $X_i^{(h)} = hv_i$ — временных интервалов между последовательными успехами — получаем при $x_1, x_2, \dots, x_m > 0$

$$\mathcal{P}(X_1^{(h)} > x_1, X_2^{(h)} > x_2, \dots, X_m^{(h)} > x_m) \rightarrow \prod_{i=1}^m e^{-\lambda x_i}, \quad h \rightarrow 0, \quad (3)$$

т. е. совместное предельное распределение сл. в. $X_i^{(h)}$, $i = 1, \dots, m$, соответствует независимым экспоненциально распределенным сл. в. с масштабным параметром λ^{-1} . Допустим, что X_1, X_2, \dots — последовательность независимых сл. в. с экспоненциальной плотностью $\lambda e^{-\lambda x}$, $x > 0$. Из (3) вытекает, таким образом, что при любых m , x_1, \dots, x_m и $h \rightarrow 0$

$$\mathcal{P}(X_1^{(h)} \leq x_1, \dots, X_m^{(h)} \leq x_m) \rightarrow \mathcal{P}(X_1 \leq x_1, \dots, X_m \leq x_m). \quad (4)$$

Сходимость многомерных функций распределения (4) влечет за собой сходимость вероятностей любых событий, связанных с этими величинами:

$$\mathcal{P}((X_1^{(h)}, \dots, X_m^{(h)}) \in B_m) \rightarrow \mathcal{P}((X_1, \dots, X_m) \in B_m), h \rightarrow 0, \quad (5)$$

где B_m — борелевское множество m -мерного пространства R^m . В частности, для B_m вида

$$B_m = \{(x_1, \dots, x_m) : x_1 + \dots + x_j \leq y_j, j = 1, \dots, m\}$$

соотношение (5) превращается в следующее

$$\mathcal{P}(S_1^{(h)} \leq y_1, \dots, S_m^{(h)} \leq y_m) \rightarrow \mathcal{P}(S_1 \leq y_1, \dots, S_m \leq y_m), h \rightarrow 0, \quad (6)$$

где

$$S_j^{(h)} = X_1^{(h)} + \dots + X_j^{(h)}, \quad S_j = X_1 + \dots + X_j, \quad j = 1, 2, \dots$$

Отметим на числовой оси точки

$$S_1^{(h)}, S_2^{(h)}, \dots, S_j^{(h)}, \dots \quad (7)$$

Говорят, что последовательность случайных точек (7) образует *случайный точечный процесс* на прямой. Аналогично получаем точечный процесс

$$S_1, S_2, \dots, S_j, \dots \quad (8)$$

Утверждение (6) означает, что точечный процесс (7) сходится по распределению к точечному процессу (8), образованному суммами $S_j = X_1 + \dots + X_j, j = 1, 2, \dots$, независимых экспоненциальных сл. в. с параметром λ . Процесс (8) называется *пуассоновским*.

Обозначим через $N_{(0,x]}^{(h)}$ и $N_{(0,x]}$ число точек на временном интервале $(0, x]$ в процессах (7) и (8), соответственно:

$$\begin{aligned} N_{(0,x]}^{(h)} = m &\Leftrightarrow S_m^{(h)} \leq x, S_{m+1}^{(h)} > x, \\ N_{(0,x]} = m &\Leftrightarrow S_m \leq x, S_{m+1} > x. \end{aligned} \quad (9)$$

Событие $\{N_{(0,x]}^{(h)} = m\}$ происходит тогда и только тогда, когда в последовательности из $n = [x/h]$ испытаний Бернулли происходит ровно m успехов, а его вероятность равна

$$C_n^m p^m q^{n-m}, \quad q = 1 - p, \quad p = \lambda h, \quad n = [x/h].$$

Устремляя h к нулю, находим

$$\begin{aligned} \mathcal{P}(N_{(0,x]}^{(h)} = m) &= C_n^m (\lambda h)^m (1 - \lambda h)^{n-m} = \frac{(n\lambda h)^m}{m!} (1 - \lambda h)^n \times \\ &\times \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) (1 - \lambda h)^{-m} \rightarrow (\lambda x)^m / m! \cdot e^{-\lambda x}. \end{aligned} \quad (10)$$

Принимая во внимание (6) и (9), получаем отсюда

$$\mathcal{P}(N_{(0,x]} = m) = \frac{(\lambda x)^m}{m!} e^{-\lambda x}. \quad (11)$$

Это хорошо известное распределение Пуассона с параметром λx . \blacksquare .

Пуассоновский процесс был определен как последовательность (8) моментов появления «точек», или «событий», на оси x . Формула (9) определяет случайный процесс $N_{(0, x]}$, $x > 0$, т. е. функцию, сопоставляющую каждому $x > 0$ случайную величину $N_{(0, x)}$. Процесс $N_{(0, x]}$, $x > 0$, тоже называется пуассоновским. В следующей лемме отражены важнейшие свойства процесса $N_{(0, x)}$. Положим при $x < y$

$$N_{(x, y]} = N_{(0, y]} - N_{(0, x]}.$$

Лемма 1. Пусть $x_0 = 0 < x_1 < x_2 < \dots < x_k$. Тогда сл. в. $N_{(x_{i-1}, x_i]}$, $i = 1, 2, \dots, k$, независимы и

$$\mathcal{P}(N_{(x_{i-1}, x_i]} = m) = \frac{(\lambda(x_i - x_{i-1}))^m}{m!} e^{-\lambda(x_i - x_{i-1})}. \quad (12)$$

Доказательство. Как и при выводе (11), используем предельный переход от схемы Бернулли. Положим

$$N_{(x, y]}^{(h)} = N_{(0, y]}^{(h)} - N_{(0, x]}^{(h)}, \quad x < y.$$

Тогда

$$\begin{aligned} \mathcal{P}(N_{(x_{i-1}, x_i]}^{(h)} = m_i, i = 1, 2, \dots, k) &= \mathcal{P}(S_{m_i}^{(h)} \leq x_1, \\ S_{m_1+m_2}^{(h)} > x_1, S_{m_1+m_2}^{(h)} < x_2, S_{m_1+m_2+m_3}^{(h)} > x_2, \dots, \\ S_{m_1+\dots+m_k}^{(h)} \leq x_k, S_{m_1+\dots+m_k+1}^{(h)} > x_k). \end{aligned} \quad (13)$$

Устремляя h к нулю, получаем с учетом (5), что правая часть (13) стремится к

$$\begin{aligned} \mathcal{P}(S_{m_1} \leq x_1, S_{m_1+1} > x_1, S_{m_1+m_2} \leq x_2, S_{m_1+m_2+1} > x_2, \dots, \\ S_{m_1+\dots+m_k} \leq x_k, S_{m_1+\dots+m_k+1} > x_k) &= \mathcal{P}(N_{(x_{i-1}, x_i]} = m_i, i = 1, \dots, k). \end{aligned} \quad (14)$$

С другой стороны, поскольку величины $N_{(x_{i-1}, x_i]}^{(h)}$, как числа успехов на непересекающихся отрезках испытаний Бернулли, независимы, то левая часть (13) равна

$$\prod_{i=1}^k \mathcal{P}(N_{(x_{i-1}, x_i]}^{(h)} = m_i).$$

Переходя к пределу при $h \rightarrow 0$, найдем, что левая часть (13) стремится к

$$\prod_{i=1}^k \frac{(\lambda(x_i - x_{i-1}))^{m_i}}{m_i!} e^{-\lambda(x_i - x_{i-1})}. \quad (15)$$

Приравнивая (14) и (15), получаем требуемый результат.

3. Условное распределение точек пуассоновского процесса.

Следующие утверждения о пуассоновском процессе и о независимых экспоненциально распределенных сл. в. имеют важное теоретическое и прикладное значение.

Лемма 2. Пусть X_1, X_2, \dots — независимые экспоненциально распределенные с параметром λ сл. в., $S_n = X_1 + \dots + X_n$, $n = 1, 2, \dots$, $N_{(0, x)}$ — построенный по последовательности S_n пуассоновский процесс. Тогда

(I) плотность распределения сл. в. S_n равна

$$f_{S_n}(x) = \begin{cases} \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}, & x > 0, \\ 0 & x \leq 0; \end{cases}$$

(II) условная совместная плотность сл. в. S_1, \dots, S_n при условии сл. в. S_{n+1} равна

$$f_{S_1, \dots, S_n | S_{n+1}}(y_1, \dots, y_n | y_{n+1}) = \frac{n!}{y_{n+1}^n}, \quad 0 < y_1 < \dots < y_n < y_{n+1},$$

и равна нулю для остальных значений переменных;

(III) совместная плотность сл. в. $Z_{(i)} = S_i / S_{n+1}$, $i = 1, 2, \dots, n$, равна

$$f_{Z_{(1)}, \dots, Z_{(n)}}(z_1, \dots, z_n) = n!, \quad 0 < z_1 < \dots < z_n < 1,$$

и равна нулю для остальных значений переменных;

(IV) при условии события $N_{(0, x)} = n$ совместная плотность распределения сл. в. S_i , $i = 1, 2, \dots, n$, (моментов появления точек в пуассоновском процессе) равна

$$f_{S_1, \dots, S_n | N_{(0, x)} = n}(y_1, \dots, y_n | n) = \frac{n!}{x^n}, \quad 0 < y_1 < \dots < y_n < x,$$

и равна нулю при остальных значениях переменных.

Замечание. Если U_1, U_2, \dots, U_n — независимые и равномерно распределенные на $[0, 1]$ сл. в., $U_{(i)}$, $i = 1, \dots, n$, — порядковые статистики, то, очевидно, все $n!$ перестановок $U_{i_1} < U_{i_2} < \dots < U_{i_n}$ являются равновероятными событиями и поэтому

$$\mathcal{P}(U_{(1)} \in B_1, \dots, U_{(n)} \in B_n) = n! \mathcal{P}(U_1 \in B_1, \dots, U_n \in B_n, U_1 < \dots < U_n). \quad (16)$$

Выбирая $B_i = (u_i, u_i + \delta)$ при $0 < u_1 < \dots < u_n$ и достаточно малом $\delta > 0$, получаем, что (16) будет равно $n! \delta^n$, т. е. совместная плотность величин $U_{(i)}$ равна

$$f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = n!, \quad 0 < u_1 < \dots < u_n < 1.$$

Следовательно, результаты (II) и (IV) означают, что условные плотности

$$f_{S_1, \dots, S_n | S_{n+1}} \text{ и } f_{S_1, \dots, S_n | N(0, x)}$$

совпадают с плотностями порядковых статистик из равномерного распределения на $[0, S_{n+1}]$ и $[0, x]$ соответственно. ■

Доказательство. (I) Из формул (9) и (11) и того факта, что из события $S_{n+1} \leq x$ вытекает $S_n \leq x$, получаем

$$\begin{aligned} \mathcal{P}(N_x = n) &= \mathcal{P}(S_n \leq x, S_{n+1} > x) = \mathcal{P}(S_n \leq x) - \mathcal{P}(S_n \leq x, S_{n+1} \leq x) = \\ &= \mathcal{P}(S_n \leq x) - \mathcal{P}(S_{n+1} \leq x), \end{aligned}$$

т. е.

$$\mathcal{P}(S_{n+1} \leq x) = \mathcal{P}(S_n \leq x) + \frac{(\lambda x)^n}{n!} e^{-\lambda x}. \quad (17)$$

Последовательно используя (17) при $n = 1, 2, \dots$ и принимая во внимание, что $S_1 = X_1$, имеем для ф. р. сл. в. S_n :

$$\mathcal{P}(S_n \leq x) = 1 - e^{-\lambda x} \left(1 + \frac{\lambda x}{1!} + \dots + \frac{(\lambda x)^{n-1}}{(n-1)!} \right). \quad (18)$$

Дифференцируя (18) по x и производя сокращения, находим плотность распределения сл. в. S_n :

$$\lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}, \quad x > 0. \quad (19)$$

(II) Пусть $0 < y_1 < y_2 < \dots < y_{n+1}$. Выберем $\delta > 0$ таким, что $y_1 + \delta > y_2, y_2 + \delta < y_3, \dots, y_n + \delta > y_{n+1}$. Тогда с учетом леммы 1 получим

$$\begin{aligned} \mathcal{P}(y_1 < S_1 < y_1 + \delta, y_2 < S_2 < y_2 + \delta, \dots, y_{n+1} < S_{n+1} < y_{n+1} + \delta) &= \\ := \mathcal{P}(N_{(0, y_1]} = 0, N_{(y_1, y_1 + \delta]} = 1, N_{(y_1 + \delta, y_2]} = 0, \dots, N_{(y_{n+1}, y_{n+1} + \delta]} = 1) &= \\ = e^{-\lambda y_1} (\lambda \delta e^{-\lambda \delta})^{n+1} \prod_{j=1}^n e^{-\lambda(y_{j+1} - y_j - \delta)} &= (\lambda \delta)^{n+1} e^{-\lambda \delta} e^{-\lambda y_{n+1}}. \end{aligned} \quad (20)$$

Поделив выражение (20) на δ^{n+1} и устремив δ к нулю, получим

$$f_{S_1, \dots, S_{n+1}}(y_1, \dots, y_{n+1}) = \lambda^{n+1} e^{-\lambda y_{n+1}}, \quad 0 < y_1 < \dots < y_{n+1}. \quad (21)$$

Поскольку с вероятностью единица $0 < S_1 < \dots < S_{n+1}$, то для значений y_1, \dots, y_{n+1} , не входящих в (21), совместная плотность равна нулю. Поделив плотность (21) на выражение для $f_{S_{n+1}}(x_{n+1})$, найденное в (I), получаем требуемый результат.

(III) При $0 < x_1 < x_2 < \dots < x_n < 1$ и $\delta > 0$ таком, что $x_i + \delta < x_{i+1}, \quad i = 1, \dots, n - 1$, имеем

$$\mathcal{P}\left(x_i < \frac{S_i}{S_{n+1}} < x_i + \delta, i = 1, \dots, n\right) =$$

$$= \int_{B_\delta} \dots \int f_{S_1, \dots, S_{n+1}}(y_1, \dots, y_{n+1}) dy_1 \dots dy_{n+1}. \quad (22)$$

где интегрирование ведется по $(n+1)$ -мерному множеству

$$B_\delta = \{(y_1, \dots, y_{n+1}) : x_i < y_i / y_{n+1} < x_i + \delta, i = 1, \dots, n\}.$$

Подставляя (21) в (22), легко видеть, что интеграл (22) сводится к повторному

$$\begin{aligned} & \int_0^\infty \lambda^{n+1} e^{-\lambda y_{n+1}} dy_{n+1} \int_{x_1 y_{n+1}}^{(x_1 + \delta) y_{n+1}} dy_1 \dots \int_{x_n y_{n+1}}^{(x_n + \delta) y_{n+1}} dy_n = \\ & = \delta^n \int_0^\infty \lambda^{n+1} e^{-\lambda y_{n+1}} y_{n+1}^n dy_{n+1} = \delta^n n!, \end{aligned} \quad (23)$$

где мы использовали, что интеграл от плотности $f_{S_{n+1}}(y) = \lambda^{n+1} \times \lambda e^{-\lambda y} y^n / n!$ равен 1. Из (22) и (23) вытекает, что совместная плотность $S_i / S_{n+1}, i = 1, \dots, n$, на множестве $0 < x_1 < \dots < x_n < 1$ постоянна и равна $n!$. Очевидно, она равна нулю для остальных значений переменных.

(IV) Это утверждение доказывается аналогично (II).

Из утверждения (III) леммы 2 вытекает следующий критерий проверки гипотезы об экспоненциальности случайной выборки. Допустим, наблюдаются результаты n независимых испытаний x_1, x_2, \dots, x_n и имеется предположение, что соответствующие сл. в. X_1, X_2, \dots, X_n независимы и имеют одну и ту же экспоненциальную плотность $\lambda e^{-\lambda x}$, $x > 0$ с неизвестным λ . Если предположение верно, то значения

$$z_{(i)} = \left(\sum_{j=1}^i x_j \right) / \sum_{j=1}^n x_j, \quad i = 1, \dots, n-1, \quad (24)$$

можно рассматривать, в соответствии с замечанием к лемме 2, как порядковые статистики для независимой выборки из равномерного на $[0, 1]$ распределения. Отсюда возникает возможность применить различные способы проверки согласованности наблюдений с заданной (известной) теоретической ф. р.

4. Пример.

Следующие данные представляют собой количество летних часов между последовательными отказами установки для кондиционирования воздуха на самолете типа «Боинг-720» [4]:

23	261	87	7	120	14	62	47	225	71
26	21	42	20	5	12	120	11	3	14
71	11	14	11	16	90	1	16	52	95

Считая промежутки между отказами независимыми, проверим предположение об экспоненциальном законе распределения этих промежутков с помощью графического приема, описанного в предыдущем параграфе (вероятностная бумага). На рис. 6 отмечены точки $(x_{(i)}, -\ln(1-(i-0,5)/n))$; $i=1, 2, \dots, n$; $n=30$.

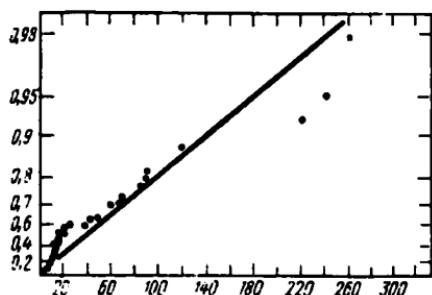


Рис. 6. Точки $(x_{(i)}, -\ln(1-(i-0,5)/n))$ и прилегающая прямая для отказов установки для кондиционирования воздуха

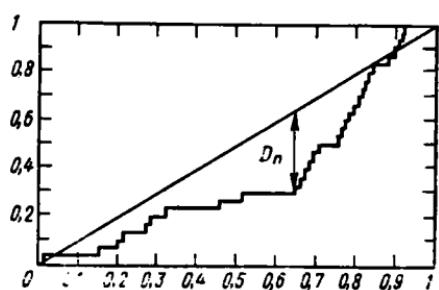


Рис. 7. Эмпирическая функция распределения для точек s_i/s_n , $s_i = x_{(i)} + \dots + x_i$; x_i — времена между отказами установки кондиционирования воздуха

Прямая $x=\hat{\sigma}y$; $\hat{\sigma}=59,1$ проведена на глаз, и согласие с экспоненциальным распределением можно считать удовлетворительным. Эмпирическое среднее равно

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 59,6,$$

так что графическая оценка $\hat{\sigma}$ параметра масштаба σ близка к оценке этого параметра с помощью выборочного среднего. Отклонение по вертикали выборочных точек от прямой достигает наибольшего значения около 0,21 для $x_{(16)}=23$ (см. рис. 6 и учтите неравномерность масштаба по вертикали).

Хотелось бы использовать статистику Колмогорова:

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Чтобы оценить значимость наблюдаемого отклонения э. ф. р. от теоретической ф. р. Однако этому мешает то обстоятельство, что теоретическая ф. р. $F(x/\sigma) = 1 - e^{-x/\sigma}$ содержит неизвестный параметр σ . Можно применить критерий Колмогорова к данной задаче, используя лемму 2. Вычислим значения $s_i = \sum_{j=1}^i x_j$, где x_i — последовательные значения времен между отказами, $j=1, \dots, 30$:

23	284	371	378	478	512	574	621	846	917
1163	1184	1226	1246	1251	1263	1383	1394	1397	1411
1482	1493	1507	1518	1534	1624	1625	1641	1693	1788

После деления на $s_{30} = 1788$ получаем следующий ряд значений $z_{(i)} = s_i / s_{30}$, $i = 1, \dots, 29$:

0,0128	0,1588	0,2074	0,2114	0,2785	0,2863	0,3210	0,3473
0,4731	0,5128	0,6504	0,6621	0,6856	0,6968	0,6996	0,7063
0,7734	0,7796	0,7813	0,7891	0,8288	0,8350	0,8428	0,8485
0,8579	0,9082	0,9088	0,9177	0,9468			

Э. ф. р. для $z_{(i)}$, $i = 1, \dots, 29$, представлена на рис. 7. Далее имеем

$$D_{29} = 0,6504 - \frac{10}{29} = 0,2056, \quad \sqrt{29} D_{29} = 1,107,$$

$$\mathcal{P}(\sqrt{n} D_n > 1,107) \approx 1 - K(1,107) = 0,173,$$

т. е. примерно в 17 случаях из 100 могло бы наблюдаться такое же или большее отклонение.

§ 6. ОЦЕНИВАНИЕ ПАРАМЕТРОВ ЭКСПОНЕНЦИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

1. Оценивание параметра сдвига при известном параметре масштаба.

Пусть X_i , $i = 1, \dots, n$, — независимые сл. в. с плотностью $\sigma^{-1}f((x-\mu)/\sigma)$, где

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

Распределение вероятностей сл. в. X_i сосредоточено на $(\mu, +\infty)$, и, как мы уже видели в § 3, п. 4, оценить μ можно с помощью статистики $X_{(1)}$. Распределение вероятностей $X_{(1)}$ легко находится

$$\mathcal{P}(X_{(1)} > x) = \mathcal{P}(X_1 > x, \dots, X_n > x) = e^{-n(x-\mu)/\sigma}, \quad x > \mu, \quad (1)$$

среднее и дисперсия равны

$$MX_{(1)} = \mu + \sigma/n, \quad DX_{(1)} = \sigma^2/n^2. \quad (2)$$

Считая параметр σ известным, смещение оценки $X_{(1)}$ можно исправить, рассматривая вместо нее $X_{(1)} - \sigma/n$. Если σ неизвестно, то можно попытаться заменить в формуле $X_{(1)} - \sigma/n$ неизвестное σ его оценкой по той же выборке.

2. Оценивание параметра масштаба при известном параметре сдвига.

Рассмотрим задачу оценивания σ , предполагая, что μ известно. Поскольку

$$MX_i = \mu + \sigma, \quad (3)$$

в качестве оценки σ можно предложить

$$T(\mathbf{X}) = \bar{X} - \mu = \frac{1}{n} \sum_{i=1}^n (X_i - \mu).$$

Оценка $T(\mathbf{X})$ — несмещенная, ее дисперсия равна σ^2/n . Так как

$$Y_i = (X_i - \mu)/\sigma, i = 1, 2, \dots, n, \quad (4)$$

— сл. в. со стандартной экспоненциальной плотностью $f(x)$, то для плотности распределения сл. в. $\sum_{i=1}^n Y_i$ можем воспользоваться формулой, выведенной в § 5:

$$g(y) = \frac{y^{n-1}}{(n-1)!} e^{-y}, \quad y > 0.$$

Распределение с плотностью $g(y/2)/2$ играет важную роль в статистике и называется *хи-квадрат распределением с $2n$ степенями свободы*; обозначение χ^2_{2n} . Отметим, что экспоненциальное распределение с плотностью $1/2e^{-x/2}$ является в то же время χ^2_2 -распределением. Подробнее о χ^2 -распределении будет идти речь в следующем параграфе.

Из сказанного следует, что статистика $2 \sum_{i=1}^n Y_i$ имеет χ^2_{2n} -распределение, и поскольку

$$T(\mathbf{X}) = \frac{\sigma}{2n} 2 \sum_{i=1}^n Y_i,$$

то, используя обозначения x_p для p -квантили распределения χ^2_{2n} , получаем

$$\mathcal{P}(2nT(\mathbf{X})/\sigma \leq x_p) = p.$$

Таким образом, α -доверительный интервал для σ можно записать в форме

$$x_{\alpha/2} \leq 2nT(\mathbf{X})/\sigma \leq x_{1-\alpha/2} \text{ или } 2nT(\mathbf{X})/x_{1-\alpha/2} \leq \sigma \leq 2nT(\mathbf{X})/x_{\alpha/2}.$$

Указанный интервал является симметричным в том смысле, что одинаково вероятно, окажется истинное значение σ слева или справа от этого интервала. Легко построить тем же способом и несимметричный α -доверительный интервал.

Функция распределения χ^2_n табулирована (см. [1]), там же указаны квантили x_p распределения χ^2_n для большого набора значений p и $n=1(1)100$ (n от 1 до 100 с шагом 1).

Пример 1. Построим искусственную выборку объема $n=20$ с параметрами $\mu=1$, $\sigma=2$ с помощью таблицы случайных чисел

[1]. Если U — равномерно распределенная сл. в., то при $0 < x < \infty$

$$\mathcal{P}(-\ln U \leq x) = \mathcal{P}(U \geq e^{-x}) = 1 - e^{-x},$$

т. е. сл. в. $-\ln U$ имеет стандартное экспоненциальное распределение. Выборку z_i , $i=1, \dots, n$, из равномерного на $[0, 1]$ распределения получим группировкой по 4 цифры табл. 7.1, а [1]. Затем строим выборку x_i , $i=1, \dots, n$, по формуле $x_i = -2 \ln z_i + 1$:

7,442	2,906	4,481	3,239	1,725	3,844	1,982
3,372	7,550	2,458	2,009	1,353	3,219	4,288
5,374	3,294	1,244	3,346	3,949	1,491	

Вычисляем

$$x_{(1)} = 1,244; \bar{x} = 3,4283.$$

Если предположить, что значение параметра $\sigma = 2$ известно, то в качестве несмешенной оценки μ получаем $x_{(1)} - \sigma/20 = 1,244 - 0,1 = 1,144$. Если предположить, что μ известно, то получаем несмешенную оценку σ : $T(x) = \bar{x} - \mu = 2,4283$. По табл. 2.2, а [1] находим значения квантилей распределения χ^2_{40} :

$$x_{0,025} = 24,433; x_{0,975} = 59,342$$

и вычисляем границы доверительного интервала с коэффициентом доверия 0,95:

$$2nT(x)/x_{0,975} = 1,636; 2nT(x)/x_{0,025} = 3,975.$$

3. Точечные оценки μ и σ .

Когда оба параметра неизвестны, оценку σ можно получить, заменив в (4) μ на его оценку $X_{(1)}$:

$$\hat{\sigma} = \bar{X} - X_{(1)}.$$

Из (1) и (3) вытекает, что $\hat{\sigma}$ — смешенная оценка:

$$M\hat{\sigma} = \mu + \sigma - \left(\mu + \frac{\sigma}{n} \right) = \frac{n-1}{n} \sigma,$$

поэтому несмешенной оценкой будет

$$\sigma^* = \frac{n}{n-1} \hat{\sigma} = \frac{1}{n-1} \sum_{i=2}^n (X_{(i)} - X_{(1)}). \quad (5)$$

Несмешенной оценкой параметра μ , очевидно, будет служить

$$\mu^* = X_{(1)} - \sigma^*/n. \quad (6)$$

4. Распределение интервалов между порядковыми статистиками.

Лемма 1. Пусть Y_i , $i=1, \dots, n$, — независимые сл. в. со стандартной экспоненциальной плотностью $f(y) = e^{-y}$, $y > 0$, а $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ — порядковые статистики. Тогда сл. в. $Z_i =$

$(n-i+1)(Y_{(i)} - Y_{(i-1)})$, $i=1, \dots, n$, $Y_{(0)}=0$, независимы и имеют ту же плотность $f(z)$.

Доказательство. Легко видеть, что события

$$Y_{i_1} < Y_{i_2} < \dots < Y_{i_n},$$

где (i_1, i_2, \dots, i_n) — перестановка из чисел $(1, 2, \dots, n)$, равновероятны. Поэтому, полагая $Y_{(0)}=Y_0=0$ и выбирая $x_i>0$, $i=1, \dots, n$, получаем

$$\begin{aligned} & \mathcal{P}((n-i+1)(Y_{(i)} - Y_{(i-1)}) > x_i, i=1, \dots, n) = \\ & = n! \mathcal{P}(Y_1 < Y_2 < \dots < Y_n, (n-i+1)(Y_i - Y_{i-1}) > x_i, i=1, \dots, n) = \\ & = n! \mathcal{P}(Y_i - Y_{i-1} > x_i / (n-i+1), i=1, \dots, n) = \\ & = n! \int_{x_1/n}^{\infty} e^{-y_1} dy_1 \int_{y_1+x_1/(n-1)}^{\infty} e^{-y_2} dy_2 \dots \int_{y_{n-2}+x_{n-1}/2}^{\infty} e^{-y_{n-1}} dy_{n-1} \int_{y_{n-1}+x_{n-1}/2}^{\infty} e^{-y_n} dy_n = \\ & = n! \int_{x_1/n}^{\infty} e^{-y_1} dy_1 \dots \int_{y_{n-2}+x_{n-1}/2}^{\infty} e^{-2y_{n-1}} dy_{n-1} \frac{1}{1} e^{-x_n} = \\ & = n! \int_{x_1/n}^{\infty} e^{-y_1} dy_1 \dots \int_{y_{n-3}+x_{n-2}/3}^{\infty} e^{-3y_{n-2}} dy_{n-2} \frac{1}{2} e^{-x_{n-1}} \frac{1}{1} e^{-x_n} = \dots = \\ & = e^{-x_1} e^{-x_2} \dots e^{-x_n}. \end{aligned}$$

Замечание. Лемму 1 можно рассматривать как одно из следствий свойства отсутствия последействия экспоненциального распределения. Представим себе, что Y_i , $i=1, \dots, n$, — времена жизни n приборов, поставленных одновременно на испытания. Тогда $Y_{(1)}$ — время до отказа самого «плохого» прибора и, как следует из (1), имеет плотность $n e^{-ny}$, $y>0$. Момент $Y_{(1)}$ ввиду свойства отсутствия последействия экспоненциального распределения можно рассматривать как начальный момент работы для оставшихся «в живых» $n-1$ приборов. Следовательно, $Y_{(2)}-Y_{(1)}$ — время до отказа самого «плохого» из этих $n-1$ приборов имеет распределение с плотностью $(n-1)e^{-(y-n+1)}$, $y>0$, и т. д. Эти рассуждения можно сделать строгими, но приведенное выше доказательство короче.

Теорема 1. Пусть сл. в. X_i , $i=1, \dots, n$, независимы и одноково распределены с плотностью $\sigma^{-1}e^{-(x-\mu)/\sigma}$, $x>\mu$. Тогда сл. в.

$$nX_{(1)}, (n-i+1)(X_{(i)} - X_{(i-1)}), i=2, \dots, n,$$

независимы, $nX_{(1)}$ имеет плотность $\sigma^{-1}e^{-(x-n\mu)/\sigma}$, $x>n\mu$, а $(n-i+1)(X_{(i)} - X_{(i-1)})$, $i=2, \dots, n$, имеют одинаковую плотность $\sigma^{-1}e^{-(x/\sigma)}$, $x>0$.

Доказательство. Из формулы (4) имеем, что сл. в.

$$Y_{(i)} = (X_{(i)} - \mu) / \sigma, \quad i = 1, \dots, n,$$

применима лемма 1. Используя ее, получаем, что сл. в.

$$n(X_{(1)} - \mu) / \sigma, \quad (n-i+1)(X_{(i)} - X_{(i-1)}) / \sigma, \quad i = 2, \dots, n, \quad (7)$$

независимы и одинаково распределены с плотностью $e^{-x}, x > 0$.

5. Дисперсии несмешанных оценок μ^* , σ^* ; интервальное оценивание μ , σ .

Легко проверить, что

$$\frac{1}{n-1} \sum_{i=2}^n (n-i+1)(X_{(i)} - X_{(i-1)}) = \frac{1}{n-1} \sum_{i=2}^n X_{(i)} - X_{(1)} = \sigma^*. \quad (8)$$

Применяя теорему 1, находим дисперсии несмешанных оценок:

$$\begin{aligned} D\sigma^* &= D \left\{ \frac{1}{n-1} \sum_{i=2}^n (n-i+1)(X_{(i)} - X_{(i-1)}) \right\} = \\ &= \frac{1}{(n-1)^2} \sum_{i=2}^n D \{(n-i+1)(X_{(i)} - X_{(i-1)})\} = \frac{\sigma^2}{n-1}, \\ D\mu^* &= D \left(X_{(1)} - \frac{1}{n} \sigma^* \right) = DX_{(1)} + \frac{1}{n^2} D\sigma^* = \frac{\sigma^2}{(n-1)n}. \end{aligned} \quad (9)$$

Из (7) вытекает, что сл. в. $2(n-1)\sigma^*/\sigma$ имеет распределение $\chi_{2(n-1)}^2$ (см. п. 2), а независящая от σ^* сл. в. $2n(X_{(1)} - \mu) / \sigma$ имеет χ_2^2 -распределение (т. е. распределение с плотностью $1/2e^{-x/2}$, $x > 0$). Отсюда вытекает следующая процедура построения доверительных интервалов для μ и σ .

Пусть x_p — p -квантиль распределения $\chi_{2(n-1)}^2$, тогда α -доверительный интервал для σ имеет вид

$$2(n-1)\sigma^*/x_{1-\alpha/2} \leq \sigma \leq 2(n-1)\sigma^*/x_{\alpha/2}. \quad (10)$$

Для построения доверительного интервала для μ рассмотрим сл. в.

$$\frac{2n(X_{(1)} - \mu)}{2(n-1)\sigma^*} = \frac{2n(X_{(1)} - \mu)/\sigma}{2(n-1)\sigma^*/\sigma}. \quad (11)$$

Сл. в. (11) представлена как отношение двух независимых сл. в., имеющих хи-квадрат распределения, домноженная на $n-1$, она имеет распределение Фишера $F_{2,2(n-1)}$ (см. п. 7 § 7). Пусть x_p —

ρ -квантиль этого распределения. Тогда получаем β -доверительный интервал для μ :

$$X_{(1)} - x_{1-\beta/2}\sigma^*/n \leq \mu \leq X_{(1)} + x_{\beta/2}\sigma^*/n. \quad (12)$$

Пример 1 (продолжение). Для экспоненциальной выборки с $\mu = 1$, $\sigma = 2$ имеем

$$x_{(1)} = 1,244; \sum_{i=2}^{20} (x_{(i)} - x_{(1)}) = 44,93;$$

$$\sigma^* = \frac{1}{19} \sum_{i=2}^{20} (x_{(i)} - x_{(1)}) = 2,364; \mu^* = x_{(1)} - \frac{1}{20} \sigma^* = 1,125.$$

С помощью таблиц квантилей χ^2_{38} находим по формуле (10) доверительный интервал для σ с $\alpha = 0,05$:

$$2 \cdot 44,98/56,895 \leq \sigma \leq 2 \cdot 44,93/22,878, \text{ или } 1,579 \leq \sigma \leq 3,927.$$

Квантили x_p распределения $F_{2,38}$ для $p_1 = 0,025$ и $p_2 = 0,975$ рассчитаны в примере § 7, п. 7:

$$x_{p_1} = 0,025; x_{p_2} = 4,071.$$

Отсюда из (12) получаем доверительный интервал для μ с $\beta = 0,05$:

$$1,244 - 4,071 \cdot 2,364 \cdot 0,05 \leq \mu \leq 1,244 - 0,025 \cdot 2,364 \cdot 0,05,$$

или

$$0,7628 \leq \mu \leq 1,2410.$$

6. Оценивание по первым r порядковым статистикам.

До сих пор рассматривались статистические процедуры, основанные на результатах x_i , $i = 1, \dots, n$, независимых одинаково распределенных наблюдений. Фактически все статистические выводы зависели от наблюдений через порядковые статистики $x_{(i)}$, $i = 1, \dots, n$. В некоторых приложениях возникает задача оценивания по некоторому подмножеству порядковых статистик. Например, при испытании изделий на надежность может быть применен следующий план испытаний: выбирается определенное $r \leq n$, и испытания продолжаются до момента появления r -го по порядку отказа. Результатами наблюдений в таком случае оказываются значения первых r порядковых статистик $x_{(i)}$, $i = 1, \dots, r$. Проведенные выше построения для полной выборки легко переносятся и на эту ситуацию.

Итак, пусть $X_{(i)}$, $i = 1, \dots, r$, $r \leq n$, — первые r порядковых статистик для случайной выборки X_i , $i = 1, \dots, n$, из экспоненциального распределения с параметрами μ и σ . Из теоремы 1 вытекает, что сл. в.

$$Z_i = (n-i+1)(X_{(i)} - X_{(i-1)}), \quad i = 2, \dots, r, \quad (13)$$

независимы и имеют плотность $\frac{1}{\sigma} e^{-x/\sigma}$. Оценкой σ , теоретического среднего, служит выборочное среднее наблюдений Z_i :

$$\bar{Z} = \frac{1}{r-1} \sum_{i=2}^r Z_i = \frac{1}{r-1} \left(\sum_{i=2}^r (X_{(i)} - X_{(1)}) + (n-r)(X_{(r)} - X_{(1)}) \right) = \\ = \frac{1}{r-1} \left(\sum_{i=1}^r X_{(i)} - (n-r)X_{(r)} - nX_{(1)} \right). \quad (14)$$

При $r=n$ формула (14) совпадает с (5). Отметим, что переход от наблюдений $X_{(i)}$, $i=1, \dots, r$, к $Z_i = (n-i+1)(X_{(i)} - X_{(i-1)})$, $i=2, \dots, r$, приводит к некоторой потере информации, так как, чтобы восстановить последовательность $X_{(i)}$, $i=1, \dots, r$, из Z_i , $i=2, \dots, r$, необходимо еще знание $X_{(1)}$. На самом деле, как мы увидим в гл. III, вся нужная для оценки σ информация содержится в $Z = (Z_2, \dots, Z_r)$.

Оценка \bar{Z} , очевидно, несмешенная, ее дисперсия равна $\sigma^2/(r-1)$. В качестве оценки μ естественно взять (ср. с. (6))

$$X_{(1)} - \frac{1}{n} \bar{Z}. \quad (15)$$

Эта оценка также несмешенная, ее дисперсия равна $\sigma^2 r / ((r-1) \times n^2)$. (ср. с (9)). Доверительные интервалы для μ и σ строятся на основании теоремы 1 и имеют тот же вид (10) при $n=r$ и (12) со следующими изменениями: вместо σ^* , определяемой формулой (5), берется \bar{Z} , определяемое в (14); квантили x_p и $x'_{p'}$ в (10) и (12) берутся соответственно для распределений $\chi^2_{2(r-1)}$ и $F_{2,2(r-1)}$.

Пример 1 (продолжение). Проведем по формулам (14), (15) вычисление оценок, построенных по $x_{(i)}$, $i=1, \dots, r$, для трех случаев $r=5$, $r=10$, $r=15$ и объединим результаты в таблицу вместе со статистикой полной выборки ($r=20$):

r	5	10	15	20
σ^*	3,161	2,914	2,728	2,364
μ^*	1,086	1,099	1,108	1,125

Теоретические значения: $\mu=1$, $\sigma=2$.

§ 7*. СВЕДЕНИЯ О ВАЖНЕЙШИХ НЕПРЕРЫВНЫХ РАСПРЕДЕЛЕНИЯХ В R^1

1. Гамма-распределение.

Стандартная плотность задается формулой

$$f(x) = \begin{cases} \frac{\Gamma(p)}{x^{p-1}} e^{-x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

где

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx \quad (1)$$

— гамма-функция Эйлера, p — параметр, называемый *параметром формы*. Обычно рассматривают гамма-распределение, содержащее еще и масштабный параметр λ . Примем обозначение $G(\lambda, p)$ для распределения с плотностью

$$\lambda f(\lambda x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}, \quad x > 0. \quad (2)$$

Центральные моменты распределения $G(1, p)$ равны

$$\int_0^\infty x^r f(x) dx = \frac{1}{\Gamma(p)} \int_0^\infty x^r x^{p-1} e^{-x} dx = \frac{\Gamma(p+r)}{\Gamma(p)}, \quad r > -p. \quad (3)$$

Из формулы (1) интегрированием по частям легко получить

$$\Gamma(p+1) = p\Gamma(p). \quad (4)$$

Отметим также известную из анализа формулу $\Gamma(1/2) = \sqrt{\pi}$. Принимая во внимание (4), из (3) находим среднее и дисперсию $G(1, p)$:

$$\int_0^\infty x f(x) dx = p, \quad (5)$$

$$\int_0^\infty (x-p)^2 f(x) dx = \int_0^\infty x^2 f(x) dx - p^2 = p.$$

Центральный момент $\int x^r f(x) dx$ порядка $r > -p$ распределения $G(\lambda, p)$ получается домножением выражения (3) на λ^{-r} ; среднее и дисперсия равны соответственно p/λ и p/λ^2 .

Лемма 1. Пусть X_i — независимые сл. в. с распределением $G(\lambda_i, p)$, $i = 1, 2, \dots, n$. Тогда $X_1 + \dots + X_n$ имеет распределение $G(\lambda_1 + \dots + \lambda_n, p)$.

* К материалу этого параграфа можно обращаться по мере необходимости.

Доказательство. Без ограничения общности положим $\lambda=1$. Достаточно рассмотреть случай $n=2$. По формуле свертки ([2], т. 2) плотность сл. в. X_1+X_2 равна при $x>0$

$$\begin{aligned} f_{x_1+x_2}(x) &= \int_{-\infty}^{+\infty} f_{X_1}(t) f_{X_2}(x-t) dt = \\ &= \frac{1}{\Gamma(p_1)\Gamma(p_2)} \int_0^x t^{p_1-1} e^{-t} (x-t)^{p_2-1} e^{-(x-t)} dt. \end{aligned}$$

Сделав замену переменных $s=xt$, приведем выражение для $f_{x_1+x_2}(x)$ к виду

$$ce^{-x} x^{p_1+p_2-1}, \quad c = \frac{1}{\Gamma(p_1)\Gamma(p_2)} \int_0^1 s^{p_1-1} (1-s)^{p_2-1} ds. \quad (6)$$

Так как интеграл от плотности равен 1, то постоянный множитель c равен $1/\Gamma(p_1+p_2)$ (см. (1)), что и доказывает лемму при $n=2$.

Отметим попутно доказанную формулу

$$\int_0^1 s^{p_1-1} (1-s)^{p_2-1} ds = \frac{\Gamma(p_1)\Gamma(p_2)}{\Gamma(p_1+p_2)}, \quad p_1, p_2 > 0. \quad (7)$$

Из леммы 1 вытекает, что распределение $G(\lambda, p)$ при больших p приближается нормальным с теми же средним и дисперсией. Действительно, если $p=n$ — натуральное число, то соответствующая сл. в. распределена как сумма n независимых сл. в. с распределением $G(\lambda, 1)$ и применима центральная предельная теорема для одинаково распределенных слагаемых. Если p не натуральное, то к сумме добавляется еще одно слагаемое с распределением $G(\lambda, p-[p])$, где $[p]$ — целая часть p , которое асимптотически пренебрежимо.

2. Экспоненциальное распределение.

Экспоненциальное распределение с плотностью $\lambda e^{-\lambda x}$, $x>0$, является частным случаем гамма-распределения с параметром формы, равным 1, и обозначается $G(\lambda, 1)$. Из леммы 1 вытекает, что сумма n независимых $G(\lambda, 1)$ -распределенных сл. в. имеет распределение $G(\lambda, n)$. Этот результат был получен ранее другим способом (§ 5, лемма 2). Среднее и дисперсия экспоненциального распределения $G(\lambda, 1)$ равны λ^{-1} и λ^{-2} соответственно.

3. Нормальное распределение.

Нормальное распределение в стандартной форме имеет плотность

$$\varphi(x) = ce^{-x^2/2}, -\infty < x < +\infty,$$

где константа $c = 1/\sqrt{2\pi}$ находится из условия нормировки:

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} \varphi(x) dx = \sqrt{2\pi} c \int_0^{\infty} e^{-x^2/2} \sqrt{\frac{2}{x^2}} d \frac{x^2}{2} = \\ &= \sqrt{2\pi} c \int_0^{\infty} u^{-\frac{1}{2}} e^{-u} du = \sqrt{2\pi} c \Gamma\left(\frac{1}{2}\right) = \sqrt{2\pi} c \sqrt{\pi}. \end{aligned}$$

Нормальное распределение с плотностью

$$\frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \quad (8)$$

обозначим через $N(\mu, \sigma^2)$. Плотность стандартного нормального распределения, как правило, будет обозначаться $\varphi(x)$, а соответствующая ф.р. — $\Phi(x)$. Центральные моменты распределения $N(0, 1)$ нечетного порядка, очевидно, равны нулю, а четного —

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-\frac{x^2}{2}} dx &= \frac{\sqrt{2} \cdot 2^k}{\sqrt{2\pi}} \int_0^{\infty} \left(\frac{x^2}{2}\right)^k e^{-\frac{x^2}{2}} \sqrt{\frac{2}{x^2}} d \frac{x^2}{2} = \\ &= \frac{2^k}{\sqrt{\pi}} \int_0^{\infty} u^{k-1/2} e^{-u} du = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right). \end{aligned}$$

Принимая во внимание формулы $\Gamma(p+1) = p\Gamma(p)$ и $\Gamma(1/2) = \sqrt{\pi}$, получаем рекуррентное соотношение

$$\begin{aligned} \Gamma(k+1/2) &= \Gamma((k-1/2)+1) = (k-1/2)\Gamma(k-1/2) = \\ &= (k-1/2)\Gamma((k-1)+1/2), \end{aligned}$$

откуда для центрального момента порядка $2k$ получаем

$$(2k-1)(2k-3) \cdots 3 \cdot 1. \blacksquare \quad (9)$$

Пусть сл. в. X имеет распределение $N(0, 1)$. Тогда в соответствии со сказанным

$$MX = 0, DX = 1.$$

Сл. в. $\sigma X + \mu$, $\sigma > 0$, имеет среднее μ , дисперсию σ^2 , а ее плотность равна $\sigma^{-1}\varphi((x-\mu)/\sigma)$, т. е. сл. в. $\sigma X + \mu$ имеет распределение $N(\mu, \sigma^2)$. Таким образом, μ и σ^2 являются соответственно средним и дисперсией распределения $N(\mu, \sigma^2)$. ■

Лемма 2. Пусть X_i — независимые сл. в. с распределением $N(\mu_i, \sigma_i^2)$, $i=1, 2, \dots, n$. Тогда сл. в. $X_1 + \dots + X_n$ имеет распределение $N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2)$.

Доказательство проведем для $n=2$. Положим $Y_i = (X_i - \mu_i)/\sigma_i$ и запишем

$$X_1 + X_2 = \sigma_1 \left(Y_1 + \frac{\sigma_2}{\sigma_1} Y_2 \right) + \mu_1 + \mu_2.$$

Легко видеть, что достаточно установить, что сл. в. $Y_1 + aY_2$, $a = \sigma_2/\sigma_1$, имеет распределение $N(0, 1+a^2)$. По формуле свертки

$$f_{Y_1+aY_2}(y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}a} e^{-\frac{(y-x)^2}{2a^2}} dx. \quad (10)$$

Проведем преобразования показателя экспоненты в интеграле (10):

$$\begin{aligned} x^2 - (y-x)^2/a^2 &= x^2 \left(1 + \frac{1}{a^2} \right) - 2xy \frac{1}{a^2} + \\ &- y^2 \frac{1}{a^2} = \left(\frac{x}{a} \sqrt{a^2+1} - \frac{y}{a} \frac{1}{\sqrt{1+a^2}} \right)^2 - \frac{y^2}{a^2(a^2+1)} + y^2 \frac{1}{a^2}. \end{aligned} \quad (11)$$

Подставляя (11) в (10), получаем, что

$$f_{Y_1+aY_2}(y) = b e^{-\frac{1}{2} \left(\frac{1}{a^2} - \frac{1}{a^2(a^2+1)} \right) y^2} = b e^{-\frac{1}{2} \frac{y^2}{a^2+1}}, \quad (12)$$

где

$$b = \frac{1}{2\pi a} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{x}{a} \sqrt{a^2+1} - \frac{y}{a} \frac{1}{\sqrt{a^2+1}} \right)^2} dx,$$

очевидно, не зависит от y , т. е. является постоянной. Сравнивая (12) с (8), получаем требуемое утверждение.

4. Распределение χ^2 с n степенями свободы.

Это распределение определим как распределение суммы n квадратов независимых сл. в. с общим распределением $N(0, 1)$. Пусть сл. в. X имеет распределение $N(0, 1)$, тогда ф. р. X^2 равна при $x > 0$

$$F_{X^2}(x) = \mathcal{P}(X^2 \leq x) = \mathcal{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}),$$

а ее плотность равна

$$\begin{aligned} f_{X^2}(x) &= \frac{1}{2\sqrt{x}} \Phi'(\sqrt{x}) + \frac{1}{2\sqrt{x}} \Phi'(-\sqrt{x}) = \\ &= \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} = \Gamma\left(\frac{1}{2}\right)^{-1} \left(\frac{1}{2}\right)^{\frac{1}{2}} x^{\frac{1}{2}-1} e^{-x/2}. \end{aligned}$$

Сравнивая полученное выражение для $f_{X^*}(x)$ с формулой (2), заключаем, что квадрат стандартной нормальной сл. в. имеет распределение $G(1/2, 1/2)$. Отсюда и из леммы 1 получаем, что хи-квадрат распределение с n степенями свободы совпадает с распределением $G(1/2, n/2)$. Для хи-квадрат распределения принято независимое обозначение χ^2_n . Отметим, что экспоненциальное распределение $G(1/2, 1)$ совпадает с χ^2_2 , а сумма n независимых $G(1/2, 1)$ -распределенных сл. в. имеет распределение χ^2_{2n} , с чем мы уже встречались в § 6.

Функция распределения χ^2_n табулирована, при больших n можно пользоваться нормальным приближением. Если сл. в. Y имеет распределение χ^2_n , то (см. (5))

$$MY = n, \quad DY = 4n/2 = 2n,$$

и ввиду центральной предельной теоремы

$$\Phi((Y-n)/\sqrt{2n}) \leq y \Rightarrow \Phi(y). \quad (13)$$

5. Бета-распределение.

Это распределение задается плотностью

$$f(x) = \begin{cases} \beta(a, b)^{-1} x^{a-1} (1-x)^{b-1}, & 0 < x < 1, \\ 0 & , x \geq 1, x \leq 0. \end{cases}$$

где

$$\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

— бета-функция Эйлера, $a > 0, b > 0$ — параметры. Функция распределения, которая называется также неполной бета-функцией, обозначается $I_x(a, b)$. Учитывая формулу связи бета- и гамма-функций (см. 7))

$$B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b),$$

запишем

$$I_x(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad 0 < x < 1. \quad (14)$$

Отметим очевидное тождество

$$I_x(a, b) = 1 - I_{(1-x)}(b, a), \quad (15)$$

которое учитывается при составлении таблиц ф. р. и квантилей. Обозначив через $x_p(I_x(a, b))$ p -квантиль распределения $I_x(a, b)$, имеем из (15)

$$1 - I_{1-x_p(I(a, b))}(b, a) = p,$$

или

$$x_{1-p}(I(b, a)) = 1 - x_p(I(a, b)). \quad \blacksquare \quad (16)$$

Между сл. в. с гамма- и бета- распределениями имеется интересная связь, используемая в статистике и устанавливаемая в следующей лемме.

Лемма 3. Пусть сл. в. X и Y независимы и имеют распределения $G(\lambda, a)$ и $G(\lambda, b)$ соответственно. Тогда сл. в. $Z=X/(X+Y)$ имеет ф. р. $I_z(a, b)$.

Доказательство. Рассмотрим отношение $U=X/Y$ и, не ограничивая общности, положим $\lambda=1$. Для ф. р. U получаем при $u>0$

$$F_U(u) = \mathbb{P}(X < uY) = \iint_{x < uy} f_X(x) f_Y(y) dx dy = \\ = \int_0^\infty f_Y(y) dy \int_0^{uy} f_X(x) dx = \int_0^\infty f_Y(y) F_X(uy) dy.$$

Дифференцируя, находим плотность

$$f_U(u) = \int_0^\infty f_Y(y) f_X(uy) y dy = \frac{1}{\Gamma(a)\Gamma(b)} \int_0^\infty y^{b-1} e^{-y} (uy)^{a-1} e^{-uy} y dy = \\ = \frac{u^{a-1}}{\Gamma(a)\Gamma(b)} \int_0^\infty y^{a+b-1} e^{-(1+u)y} dy = \frac{u^{a-1}}{\Gamma(a)\Gamma(b)} \frac{1}{(1+u)^{a+b}} \Gamma(a+b). \quad (17)$$

Так как

$$F_Z(z) = \mathbb{P}\left(\frac{U}{1+U} < z\right) = \mathbb{P}\left(U < \frac{z}{1-z}\right) = F_U\left(\frac{z}{1-z}\right), \quad (18)$$

то

$$f_Z(z) = f_U\left(\frac{z}{1-z}\right) (1-z)^{-2} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{z}{1-z}\right)^{a-1} \left(1 + \frac{z}{1-z}\right)^{-a-b} \times \\ \times (1-z)^{-2} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1} (1-z)^{b-1}. \quad (19)$$

6. Равномерное распределение.

Равномерное распределение на $(0, 1)$ является частным случаем бета-распределения с параметрами $a=b=1$; это распределение подробно рассматривалось в § 2.

7. Распределение Фишера, или F-распределение.

Это распределение определяется как распределение сл. в.

$$V = \frac{X/m}{Y/n} = \frac{nX}{mY}, \quad (20)$$

где X и Y независимы и имеют распределения χ^2_m и χ^2_n соответственно. Для F-распределения принято обозначение $F_{m,n}$, парамет-

ры m и n называются степенями свободы. Так как $\chi^2_k = G(1/2, k/2)$, то плотность сл. в. $U=X/Y$ получается из формулы (17) при $a=m/2$, $b=n/2$, откуда получаем, что при $v>0$

$$f_V(v) = \frac{m}{n} f_U\left(\frac{m}{n} v\right) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2) \Gamma(n/2)} m^{m/2} n^{n/2} v^{m/2-1} (n+mv)^{-(m+n)/2} \quad (21)$$

Из (21) и (17) легко получаем значения для центральных моментов $F_{2m,2n}$ распределения:

$$\begin{aligned} \int_0^\infty u^l f_V(v) dv &= \left(\frac{n}{m}\right)^l \int_0^\infty u^l f_U(u) du = \left(\frac{n}{m}\right)^l \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} \times \\ &\times \int_0^\infty u^{m+l-1} (1+u)^{-m-n} du = \left(\frac{n}{m}\right)^l \frac{\Gamma(m+l)\Gamma(n-l)}{\Gamma(m)\Gamma(n)} \times \\ &\times \frac{\Gamma(m+n)}{\Gamma(m+l)\Gamma(n-l)} \int_0^\infty u^{(m+l)-1} (1+u)^{-(m+l)-(n-l)} du = \\ &= \left(\frac{n}{m}\right)^l \frac{\Gamma(m+l)\Gamma(n-l)}{\Gamma(m)\Gamma(n)}, \quad l < n. \end{aligned}$$

Из (18), (19) имеем

$$F_U(u) = F_Z\left(\frac{u}{u+1}\right) = I_{u/(u+1)}\left(\frac{m}{2}, \frac{n}{2}\right),$$

откуда устанавливаем связь распределения Фишера с бета-распределением

$$F_V(v) = F_{nU/m}(v) = I_{mv/(n+mv)}(m/2, n/2). \blacksquare$$

Из формулы (20) имеем

$$\mathcal{P}\left(\frac{nX}{mY} < v\right) = 1 - \mathcal{P}\left(\frac{mY}{nX} < \frac{1}{v}\right). \quad (22)$$

Обозначив через $x_p(F_{m,n})$ p -квантиль распределения $F_{m,n}$, получаем из (22) следующее тождество:

$$x_p(F_{m,n}) = 1/x_{1-p}(F_{n,m}), \quad (23)$$

которое учитывается при составлении таблиц квантилей F -распределения. ■

В книге [1] приведены таблицы квантилей $x_p(F_{m,n})$ для ряда значений p и

$$m=1(1)10, 12, 15, 20, 24, 30, 40, 60, 120, \infty,$$

$$n=1(1), 30, 40, 60, 120, \infty.$$

Интерполяцию рекомендуется проводить по аргументам $1/m$, $1/n$,

для чего значения аргументов $m \geq 10$, $n \geq 30$ выбраны из условия постоянства шага по переменным $1/m$, $1/n$, что упрощает вычисления.

Пример. Найдем $x_{p_1}(F_{2,38})$ для значений $p_1=0.025$ и $p_2=0.975$. По табл. 3.5 [1] находим

$$x_{p_1}(F_{2,30}) = 4,1821; \quad x_{p_1}(F_{2,40}) = 4,0510,$$

$$x_{p_2}(F_{30,2}) = 39,465; \quad x_{p_2}(F_{40,2}) = 39,473.$$

Линейной интерполяцией по $1/n$ находим

$$\begin{aligned} x_{p_1}(F_{2,38}) &= x_{p_1}(F_{2,30}) + (x_{p_1}(F_{2,40}) - x_{p_1}(F_{2,30})) \left(\frac{1}{40} - \frac{1}{30} \right)^{-1} \times \\ &\times \left(\frac{1}{38} - \frac{1}{30} \right) = 4,1821 + (-0,1311)(-0,00833)^{-1}(-0,007017) = \\ &= 4,071; \\ x_{p_2}(F_{38,2}) &= 39,465 + (-0,00833)^{-1}(-0,007017)0,008 = 39,4717; \\ x_{p_1}(F_{2,38}) &= (x_{p_2}(F_{38,2}))^{-1} = (39,4717)^{-1} = 0,025. \end{aligned}$$

8. Распределение Стьюдента, или t-распределение.

Это распределение определяется как распределение сл. в.

$$T = X / \sqrt{Y/n}, \quad (24)$$

где сл. в. X и Y независимы и имеют распределения $N(0, 1)$ и χ^2_n соответственно; параметр n называют числом степеней свободы t-распределения. Плотность сл. в. T , очевидно, симметричная:

$$f_T(t) = f_T(-t).$$

Замечая, что T^2 имеет $F_{1,n}$ -распределение и что при $t > 0$

$$F_T(t) - F_T(-t) = F_{T^2}(t^2), \quad 2f_T(t) = 2tf_{T^2}(t^2),$$

получаем из (21) при $m=1$, $t > 0$:

$$f_T(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} (1+t^2/n)^{-\frac{n+1}{2}}$$

С ростом n распределение сл. в. T приближается к $N(0, 1)$.

§ 8. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ: ОЦЕНИВАНИЕ ПАРАМЕТРОВ, СРАВНЕНИЕ ДВУХ ВЫБОРОК

1. Доверительные оценки для μ и σ^2 , когда один из параметров известен.

Рассмотрим выборку X_i , $i=1, \dots, n$, из нормального распределения $N(\mu, \sigma^2)$ со средним μ и дисперсией σ^2 . Точечными оценками μ и σ^2 служат выборочные среднее и дисперсия

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Предположим вначале, что один из параметров μ или σ^2 известен. В случае известного σ величины

$$\sqrt{n}(\bar{X} - \mu)/\sigma \quad (1)$$

имеет нормальное распределение $N(0, 1)$ (см. § 7, п. 3), и α -доверительный интервал для μ получаем из соотношения

$$\mathcal{P}(x_{\alpha/2} < \sqrt{n}(\bar{X} - \mu)/\sigma \leq x_{1-\alpha/2}) = \alpha,$$

или

$$\mathcal{P}(\bar{X} - \sigma x_{1-\alpha/2}/\sqrt{n} \leq \mu \leq \bar{X} + \sigma x_{\alpha/2}/\sqrt{n}) = \alpha,$$

где x_α — α -квантиль распределения $N(0, 1)$. Ввиду симметрии $N(0, 1)$ имеем $x_\alpha = -x_{1-\alpha}$, следовательно, α -доверительный интервал можно представить в виде

$$\bar{X} - \sigma x_{1-\alpha/2}/\sqrt{n} \leq \mu \leq \bar{X} + \sigma x_{\alpha/2}/\sqrt{n}. \blacksquare$$

Допустим, что μ известно, а требуется оценить σ^2 . Для точечного оценивания σ^2 естественно вместо выборочной дисперсии взять статистику

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2,$$

которая отличается от S_0^2 заменой выборочного среднего на теоретическое. Для построения интервальной оценки образуем величину

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2, \quad (2)$$

имеющую распределение χ^2_n (см. § 7, п. 4). Отсюда получаем α -доверительный интервал для σ^2 в виде

$$x_{1-\alpha/2}^{-1} \sum_{i=1}^n (X_i - \mu)^2 \leq \sigma^2 \leq x_{\alpha/2}^{-1} \sum_{i=1}^n (X_i - \mu)^2,$$

где x_α — α -квантиль распределения χ^2_n .

2. Несмещенная оценка дисперсии.

В случае независимой выборки из произвольного распределения выборочное среднее является несмещенной оценкой теоретического среднего

$$\mathbf{M}\bar{X} = \mathbf{M}X_1.$$

Однако выборочная дисперсия имеет смещение

$$\begin{aligned} MS_0^2 &= M \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \sum_{i,j=1}^n X_i X_j \right) = MX_1^2 - \\ &- \frac{1}{n} MX_1^2 - \frac{(n-1)}{n} (MX_1)^2 = \left(1 - \frac{1}{n} \right) DX_1 \end{aligned}$$

Поэтому обычно вместо S_0^2 рассматривают несмешенную оценку

$$S^2 = \frac{n}{n-1} S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

3. Доверительные интервалы для μ и σ^2 , когда оба параметра неизвестны.

Допустим, что оба параметра μ и σ^2 распределения $N(\mu, \sigma^2)$ неизвестны. Для построения интервальных оценок вместо величин (1), (2) образуем следующие:

$$\sqrt{n}(\bar{X} - \mu)/S, \quad \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (3)$$

Введем нормированные сл. в.

$$Y_i = (X_i - \mu)/\sigma, \quad i = 1, \dots, n,$$

которые независимы и $N(0, 1)$ -распределены. Так как

$$\frac{1}{\sigma} (\bar{X} - \mu) = \bar{Y},$$

то величины (3) можно переписать в виде

$$\bar{Y} / \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad \sum_{i=1}^n (Y_i - \bar{Y})^2. \quad (4)$$

Таким образом, распределения вероятностей величин (3) не зависят от неизвестных параметров μ и σ^2 , и для построения доверительных интервалов остается лишь найти эти распределения.

Лемма 1. Пусть $Y_i, i = 1, \dots, n$, независимы, $N(0, 1)$ -распределены, $C = [c_{ij}, i, j = 1, \dots, n]$ — ортогональная $n \times n$ -матрица:

$\sum c_{ii} c_{ik} = 0$ при $j \neq k$, $\sum_{i=1}^n c_{ij}^2 = 1$. Положим

$$Z_i = \sum_{j=1}^n c_{ij} Y_j, \quad i = 1, \dots, n. \quad (5)$$

Тогда сл. в. Z_i , $i=1, \dots, n$, являются независимыми, $N(0, 1)$ -распределенными. При этом

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n Y_i^2. \quad (6)$$

Доказательство. Совместная плотность сл. в. Y_i , $i=1, \dots, n$, равна

$$f_Y(y_1, \dots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n y_i^2\right). \quad (7)$$

Совместная ф. р. сл. в. Z_i , $i=1, \dots, n$,

$$\begin{aligned} F_{Z(z_1, \dots, z_n)} &= P(Z_i \leq z_i, i=1, \dots, n) = \\ &= P\left(\sum_{i=1}^n c_{ij} Y_j \leq z_i, i=1, \dots, n\right) \end{aligned}$$

равна интегралу от плотности (7) по области

$$\left\{y_1, \dots, y_n : \sum_{i=1}^n c_{ij} y_j \leq z_i, i=1, \dots, n\right\}.$$

Произведем в этом интеграле замену переменных по формулам

$$v_i = \sum_{j=1}^n c_{ij} y_j, \quad i=1, \dots, n. \quad (8)$$

Так как матрица C ортогональна, то из (8) вытекает

$$\sum_{i=1}^n v_i^2 = \sum_{j=1}^n y_j^2. \quad (9)$$

[Ортогональное преобразование сохраняет длины векторов:

$$\begin{aligned} \sum_{i=1}^n v_i^2 &= \sum_{i=1}^n \left(\sum_{j=1}^n c_{ij} y_j\right)^2 = \sum_{i=1}^n \sum_{j,k=1}^n c_{ij} c_{ik} y_j y_k = \\ &= \sum_{j,k=1}^n y_j y_k \sum_{i=1}^n c_{ij} c_{ik} = \sum_{j=1}^n y_j^2. \end{aligned}$$

Следовательно, подынтегральная функция (7) в новых переменных сохраняет прежний вид. Якобиан перехода от переменных y_i к v_i ввиду ортогональности матрицы $[c_{ij}]$ равен 1, так что

$$F_{Z(z_1, \dots, z_n)} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{v_1 \leq z_1, i=1, \dots, n} \dots \int \exp\left(-\frac{1}{2} \sum_{i=1}^n v_i^2\right) dv_1 \dots dv_n.$$

что доказывает первое утверждение леммы 1. Равенство (6) вытекает из (8), (9). ■

Выберем в условиях леммы 1 ортогональную $n \times n$ -матрицу $C = [c_{ij}]$ так, чтобы $c_{nj} = 1/\sqrt{n}$, $j = 1, \dots, n$. Значения c_{ij} при $i \neq n$ нам не важны, лишь бы матрица C была ортогональной (например, легко проверить, что годятся c_{ij} в форме

$$c_{ij} = 1/\sqrt{i(i+1)}, \quad j \leq i < n; \quad c_{i,i+1} = -i/\sqrt{i(i+1)}; \quad c_{ii} = 0,$$

$i+1 < j \leq n$). Тогда

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{\sqrt{n}} Z_n, \quad \sum_{i=1}^n Y_i^2 = \sum_{i=1}^n Z_i^2,$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = \sum_{i=1}^n Z_i^2 - Z_n^2 = \sum_{i=1}^{n-1} Z_i^2.$$

Учитывая (3), (4) и определения χ^2_n и t_{n-1} (см. § 7), приходим к следующему утверждению.

Лемма 2. Пусть X_i , $i = 1, \dots, n$, независимы и $N(\mu, \sigma^2)$ -распределены. Тогда

$$\sqrt{n}(\bar{X} - \mu)/\sigma = \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \mu) \sim \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

независимы и распределены соответственно $N(0, 1)$ и χ^2_{n-1} , а отношение

$$\sqrt{n}(\bar{X} - \mu)/S$$

имеет распределение t_{n-1} . ■.

Из леммы 2 получаем α -доверительный интервал для σ^2 :

$$(n-1)S^2/x_{1-\alpha/2}(\chi^2_{n-1}) \leq \sigma^2 \leq (n-1)S^2/x_{\alpha/2}(\chi^2_{n-1}), \quad (10)$$

где $x_\alpha(\chi^2_{n-1})$ — α -квантиль χ^2_{n-1} -распределения, и α -доверительный интервал для μ :

$$\bar{X} - x_{1-\alpha/2}(t_{n-1})S/\sqrt{n} \leq \mu \leq \bar{X} - x_{\alpha/2}(t_{n-1})S/\sqrt{n}, \quad (11)$$

где $x_\alpha(t_{n-1})$ — α -квантиль t_{n-1} -распределения. Если в (10) или (11) оставить лишь одно неравенство, то получим односторонние $\alpha/2$ -доверительные границы, верхние или нижние.

4. Сравнение дисперсии в двух выборках.

Одно из важных приложений изложенной выше теории относится к так называемой задаче о двух выборках. Допустим, имеются две независимые нормальные выборки:

$$X_{k1}, X_{k2}, \dots, X_{k_k}, \quad k = 1, 2. \quad (12)$$

Обозначим через μ_k и σ_k^2 , $k=1, 2$, средние и дисперсии сл. в. (12). Во многих экспериментальных исследованиях возникает вопрос о сравнении по результатам испытаний (12) значений σ_1^2 и σ_2^2 , μ_1 и μ_2 . Это сравнение естественно проводить, опираясь на соответствующие выборочные характеристики

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ki}, \quad S_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k-1} (X_{ki} - \bar{X}_k)^2. \quad (13)$$

Отношение

$$V = S_1^2 / S_2^2$$

выборочных дисперсий имеет распределение, зависящее лишь от отношения $\lambda = \sigma_1^2 / \sigma_2^2$, но не от μ_1 и μ_2 . Таким образом, если речь идет о сравнении σ_1^2 и σ_2^2 , а параметры μ_1 и μ_2 не представляют интереса и являются, как говорят, *мешающими параметрами*, то переход к статистике V полностью устраниет влияние этих мешающих параметров. Величина $\lambda^{-1}V$ имеет в силу леммы 2 распределение F_{n_1-1, n_2-1} , откуда для отношения λ неизвестных дисперсий можно написать α -доверительный интервал

$$V/x_{1-\alpha/2}(F_{n_1-1, n_2-1}) \leq \lambda \leq V/x_{\alpha/2}(F_{n_1-1, n_2-1}). \quad (14)$$

Если здесь оставить лишь одно неравенство, то получается односторонний $\alpha/2$ -доверительный интервал.

Особенно часто в приложениях возникает вопрос, верно ли, что $\sigma_1^2 = \sigma_2^2$, т. е. $\lambda = 1$. Говорят в этом случае, что необходимо проверить гипотезу $H_0 = \{\lambda = 1\}$. Наиболее широкая альтернатива к H_0 — это противоположное допущение, что $\lambda \neq 1$, однако значительный интерес представляют и так называемые односторонние альтернативы $\lambda > 1$ или $\lambda < 1$ (случай $\lambda < 1$ сводится к $\lambda > 1$ переменой местами σ_1^2 и σ_2^2). Например, может быть известно, что точность некоторого усовершенствованного измерительного прибора или некоторой методики по крайней мере не хуже прежней, и задача состоит в том, чтобы по результатам наблюдений (12) принять или отклонить гипотезу $\lambda = 1$ против альтернативы $\lambda > 1$. Наличие доверительного интервала (14) наводит на мысль сформулировать следующий критерий: подставим в (14) $\lambda = 1$, и если результаты наблюдений

$$x_{k1}, x_{k2}, \dots, x_{kn_k}, \quad k = 1, 2, \quad (15)$$

приводят к верному неравенству

$$v/x_{1-\alpha/2} \leq 1 \leq v/x_{\alpha/2}, \quad (16)$$

где

$$v = S_1^2 / S_2^2, \quad S_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k-1} (x_{ki} - \bar{x}_k)^2,$$

$$\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki}, \quad k = 1, 2, \quad (17)$$

то считать гипотезу $\lambda=1$ против двусторонней альтернативы $\lambda \neq 1$ верной; в противном случае гипотезу отвергнуть.

Как мы уже не раз отмечали, статистические выводы о вероятностных моделях обладают принципиальной неопределенностью. Так, указывая α -доверительный интервал (14) для λ , следует помнить, что утверждение (14) о границах, в которых лежит неизвестное λ , может оказаться верным при одной реализации опыта (12) и неверным при другой. Увеличивая в (14) значение α , мы тем самым уменьшаем допуск для неизвестного λ , но за это расплачиваемся уменьшением коэффициента доверия $1-\alpha$ к указываемым границам, т. е. уменьшаем долю верных заключений при многократном использовании данной процедуры. Это замечание равным образом относится и к задаче проверки статистической гипотезы $\lambda=1$. Если предположить, что на самом деле гипотеза $\lambda=1$ верна, то тем не менее остается вероятность α , что выборочная точка (15) попадает за пределы области (16) и тем самым верная гипотеза будет отвергнута. Значение α называется *уровнем значимости* критерия. Обычно критерий задают не областью (16), принятия гипотезы, а дополнительной областью, которую называют *критической областью*. Для рассматриваемого критерия эта область получается объединением выборочных точек (15), удовлетворяющих одному из неравенств:

$$v < x_{\alpha/2} \text{ или } v > x_{1-\alpha/2}. \quad (18)$$

Совокупность критических множеств (18), зависящая от параметра α , собственно и определяет данный критерий. Иначе критерий (18) можно задать с помощью статистики критерия $v = -s_1^2/s_2^2$ и указания того, какие значения критерия приводят к отвержению гипотезы. Введенный критерий называют *F-критерием*. ■

5. Задача сравнения средних μ_1 и μ_2 .

В начале нашего века англичанин Госсет, работая на фирме, занимавшейся пивоварением, столкнулся со следующей проблемой. Было необходимо сравнить теоретические средние μ_i в некотором числе $k \geq 2$ малых выборок. Затруднения заключались в малости выборок, ибо в противном случае можно было рассчитывать на приближенное равенство $s_i^2 \approx \sigma_i^2$ и полагать \bar{x}_i/s_i , распределенными приближенно нормально со средними μ_i и дисперсиями $1/n_i$ — такая техника работы в те времена широко применялась. Госсет нашел точное распределение \bar{x}_i/s_i , что совершило, по общему признанию, переворот в статистике. Идея избавиться от мешающих параметров в последующем привела Фишера к созданию дисперсионного анализа, простейшим примером которого является

изложенная выше методика сравнения дисперсий с помощью F -критерия.

Рассмотрим задачу сравнения μ_1 и μ_2 в двух независимых нормальных выборках в предположении, что дисперсии одинаковые: $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Статистика $\bar{X}_1 - \bar{X}_2$ в таком случае имеет нормальное распределение со средним $\mu_1 - \mu_2$ и дисперсией $\sigma^2/n_1 + \sigma^2/n_2 = \sigma^2(1/n_1 + 1/n_2)$ (см. § 7, п. 3), зависящей от неизвестного σ . Для оценивания σ^2 можно было бы воспользоваться любой из статистик S_k^2 , $k=1, 2$. Ясно, однако, что более точной будет оценка, использующая всю выборку. Именно ввиду леммы 2 и свойств распределения χ^2 (§ 7, п. 4) величина

$$(n_1-1)S_1^2 + (n_2-1)S_2^2)/\sigma^2$$

имеет $\chi_{n_1+n_2-2}^2$ -распределение и, следовательно, статистика

$$S^2 = ((n_1-1)S_1^2 + (n_2-1)S_2^2)/(n_1+n_2-2) \quad (19)$$

является несмещенной оценкой σ^2 , построенной по полной выборке. Вспоминая, что дисперсия χ_n^2 -распределения равна $2n$, получаем, что $DS^2 = (n_1+n_2-1)^{-1}\sigma^2$, в то время как $DS_i^2 = (n_i-1)^{-1}\sigma^2$.

Таким образом, отношение

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S \sqrt{1/n_1 + 1/n_2}} \quad (20)$$

имеет распределение $t_{n_1+n_2-2}$. Отсюда получаем α -доверительный интервал для разности средних (ср. (11))

$$|(\mu_1 - \mu_2) - (\bar{X}_1 - \bar{X}_2)| \leq x_{1-\alpha/2}(t_{n_1+n_2-2}) S \sqrt{1/n_1 + 1/n_2}.$$

Обычным образом строятся односторонние доверительные границы. Статистика критерия Стьюдента*) для проверки гипотезы $H_0 = \{\mu_1 = \mu_2\}$ имеет вид

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{1/n_1 + 1/n_2}},$$

где $s^2 = ((n_1-1)s_1^2 + (n_2-1)s_2^2)/(n_1+n_2-2)$. Если альтернативой к H_0 служит $\mu_1 > \mu_2$, то в критическое множество зачисляются все выборочные точки (15), для которых статистика t принимает значения, большие некоторого заданного:

$$\{x_{11}, x_{12}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2} : t > c\}. \quad (21)$$

Выбор числа c определяет *уровень значимости*, или, как еще говорят, *размер критической области*. Полагая $c = x_\alpha(t_{n_1+n_2-2})$, получаем одностороннюю критическую область размера α . При

* Условия контракта, по которому Госсст работал с фирмой по пивоварению, не давали ему свободы публикации, и его знаменитая работа была опубликована под скромным псевдонимом Student.

альтернативе $\mu_1 \neq \mu_2$ критическая область размера α получается объединением двух областей $t > c_1$ и $t < c_2$.

Дисперсионный, или F -критерий, (18) и t -критерий, или критерий Стьюдента, (21) относят к числу так называемых *критериев значимости* (их еще иначе называют *критериями согласия*). Задача проверки гипотезы ставится в форме поиска функции от результатов наблюдений — статистики критерия, — которая служила бы подходящей мерой согласия наблюденных значений со статистической моделью. По идеи в зависимости от задачи слишком большие или слишком маленькие, или и те и другие значения статистики критерия должны рассматриваться как противоречащие нулевой гипотезе H_0 . В предположении, что гипотеза H_0 верна, эти значения могут быть наблюдены с вероятностью α — уровнем значимости нашего вывода. Итак, отвергая гипотезу H_0 , мы должны считаться с возможностью, что на самом деле гипотеза верна и тем самым сделан ошибочный вывод. Значение α является вероятностью этой ошибки. Разумеется, есть и другая возможность ошибиться — принять гипотезу H_0 , когда на самом деле она неверна. К этому вопросу мы вернемся в гл. IV.

6. Пример [5].

В следующей таблице приведены результаты двух опытов над мухами. Мухи подвергались действию яда в течение 30 с в первом опыте и 60 с — во втором. Измерялось время от момента соприкосновения мухи с ядом до момента, называемого временем реакции, когда вызываемый ядом паралич распространится настолько, что муха уже не может стоять и падает. Количество мух

Таблица 4

Времена реакции y_{ki} , $k = 1, 2$, двух групп мух на действие яда,
 $x_{ki} = \ln y_{ki}$, $v_{ki} = \Phi^{-1}((i - 0,5)/n_k)$, $n_1 = 15$, $n_2 = 16$

y_{1i}	$p_{1i} \times 100\%$	v_{1i}	x_{1i}	y_{2i}	$p_{2i} \times 100\%$	v_{2i}	x_{2i}
3	3,1	-1,866	0,477	2	3,3	-1,838	0,301
5	9,4	-1,316	0,699	5	10,0	-1,281	0,699
5	15,6	-0,911	0,699	5	16,7	-0,966	0,699
7	21,9	-0,775	0,845	7	23,3	-0,729	0,845
9	28,1	-0,579	0,954	8	30,0	-0,524	0,903
9	34,4	-0,401	0,954	9	36,7	-0,339	0,954
10	40,6	-0,237	1,000	14	43,3	-0,168	1,146
12	46,9	-0,077	1,079	18	50,0	0,000	1,255
20	53,1	0,077	1,301	24	56,7	0,168	1,380
24	59,4	0,237	1,380	26	63,3	0,339	1,415
24	65,6	0,401	1,380	26	70,0	0,524	1,415
34	71,9	0,579	1,532	34	76,7	0,729	1,532
43	78,1	0,775	1,634	37	83,3	0,966	1,568
46	84,4	0,911	1,663	42	90,0	1,281	1,623
58	90,6	1,316	1,763	90	96,7	1,838	1,954
140	96,9	1,866	2,146				

в первом опыте $n_1=16$, а во втором $n_2=15$. В табл. 4 приведены значения y_{ik} , $k=1, 2$, $i=1, 2, \dots, n_k$, — времена реакции в минутах в порядке возрастания (порядковые статистики), а также значения $p_{ki}=(i-0,5)/n_k$ и $v_{ki}=\Phi^{-1}(p_{ki})$, $k=1, 2$, $i=1, 2, \dots, n_k$, с целью проверки данных на нормальность. Точки (y_{ii}, v_{ii}) отмечены на рис. 8. Как видно, зависимость v от y далека от линейной

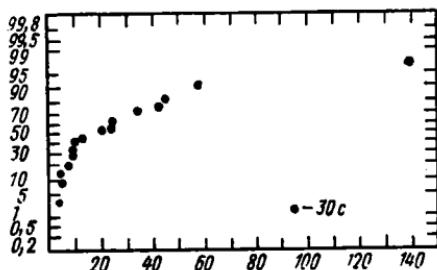


Рис. 8. Времена реакции на действие яда группы из 16 мух на нормальной бумаге

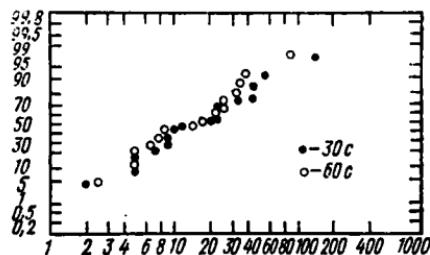


Рис. 9. Логарифм времен реакции на действие яда двух групп из 15 и 16 мух на нормальной бумаге

и похоже, что эти точки хорошо укладываются на логарифмическую кривую вида $v=a+b \ln y$. Значения $x_{ki}=\ln y_{ki}$, $k=1, 2$, $i=1, 2, \dots, n_k$, также приведены в табл. 4, а на рис. 9 нанесены точки $(x_{ki}, \Phi^{-1}(p_{ki}))$, $i=1, 2, \dots, n_k$. Графический анализ показывает, что наблюденные значения логарифма времен реакций, которые мы по условиям эксперимента считаем независимыми и одинаково распределенными в пределах каждой выборки, могут рассматриваться как наблюдения над нормальными сл. в. Основываясь на этих допущениях, применим изложенную выше теорию для сравнения этих выборок с целью выяснения влияния разницы в длительности воздействия ядом на время реакции. Выборочные средние и дисперсии равны

$$\bar{x}_1=1,219; \bar{x}_2=1,179; s_1^2=0,2080; s_2^2=0,1930.$$

Начнем с вопроса о значимости различия выборочных дисперсий. Наблюденное значение статистики F -критерия равно

$$v_{\text{набл}} = s_1^2/s_2^2 = 1,08 \approx x_{0,5} (F_{15,14}) = 1,0033,$$

т. е. гипотеза $\sigma_1^2=\sigma_2^2$ при любом α , чуть меньшем 0,5, должна быть принята. Итак, согласие наблюденных значений с предположением $\sigma_1^2=\sigma_2^2$ очень хорошее, и мы принимаем эту гипотезу.

Считая $\sigma_1^2=\sigma_2^2$, проверим с помощью t -критерия гипотезу о равенстве средних. Имеем

$$\bar{x}_1-\bar{x}_2=0,040; \quad s^2=(15s_1^2+14s_2^2)/29=0,2008;$$

$$s \sqrt{1/16 + 1/15} = 0,02594.$$

Наблюденное значение статистики t -критерия равно

$$t = (\bar{x}_1 - \bar{x}_2) / (s \sqrt{1/n_1 + 1/n_2}) = 0,040 / \sqrt{0,02594} = \\ = 0,040 / 0,161 = 0,25 \approx x_{0.6}(t_{29}) = 0,2557,$$

и, следовательно, гипотеза $\mu_1 = \mu_2$ принимается на уровне значимости $\alpha < 0,4$, что также свидетельствует о хорошем согласии данных с гипотезой.

ГЛАВА II

ЛИНЕЙНАЯ СТАТИСТИЧЕСКАЯ МОДЕЛЬ

§ 9. ОЦЕНИВАНИЕ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ МОДЕЛИ

1. Введение.

Модели, которые рассматривались до сих пор, включали предположение о независимости и одинаковой распределенности наблюдений. Для построения оценок функционалов от теоретической ф. р. или ее параметров, имеющих вероятностный смысл, мы обращались к э. ф. р. Если для оценивания одной и той же величины оказывалось возможным предложить более одной статистики, то предпочтение отдавалось той, распределение которой теснее сконцентрировано вокруг оцениваемой величины. Если оценка несмещенная, то мерой рассеивания служила ее дисперсия. Сравнение несмешенных оценок по величине их дисперсий является одним из наиболее часто используемых в математической статистике критериев. В некоторых статистических моделях оказывается возможным поставить и решить задачу нахождения наилучших по данному критерию оценок — оценок, имеющих минимальную дисперсию в классе всех несмешенных. В этом и следующем параграфах рассматривается более простая задача отыскания наилучших несмешенных оценок в классе линейных функций от наблюдений или от их порядковых статистик.

2. Примеры линейных моделей.

Пример 1. Пусть

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$$

независимая выборка из $N(\mu, \sigma^2)$ с неизвестными μ и σ . Представим наблюдения в форме

$$Y_i = \mu + \sigma \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

где ε_i — независимые $N(0, 1)$ -распределенные сл. в. Вводя n -мерные векторы-строки $\mathbf{1} = (1, 1, \dots, 1)$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$, запишем (1) в векторной форме

$$\mathbf{Y} = \mu \cdot \mathbf{1} + \sigma \boldsymbol{\varepsilon}. \quad (2)$$

Поставим задачу отыскания линейной несмешенной оценки пара-

метра μ с минимальной дисперсией. Пусть $\mathbf{b} = (b_1, \dots, b_n)$ — числовой вектор, такой, что статистика

$$\mathbf{b}\mathbf{Y}' = b_1 Y_1 + \dots + b_n Y_n \quad (3)$$

несмешенно оценивает μ (знак транспонирования ' означает переход к вектору-столбцу):

$$\mu = \mathbf{M}(\mathbf{b}\mathbf{Y}') = b_1 \mathbf{M}X_1 + \dots + b_n \mathbf{M}X_n = (b_1 + \dots + b_n) \mu.$$

Отсюда следует, что вектор \mathbf{b} должен удовлетворять условию

$$b_1 + \dots + b_n = 1.$$

Дисперсия оценки (3) равна

$$\mathbf{D}(\mathbf{b}\mathbf{Y}') = b_1^2 \mathbf{D}Y_1 + \dots + b_n^2 \mathbf{D}Y_n = (b_1^2 + \dots + b_n^2) \sigma^2 = \|\mathbf{b}\|^2 \sigma^2.$$

Таким образом, оценка (3) будет наилучшей, если вектор \mathbf{b} имеет наименьшую возможную длину, а сумма его координат равна 1. Хорошо известно, что таким вектором является $\mathbf{b} = n^{-1} \mathbf{1}$, т. е. наилучшая линейная несмешенная оценка μ совпадает с выборочным средним

$$\hat{\mu} = \frac{1}{n} \mathbf{1} \mathbf{Y}' = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}. \quad (4)$$

Подчеркнем следующее важное обстоятельство: при решении задачи отыскания наилучшей линейной несмешенной оценки были использованы лишь свойства моментов сл. вектора \mathbf{e} :

$$\mathbf{M}\mathbf{e}_i = 0, \quad i = 1, \dots, n, \quad [\mathbf{M}\mathbf{e}_i \mathbf{e}_j, \quad i, j = 1, \dots, n] = I, \quad (5)$$

где I — единичная матрица. Если выполнены соотношения (5), то рассуждения, приводящие к оценке (4), сохраняют свою силу и когда сл. в. $\mathbf{e}_i, \quad i = 1, \dots, n$, имеют произвольное совместное распределение вероятностей.

Пример 2 [6]. В следующей таблице приведены результаты $n=15$ измерений влажности некоторого материала, проведенных для градуирования прибора. Значения x_i — влажность в процентах, измеренная некоторым методом достаточно точно, так что можно сказать, что ошибка отсутствует, y_i — показания прибора для материала с влажностью x_i .

x_i	6,0	6,3	6,5	6,8	7,0	7,1	7,5	7,5	7,6	7,8	8,0	8,2	8,4	8,4	8,9
y_i	39	58	49	53	80	86	115	124	104	131	147	160	156	172	180

Предположим, что прибор осуществляет линейную связь между измеряемой величиной x и показанием y , но что каждое измерение подвержено случайной ошибке, а ошибки от опыта к опыту не-

коррелированы и имеют одинаковые дисперсии σ^2 . Тогда для описания рассматриваемого опыта можно предложить модель

$$Y_i = \theta_1 + \theta_2 x_i + \sigma e_i, \quad i = 1, \dots, n,$$

или

$$\mathbf{Y} = \theta_1 \mathbf{1} + \theta_2 \mathbf{x} + \sigma \mathbf{e},$$

где сл. вектор e подчиняется условию (5), σ , θ_1 и θ_2 — неизвестные параметры, $x = (x_1, x_2, \dots, x_n)$. Пусть $a = (a_1, a_2)$ — произвольный числовой вектор. Будем искать линейную оценку

$$\mathbf{b} \mathbf{Y}' = b_1 Y_1 + \dots + b_n Y_n,$$

несмешенно оценивающую линейную комбинацию неизвестных параметров

$$a_1 \theta_1 + a_2 \theta_2 = a \theta'.$$

В частных случаях $a = (1, 0)$ и $a = (0, 1)$ получим соответствующие оценки для θ_1 и θ_2 . По условию несмешенности

$$\begin{aligned} \mathbf{M} \mathbf{b} \mathbf{Y}' &= \sum_{i=1}^n b_i \mathbf{M} Y_i = \sum_{i=1}^n b_i (\theta_1 + \theta_2 x_i) = \\ &= \theta_1 \sum_{i=1}^n b_i + \theta_2 \sum_{i=1}^n b_i x_i = \theta_1 a_1 + \theta_2 a_2, \end{aligned}$$

каковы бы ни были θ_1 и θ_2 . В таком случае

$$\sum_{i=1}^n b_i = a_1, \quad \sum_{i=1}^n b_i x_i = a_2$$

или

$$(\mathbf{b} \mathbf{1}', \mathbf{b} \mathbf{x}') = \mathbf{b} [\mathbf{1}', \mathbf{x}] = (a_1, a_2). \quad (6)$$

Дисперсия оценки $\mathbf{b} \mathbf{Y}'$ равна

$$\mathbf{D} \mathbf{b} \mathbf{Y}' = \sum_{i=1}^n b_i^2 \mathbf{D} Y_i = \sum_{i=1}^n b_i^2 \sigma^2 = \sigma^2 \| \mathbf{b} \|^2. \quad (7)$$

Таким образом, среди всех векторов, удовлетворяющих (6), надо найти тот, который имеет наименьшую длину. Заметим, что если \mathbf{b} — любой вектор, удовлетворяющий (6), то его проекция на линейное подпространство V , порожденное векторами 1 , \mathbf{x} , очевидно, также удовлетворяет (6), а длина проекции не превосходит длины вектора. Поэтому при поиске вектора наименьшей длины, удовлетворяющего (6), мы ничего не потеряем, если ограничимся лишь теми векторами \mathbf{b} , которые лежат в подпространстве V :

$$\mathbf{b} = c_1 \mathbf{1} + c_2 \mathbf{x}. \quad (8)$$

Подставляя (8) в условие несмещенності (6), получаем

$$\begin{aligned} c_1 \mathbf{1}' + c_2 \mathbf{x}' &= a_1, \\ c_1 \mathbf{x}' + c_2 \mathbf{x} \mathbf{x}' &= a_2. \end{aligned} \quad (9)$$

Определитель системы (9) равен

$$n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = n \sum_{i=1}^n (x_i - \bar{x})^2$$

и отличен от нуля, если только $\mathbf{x} \neq \mathbf{1}$, и в этом случае существует единственный вектор \mathbf{b} в форме (8), дающий несмешенную оценку, а следовательно, соответствующая оценка $\mathbf{b} \mathbf{Y}'$ имеет минимальную дисперсию.

Вычисляем матрицу, обратную к матрице системы (9):

$$\left(n \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-1} \times \begin{bmatrix} \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \\ - \sum_{i=1}^n x_i & n \end{bmatrix} \equiv H,$$

так что решение системы (9) представимо в виде

$$(c_1, c_2) = (a_1, a_2) H.$$

Находим оценки параметров θ_1, θ_2 (см. (8)):

$$\begin{aligned} \widehat{\theta}_1 &= \mathbf{b} \mathbf{Y}' = (c_1, c_2) [\mathbf{1}', \mathbf{x}']' \mathbf{Y}' = (a_1, a_2) H [\mathbf{1}', \mathbf{x}']' \mathbf{Y}' = (1, 0) H [\mathbf{1}', \mathbf{x}']' \mathbf{Y}', \\ \widehat{\theta}_2 &= (0, 1) H [\mathbf{1}', \mathbf{x}']' \mathbf{Y}'. \end{aligned}$$

Проведя элементарные выкладки, найдем окончательные формулы для $\widehat{\theta}_1, \widehat{\theta}_2$:

$$\begin{aligned} [\mathbf{1}', \mathbf{x}']' \mathbf{Y}' &= \left(\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i x_i \right)', \\ H [\mathbf{1}', \mathbf{x}']' \mathbf{Y}' &= \left(n \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-1} \times \\ &\times \left(\sum_{i=1}^n x_i^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i + n \sum_{i=1}^n Y_i x_i \right)', \\ \theta_2 &= \sum_{i=1}^n Y_i (x_i - \bar{x}) / \sum_{i=1}^n (x_i - \bar{x})^2, \end{aligned}$$

$$\theta_1 = \left(n \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n Y_i + n\bar{x}^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n x_i \sum_{i=1}^n Y_i x_i \right) =$$

$$= \bar{Y} + \bar{x} \left(n\bar{x} \sum_{i=1}^n Y_i - n \sum_{i=1}^n Y_i x_i \right) \left(n \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-1} = \bar{Y} - \hat{\theta}_2 \bar{x}.$$

Для числовых данных нашего примера получаем

$$\bar{x} = 7,466; \quad \bar{y} = 100,266;$$

$$\sum_{i=1}^n x_i^2 = 846,260; \quad \bar{x}^2 = 55,741;$$

$$\sum_{i=1}^n x_i^2 - n\bar{x}^2 = 10,154;$$

$$\sum_{i=1}^n y_i x_i = 12898,8;$$

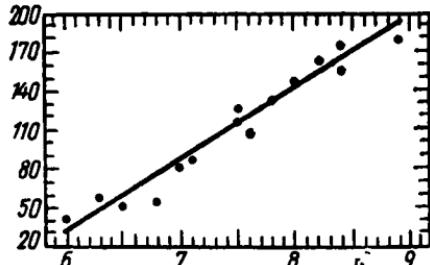


Рис. 10. Градиуровка прибора для измерения влажности

$$\sum_{i=1}^n y_i x_i - \bar{x} \sum_{i=1}^n y_i = 550,11; \quad \hat{\theta}_2 = 54,176; \quad \hat{\theta}_1 = \bar{y} - \hat{\theta}_2 \bar{x} = -294,21.$$

Пары (x_i, y_i) и прямая $y = \hat{\theta}_1 + \hat{\theta}_2 x$ изображены на рис. 10.

3. Линейная статистическая модель.

Определим линейную статистическую модель, полагая, что вектор наблюдений $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ имеет следующую структуру:

$$\mathbf{Y}' = \mathbf{X}\boldsymbol{\theta}' + \sigma\boldsymbol{\varepsilon}', \text{ или } \mathbf{Y} = \mathbf{X}'\boldsymbol{\theta} + \sigma\boldsymbol{\varepsilon}, \quad (10)$$

где $\mathbf{X} = [x_{ij}, i=1, \dots, n; j=1, \dots, r]$, $r \leq n$, — известная числовая матрица, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ — вектор неизвестных коэффициентов, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ — случайный вектор, такой, что

$$\mathbf{M}\boldsymbol{\varepsilon} = 0, \quad i=1, \dots, n, \quad [\mathbf{M}\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T, i, j=1, \dots, n] = I, \quad (11)$$

$\sigma > 0$ — неизвестный параметр, I — единичная матрица.

Из (10) получаем, что вектор математических ожиданий наблюдений равен

$$\boldsymbol{\eta} = (\mathbf{M}Y_1, \dots, \mathbf{M}Y_n) = \mathbf{X}'\boldsymbol{\theta},$$

а матрица ковариаций

$$\mathbf{R}_{\mathbf{Y}} = [\text{cov}(Y_i, Y_j), i, j=1, \dots, n] = \mathbf{R}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T} = \sigma^2 I.$$

Введем векторы

$$\mathbf{x}_j = (x_{1j}, \dots, x_{nj}), \quad j=1, \dots, r,$$

так что

$$X = [x_1', x_2', \dots, x_r'],$$

и (10) можно записать в виде

$$Y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_r x_r + \sigma \varepsilon. \quad (12)$$

Свое название модель (12) получила за то, что неизвестный вектор среднего

$$\eta = \theta_1 x_1 + \dots + \theta_r x_r$$

лежит в линейном подпространстве $V \subset R^n$, порожденном векторами x_1, \dots, x_r . Модель (12) можно задать в непараметризованном виде:

$$Y = \eta + \sigma \varepsilon, \eta \in V, \quad (13)$$

где V — заданное каким-либо способом линейное подпространство. Выбирая ту или иную систему векторов, порождающих V , приходим к форме (12). Нередко бывает полезно от уже заданной в форме (12) модели перейти к другой, выбирая иначе систему порождающих векторов.

4. Несмешенные оценки с минимальной дисперсией: матрица полного ранга.

Пусть

$$bY' = b_1 Y_1 + \dots + b_n Y_n$$

— несмешенная оценка для

$$a\theta' = a_1 \theta_1 + \dots + a_r \theta_r,$$

где $a = (a_1, \dots, a_r)$ — произвольный числовой вектор. Это значит, что математическое ожидание оценки

$$M(bY') = \sum_{i=1}^n b_i M Y_i = b(\theta_1 x_1' + \dots + \theta_r x_r') = \theta_1 b x_1' + \dots + \theta_r b x_r' \quad (14)$$

тождественно по $\theta_1, \dots, \theta_r$ совпадает с $a\theta'$, откуда получаем

$$b x_1' = a_1, \dots, b x_r' = a_r, \text{ или } b X = a. \quad (15)$$

Условие (15) является необходимым и достаточным, чтобы статистика bY' несмешенно оценивала $a\theta'$.

Дисперсия оценки bY' равна

$$D(bY') = \sum_{i=1}^n b_i^2 D Y_i = \sigma^2 \sum_{i=1}^n b_i^2 = \sigma^2 \|b\|^2.$$

Если b — произвольный вектор, подчиненный условию (15), то вектор проекции $\text{pr}_V b$, очевидно, также удовлетворяет соотношениям (15), т. е. статистика

$$(\text{pr}_V b) Y'$$

несмешенно оценивает величину $a\theta'$, и в то же время ее дисперсия с учетом теоремы Пифагора удовлетворяет неравенству

$$D((\text{пр}_{\mathbf{b}} \mathbf{Y}') = \sigma^2 \|\text{пр}_{\mathbf{b}} \mathbf{b}\|^2 \leq \sigma^2 \|\mathbf{b}\|^2 = D(\mathbf{b} \mathbf{Y}').$$

Отсюда следует, что в поисках линейной несмешенной оценки с минимальной дисперсией можно ограничиться векторами \mathbf{b} , лежащими в подпространстве V :

$$\mathbf{b} = c_1 \mathbf{x}_1 + \dots + c_r \mathbf{x}_r = \mathbf{c} \mathbf{X}' . \quad (16)$$

Подставляя (16) в (15), приходим к системе уравнений для определения \mathbf{c} (ср. (8), (9)):

$$\mathbf{c} \mathbf{X}' \mathbf{X} = \mathbf{a} . \quad (17)$$

Допустим, матрица X имеет полный ранг r . Тогда матрица $X'X$ невырождена, (17) имеет единственное решение при любом векторе \mathbf{a} :

$$\mathbf{c} = \mathbf{a} (X'X)^{-1},$$

и наилучшая линейная несмешенная оценка величины $a\theta'$ равна

$$\mathbf{b}' \mathbf{Y}' = \mathbf{c} \mathbf{X}' \mathbf{Y}' = \mathbf{a} (X'X)^{-1} X' \mathbf{Y}' . \quad (18)$$

Выбирая $\mathbf{a} = \delta_i = (\delta_{i1}, \delta_{i2}, \dots, \delta_{in})$, где $\delta_{ij} = 0$ при $i \neq j$ и $\delta_{ii} = 1$ (символ Кронекера), получим несмешенную линейную оценку для θ_i с минимальной дисперсией:

$$\hat{\theta}_i = \delta_i (X'X)^{-1} X' \mathbf{Y}' . \quad (19)$$

Организуя оценки $\hat{\theta}_i$ в вектор-строку $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$ и замечая, что $[\delta_1', \dots, \delta_n'] = I$, запишем вектор наилучших линейных несмешенных оценок:

$$\hat{\theta}' = (X'X)^{-1} X' \mathbf{Y}' . \quad (20)$$

Наилучшая несмешенная оценка (18) параметрической функции $a\theta'$ может быть записана в виде $a\hat{\theta}'$. Таким образом, доказана

Теорема 1. (Гаусс, Марков). Если матрица X линейной модели имеет полный ранг, то наилучшие линейные несмешенные оценки параметров θ_i даются формулой (20). Если $\mathbf{a} = (a_1, \dots, a_r)$ — произвольный числовой вектор, то линейная несмешенная оценка параметрической функции $a\theta'$ с минимальной дисперсией единственна и дается формулой $a\hat{\theta}'$.

5. Наилучшие оценки в случае матрицы неполного ранга.

Случай матрицы X неполного ранга возникает в моделях дисперсионного анализа. Векторы $\mathbf{x}_1, \dots, \mathbf{x}_r$ в этом случае линейно зависимы, и разложение вектора среднего $\mathbf{M}\mathbf{Y}$ по векторам системы $\mathbf{x}_1 \dots \mathbf{x}_r$ не единственны. Можно, однако, образовать такие линейные комбинации неизвестных коэффициентов θ_i , которые по

вектору MY восстанавливаются однозначно. Поясним на примере. Пусть $r=3$, x_1, x_2 линейно независимы, $x_3 = x_1 + x_2$. Тогда

$$\eta = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = (\theta_1 + \theta_3) x_1 + (\theta_2 + \theta_3) x_2.$$

Поскольку разложение вектора η по подсистеме x_1, x_2 единственны, величины $\theta_1 + \theta_3$ и $\theta_2 + \theta_3$ по вектору η восстанавливаются однозначно. Именно такого рода линейные комбинации неизвестных параметров и оцениваются в линейной модели с матрицей X неполного ранга.

Параметрическая функция $a\theta'$ называется *допускающей оценкой* (или *оцениваемой*), если ее значение восстанавливается по вектору $\eta = \theta_1 x_1 + \dots + \theta_r x_r$, т. е. она является линейной функцией от координат вектора η :

$$a\theta' = b\eta' = bX\theta'.$$

Отсюда следует, что выполняются соотношения (15), т. е. для допускающих оценку параметрических функций $a\theta'$ и только для них существуют линейные несмешанные оценки. Как и при выводе теоремы 1, при поиске наилучшей несмешанной оценки можно ограничиться такими линейными функциями bY' , что вектор $b \in V$, т. е. имеет место представление (16) и при этом выполнено условие несмешанности (15). Такой вектор b существует и притом только один. Действительно, если b_1 и b_2 удовлетворяют (15), то для их разности получаем

$$(b_1 - b_2)x_i' = 0, \quad i = 1, \dots, r. \quad (21)$$

Но вектора b_1, b_2 линейно выражаются через систему x_1, \dots, x_r , поэтому условие ортогональности (21) приводит к выводу $b_1 = b_2$.

Следовательно, взяв любое решение системы (17) и подставляя его в (16), мы найдем вектор b , для которого bY' является линейной несмешанной оценкой $a\theta'$ с минимальной дисперсией. Итак, получена следующая

Теорема 2 (Гаусс, Марков). Если матрица X линейной модели имеет неполный ранг, то линейные несмешанные оценки существуют только для таких параметрических функций $a\theta'$, что вектор a является линейной комбинацией строк матрицы X ; в этом случае существует единственная несмешанная оценка b^*Y' с минимальной дисперсией. Вектор b^* может быть получен проектированием на V вектора b , отвечающего произвольной несмешанной оценке, либо подстановкой любого решения системы (17) в формулу (16).

6. Пример: модель с матрицей неполного ранга.

В табл. 5 приведены годовые урожаи пшеницы (в центнерах на акр) на шести английских сельскохозяйственных станциях за три года [6]. Естественно, что годовой урожай на данной станции

есть величина, зависящая от многих неподдающихся учету обстоятельств. Табличные данные позволяют заняться исследованием вопроса о том, как зависел урожай от года и расположения станции. Предположим, что реализация факторов, связанных с условиями данного года i и данной местности j , при некоторых средних характеристиках всех других случайных факторов при-

Таблица 5
Годовые урожаи пшеницы (в центнерах на акр)
на шести английских сельскохозяйственных станциях

Год \ № станции	1	2	3	4	5	6
Год	1 (1933)	22,2	35,3	32,8	25,3	35,8
	2 (1934)	32,4	32,2	43,7	35,7	28,3
	3 (1935)	26,2	34,7	40,0	29,6	20,6
						47,2

водит к урожаю μ_{ij} — некоторому неизвестному числу, которое желательно оценить, а что наблюдается $y_{ij} = \mu_{ij} + \sigma e_{ij}$, где составляющая σe_{ij} представляет собой вклад в урожай неучитываемых в данной модели факторов. Переходя к случайным величинам, запишем

$$Y_{ij} = \mu_{ij} + \sigma e_{ij}, \quad i=1, \dots, p, \quad j=1, \dots, q, \quad p=3, \quad q=6.$$

Предполагая, что сл. в. e_{ij} некоррелированы, $M e_{ij} = 0$, $M e^2_{ij} = 1$, приходим к линейной статистической модели. В данной модели, не считая σ , имеется $pq = 3 \cdot 6 = 18$ неизвестных параметров. Число наблюдений также 18, и оценить μ_{ij} можно, только положив $\mu_{ij} = Y_{ij}$. Упростим модель, считая вклад в урожай, связанный с факторами года и местности аддитивным:

$$\mu_{ij} = \mu + \alpha_i + \beta_j, \quad i=1, \dots, p, \quad j=1, \dots, q,$$

где μ — некоторый «средний» по годам и участкам урожай, α_i , β_j — добавки (со знаком) в урожай, определяемые годом и местностью. Конечно, адекватность упрощенной модели (как, впрочем, и исходной) подлежит проверке, но данного статистического материала для этой цели недостаточно, и мы оставим этот вопрос в стороне. Образуя вектор \mathbf{Y} :

$$\mathbf{Y} = (Y_1, \dots, Y_p), \quad Y_i = (Y_{i1}, \dots, Y_{iq}), \quad i=1, \dots, p,$$

представим линейную модель в виде

$$\begin{pmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_p' \end{pmatrix} = \begin{pmatrix} 1_q & 1_q & 0_q & \dots & 0_q & I_q \\ 1_q & 0_q & 1_q & \dots & 0_q & I_q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1_q & 0_q & 0_q & \dots & 1_q & I_q \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_p \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} + \sigma \varepsilon',$$

или

$$Y' = X\theta' + \sigma \varepsilon',$$

где $\theta = (\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$, $X = [x_1', x_2', \dots, x_{p+q+1}']$ — $pq \times (p+q+1)$ -матрица линейной модели ранга $p+q-1$, $1_q = (1, \dots, 1)$, $0_q = (0, \dots, 0)$ — q -координатные векторы, I_q — единичная $q \times q$ -матрица.

Матрица X не имеет полного ранга, и параметры μ, α_i, β_j не являются оцениваемыми: они не могут быть линейно выражены через координаты вектора среднего:

$$\eta = (\mu + \alpha_1 + \beta_1, \dots, \mu + \alpha_1 + \beta_q, \mu + \alpha_2 + \beta_1, \dots, \mu + \alpha_2 + \beta_q, \dots, \mu + \alpha_p + \beta_1, \dots, \mu + \alpha_p + \beta_q).$$

Зато оценку допускают, например, разности $\alpha_i - \alpha_1, i=2, \dots, p$, а вектор $(\alpha_2 - \alpha_1, \alpha_3 - \alpha_1, \dots, \alpha_p - \alpha_1)$, очевидно, вполне характеризует влияние фактора года на урожай, что, собственно, и интересует экспериментатора. Поскольку $\alpha_2 - \alpha_1 = a\theta'$, где $a = (0, 1, -1, 0, \dots, 0)$, то наилучшая оценка для $a\theta'$ имеет вид bY' , где b — вектор из линейной оболочки столбцов матрицы X , такой, что $a = bX$. Легко проверить, что этим условиям удовлетворяет вектор

$$b = q^{-1}(-1_q, 1_q, 0_q, \dots, 0_q)$$

и, следовательно,

$$\hat{\alpha}_2 - \hat{\alpha}_1 = q^{-1} \sum_{j=1}^q (Y_{2j} - Y_{1j}) = q^{-1} \sum_{j=1}^q Y_{2j} - q^{-1} \sum_{j=1}^q Y_{1j}.$$

Аналогично оцениваются остальные $\hat{\alpha}_i - \hat{\alpha}_1, i=3, \dots, p$.

Для представленных в табл. 5 данных находим

i	1	2	3
$q^{-1} \sum_{j=1}^q y_{ij}$	28,4	34,6	33,0

Таким образом, $\hat{\alpha}_2 - \hat{\alpha}_1 = 6,2$; $\hat{\alpha}_3 - \hat{\alpha}_1 = 4,6$.

7. Наилучшие оценки как оценки наименьших квадратов.

К изложению теории несмешенного линейного оценивания имеется другой подход, основанный на следующем соображении двойственности. Пусть bY' — произвольная несмешенная оценка своего математического ожидания (см. (14)):

$$MbY' = bX\theta' = a\theta'.$$

Тогда наилучшая оценка, как мы знаем, имеет вид $\text{pr}_V bY'$. Обозначая через $R^n \ominus V$ ортогональное дополнение V в R^n , имеем

$$\begin{aligned} (\text{pr}_V b) Y' &= (\text{pr}_V b)(\text{pr}_V Y + \text{pr}_{R^n \ominus V} Y)' = \text{pr}_V b(\text{pr}_V Y)' = \\ &= (\text{pr}_V b + \text{pr}_{R^n \ominus V} b)(\text{pr}_V Y)' = b(\text{pr}_V Y)'. \end{aligned}$$

Таким образом, чтобы перейти от оценки bY' к соответствующей наилучшей, надо найти вектор $\text{pr}_V Y$. Условия проектирования выражаются соотношением

$$(Y - \text{pr}_V Y) X = 0. \quad (22)$$

Записывая разложение

$$\text{pr}_V Y = \hat{\theta}_1 x_1 + \dots + \hat{\theta}_r x_r = \hat{\theta} X' \quad (23)$$

и подставляя его в (22), получаем так называемую нормальную систему уравнений относительно $\hat{\theta}$:

$$\hat{\theta} X' X = Y X. \quad (24)$$

Если матрица X полного ранга, то от уравнения (24) приходим к ранее полученным формулам (20) для вектора наилучших оценок $\hat{\theta}$. В общем случае любое решение $\hat{\theta}$ системы (24) определяет по формулам (23) вектор $\text{pr}_V Y$ и, следовательно, наилучшую оценку $b(\text{pr}_V Y)'$ с математическим ожиданием $bX\theta' = a\theta'$.

Полезно перенести этот подход на модель в форме (13):

$$Y = \eta + \sigma \varepsilon, \quad \eta \in V.$$

Положим

$$\hat{\eta} = \text{pr}_V Y.$$

Тогда $\hat{\eta}$ несмешенно оценивает η и является наилучшей оценкой η в том смысле, что для любого вектора b дисперсия статистики $b\hat{\eta}'$, несмешенно оценивающей $b\eta'$, минимальна в классе линейных несмешенных оценок. Так как

$$D(b\hat{\eta}') = \sum_{i,j=1}^n b_i b_j \text{cov}(\hat{\eta}_i, \hat{\eta}_j) = b R_{\hat{\eta}} b',$$

то, используя частичную упорядоченность матриц

$$A \geqslant B \quad (A > B),$$

если $A-B$ — неотрицательно (положительно) определенная, заключаем, что в классе линейных несмещенных оценок вектора η имеется оценка (и она единственна) с наименьшей матрицей ковариаций. ■

Вектор проекции (23), как известно, дает решение следующей экстремальной задачи:

$$\begin{aligned} \|Y - \text{pr}_V Y\|^2 &= \min_{\theta_1, \dots, \theta_r} \|Y - (\theta_1 x_1 + \dots + \theta_r x_r)\|^2 = \\ &= \min_{\theta_1, \dots, \theta_r} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^r \theta_j x_{ij} \right)^2. \end{aligned} \quad (25)$$

Приравнивая к нулю производные от выражения под знаком \min в (25), легко получить отсюда нормальную систему уравнений (24). Метод отыскания неизвестных θ_j , $j=1, \dots, r$, из условия (25) называется *методом наименьших квадратов*. Определенные нами выше оценки $\hat{\theta}_j$, $j=1, \dots, r$, и $\hat{\theta}'$ могут быть также названы оценками метода наименьших квадратов, или м.н.к.-оценками. В такой форме эти оценки были использованы независимо Лежандром (1806) и Гауссом (1809), который впоследствии вывел оптимальные свойства этих оценок.

§ 10. КОВАРИАЦИИ, КАНОНИЧЕСКАЯ ФОРМА ЛИНЕЙНОЙ МОДЕЛИ, ОБОБЩЕНИЯ

1. Матрицы из случайных элементов.

Пусть

$$Z = [Z_{ij}, i=1, \dots, m, j=1, \dots, n]$$

— $m \times n$ -матрица из сл. в. Z_{ij} (при $m=1$ получаем вектор-строку, при $n=1$ — вектор-столбец; 1×1 -матрицу не будем отличать от ее элемента). Математическим ожиданием Z назовем матрицу

$$MZ = [MZ_{ij}, i=1, \dots, m, j=1, \dots, n],$$

составленную из математических ожиданий ее компонент (в предположении, что они существуют). Оператор математического ожидания, очевидно, перестановочен с операцией транспонирования:

$$(MZ)' = MZ',$$

математическое ожидание суммы матриц равно сумме математических ожиданий:

$$M(Z^{(1)} + Z^{(2)}) = MZ^{(1)} + MZ^{(2)},$$

а постоянные матричные множители выносятся за знак математического ожидания:

$$M(AZ) = AMZ, M(ZB) = (MZ)B.$$

Проверим последнее свойство:

$$\mathbf{M}(AZ) = \left[\mathbf{M} \sum_k a_{ik} Z_{kj} \right] = \left[\sum_k a_{ik} \mathbf{M} Z_{kj} \right] = A [\mathbf{M} Z_{kj}] = A \mathbf{M} Z.$$

2. Матрица ковариаций.

Пусть

$$\mathbf{U} = (U_1, U_2, \dots, U_n)$$

— случайный вектор. Используя представление

$$[(U_i - \mathbf{M} U_i)(U_j - \mathbf{M} U_j), i, j=1, \dots, n] = (\mathbf{U} - \mathbf{M} \mathbf{U})'(\mathbf{U} - \mathbf{M} \mathbf{U})$$

и приведенные выше правила вычисления математического ожидания, получаем для матрицы ковариаций вектора \mathbf{U} формулу

$$R_{\mathbf{U}} = \mathbf{M}((\mathbf{U} - \mathbf{M} \mathbf{U})'(\mathbf{U} - \mathbf{M} \mathbf{U})) = \mathbf{M} \mathbf{U}' \mathbf{U} - \mathbf{M} \mathbf{U}' \mathbf{M} \mathbf{U}.$$

Отметим некоторые свойства матрицы ковариаций. Очевидно, что если a — постоянный вектор, то

$$R_{\mathbf{U} + a} = R_{\mathbf{U}}.$$

При умножении случайного вектора на постоянную матрицу матрица ковариаций преобразуется по формуле

$$R_{\mathbf{U}A} = A' R_{\mathbf{U}} A. \quad (1)$$

Действительно,

$$\begin{aligned} R_{\mathbf{U}A} &= \mathbf{M}(A' \mathbf{U}' \mathbf{U} A) - \mathbf{M}(A' \mathbf{U}') \mathbf{M}(\mathbf{U} A) = A' (\mathbf{M} \mathbf{U}' \mathbf{U}) A - \\ &- A' (\mathbf{M} \mathbf{U}') (\mathbf{M} \mathbf{U}) A = A' (\mathbf{M} \mathbf{U}' \mathbf{U} - \mathbf{M} \mathbf{U}' \mathbf{M} \mathbf{U}) A = A' R_{\mathbf{U}} A. \blacksquare \end{aligned}$$

Взяв в (1) в качестве A матрицу-столбец a' , запишем для дисперсии сл. в. $\mathbf{U}a'$

$$\mathbf{D} \mathbf{U} a' = R_{\mathbf{U} a'} = a' R_{\mathbf{U}} a'. \quad (2)$$

Так как дисперсия неотрицательна, то из (2) вытекает, что $R_{\mathbf{U}}$ — неотрицательно определенная матрица.

Как известно из линейной алгебры, симметричную неотрицательно определенную матрицу можно привести к диагональной форме:

$$C' R_{\mathbf{U}} C = \Lambda \quad (3)$$

при помощи подходящей ортогональной $n \times n$ -матрицы

$$C : CC' = C'C = I.$$

При этом матрица Λ имеет на главной диагонали неотрицательные числа λ_i , $i=1, \dots, n$, а если матрица $R_{\mathbf{U}}$ положительно опре-

делена, то все $\lambda_i > 0$. Домножим (3) слева и справа на матрицу C и C' соответственно:

$$R_U = C \Lambda C' = (C \Lambda^{1/2})(C \Lambda^{1/2})',$$

где $\Lambda^{1/2}$ — диагональная матрица с элементами $\lambda_i^{1/2}$ на главной диагонали. Полагая $C \Lambda^{1/2} = B'$, имеем

$$R_U = B' B, \quad (4)$$

причем ранг матрицы B совпадает с рангом R_U . Представление (4) будет неоднократно использоваться в последующем.

3. Ковариации наилучших оценок.

Воспользуемся формулой (1), чтобы найти матрицу ковариаций вектора наилучших несмешанных оценок $\hat{\theta}$ в модели с матрицей X полного ранга. Учитывая, что $R_Y = R_{\sigma e} = \sigma^2 I$, получим

$$\hat{\theta} = YX(X'X)^{-1}, \quad R_{\hat{\theta}} = (X'X)^{-1}X'R_YX(X'X)^{-1} = \sigma^2(X'X)^{-1}. \quad (5)$$

Отметим, что если матрица

$$X = [x_1', \dots, x_r']$$

состоит из ортогональных векторов x_i , $i=1, \dots, r$, то матрицы $X'X$ и $(X'X)^{-1}$ диагональны и оценки $\hat{\theta}_i$, $i=1, \dots, r$, некоррелированы. Случай ортогональной матрицы X оказывается интересным и в некоторых других отношениях. Во-первых, нахождение $\hat{\theta}_i$ не требует обращения матрицы. Затем поскольку $\hat{\theta}_i$ находится в результате проектирования вектора наблюдений на направление x_i , то в случае, когда в процессе анализа модели решаем отбросить некоторые x_i , полагая, что истинные значения коэффициентов при них равны нулю, нет необходимости заново проводить вычисления оценок оставшихся коэффициентов, так как при этой процедуре они не изменяются.

4. Пример оптимального выбора матрицы модели.

Рассмотрим пример, когда экспериментатор не привязан жестко к данной линейной модели, а имеет некоторую свободу в проведении эксперимента и, следовательно, в выборе матрицы X . Допустим, что требуется найти веса θ_i , $i=1, \dots, r$, предметов [7]. Если каждый из предметов взвешивается k раз, то наилучшей несмешанной оценкой служит выборочное среднее. Дисперсия каждой из оценок равна σ^2/k , где σ^2 — дисперсия отдельного взвешивания, а всего производится $n=rk$ взвешиваний. По другому методу при каждом взвешивании выбирают две группы предметов и размещают на двух чашечках весов, а затем уравновешивают

вают гилями. Эксперимент может быть описан линейной моделью:

$$Y_i = \theta_1 x_{i1} + \dots + \theta_r x_{ir} + \sigma e_i, \quad i=1, \dots, n,$$

где $x_{ij} = -1, +1$ или 0 в зависимости от того, положен ли j -й предмет на левую, правую чашку весов или не участвует в i -м взвешивании; Y_i — величина уравновешивающих гирь со знаком, указывающим, на какую из двух чашек добавлен разновес. Предположим, что X имеет полный ранг, и найдем дисперсию оценки $\hat{\theta}_1$, для чего разобьем матрицу $X'X$ на блоки, выделив угловой элемент:

$$X'X = \begin{bmatrix} \|x_1\|^2 & b \\ b' & F \end{bmatrix},$$

где $b = (x_1 \cdot x_2', \dots, x_1 \cdot x_r')$, F — квадратная $(n-1) \times (n-1)$ -матрица, так что ввиду (5)

$$D\hat{\theta}_1 = \sigma^2 \det F / \det X'X. \quad (6)$$

Выразим определитель матрицы $X'X$ через компоненты разбиения. Составим линейную комбинацию последних $r-1$ столбцов $r \times r$ -матрицы $X'X$ с коэффициентами — компонентами вектора $F^{-1}b'$. В результате получаем матрицу-столбец:

$$\begin{bmatrix} b \\ F \end{bmatrix} \cdot F^{-1}b' = \begin{bmatrix} bF^{-1}b' \\ FF^{-1}b' \end{bmatrix} = \begin{bmatrix} bF^{-1}b' \\ b' \end{bmatrix}.$$

Вычитаем ее из первого столбца $X'X$. Так как определитель при таких преобразованиях не изменяется, то, разлагая преобразованный определитель по первой строке, получим

$$\det X'X = \det \begin{bmatrix} \|x_1\|^2 - bF^{-1}b' & b \\ 0 & F \end{bmatrix} = (\|x_1\|^2 - bF^{-1}b') \det F.$$

Подставляя полученное выражение в (6), находим

$$D\hat{\theta}_1 = \sigma^2 (\|x_1\|^2 - bF^{-1}b')^{-1}.$$

Минимальное значение $D\hat{\theta}_1$ приобретает, когда вектор x_1 имеет наибольшую возможную длину, а вектор $b=0$, так как квадратичная форма $bF^{-1}b'$ положительна, если только $b \neq 0$.

[Проверим, что матрица F^{-1} положительно определена. Квадратичная форма

$$aX'Xa' = aX'(aX')' = \|aX'\|^2$$

положительна, если только $a \neq 0$, так как X имеет полный ранг. Для вектора вида $a = (0, a_2, \dots, a_r)$

$$aX'Xa' = (a_2, \dots, a_r) F (a_2, \dots, a_r)',$$

откуда следует, что матрица F — положительно определена.

Пусть C — ортогональная $(n-1) \times (n-1)$ -матрица ($CC' = C'C = I$), такая, что

$$C'FC = \Lambda,$$

где Λ — диагональная $(n-1) \times (n-1)$ -матрица с положительными элементами на главной диагонали. Домножая это соотношение на C слева и на C' справа, получим

$$F = C\Lambda C'.$$

Очевидно,

$$F^{-1} = C\Lambda^{-1}C',$$

откуда получаем

$$\mathbf{b}F^{-1}\mathbf{b}' = \mathbf{b}C\Lambda^{-1}C'\mathbf{b}' = \mathbf{b}C\Lambda^{-1/2}\Lambda^{-1/2}C'\mathbf{b}' = \|\mathbf{b}C\Lambda^{1/2}\|^2,$$

что и требовалось.]

Те же рассуждения применимы и к дисперсии любой другой оценки $\hat{\theta}_i$, $i=2, \dots, r$. Итак, чтобы дисперсия оценок коэффициентов $\hat{\theta}_i$, $i=1, \dots, r$, была минимальна, надо, если это выполнимо, организовать взвешивание так, чтобы, во-первых, вектора x_i не содержали нулевых компонент, т. е. при каждом взвешивании участвовали все предметы, и, во-вторых, матрица $X'X$ была диагональна, т. е. столбцы X ортогональны друг другу. Минимальное значение дисперсии одинаково для всех оценок и равно σ^2/n . Если $n=rk$, то дисперсия взвешиваний σ^2/n в r раз меньше, чем дисперсия σ^2/k при k -кратном взвешивании отдельных предметов и том же общем числе взвешиваний n .

Приведем пример ортогональной матрицы из элементов ± 1 :

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

При $r=n=4$ эксперимент с $X=A$ приводит к дисперсии оценок $\sigma^2/4$; при $r=4$, $n=8$, взяв $X'=[A', A']$, получим дисперсию $\sigma^2/8$, в то время как при взвешивании по отдельности эти дисперсии равны соответственно σ^2 и $\sigma^2/2$.

5. Каноническая форма линейной модели.

Рассмотрим линейную модель следующего частного вида:

$$U_i = \psi_i + \sigma \delta_i, \quad i=1, \dots, s; \quad U_i = \sigma \delta_i, \quad i=s+1, \dots, n, \\ M\delta_i = 0, \quad i=1, \dots, n; \quad R_{\delta} = I, \quad (7)$$

где $U = (U_1, \dots, U_n)$ — вектор наблюдений, $\Psi = (\psi_1, \dots, \psi_s)$ — вектор неизвестных коэффициентов. Теория наилучшего несмешленного

го оценивания для модели (7) сильно упрощается по сравнению с общей моделью предыдущего параграфа. Чтобы статистика

$$\sum_{i=1}^n b_i U_i = \sum_{i=1}^s b_i \psi_i + \sigma \sum_{i=1}^n b_i \delta_i$$

несмешенно оценивала, скажем, ψ_1 , очевидно, необходимо и достаточно выполнение условия $b_1=1$, $b_2=\dots=b_s=0$. Чтобы ее дисперсия $\sigma^2 \sum_{i=1}^n b_i^2$ была минимальна, надо положить еще $b_{s+1}=\dots=b_n=0$. Таким образом, наилучшие несмешенные оценки коэффициентов ψ_i равны

$$\hat{\psi}_i = U_i, \quad i = 1, \dots, s.$$

Такие же рассуждения показывают, что наилучшей оценкой параметрической функции $a_1\psi_1+\dots+a_s\psi_s$ будет $a_1\hat{\psi}_1+\dots+a_s\hat{\psi}_s$. Из второй группы соотношений (7) сама собой напрашивается оценка дисперсии

$$\hat{\sigma}^2 = \frac{1}{n-s} \sum_{i=s+1}^n U_i^2, \quad (8)$$

которая, очевидно, является несмешенной.

Покажем, что линейную модель общего вида:

$$Y = \theta X' + \sigma \varepsilon, \quad M\varepsilon_i = 0, \quad i = 1, \dots, n, \quad R_\varepsilon = I,$$

$$X = [x_1, x_2, \dots, x_r], \quad \theta = (\theta_1, \theta_2, \dots, \theta_r), \quad (9)$$

ортогональной заменой переменных можно привести к виду (7), который называют канонической формой модели (9), и получить заново оценки $\hat{\theta}$ исходя из оценок $\hat{\psi}$. Обозначим через s ранг матрицы X . Выберем ортонормированный базис в пространстве V размерности s , порожденном столбцами матрицы X , и дополним его до ортонормированного базиса всего пространства R^n . Обозначим через C ортогональную матрицу перехода к новым координатам:

$$u = yC, \quad u = (u_1, \dots, u_n), \quad y = (y_1, \dots, y_n). \quad (10)$$

Домножим обе части соотношения (9) на матрицу C , положим $U = YC$, $\delta = \varepsilon C$. Так как вектор среднего $\eta = \theta X'$ лежит в пространстве V , то его новые координаты можно представить в виде $(\psi_1, \dots, \psi_s, 0, \dots, 0)$, а новые параметры ψ_1, \dots, ψ_s связаны со старыми соотношением

$$\eta C = \theta X' C = (\psi_1, \dots, \psi_s, 0, \dots, 0). \quad (11)$$

Замечая, что (см. (1))

$$R_{\delta} = R_{eC} = C'R_eC = C'IC = I,$$

от модели (9) приходим к (7).

Подставляя $\widehat{\theta}_i = U_i$, $i = 1, \dots, s$, вместо θ_i в (11), получим следующие соотношения для нахождения $\widehat{\theta}$:

$$\widehat{\theta}X'C = (U_1, \dots, U_s, 0, \dots, 0). \quad (12)$$

Умножая обе части равенства (12) на $C'X$, имеем

$$\begin{aligned} \widehat{\theta}X'X &= (U_1, \dots, U_s, 0, \dots, 0)C'X = \\ &= (U_1, \dots, U_s, 0, \dots, 0)[C'x'_1, \dots, C'x'_r]. \end{aligned} \quad (13)$$

По условию выбора нового ортонормированного базиса новые координаты любого вектора y из V начиная с $s+1$ -й нулевые:

$$yC = (u_1, \dots, u_s, 0, \dots, 0).$$

Следовательно, все векторы x_iC , $i = 1, \dots, r$, имеют последние $n-s$ координат нулевые, и потому (13) не изменится, если вектор $(U_1, \dots, U_s, 0, \dots, 0)$ заменить на (U_1, \dots, U_n) :

$$\widehat{\theta}X'X = (U_1, \dots, U_n)C'X = YCC'X = YX. \quad (14)$$

Матричное уравнение (14) совпадает с нормальной системой уравнений (24) § 9.

6. Оценивание дисперсии.

Пользуясь канонической формой (7) модели (9), выразим оценку $\widehat{\sigma}^2$ из соотношения (8) через исходные величины Y_i , $i = 1, \dots, n$. Из ортогональности матрицы C вытекает

$$\sum_{i=1}^n U_i^2 = \sum_{i=1}^n Y_i^2,$$

а из (12) получаем, что

$$\sum_{i=1}^s U_i^2 = \|\widehat{\theta}X'C\|^2 = \widehat{\theta}X'CC'X\widehat{\theta}' = \widehat{\theta}X'X\widehat{\theta}'.$$

Таким образом,

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n-s} \left(\sum_{i=1}^n U_i^2 - \sum_{i=1}^s U_i^2 \right) = \\ &= \frac{1}{n-s} \left(\sum_{i=1}^s Y_i^2 - \widehat{\theta}X'X\widehat{\theta}' \right) = \frac{1}{n-s} (\|Y\|^2 - \|\widehat{\theta}X'\|^2). \end{aligned} \quad (15)$$

Воспользовавшись равенствами

$$\hat{\theta}X' = \text{пр}_V Y, \quad Y = \text{пр}_V Y + \text{пр}_{R^n \ominus V} Y,$$

$$\|Y\|^2 = \|\text{пр}_V Y\|^2 + \|\text{пр}_{R^n \ominus V} Y\|^2,$$

перепишем (15) в следующей форме:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-s} \|\text{пр}_{R^n \ominus V} Y\|^2 = \frac{1}{n-s} \|Y - \hat{\theta}X'\|^2 = \\ &= \frac{1}{n-s} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^r \hat{\theta}_j x_{ij} \right)^2.\end{aligned}\quad (16)$$

Отметим, что оценка $\hat{\sigma}^2$ — несмешенная (см. (8)).

7. Обобщенная линейная модель.

Обобщим линейную модель, допуская, что компоненты вектора e коррелированы:

$$Y = \theta X' + \sigma e, \quad M_e = 0, \quad R_e = Q = [q_{ij}, \quad i, j = 1, \dots, n], \quad (17)$$

но при этом будем предполагать, что матрица Q известна и положительно определена. Рассмотрим только случай, когда матрица X полного ранга. Представим матрицу ковариации вектора e в виде (см. (4))

$$Q = B'B, \quad (18)$$

где B — невырожденная $n \times n$ -матрица. Умножая обе части (17) на B^{-1} , получаем

$$YB^{-1} = \theta X'B^{-1} + \sigma eB^{-1}.$$

Вектор $\delta = eB^{-1}$ имеет единичную ковариационную матрицу (см. (1)):

$$R_\delta = (B^{-1})' R_e B^{-1} = (B')^{-1} (B'B) B^{-1} = I.$$

Полагая

$$U = YB^{-1}, \quad W' = X'B^{-1},$$

приходим к обычной линейной модели:

$$U = \theta W' + \sigma \delta, \quad M\delta = 0, \quad R_\delta = I, \quad (19)$$

причем матрица W' имеет полный ранг. Покажем, как получить наилучшие несмешенные оценки в обобщенной модели (17), пользуясь уже известными нам результатами для обычной линейной модели (19). Пусть d^*U' — оценка с минимальной дисперсией, несмешенно оценивающая $a\theta'$. Тогда

$$d^*(B^{-1})' Y = d^* U'$$

является линейной несмешенной оценкой $\mathbf{a}\theta'$ в модели (17), и мы покажем, что она имеет минимальную дисперсию. Допустим, что $\mathbf{b}_1\mathbf{Y}'$ — такая несмешенная оценка $\mathbf{a}\theta'$, что

$$\mathbf{D}\mathbf{b}_1\mathbf{Y}' < \mathbf{D}\mathbf{d}^*(\mathbf{B}^{-1})'\mathbf{Y}' = \mathbf{D}\mathbf{d}^*\mathbf{U}', \quad (20)$$

Поскольку

$$\mathbf{b}_1\mathbf{B}'\mathbf{U} = \mathbf{b}_1\mathbf{Y}'$$

несмешенно оценивает $\mathbf{a}\theta'$, а ввиду (20)

$$\mathbf{D}\mathbf{b}_1\mathbf{B}'\mathbf{U} = \mathbf{D}\mathbf{b}_1\mathbf{Y}' < \mathbf{D}\mathbf{d}^*\mathbf{U},$$

то оценка $(\mathbf{b}_1\mathbf{B}')\mathbf{U}$ имеет дисперсию, меньшую минимально возможной. Полученное противоречие доказывает, что наилучшая несмешенная оценка $\mathbf{a}\theta'$ в модели (17) имеет вид

$$\mathbf{d}^*(\mathbf{B}^{-1})'\mathbf{Y}, \quad (21)$$

где \mathbf{d}^* находится из условия, что $\mathbf{d}^*\mathbf{U}'$ является наилучшей несмешенной оценкой $\mathbf{a}\theta'$ в модели (19). Используя формулу (5) наилучших несмешенных оценок, получаем

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &= \mathbf{U}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} = \mathbf{Y}\mathbf{B}^{-1}(\mathbf{B}^{-1})'\mathbf{X}(\mathbf{X}'\mathbf{B}^{-1}(\mathbf{B}^{-1})'\mathbf{X})^{-1} = \\ &= \mathbf{Y}\mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{Q}^{-1}\mathbf{X})^{-1}. \end{aligned} \quad (22)$$

Матрица ковариаций оценок $\widehat{\boldsymbol{\theta}}$ равна (см. (5))

$$R_{\widehat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{W}'\mathbf{W})^{-1} = \sigma^2(\mathbf{X}'\mathbf{B}^{-1}(\mathbf{B}^{-1})'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{Q}^{-1}\mathbf{X})^{-1}. \quad (23)$$

Оценка дисперсии принимает вид (см. (16))

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n-s}(\mathbf{U} - \widehat{\boldsymbol{\theta}}\mathbf{W}')(\mathbf{U} - \widehat{\boldsymbol{\theta}}\mathbf{W}')' = \\ &= \frac{1}{n-s}(\mathbf{Y} - \widehat{\boldsymbol{\theta}}\mathbf{X}')\mathbf{B}^{-1}(\mathbf{B}^{-1})'(\mathbf{Y} - \widehat{\boldsymbol{\theta}}\mathbf{X}')' = \\ &= \frac{1}{n-s}(\mathbf{Y} - \widehat{\boldsymbol{\theta}}\mathbf{X}')\mathbf{Q}^{-1}(\mathbf{Y} - \widehat{\boldsymbol{\theta}}\mathbf{X})'. \blacksquare \end{aligned} \quad (24)$$

Представляется полезным дать еще и независимое изложение теории наилучшего несмешенного оценивания для обобщенной линейной модели (17). Используем при этом геометрический подход. Итак, необходимо минимизировать дисперсию линейной оценки $\mathbf{b}\mathbf{Y}'$ параметрической функции $\mathbf{a}\theta'$:

$$\mathbf{D}\mathbf{b}\mathbf{Y}' = \mathbf{b}\mathbf{R}_{\mathbf{Y}}\mathbf{b}' = \sigma^2 \mathbf{b}\mathbf{Q}\mathbf{b}' \quad (25)$$

при условии несмешенности

$$\mathbf{b}\mathbf{X} = \mathbf{a}. \quad (26)$$

Определим в пространстве R^n скалярное произведение по формуле

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{Q}\mathbf{y}' \quad (27)$$

и перепишем условие (26) в виде

$$\mathbf{b}Q(Q^{-1}\mathbf{X}) = \mathbf{a}. \quad (28)$$

Задача минимизации квадратичной формы (25) при линейных ограничениях (28) решается в полной аналогии со случаем обычной линейной модели: $Q = I$. Пусть вектор \mathbf{b} удовлетворяет (28), тогда вектор \mathbf{b}^* , получаемый проектированием \mathbf{b} в смысле скалярного произведения (27) на подпространство, порожденное столбцами матрицы $Q^{-1}\mathbf{X}$, также удовлетворяет (28) и в то же время имеет минимальный квадрат длины:

$$\langle \mathbf{b}^*, \mathbf{b}^* \rangle = \mathbf{b}^* Q \mathbf{b}^*.$$

Следовательно, вектор \mathbf{b} , приводящий к наилучшей несмешенной оценке, лежит в линейном подпространстве, порожденном столбцами $Q^{-1}\mathbf{X}$ (ср. (16) § 9):

$$\mathbf{b} = \mathbf{c}(Q^{-1}\mathbf{X})' = \mathbf{c}\mathbf{X}'(Q^{-1})'. \quad (29)$$

Подставляя (29) в (28), находим с:

$$\mathbf{c}\mathbf{X}'(Q^{-1})'\mathbf{X} = \mathbf{a}, \quad \mathbf{c} = \mathbf{a}(\mathbf{X}'(Q^{-1})'\mathbf{X})^{-1},$$

таким образом, наилучшая несмешенная оценка имеет вид

$$\mathbf{b}\mathbf{Y}' = \mathbf{Y}\mathbf{b}' = \mathbf{Y}Q^{-1}\mathbf{X}(\mathbf{X}'Q^{-1}\mathbf{X})^{-1}\mathbf{a}',$$

что совпадает с (22).

§ 11. ПОРЯДКОВЫЕ ЛИНЕЙНЫЕ ОЦЕНКИ ДЛЯ ПАРАМЕТРОВ СДВИГА И МАСШТАБА

1. Обобщенная линейная модель для семейства сдвига-масштаба.

Рассмотрим независимую выборку

$$X_1, X_2, \dots, X_n \quad (1)$$

с общей ф. р. $F((x-\mu)/\sigma)$; ф. р. $F(u)$ стандартизованных сл. в.

$$U_i = (X_i - \mu)/\sigma, \quad i = 1, \dots, n, \quad (2)$$

считается известной, параметры μ и $\sigma > 0$ неизвестны. Порядковые статистики сл. в. (1) и (2) связаны формулой

$$U_{(i)} = (X_{(i)} - \mu)/\sigma, \quad i = 1, \dots, n, \quad (3)$$

и так как ф. р. $F(u)$ сл. в. U_i известна, то в принципе любые характеристики $U_{(i)}$ могут быть рассчитаны. Положим

$$\mathbf{M}U_{(i)} = \alpha_i, \quad \text{cov}(U_{(i)}, U_{(j)}) = q_{ij}, \quad i, j = 1, \dots, n,$$

и, учитывая (3), запишем

$$X_{(i)} = \mu + \sigma U_{(i)} = \mu + \sigma \alpha_i + \sigma (U_{(i)} - \alpha_i), \quad i = 1, \dots, n. \quad (4)$$

Полагая

$$X_{(i)} = Y_i; \quad U_{(i)} - \alpha_i = \varepsilon_i, \quad i = 1, \dots, n, \quad (5)$$

представим (4) в виде обобщенной линейной модели:

$$Y_i = \mu + \sigma \alpha_i + \sigma \varepsilon_i, \quad i = 1, \dots, n, \quad (6)$$

или в векторной форме:

$$\mathbf{Y} = (\mu, \sigma) \begin{bmatrix} 1 \\ \alpha \end{bmatrix} + \sigma \mathbf{\varepsilon},$$

$$\begin{aligned} \mathbf{1} &= (1, \dots, 1), \quad \alpha = (\alpha_1, \dots, \alpha_n), \quad M\mathbf{\varepsilon} = \mathbf{0}, \quad R_{\mathbf{\varepsilon}} = \\ &= Q = [q_{ij}, \quad i, j = 1, \dots, n]. \end{aligned} \quad (7)$$

2. Наилучшие несмешанные оценки.

Оценки $\hat{\mu}$, $\hat{\sigma}$ находятся по формулам (22) § 10:

$$(\hat{\mu}, \hat{\sigma}) = \mathbf{Y} Q^{-1} [\mathbf{1}', \alpha'] \left(\begin{bmatrix} 1 \\ \alpha \end{bmatrix} Q^{-1} [\mathbf{1}', \alpha'] \right)^{-1}. \quad (8)$$

Производя перемножение, преобразуем последний сомножитель в (8) к 2×2 -матрице и найдем обратную матрицу:

$$\begin{bmatrix} 1Q^{-1}\mathbf{1}' & 1Q^{-1}\alpha' \\ \alpha Q^{-1}\mathbf{1}' & \alpha Q^{-1}\alpha' \end{bmatrix}^{-1} = \frac{1}{\Delta} \begin{bmatrix} \alpha Q^{-1}\alpha' & -\alpha Q^{-1}\mathbf{1}' \\ -1Q^{-1}\alpha' & 1Q^{-1}\mathbf{1}' \end{bmatrix}, \quad (9)$$

где

$$\Delta = (1Q^{-1}\mathbf{1}') (\alpha Q^{-1}\alpha') - (1Q^{-1}\alpha')^2. \quad (10)$$

Матричный множитель при векторе \mathbf{Y} в формуле (8) равен произведению $n \times 2$ -матрицы

$$Q^{-1} [\mathbf{1}', \alpha'] = [Q^{-1}\mathbf{1}', Q^{-1}\alpha'] \quad (11)$$

на 2×2 -матрицу (9). Результатом их перемножения будет $n \times 2$ -матрица, столбцы которой являются линейной комбинацией столбцов матрицы (11):

$$\begin{aligned} &\frac{1}{\Delta} [Q^{-1} (\alpha Q^{-1}\alpha') + Q^{-1}\alpha' (-1Q^{-1}\alpha'), \\ &Q^{-1}\mathbf{1}' (-\alpha Q^{-1}\mathbf{1}') + Q^{-1}\alpha' (1Q^{-1}\mathbf{1}')] = \\ &= \frac{1}{\Delta} [Q^{-1} (\mathbf{1}'\alpha - \alpha'\mathbf{1}) Q^{-1}\alpha', Q^{-1} (-\mathbf{1}'\alpha + \alpha'\mathbf{1}) Q^{-1}\mathbf{1}']. \end{aligned}$$

Таким образом,

$$(\hat{\mu}, \hat{\sigma}) = \frac{1}{\Delta} (\Upsilon \Gamma \alpha', -\Upsilon \Gamma \mathbf{1}'), \quad \Gamma = Q^{-1} (\mathbf{1}' \alpha - \alpha' \mathbf{1}) Q^{-1}. \quad (12)$$

Ковариационная матрица вектора оценок $(\hat{\mu}, \hat{\sigma})$ равна

$$\sigma^2 \left(\begin{bmatrix} 1 \\ \alpha \end{bmatrix} Q^{-1} [\mathbf{1}', \alpha'] \right)^{-1} = \frac{\sigma^2}{\Delta} \begin{bmatrix} \alpha Q^{-1} \alpha' & -\alpha Q^{-1} \mathbf{1}' \\ -\mathbf{1}' Q^{-1} \alpha' & \mathbf{1}' Q^{-1} \mathbf{1}' \end{bmatrix} \quad (13)$$

(см. (23) § 10). ■■■

Для дальнейшего полезно иметь в виду равенство

$$\sum_{i=1}^n \mathbf{U}_{(i)} = \sum_{i=1}^n \mathbf{U}_i.$$

Взяв математические ожидания и дисперсии от обеих частей равенства, получаем

$$\mathbf{1}\alpha' = n \mathbf{M} \mathbf{U}_1, \quad \mathbf{1}' Q \mathbf{1}' = n \mathbf{D} \mathbf{U}_1 = n. \quad (14)$$

3. Симметричное распределение.

Рассмотрим случай симметричного распределения выборки (1), и пусть $\mu = M \mathbf{X}_i$. Стандартизованные сл. в. \mathbf{U}_i , $i=1, \dots, n$, в этом случае имеют распределение, симметричное относительно нуля, так что векторы $(\mathbf{U}_1, \dots, \mathbf{U}_n)$ и $(-\mathbf{U}_1, \dots, -\mathbf{U}_n)$ одинаково распределены. Следовательно, одинаково распределены и векторы, составленные из их порядковых статистик, т. е. $\mathbf{V} = (\mathbf{U}_{(1)}, \dots, \mathbf{U}_{(n)})$ и $(-\mathbf{U}_{(n)}, \dots, -\mathbf{U}_{(1)})$. Введем матрицу

$$H = [\delta'_{\mathbf{n}}, \delta'_{\mathbf{n}-1}, \dots, \delta'_1], \quad \delta_i = (\delta_{i1}, \dots, \delta_{in}),$$

где δ_{ij} — символ Кронекера: $\delta_{ii}=1$; $\delta_{ij}=0$ при $i \neq j$, и запишем

$$(-\mathbf{U}_{(n)}, \dots, -\mathbf{U}_{(1)}) = -(\mathbf{U}_{(1)}, \dots, \mathbf{U}_{(n)}) H = -\mathbf{V} H.$$

Таким образом, векторы \mathbf{V} и $-\mathbf{V} H$ одинаково распределены. В частности, совпадают их математические ожидания и матрицы ковариаций

$$M\mathbf{V} = -M\mathbf{V} H, \quad R_{\mathbf{V}} = R_{-\mathbf{V} H},$$

откуда получаем

$$\alpha = M\mathbf{V} = -\alpha H, \quad Q = R_{\mathbf{V}} = H' R_{\mathbf{V}} H = H' Q H.$$

Заметим, что $H' = H = H^{-1}$ и $\mathbf{1} H = \mathbf{1}$, получаем отсюда

$$Q^{-1} = (H Q H)^{-1} = H Q^{-1} H,$$

$$\mathbf{1} Q^{-1} \alpha' = \mathbf{1} H Q^{-1} H \alpha' = \mathbf{1} Q^{-1} H \alpha' = -\mathbf{1} Q^{-1} \alpha',$$

т. е.

$$\mathbf{1} Q^{-1} \alpha' = 0 = \alpha Q^{-1} \mathbf{1}'. \quad (15)$$

Подставляя (15) в (12) и (10), находим

$$\begin{aligned}\Gamma\alpha' &= Q^{-1}(1'\alpha - \alpha'1)Q^{-1}\alpha' = Q^{-1}1'(\alpha Q^{-1}\alpha'), \\ \Gamma 1' &= Q^{-1}(1\alpha - \alpha'1)Q^{-1}1' = -Q^{-1}\alpha'(1Q^{-1}1'),\end{aligned}$$

$$\Delta = (1Q^{-1}1')(\alpha Q^{-1}\alpha'), \quad (16)$$

$$\begin{aligned}\hat{\mu} &= 1/\Delta Y\Gamma\alpha' = YQ^{-1}1'/(1Q^{-1}1'), \\ \hat{\sigma} &= 1/\Delta Y\Gamma 1' = YQ^{-1}\alpha'/(1Q^{-1}\alpha').\end{aligned} \quad (17)$$

Подставляя (15) и (16) в матрицу ковариаций (13), получаем, что $\hat{\mu}$ и $\hat{\sigma}$ некоррелированы, а их дисперсии равны

$$\begin{aligned}D\hat{\mu} &= \sigma^2/(1Q^{-1}1'), \\ D\hat{\sigma} &= \sigma^2/(\alpha Q^{-1}\alpha').\end{aligned} \quad (18)$$

Наилучшая линейная несмешенная оценка $\hat{\mu}$ имеет дисперсию, не превосходящую дисперсии выборочного среднего:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n X_{(i)} = \bar{Y},$$

поскольку \bar{X} представима как линейная функция порядковых статистик и является несмешенной оценкой μ . Покажем, что равенство

$$D\hat{\mu} = D\bar{X}$$

осуществляется лишь при условии

$$Q1' = 1', \quad (19)$$

и при этом $\hat{\mu} = \bar{X}$. Для доказательства воспользуемся тем, что неравенство Коши—Буняковского—Шварца

$$(aQa')(bQb') \geq (aQb')^2, \quad (20)$$

обращается в равенство, лишь когда векторы пропорциональны:

$$a = \lambda b \text{ или } b = \lambda a.$$

Выбирая $b = b_1 Q^{-1}$ и подставляя в (20), получаем

$$(aQa')(b_1 Q^{-1}b_1') \geq (ab_1')^2,$$

где неравенство выполняется при

$$a = \lambda b_1 Q^{-1} \text{ или } b_1 Q^{-1} = \lambda a.$$

Полагая здесь $a = b_1 = 1$, получаем

$$(1Q1')(1Q^{-1}1') \geq (11')^2.$$

Но (см. (14))

$$1Q1' = n,$$

так что

$$1Q^{-1}1' \geq n,$$

а равенство имеет место, лишь когда

$$Q1' = \lambda 1'.$$

При этом из равенства $1Q1' = n$ вытекает, что $\lambda = 1$. Таким образом,

$$\widehat{D\mu} = \sigma^2 / (1Q1') < \sigma^2 / n = D\bar{X},$$

причем равенство имеет место лишь при условии (19). Наконец, из (19) вытекает, что $Q^{-1}1' = 1'$, и из (17) находим, что при этом $\widehat{\mu} = \bar{Y} = \bar{X}$.

4. Пример: равномерное распределение на интервале $(\mu, \mu + \sigma)$.

Используя формулу (21) § 3 для плотности распределения k -й порядковой статистики, находим в случае равномерного на $(\mu, \mu + \sigma)$ распределения сл. в. (1):

$$\begin{aligned} a_k = MU_{(k)} &= \frac{n!}{(k-1)! (n-k)!} \int_0^1 t \cdot t^{k-1} (1-t)^{n-k} dt = \\ &= \frac{(n+1)!}{k! (n+1-k)!} \int_0^1 t^{(k+1)-1} (1-t)^{(n+1)-(k+1)} dt - \frac{k}{n+1} = \frac{k}{n+1}. \end{aligned}$$

и аналогично

$$\begin{aligned} q_{kk} = DU_{(k)} &= MU_{(k)}^2 - a_k^2 = \frac{k(k+1)}{(n+1)(n+2)} - \left(\frac{k}{n+1}\right)^2 = \\ &= \frac{k(n-k+1)}{(n+1)^2(n+2)}. \end{aligned} \quad (21)$$

Для вычисления

$$q_{kl} = \text{cov}(U_{(k)}, U_{(l)}) = MU_{(k)}U_{(l)} - a_k a_l, \quad k < l,$$

используем формулу (35) § 3 для совместной плотности пары порядковых статистик и запишем

$$\begin{aligned} MU_{(k)}U_{(l)} &= \frac{n!}{(k-1)! (l-k-1)! (n-l)!} \iint_{0 < x < y < 1} xyx^{k-1} \times \\ &\times (y-x)^{l-k-1} (1-y)^{n-l} dx dy. \end{aligned} \quad (22)$$

Сделав замену переменных $x = zv$, $y = v$ и заметив, что якобиан перехода к новым переменным равен v , преобразуем интеграл в (22) к виду

$$\begin{aligned}
& \int_0^1 \int_0^1 zvv(zv)^{k-1} v^{l-k-1} (1-z)^{l-k-1} (1-v)^{n-l} v dz dv = \\
& = \int_0^1 z^k (1-z)^{l-k-1} dz \int_0^1 v^{l+1} (1-v)^{n-l} dv = \\
& = \beta(k+1, l-k) \beta(l+2, n-l+1).
\end{aligned}$$

С учетом проведенной выкладки из (22) получаем при $k < l$

$$MU_k U_{(l)} = \frac{n! k! (l-k-1)! (l+1)! (n-l)!}{(k-1)! (l-k-1)! (n-l)! l! (n+2)!} = \frac{k(l+1)}{(n+1)(n+2)},$$

откуда

$$q_{kl} = \frac{k(l+1)}{(n+1)(n+2)} - \frac{kl}{(n+1)^2} = \frac{k(n-l+1)}{(n+1)^2(n+2)}, \quad k < l. \quad (23)$$

Сравнив (21) с (23), замечаем, что формула (23) годится и при $k=l$.

Проверим, что матрица

$$\begin{aligned}
P &= (n+1)(n+2) [p_{kl}, \quad k, l = 1, \dots, n], \\
p_{kk} &= 2, \quad p_{k-1,k} = p_{k,k-1} = -1, \quad p_{kl} = 0, \quad |k-l| > 1,
\end{aligned}$$

является обратной к Q . Положим $QP = [a_{kj}, \quad k, j = 1, \dots, n]$, тогда

$$a_{kj} = (n+1)(n+2) \sum_{l=1}^n q_{kl} p_{lj}.$$

Доопределим элементы $q_{k,l}$ при k и l принимающих значения 0 и $n+1$ с помощью формулы (23). Легко проверить, что при $k < j$

$$a_{kj} = (n+1)(n+2) (-q_{k,j-1} + 2q_{k,j} - q_{k,j+1}).$$

Отсюда при $k < j$ находим

$$a_{kj} = \frac{1}{n+1} (-k(n-j+2) + 2k(n-j+1) - k(n-j)) = 0,$$

а при $k=j$, воспользовавшись симметрией матрицы Q , получаем

$$\begin{aligned}
a_{jj} &= \frac{1}{n+1} (-q_{j-1,j} + 2q_{jj} - q_{j,j+1}) = \\
&= \frac{1}{n+1} ((-j+1)(n-j+1) + 2j(n-j+1) + j(n-j)) = 1.
\end{aligned}$$

Таким образом, $Q^{-1} = P$. ■

Оценки μ и σ могут быть получены применением формул (12). Мы воспользуемся более простыми формулами (17), так как распределение элементов выборки симметричное, но для этого надо перейти к параметру $\mu^* = \mu + \sigma/2$, равному среднему значению рас-

пределения элементов выборки. В новой параметризации получаем выборку из равномерного распределения на интервале $(\mu^* - \sigma/2; \mu^* + \sigma/2)$, и соответствующая модель имеет вид

$$Y_i = \mu^* + \sigma \alpha_i^* + \sigma \varepsilon_i, \quad i = 1, \dots, n,$$

где матрица ковариаций вектора ε остается прежней, $\alpha_i^* = \alpha_i - 1/2$. Проводя вычисления по формулам (17), имеем

$$Q^{-1}1' = (n+1)(n+2)(1, 0, \dots, 0, 1)', \quad 1Q^{-1}1' = 2(n+1)(n+2),$$

$$Q^{-1}\alpha^* = (n+1)(n+2)(-1/2, 0, \dots, 0, 1/2),$$

$$\alpha^* Q^{-1} \alpha^* = (n+1)(n+2)(n-1)/2(n+1),$$

$$\hat{\mu}^* = YQ^{-1}1' / (1Q^{-1}1') = 1/2(Y_1 + Y_n) = 1/2(X_{(1)} + X_{(n)}), \quad (24)$$

$$\begin{aligned} \hat{\sigma} &= YQ^{-1}1' / (\alpha^* Q^{-1} \alpha^*) = (n+1)(Y_n - Y_1)/(n-1) = \\ &= (n+1)(X_{(n)} - X_{(1)})/(n-1). \end{aligned} \quad (25)$$

Так как $\mu = \mu^* - \sigma/2$, то наилучшая несмешенная оценка равна (см. теорему 1 § 9)

$$\hat{\mu} = \hat{\mu}^* - \hat{\sigma}/2 = \frac{n}{n-1} X_{(1)} - \frac{1}{n-1} X_{(n)}. \quad (26)$$

Оценки $\hat{\mu}^*$ и $\hat{\sigma}$ некоррелированы, дисперсии оценок равны (см. (18))

$$\begin{aligned} D\hat{\mu}^* &= \sigma^2((n+1)(n+2))^{-1}, \quad D\hat{\sigma} = 2\sigma^2((n-1)(n+2))^{-1}, \\ D\hat{\mu} &\doteq D\hat{\mu}^* + D(\hat{\sigma}/2) = \sigma^2(3n-1)(2(n^2-1)(n+2))^{-1}. \end{aligned} \quad (27)$$

5. Пример: экспоненциальное распределение.

В случае экспоненциальной ф. р. $1 - \exp(-(x-\mu)/\sigma)$, $x > \mu$, стандартизованные сл. в. U_i имеют плотность e^{-x} , $x > 0$, и, в соответствии с леммой 1 § 6, сл. в.

$$Z_i = (n-i+1)(U_{(i)} - U_{(i-1)}), \quad U_{(0)} = 0, \quad i = 1, \dots, n, \quad (28)$$

независимы и имеют ту же экспоненциальную плотность e^{-x} , $x > 0$. Из (28) имеем

$$U_{(k)} = \sum_{i=1}^k Z_i / (n-i+1), \quad k = 1, \dots, n,$$

откуда

$$\alpha_k = MU_{(k)} = \sum_{i=1}^k (n-i+1)^{-1}, \quad (29)$$

$$q_{kl} = \text{cov}(U_{(k)}, U_{(l)}) = \sum_{i=1}^k \sum_{j=1}^l \text{cov}(Z_i, Z_j) \times \\ \times (n-i+1)^{-1} (n-j+1)^{-1} = \sum_{i=1}^{\min(k, l)} (n-i+1)^{-2}. \quad (30)$$

Проверим, что матрица

$$P = [p_{kl}, k, l = 1, \dots, n],$$

$$p_{k,k-1} = p_{k-1,k} = -(n-k+1)^2, \quad p_{kk} = (n-k+1)^2 + (n-k)^2 \quad (31)$$

является обратной к Q . Доопределим элементы p_{kl} и q_{kl} при k и l , равных 0 или $n+1$, с помощью формул (31) и (30) (считая там сумму по пустому множеству индексов равной нулю). Нетрудно видеть, что при этом для $1 \leq k < j < n$ справедлива формула

$$\sum_{l=1}^n q_{kl} p_{lj} = q_{k,j-1} p_{j-1,j} + q_{kj} p_{jj} + q_{k,j+1} p_{j,j+1}. \quad (32)$$

Поскольку $p_{j-1,j} + p_{jj} + p_{j,j+1} = 0$, а при $k < j$

$$q_{k,j-1} + q_{kj} = q_{k,j+1} = \sum_{i=1}^k (n-i+1)^{-2},$$

то из (32) получаем, что внедиагональные элементы матрицы QP — нулевые. Диагональные элементы QP также находим по формуле (32):

$$-(n-j+1)^2 \sum_{i=1}^{j-1} (n-i+1)^{-2} + ((n-j+1)^2 + (n-j)^2) \times \\ \times \sum_{i=1}^j (n-i+1)^{-2} - (n-j)^2 \sum_{i=1}^j (n-i+1)^{-2} = 1.$$

Таким образом, $Q^{-1} = P$.

Перейдем к вычислениям наилучших несмешанных оценок $\hat{\mu}$, σ по формулам (12), (10). Имеем

$$\sum_{l=1}^n p_{1l} = p_{11} + p_{12} = n^2 + (n-1)^2 - (n-1)^2 = n^2,$$

а при $k > 1$

$$\sum_{l=1}^n p_{kl} = p_{k,k-1} + p_{kk} + p_{k,k+1} = 0,$$

так что

$$Q^{-1} \mathbf{1}' = (n^2, 0, \dots, 0)' \quad (33)$$

Далее,

$$\sum_{l=1}^n p_{kl} a_l = p_{k,k-1} a_{k-1} + p_{kk} a_k + p_{k,k+1} a_{k+1}, \quad (34)$$

где a_j при всех j определены формулой (29). Подставляя в (34) значения входящих туда величин, получаем

$$-(n-k+1) \sum_{i=1}^{k-1} (n-i+1)^{-1} + ((n-k+1)^2 + (n-k)^2) \times \\ \times \sum_{i=1}^k (n-i+1)^{-1} - (n-k)^2 \sum_{i=1}^{k+1} (n-i+1)^{-1} = (n-k+1) - (n-k) = 1,$$

так что

$$Q^{-1} \alpha' = 1'. \quad (35)$$

Из (33) и (35) находим

$$1Q^{-1}1' = n^2, \quad \alpha Q^{-1} \alpha' = \alpha 1' = \sum_{k=1}^n a_k = M \sum_{k=1}^n U_{(k)} = \\ = M \sum_{k=1}^n U_k = n, \quad 1Q^{-1} \alpha' = 1 \cdot 1' = n, \quad \Delta = n^2 n - n^2 = n^2(n-1), \\ Q^{-1} 1' \alpha Q^{-1} = (n^2, 0, \dots, 0)' 1, \quad Q^{-1} \alpha' 1Q^{-1} = 1' (n^2, 0, \dots, 0), \\ \hat{\mu} = \frac{1}{\Delta} Y((n^2, 0, \dots, 0)' 1 \alpha' - 1' (n^2, 0, \dots, 0) \alpha') = \\ = \frac{1}{n^2(n-1)} Y((n^3, 0, \dots, 0) - 1' n^2 n^{-1}) = \frac{n}{n-1} Y_1 - \\ - \frac{1}{n(n-1)} \sum_{i=1}^n Y_i = (n X_{(1)} - \bar{X})/(n-1), \quad (36)$$

$$\hat{\sigma} = - \frac{1}{n^2(n-1)} Y((n^2, 0, \dots, 0) 1 1' - 1' (n^2, 0, \dots, 0) 1') = \\ = - \frac{1}{n^2(n-1)} Y((n^3, 0, \dots, 0) - n^2 1') = - \frac{n}{n+1} Y_1 + \frac{1}{n-1} \sum_{i=1}^n Y_i = \\ = n(\bar{X} - X_{(1)})/(n-1). \quad (37)$$

Выражение (13) для ковариационной матрицы вектора оценок $(\hat{\mu}, \hat{\sigma})$ принимает вид

$$\sigma^2 \begin{bmatrix} (n(n-1))^{-1} & -(n(n-1))^{-1} \\ -(n(n-1))^{-1} & (n-1)^{-1} \end{bmatrix}. \quad (38)$$

Оценки $\hat{\mu}$ и $\hat{\sigma}$ совпадают с оценками μ^* и σ^* § 6.

6. Пример: нормальное распределение.

Для моментов порядковых статистик стандартизованных сл. в. $U_{(k)}$ в случае нормального распределения нет простых формул. Покажем, однако, что в этом случае выполняется условие (19):

$$Q\mathbf{1}' = \mathbf{1}', \quad (39)$$

и, следовательно, выборочное среднее является наилучшей порядковой линейной несмешенной оценкой теоретического среднего μ . Как доказано в § 8, сл. в.

$$\bar{U} = \frac{1}{n} \sum_{i=1}^n U_i \text{ и } S_0^2 = \frac{1}{n} \sum_{i=1}^n (U_i - \bar{U})^2$$

независимы. В следующем параграфе будет показано (см. пример в п. 3 § 12), что на самом деле \bar{U} не зависит от вектора $(U_1 - \bar{U}, \dots, U_n - \bar{U})$, а не только от его длины. Воспользовавшись этим, заметим, что порядковые статистики набора сл. в. $U_1 - \bar{U}, \dots, U_n - \bar{U}$ имеют вид $U_{(1)} - \bar{U}, \dots, U_{(n)} - \bar{U}$ и как функции вектора $(U_1 - \bar{U}, \dots, U_n - \bar{U})$ не зависят от \bar{U} . В таком случае

$$0 = M\bar{U}(U_{(k)} - \bar{U}) = M\bar{U}U_{(k)} - 1/n,$$

откуда

$$\frac{1}{n} = M\bar{U}U_{(k)} = \frac{1}{n} M \sum_{i=1}^n U_{(i)}U_{(k)} = \frac{1}{n} \sum_{i=1}^n q_{ik},$$

что и дает (39).

7. Цензурированная выборка.

В технике и особенно в биомедицинских исследованиях типичной является ситуация, когда из ряда порядковых статистик $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ для измерения доступна лишь часть. Мы уже встречались с примером подобного рода в § 6, где рассматривалась задача одновременного испытания на надежность n изделий до момента выхода некоторого заранее установленного числа $n-r$ из них, а статистические выводы делались по значениям $X_{(1)} < \dots < X_{(r_1)} < \dots < X_{(n-r)}$. Другой пример возникает в случае, когда в ряду наблюдений X_1, X_2, \dots, X_n наибольшие, наименьшие или и те и другие значения являются ненадежными и в расчет принимают лишь значения

$$x_{(r_1+1)} \leq x_{(r_1+2)} \leq \dots \leq x_{(n-r_2)}, r_1, r_2 \geq 0. \quad (40)$$

Если наблюдению доступен лишь ряд (40), где хотя бы одно из чисел r_1, r_2 нулевое, то говорят, что выборка цензурирована (II тип цензуринга). Теория наилучшего несмешенного оценивания через линейные функции от цензурированных наблюде-

ний (40) переносится автоматически на этот случай: достаточно принять за новый вектор \mathbf{Y} . Более того, именно в случае цензурированной выборки этот метод оказывается особенно полезным, так как для полной выборки хорошие оценки во многих случаях строятся проще другими методами.

Практическое значение теории линейных порядковых оценок зависит от знания средних и ковариаций порядковых статистик. Явные выражения для μ и Q , а тем более для Q^{-1} можно получить в исключительных случаях, поэтому важное значение приобретает расчет таблиц для $MU_{(i)}$ и $\text{cov}(U_{(i)}, U_{(j)})$ и коэффициентов a_{1i}, a_{2i} наилучших оценок

$$\hat{\mu} = \sum_{i=r_1+1}^{n-r_2} a_{1i} X_{(i)}, \quad \hat{\sigma} = \sum_{i=r_1+1}^{n-r_2} a_{2i} X_{(i)} \quad (41)$$

для наиболее часто используемых распределений. Большой набор таких таблиц содержится у Сархана и Гринберга [8].

Пример 1 [9]. Ниже приведены значения $x_{(i)}$, $x_{(i)} = \log_{10} x_{(i)}$, $i = 1, 2, \dots, 7$, дней смерти и их логарифмов первых 7 из $n = 10$ мышей после вакцинации культуры туберкулеза.

$x'_{(i)}$	41	44	46	54	55	58	60
$x_{(i)}$	1,613	1,644	1,663	1,732	1,740	1,763	1,778

На рис. 11 нанесены точки $(x_{(i)}, \Phi^{-1}((i-0,5)/10))$, $i=1, 2, \dots, 7$, и проведена на глаз прямая $x = \sigma^* y + \mu^*$; $\mu^* = 1,747$; $\sigma^* = 0,090$. Согласие с нормальным законом логарифмов времен жизни можно считать удовлетворительным. Коэффициенты a_{1i}, a_{2i} находим из табл. 10.C.1 Сархана и Гринберга:

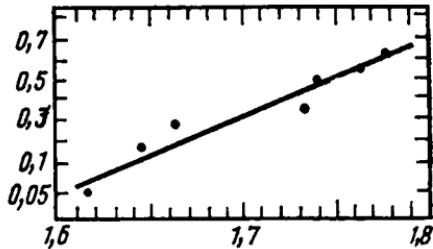


Рис. 11. Логарифм дней смерти первых 7 из 10 мышей после вакцинации культуры туберкулеза на нормальной бумаге

i	1	2	3	4	5	6	7
a_{1i}	0,0244	0,0636	0,0818	0,0962	0,1089	0,1207	0,5045
a_{2i}	-0,3252	-0,1758	-0,1058	-0,0502	-0,0006	0,0469	0,6107

Значения $\hat{\mu}$ и $\hat{\sigma}$ находятся по формулам (41):

$$\hat{\mu} = 1,746; \hat{\sigma} = 0,091$$

и, как видно, близки к значениям μ^* , σ^* , полученным графически.

8. Упрощенные линейные оценки.

Расчет матрицы ковариаций порядковых статистик $U_{(i)}$ представляет значительные трудности. Поэтому был предложен упрощенный метод, дающий хорошие результаты по крайней мере для нормальной выборки. Положим в формулах (10), (12) $Q=I$. В результате

$$\Delta = (\mathbf{1}' \mathbf{1}')(\boldsymbol{\alpha} \boldsymbol{\alpha}') - (\mathbf{1}' \boldsymbol{\alpha}')^2 = n \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2 = n \sum_{i=1}^n (a_i - \bar{a})^2.$$

$$\Gamma \boldsymbol{\alpha}' = \mathbf{1}' \boldsymbol{\alpha} \boldsymbol{\alpha}' - \boldsymbol{\alpha}' \mathbf{1}' \boldsymbol{\alpha}' = \sum_{i=1}^n a_i^2 \mathbf{1}' - \sum_{i=1}^n a_i \boldsymbol{\alpha}',$$

$$\Gamma \mathbf{1}' = \mathbf{1}' \boldsymbol{\alpha} \mathbf{1}' - \boldsymbol{\alpha}' \mathbf{1}' \mathbf{1}' = \sum_{i=1}^n a_i \mathbf{1}' - n \boldsymbol{\alpha}',$$

а оценки параметров μ и σ принимают вид

$$\begin{aligned}\check{\mu} &= \frac{1}{\Delta} \left(\sum_{i=1}^n a_i^2 \mathbf{Y} \mathbf{1}' - \sum_{i=1}^n a_i \mathbf{Y} \boldsymbol{\alpha}' \right) = \sum_{i=1}^n \left(\frac{1}{n} - \bar{a} c_i \right) Y_i, \\ \check{\sigma} &= -\frac{1}{\Delta} \left(\sum_{i=1}^n a_i \mathbf{Y} \mathbf{1}' - n \mathbf{Y} \boldsymbol{\alpha}' \right) = \sum_{i=1}^n c_i Y_i,\end{aligned}\quad (42)$$

где $c_i = (a_i - \bar{a}) / \sum_{i=1}^n (a_i - \bar{a})^2$. Формулы (42) применимы и к цензурированным наблюдениям (40), если под \mathbf{Y} понимать вектор соответствующих порядковых статистик ($\boldsymbol{\alpha} = \mathbf{M} \mathbf{Y}$, n — размерность вектора \mathbf{Y}).

Пример 1 (продолжение). По табл. 10.В.1 Сархана и Гринберга находим значения $\alpha_i = \mathbf{M} U_{(i)}$, $i=1, \dots, 5$, $\alpha_6 = -\alpha_5$, $\alpha_7 = -\alpha_4$:

i	1	2	3	4	5
α_i	1,538	1,001	0,656	0,375	0,122

Рассчитываем $\tilde{\mu}$, $\tilde{\sigma}$ по формулам (42) с $Y = (X_{(1)}, \dots, X_{(7)})$:

$$\tilde{\mu} = 1,748; \tilde{\sigma} = 0,094.$$

Результаты мало отличаются от наилучших несмешанных оценок:

$$\hat{\mu} = 1,746; \hat{\sigma} = 0,091.$$

§ 12. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

1. Невырожденное нормальное распределение.

Нормальная сл. в. U_1 с плотностью распределения $\Phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ называется *стандартной*. Сл. в. $X_1 = \sigma U_1 + \mu$ имеет плотность $\sigma^{-1}\phi((x-\mu)/\sigma)$ общего нормального закона $N(\mu, \sigma)$. Связь между U_1 и X_1 послужит нам отправной точкой для введения многомерного нормального распределения.

Определение 1. *Стандартным нормальным случайным вектором* назовем вектор

$$\mathbf{U} = (U_1, U_2, \dots, U_n)$$

из независимых стандартных нормальных компонент.

Вектор \mathbf{U} имеет математическое ожидание $\mathbf{0}$, матрицу ковариаций I , его распределение обозначим $N_n(\mathbf{0}, I)$.

Определение 2. Пусть $\mathbf{a} = (a_1, a_2, \dots, a_n)$ — произвольный числовой вектор, $B = [b_{ij}, i, j = 1, \dots, n]$ — невырожденная числовая матрица. Распределение случайного вектора

$$\mathbf{X} = \mathbf{UB} + \mathbf{a}, \quad (1)$$

где \mathbf{U} — стандартный нормальный вектор, назовем *невырожденным n -мерным нормальным распределением*.

Лемма 1. *Нормальный вектор \mathbf{X} , определенный формулой (1), имеет математическое ожидание \mathbf{a} , матрицу ковариаций $Q = B'B$ и плотность распределения вероятностей*

$$f(\mathbf{x}) = (\sqrt{2\pi})^n |\det Q|^{-1/2} \exp(-\frac{1}{2}(\mathbf{x}-\mathbf{a})' Q^{-1} (\mathbf{x}-\mathbf{a})). \quad (2)$$

Доказательство. Следующая формула

$$f_{UB}(\mathbf{x}) = |\det B|^{-1} f_U(\mathbf{x}B^{-1}) \quad (3)$$

преобразования плотности справедлива при любом распределении вектора \mathbf{U} . Действительно, взяв $A = \{\mathbf{x} : x_1 < a_1, \dots, x_n < a_n\}$ и используя замену переменных $\mathbf{x}B^{-1} = \mathbf{U}$, запишем

$$\int \dots \int_A |\det B|^{-1} f_U(\mathbf{x}B^{-1}) dx_1 \dots dx_n = \int \dots \int_{AB^{-1}} f_U(\mathbf{u}) du_1 \dots du_n =$$

$$= P(\mathbf{U} \in AB^{-1}) = P(\mathbf{UB} \in A),$$

т. е. ф. р. $P(\mathbf{UB} \in A)$ вектора \mathbf{UB} представлена как n -мерный ин-

теграл по множеству A от правой части (3), которая, следовательно, является плотностью распределения вектора UB .

Воспользуемся формулой (3), заметив, что плотность распределения вектора U в (1) равна

$$f_U(u) = f_{U_1}(u_1) \dots f_{U_n}(u_n) = (\sqrt{2\pi})^{-n} \exp(-1/2 uu'). \quad (4)$$

В результате получаем, что плотность $f_{UB}(x)$ дается формулой (2) при $a=0$. Переход к случаю $a \neq 0$ не вызывает затруднений.

Следствие 1. Если матрица B ортогональна: $B'B=I$, то вектор X , определенный формулой (1) при $a=0$, имеет распределение $N_n(0, I)$.

Следствие 2. Функция $f(x)$, заданная формулой (2) при любых a и положительно определенной Q , является плотностью некоторого нормального вектора X . Действительно, представив Q в виде $B'B$, где B — невырожденная матрица (см. (4) § 10), требуемый вектор X можно определить по формуле (1) с данными B и a . ■

Таким образом, можно дать эквивалентное определение n -мерного невырожденного нормального распределения как распределения с плотностью (2). При этом a является вектором среднего, а положительно определенная матрица Q — матрицей ковариаций этого распределения. Примем для него обозначение $N_n(a, Q)$. ■

Отметим частный случай двумерного невырожденного нормального распределения, плотность которого записывают обычно в виде

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-a_1}{\sigma_1} \right)^2 - 2\rho \frac{x_1-a_1}{\sigma_1} \frac{x_2-a_2}{\sigma_2} + \left(\frac{x_2-a_2}{\sigma_2} \right)^2 \right] \right\};$$

его матрица ковариаций равна

$$Q = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Лемма 2. Если вектор X имеет плотность (2) с положительно определенной матрицей Q , то он представим в форме (1).

Доказательство. Представив Q в виде $B'B$ (см. (4) § 10) и положив

$$U = (X-a)B^{-1},$$

найдем по формуле (3)

$$\begin{aligned} f_U(u) &= f_{(X-a)B^{-1}}(u) = |\det B^{-1}|^{-1} f_{X-a}(uB) = \\ &= |\det B|(\sqrt{2\pi})^{-n} |\det Q|^{-1/2} \exp(-1/2 uBQ^{-1}B'u') = \\ &= (\sqrt{2\pi})^{-n} \exp(-1/2 uu'), \end{aligned}$$

так что вектор \mathbf{U} имеет распределение $N_n(\mathbf{0}, I)$, а \mathbf{X} выражается через него по формуле (1). ■

В следующей лемме, а также всюду, где это полезно, будем отмечать нижним индексом размерность вектора и квадратной матрицы.

Лемма 3. Если $\mathbf{X}_n = (X_1, \dots, X_n)$ имеет распределение $N_n(\mathbf{0}_n, Q_n)$ с невырожденной матрицей $Q_n = [q_{ij}, i, j=1, \dots, n]$ и $\mathbf{Y}_k = (X_{i_1}, \dots, X_{i_k})$, $1 \leq i_1 < i_2 < \dots < i_k \leq n$, $k \leq n$, то вектор \mathbf{Y}_k имеет распределение $N_k(\mathbf{0}_k, Q_k)$, где Q_k получается вычеркиванием из Q_n строк и столбцов с номерами, отличными от i_1, \dots, i_k . В частности, компоненты X_i , $i=1, \dots, n$, являются $N(0, q_{ii})$ -распределенными сл. в.

Доказательство. Не ограничивая общности, положим $\mathbf{Y}_k = (X_1, \dots, X_k)$. Пусть

$$\mathbf{X}_n = \mathbf{U}_n \mathbf{B}, \quad \mathbf{B} = [b'_1, \dots, b'_n], \quad b_j = (b_{1j}, \dots, b_{nj}), \quad j=1, \dots, n, \quad (5)$$

и \mathbf{U}_n — $N_n(\mathbf{0}_n, I_n)$ -распределенный вектор. Приведем систему векторов $\mathbf{b}_1, \dots, \mathbf{b}_k$ к ортонормированной системе $\mathbf{c}_1, \dots, \mathbf{c}_n$ таким образом, что

$$\mathbf{b}_j = \lambda_{j1}\mathbf{c}_1 + \dots + \lambda_{jj}\mathbf{c}_j, \quad \lambda_{jj} \neq 0, \quad j=1, \dots, n. \quad (6)$$

Матрица $C = [\mathbf{c}'_1, \dots, \mathbf{c}'_n]$ является ортогональной, и потому вектор

$$\mathbf{Z}_n = \mathbf{U}_n \mathbf{C} \quad (7)$$

имеет распределение $N_n(\mathbf{0}_n, I_n)$. Из (5), (6) и (7) получаем

$$\begin{aligned} \mathbf{X}_j &= \mathbf{U}_n \mathbf{b}'_j = \mathbf{U}_n (\lambda_{j1}\mathbf{c}_1 + \dots + \lambda_{jj}\mathbf{c}_j)' = \lambda_{j1}\mathbf{U}_n \mathbf{c}'_1 + \dots \\ &\dots + \lambda_{jj}\mathbf{U}_n \mathbf{c}'_j = \lambda_{j1} Z_1 + \dots + \lambda_{jj} Z_j, \quad j=1, 2, \dots, n. \end{aligned} \quad (8)$$

Ограничиваая индекс j в соотношениях (8) значениями от 1 до k , получаем, что вектор $\mathbf{Y}_k = (X_1, \dots, X_k)$ получен невырожденным линейным преобразованием из вектора (Z_1, \dots, Z_k) , имеющего $N_k(\mathbf{0}_k, I)$ -распределение, и, следовательно, \mathbf{Y}_k — нормальный вектор. Его среднее и матрица ковариаций, очевидно, таковы, как это утверждается в лемме.

Следствие. Пусть $\mathbf{B} = [b_{ij}, i=1, \dots, n, j=1, \dots, m]$ — $n \times m$ -матрица ранга m , $m < n$, $Q_m = \mathbf{B}'\mathbf{B}$, \mathbf{U}_n — $N(\mathbf{0}_n, I_n)$ -распределенный вектор. Тогда вектор

$$\mathbf{X}_m = \mathbf{U}_n \mathbf{B} + \mathbf{a}_m$$

имеет m -мерное нормальное распределение $N_m(\mathbf{a}_m, Q_m)$ с невырожденной матрицей ковариаций Q_m . Действительно, достаточно дополнить матрицу \mathbf{B} до невырожденной $n \times n$ -матрицы и воспользоваться леммой 3.

Лемма 4. Пусть вектор $\mathbf{X} = (X_1, \dots, X_n)$ имеет невырожденное нормальное распределение, $\mathbf{Y} = (X_{i_1}, \dots, X_{i_m})$, $\mathbf{Z} = (X_{i_m+1}, \dots, X_{i_n})$, где (i_1, i_2, \dots, i_n) — некоторая перестановка из чисел

$(1, 2, \dots, n)$. Если $\text{cov}(X_{i_k}, X_{i_l}) = 0$, $k=1, \dots, m$; $l=m+1, \dots, n$, то векторы \mathbf{Y} и \mathbf{Z} независимы.

Доказательство. Вектор (\mathbf{Y}, \mathbf{Z}) , полученный из нормального вектора \mathbf{X} перестановкой координат, очевидно, нормален. Матрица ковариаций вектора (\mathbf{Y}, \mathbf{Z}) имеет по условию вид

$$R_{\mathbf{Y}, \mathbf{Z}} = \begin{bmatrix} R_{\mathbf{Y}} & \mathbf{0}_{m, n-m} \\ \mathbf{0}_{n-m, m} & R_{\mathbf{Z}} \end{bmatrix},$$

где $\mathbf{0}_{i,j}$ обозначает $i \times j$ -матрицу, целиком состоящую из нулей. В таком случае

$$R_{(\mathbf{Y}, \mathbf{Z})}^{-1} = \begin{bmatrix} R_{\mathbf{Y}}^{-1} & \mathbf{0}_{m, n-m} \\ \mathbf{0}_{n-m, m} & R_{\mathbf{Z}}^{-1} \end{bmatrix}$$

и плотность вектора (\mathbf{Y}, \mathbf{Z}) распадается в произведение двух нормальных плотностей: $N_m(\mathbf{0}_m, R_{\mathbf{Y}})$ и $N_{n-m}(\mathbf{0}_{n-m}, R_{\mathbf{Z}})$, что и доказывает утверждение.

2. Случайные векторы с вырожденной матрицей ковариаций.

Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — случайный вектор с произвольным законом распределения, $\mathbf{a} = M\mathbf{X}$ и $Q = R_{\mathbf{X}}$ существуют, причем матрица ковариаций Q имеет ранг $m < n$. Покажем, что в этом случае распределение вероятностного вектора $\mathbf{X} - \mathbf{a}$ сосредоточено в некотором m -мерном линейном подпространстве пространства R^n .

Запишем представление (см. (4) § 10)

$$\mathbf{v} Q \mathbf{v}' = \mathbf{v} B' B \mathbf{v}' = (\mathbf{v} B')^2$$

где $n \times n$ -матрица $B' = [\mathbf{b}'_1, \dots, \mathbf{b}'_n]$ имеет ранг m . Для любого вектора $\mathbf{v} = (v_1, \dots, v_n)$ квадратичная форма

$$\mathbf{v} Q \mathbf{v}' = \mathbf{v} B' B \mathbf{v}' = (\mathbf{v} B')^2$$

обращается в нуль на $(n-m)$ -мерном линейном подпространстве

$$V = \{\mathbf{v} : \mathbf{v} Q \mathbf{v}' = 0\} = \{\mathbf{v} : \mathbf{v} B' = 0\}. \quad (9)$$

Следовательно, для $v \in V$

$$\mathbf{D}\mathbf{v}\mathbf{X}' = \mathbf{v} Q \mathbf{v}' = 0$$

и с вероятностью 1

$$\mathbf{v} \mathbf{X}' = M \mathbf{v} \mathbf{X}' = \mathbf{v} \mathbf{a}', \text{ т. е. } \mathbf{v} (\mathbf{X}' - \mathbf{a}') = 0.$$

Полученное соотношение показывает, что вектор $\mathbf{X}' - \mathbf{a}'$ с вероятностью 1 лежит в m -мерном подпространстве L , являющимся ортогональным дополнением к V . Как видно из (9), подпространство L порождается столбцами матрицы B' .

Нетрудно понять, что распределение вектора $\mathbf{X} - \mathbf{a}$ не будет сосредоточено ни в каком собственном подпространстве $L^* \subset L$. Действительно, если допустить, что с вероятностью 1 $\mathbf{X} - \mathbf{a} \in L^*$,

то найдется ненулевой вектор $\mathbf{U} \in L$ и ортогональный к L^* , такой, что с вероятностью 1

$$\mathbf{u}(\mathbf{X}' - \mathbf{a}') = 0, \quad \mathbf{u}\mathbf{X}' = \mathbf{u}\mathbf{a}'$$

и, следовательно,

$$\mathbf{u}Q\mathbf{u}' = \mathbf{D}\mathbf{u}\mathbf{X}' = 0,$$

т. е. $\mathbf{u} \in V$, что приводит к противоречию, так как V и L ортогональны.

Замечание. Подчеркнем, что подпространство L , где сосредоточено распределение вероятностей вектора \mathbf{X} , определяется матрицей Q и не зависит от выбора матрицы B в представлении $Q = B'B$.

3. Вырожденное нормальное распределение.

Расширим понятие многомерного нормального распределения, допуская, что в определении (1) матрица B — любая ненулевая $n \times m$ -матрица. Таким образом, распределение вектора

$$\mathbf{X}_m = \mathbf{U}_n B + \mathbf{a}_m \quad (10)$$

при любых m, n , ненулевой матрице B и $\mathbf{U}_n = (U_1, \dots, U_n)$ — стандартном нормальном векторе называется m -мерным нормальным распределением. Невырожденное m -мерное нормальное распределение получается из (10), когда матрица B имеет ранг m (см. следствие к лемме 3). Если матрица B имеет ранг $k < m$, то матрица ковариаций $Q = B'B$ вектора \mathbf{X}_m имеет ранг k и, как доказано в предыдущем разделе, распределение вероятностей вектора $\mathbf{X}_m - \mathbf{a}_m$ сосредоточено в k -мерном подпространстве $L \subset R^m$, являющемся ортогональным дополнением к подпространству

$$V = \{\mathbf{v} : \mathbf{v}Q\mathbf{v}' = 0\}. \quad (11)$$

Покажем, что распределение вероятностей в подпространстве L задается k -мерной нормальной плотностью. Выберем в L ортонормированный базис $\mathbf{e}_1, \dots, \mathbf{e}_k$ и дополним его векторами $\mathbf{e}_{k+1}, \dots, \mathbf{e}_m$ до ортонормированного базиса всего R^m . Пусть старые координаты $(x_1, \dots, x_m) = \mathbf{x}$ пространства R^m связаны с новыми $(y_1, \dots, y_m) = \mathbf{y}$ формулой

$$\mathbf{x}C = \mathbf{y},$$

где C — ортогональная матрица. Для $\mathbf{x} \in L$ имеем $y_{k+1} = \dots = y_m = 0$. Поэтому вектор

$$\mathbf{Y}_m = (\mathbf{X}_m - \mathbf{a}_m)C$$

имеет координаты $Y_{k+1} = \dots = Y_m = 0$. Умножая (10) на C , запишем

$$\mathbf{Y}_m = \mathbf{U}_n BC.$$

Отбросив в матрице BC последние $m-k$ (нулевых) столбцов и обозначив полученную $n \times k$ -матрицу через A , имеем

$$(Y_1, \dots, Y_k) = \mathbf{U}_n A, \quad (12)$$

где $n \times k$ -матрица A имеет ранг k . Итак, в базисе e_1, \dots, e_m вектор $X_m - a_m$ имеет координаты $(Y_1, \dots, Y_k, 0, \dots, 0)$, причем ввиду (12) вектор (Y_1, \dots, Y_k) имеет k -мерную нормальную плотность $N_k(0_k, C'B'BC) = N_k(0_k, C'QC)$. ■

Проведенные рассуждения показывают, в частности, что распределение вероятностей вектора $X_m - a_m$ от матрицы B зависит только через матрицу ковариаций $Q = B'B$. Поэтому мы сохраним за нормальными распределениями общего вида обозначение $N_m(a, Q)$. Аналогично тому, как это сделано в следствии 2 из леммы 1, легко показать, используя формулу (4а) § 10, что для любого вектора a и любой ненулевой неотрицательно определенной матрицы Q существует вектор X с распределением $N(a, Q)$. ■

Отметим еще одну полезную связь нормального распределения общего вида с невырожденным нормальным. Выберем среди столбцов $n \times m$ -матрицы

$$B = [b'_1, \dots, b'_m]$$

ранга k в представлении (10) какие-либо k линейно независимых столбцов $b'_{i_1}, \dots, b'_{i_k}$, так что остальные столбцы через них линейно выражаются

$$b_j = \lambda_{j1}b_{i_1} + \dots + \lambda_{jk}b_{i_k}.$$

Тогда $n \times k$ -матрица $[b'_{i_1}, \dots, b'_{i_k}]$ имеет ранг k , поэтому вектор

$$(X_{i_1}, \dots, X_{i_k}) = U_n[b'_{i_1}, \dots, b'_{i_k}] + (a_{i_1}, \dots, a_{i_m})$$

имеет невырожденное k -мерное нормальное распределение, а остальные компоненты вектора X линейно выражаются через X_{i_1}, \dots, X_{i_k} :

$$\begin{aligned} X_j - a_j &= U_n b'_j = U_n (\lambda_{j1}b_{i_1} + \dots + \lambda_{jk}b_{i_k}) = \\ &= (\lambda_{j1}, \dots, \lambda_{jk}) U_n [b'_{i_1}, \dots, b'_{i_k}] = \\ &= (\lambda_{j1}, \dots, \lambda_{jk})(X_{i_1} - a_{i_1}, \dots, X_{i_k} - a_{i_k})'. \quad ■ \end{aligned}$$

Лемма 5. Пусть $X_m = (X_1, \dots, X_m)$ — нормально распределенный вектор, заданный формулой (10),

$$Y = (X_{i_1}, \dots, X_{i_k}), \quad Z = (X_{j_1}, \dots, X_{j_l}),$$

где (i_1, \dots, i_k) и (j_1, \dots, j_l) — две непересекающиеся последовательности индексов. Если

$$\text{cov}(X_{i_s}, X_{j_r}) = 0, \quad s = 1, \dots, k; \quad r = 1, \dots, l,$$

то векторы Y и Z независимы.

Доказательство. Из определения (10) вытекает, что вектор, составленный из части компонент X_m , также имеет нормальное

распределение. Поэтому, не ограничивая общности, предположим, что

$$\mathbf{Y} = (X_1, \dots, X_k), \quad \mathbf{Z} = (X_{k+1}, \dots, X_m),$$

$$\text{cov}(X_i, X_j) = 0, \quad i = 1, \dots, k; \quad j = k+1, \dots, m.$$

Распишем матрицу B в формуле (10) в виде

$$B = [\mathbf{b}'_1, \dots, \mathbf{b}'_m],$$

так что

$$X_j - a_j = \mathbf{U}_n \mathbf{b}'_j = \mathbf{b}'_j \mathbf{U}'_n,$$

$$\text{cov}(X_i, X_j) = M(\mathbf{b}_i \mathbf{U}'_n \mathbf{U}_n \mathbf{b}'_j) = \mathbf{b}_i \mathbf{b}'_j, \quad i = 1, \dots, k; \quad j = k+1, \dots, m.$$

Используя некоррелированность компонент векторов \mathbf{Y} и \mathbf{Z} , получаем отсюда, что системы векторов $(\mathbf{b}_1, \dots, \mathbf{b}_k)$ и $(\mathbf{b}_{k+1}, \dots, \mathbf{b}_m)$ взаимно ортогональны. Выберем в каждой из этих систем максимальные линейно независимые подсистемы $(\mathbf{b}_{i_1}, \dots, \mathbf{b}_{i_p})$ и $(\mathbf{b}_{j_1}, \dots, \mathbf{b}_{j_q})$. Объединенная система $(\mathbf{b}_{i_1}, \dots, \mathbf{b}_{i_p}, \mathbf{b}_{j_1}, \dots, \mathbf{b}_{j_q})$, очевидно, линейно независима. Поэтому вектор

$$(X_{i_1}, \dots, X_{i_p}, X_{j_1}, \dots, X_{j_q}) = \mathbf{U}_n [\mathbf{b}_{i_1}, \dots, \mathbf{b}_{i_p}, \mathbf{b}_{j_1}, \dots, \mathbf{b}_{j_q}]$$

имеет невырожденное нормальное распределение. И, применяя лемму 4, заключаем, что случайные векторы $(X_{i_1}, \dots, X_{i_p})$ и $(X_{j_1}, \dots, X_{j_q})$ независимы. Для остальных компонент вектора \mathbf{Y} из разложения

$$\mathbf{b}_j = \lambda_{j_1} \mathbf{b}_{j_1} + \dots + \lambda_{j_p} \mathbf{b}_{j_p}, \quad j \leq k,$$

вытекает, что они линейно выражаются через X_{i_1}, \dots, X_{i_p} :

$$X_j - a_j = \mathbf{U}_n \mathbf{b}'_j = \lambda_{j_1} X_{i_1} + \dots + \lambda_{j_p} X_{i_p}.$$

Аналогично оставшиеся компоненты вектора \mathbf{Z} линейно выражаются через X_{j_1}, \dots, X_{j_q} . Таким образом, компоненты векторов \mathbf{Y} и \mathbf{Z} являются функциями двух независимых между собой систем: $(X_{i_1}, \dots, X_{i_p})$ и $(X_{j_1}, \dots, X_{j_q})$, и потому \mathbf{Y} и \mathbf{Z} сами независимы.

Лемма 6. Пусть \mathbf{X}_n — нормальный вектор, $\mathbf{Y}_k = \mathbf{X}_n A$, где A — $n \times k$ -матрица. Тогда вектор \mathbf{Y}_k является либо нормальным, либо постоянным.

Доказательство. Используя представление (10), получаем

$$\mathbf{Y}_k = (\mathbf{U}_n B + \mathbf{a}_m) A = \mathbf{U}_n (BA) + \mathbf{a}_m A,$$

и если матрица BA ненулевая, то \mathbf{Y}_k — нормальный вектор. ■

Пример. Пусть X_1, \dots, X_n независимые сл. в. с общим нормальным распределением $N(\mu, \sigma^2)$. Образуем случайный вектор $\mathbf{Y} = (Y_1, Y_2, \dots, Y_{n+1}) = (X_1 - \bar{X}, \dots, X_n - \bar{X}, \bar{X})$, где \bar{X} — выборочное среднее. По лемме 6 \mathbf{Y} — нормальный вектор, и так как

$$\text{cov}(Y_i, Y_{n+1}) = \text{cov}(X_i - \bar{X}, \bar{X}) = 0, \quad i = 1, \dots, n,$$

то $Y_{n+1} = \bar{X}$ не зависит от $(Y_1, \dots, Y_n) = (X_1 - \bar{X}, \dots, X_n - \bar{X})$. Ранее (см. § 8) было установлено лишь, что \bar{X} не зависит от $\sum_{i=1}^n (X_i - \bar{X})^2$.

4. Распределение проекций стандартного нормального вектора.

Пусть пространство R^n разложено в прямую сумму ортогональных подпространств $V_1, V_2, \dots; U$ — стандартный нормальный вектор. Нас будет интересовать распределение векторов проекций $\text{pr}_{V_i} U, i = 1, 2, \dots$. Выберем ортонормированные базисы: e_1, \dots, e_k в V_1 , e_{k+1}, \dots, e_{k+l} в V_2, \dots и объединим их в базис e_1, \dots, e_n всего пространства R^n .

Пусть C — ортогональная матрица перехода от старых координат $x = (x_1, \dots, x_n)$ к новым $y = (y_1, \dots, y_n)$:

$$xC = y, \quad x = yC'. \quad (13)$$

Полагая $C = [c'_1, c'_2, \dots, c'_n]$, $c_j = (c_{1j}, \dots, c_{nj})$, отметим, что c_{1j}, \dots, c_{nj} — координаты вектора e_j в исходной системе координат. Вектор

$$Y = UC = U[c'_1, \dots, c'_n] \quad (14)$$

ввиду ортогональности матрицы C является стандартным нормальным (см. следствие 1 леммы 1). Поскольку

$$\begin{aligned} \text{pr}_{V_1} U &= Y_1 e_1 + \dots + Y_k e_k, \\ \text{pr}_{V_1} U &= Y_{k+1} e_{k+1} + \dots + Y_{k+l} e_{k+l}, \dots, \end{aligned} \quad (15)$$

то по лемме 6 вектор $(\text{pr}_{V_1} U, \text{pr}_{V_2} U, \dots)$ является нормальным. Чтобы найти координаты $\text{pr}_{V_i} U, i = 1, 2, \dots$, в исходной системе координат, подставим в (15) вместо e_j вектор из их координат c_j в старой системе, $j = 1, 2, \dots, n$, и, принимая во внимание (14), запишем

$$\begin{aligned} (\text{pr}_{V_1} U)' &= [c'_1, \dots, c'_k] (Y_1, \dots, Y_k)' = [c'_1, \dots, c'_k] [c'_1, \dots, c'_k]' U', \\ (\text{pr}_{V_2} U)' &= [c'_{k+1}, \dots, c'_{k+l}] [c'_{k+1}, \dots, c'_{k+l}]' U'. \end{aligned}$$

Взаимные ковариации векторов проекций

$$\begin{aligned} M(\text{pr}_{V_1} U)' (\text{pr}_{V_2} U) &= \\ &= [c'_1, \dots, c'_k] [c'_1, \dots, c'_k] [c'_{k+1}, \dots, c'_{k+l}] [c'_{k+1}, \dots, c'_{k+l}]', \dots \end{aligned}$$

равны нулю из-за ортогональности системы векторов c_1, c_2, \dots . Поэтому случайные вектора $\text{pr}_{V_i} U, i = 1, 2, \dots$ — взаимно независимы. Наконец отметим, что

$$\|\text{pr}_{V_1} U\|^2 = Y_1^2 + \dots + Y_k^2, \quad \|\text{pr}_{V_2} U\|^2 = Y_{k+1}^2 + \dots + Y_{k+l}^2, \dots,$$

и, значит, квадраты длин проекций имеют χ^2 -распределения с числами степеней свободы, равными размерностям подпространств, на которые производятся проектирования. ■

Для дальнейшего полезно отметить, что распределение сл. в. $\|U + a\|^2$, где a — произвольный числового вектор, зависит от a лишь через его длину $\|a\|$. Действительно, выберем в R^n новый ортонормированный базис, приняв за $e_1 = a/\|a\|$. Пусть в новом базисе вектор U имеет координаты $(Z_1, Z_2, \dots, Z_n) = Z$, причем Z — стандартный нормальный вектор; вектор a имеет координаты $(\|a\|, 0, \dots, 0)$. Таким образом,

$$\|U + a\|^2 = (Z_1 + \|a\|)^2 + \sum_{i=2}^n Z_i^2. \quad (16)$$

Распределение сл. в. (16) называется *нецентральным χ^2 -распределением* с n степенями свободы и параметром нецентральности $\delta = \|a\|^2$. ■

Для последующих применений сделаем еще одно замечание. Между распределениями вероятностей величин $\|U\|^2$ и $\|U+a\|^2$ имеет место при $\|a\| \neq 0$ неравенство

$$\mathcal{P}(\|U+a\|^2 > u) > \mathcal{P}(\|U\|^2 > u), \quad (17)$$

которое легко вывести из (16), учитывая следующее очевидное неравенство:

$$\mathcal{P}(|Z_1 + \|a\|| > z) > \mathcal{P}(|Z_1| > z), \quad z > 0.$$

Про случайные величины, связанные знаком неравенства (17), говорят, что одна из них *стохастически больше* другой.

§ 13. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ И ПРОВЕРКА ГИПОТЕЗ В ЛИНЕЙНОЙ МОДЕЛИ С НОРМАЛЬНЫМИ НАБЛЮДЕНИЯМИ

1. Распределение вектора оценок.

Мы будем рассматривать линейную модель:

$$Y = \theta X' + \sigma e, \quad M\epsilon = 0, \quad R_e = I,$$

$$Y = (Y_1, \dots, Y_n), \quad \theta = (\theta_1, \dots, \theta_r), \quad r \leq n, \quad (1)$$

$$e = (e_1, \dots, e_n), \quad X = [x'_1, \dots, x'_r], \quad x_j = (x_{1j}, \dots, x_{nj}), \\ j = 1, \dots, r.$$

Обозначим через V пространство, порожденное столбцами матрицы X , в котором лежит неизвестный вектор среднего:

$$\eta = MY = \theta_1 x_1 + \dots + \theta_r x_r = \theta X'. \quad (2)$$

Размерность V обозначим через $d = \dim V$. Наилучшая линейная несмещенная оценка вектора среднего равна

$$\hat{\eta} = \hat{\theta}_1 x_1 + \dots + \hat{\theta}_r x_r = \hat{\theta} X' = \text{пр}_{V'} Y.$$

Оценка дисперсии равна (см. (16) § 10)

$$\hat{\sigma}^2 = (n - d)^{-1} \| \text{пр}_{(R^n \ominus V)} Y \|^2. \quad (3)$$

Если векторы x_1, \dots, x_n линейно независимы, то $\hat{\theta}_i, i = 1, \dots, r$, однозначно определены и могут быть представлены формулой

$$\hat{\theta} = YX(X'X)^{-1}.$$

В противном случае разложение (2) однозначно определяет только такие параметрические функции $a_1\theta_1 + \dots + a_r\theta_r$, которые являются линейными функциями от координат вектора η :

$$a\theta = b\eta' = bX\theta', \quad b = (b_1, \dots, b_n),$$

т. е. параметрические функции с коэффициентами, подчиненными условию

$$a = bX. \quad (4)$$

Наилучшей для $a\theta'$ является оценка bY' , где $b \in V$ и удовлетворяет (4).

В этом параграфе мы дополнительно предположим, что вектор e является стандартным нормальным. В таком случае нормальными являются векторы оценок $\hat{\theta}, \hat{\eta}$ (при этом

$$M\hat{\theta} = \theta, \quad R_{\hat{\theta}} = \sigma^2 (X'X)^{-1},$$

$$M\hat{\eta} = \eta, \quad R_{\hat{\eta}} = \sigma^2 X (X'X)^{-1} X',$$

а также любой набор наилучших оценок параметрических функций

$$b_1 \hat{\eta}', \dots, b_k \hat{\eta}',$$

где $b_i \in V, i = 1, \dots, k$. Оценка дисперсии (3) является независимой от $\hat{\eta}, \hat{\theta}$ и от любой совокупности оценок параметрических функций (см. п. 4 § 12), а ее распределение с точностью до параметра масштаба совпадает с χ^2 -распределением:

$$\text{пр}_{R^n \ominus V} Y = \text{пр}_{R^n \ominus V} \eta + \text{пр}_{R^n \ominus V} \sigma e = \text{пр}_{R^n \ominus V} \sigma e,$$

так что

$$(n - \dim V) \hat{\sigma}^2 / \sigma^2 = \| \text{пр}_{R^n \ominus V} e \|^2$$

имеет распределение χ^2 с числом степеней свободы $n - \dim V$.

2. Доверительные области для параметров и параметрических функций.

Обозначим через q_{ij} элементы матрицы $(X'X)^{-1}$. Тогда $\hat{\theta}_i$ имеет распределение $N(\theta_i, \sigma^2 q_{ii})$, а сл. в.

$$(\hat{\theta}_i - \theta_i)/(\hat{\sigma} \sqrt{q_{ii}}) = ((\hat{\theta}_i - \theta_i)/(\hat{\sigma} \sqrt{q_{ii}}))/(\hat{\sigma}/\sigma)$$

имеет t -распределение с числом степеней свободы $n-d$, где $d = \dim V$, что позволяет строить доверительные интервалы для θ_i в виде

$$\hat{\theta}_i - x_{1-\alpha/2}(t_{n-d}) \hat{\sigma} \sqrt{q_{ii}} < \theta_i < \hat{\theta}_i + x_{1-\alpha/2}(t_{n-d}) \hat{\sigma} \sqrt{q_{ii}}, \quad (5)$$

где $x_\alpha(t_m)$ — α -квантиль распределения t_m .

Обычно представляет интерес одновременное доверительное оценивание ряда параметров или параметрических функций. Рассчитать вероятность одновременного выполнения нескольких неравенств типа (5) хотя и возможно, но весьма затруднительно. Иной способ одновременного доверительного оценивания коэффициентов θ (или параметрических функций) заключается в построении доверительного эллипса. Рассмотрим этот метод в применении к оценке всего набора $\theta = (\theta_1, \dots, \theta_r)$. Воспользуемся тем, что

$$\hat{\eta} - \eta = \text{pr}_V Y - \eta = \text{pr}_V (Y - \eta) = \sigma \text{pr}_V e,$$

и, следовательно, величина

$$\sigma^{-2} \|\hat{\eta} - \eta\|^2 = \sigma^{-2} \|(\hat{\theta} - \theta) X'\|^2 = \sigma^{-2} (\hat{\theta} - \theta)' X' X (\hat{\theta} - \theta)$$

имеет распределение χ^2 с числом степеней свободы $d = \dim V$ и не зависит от оценки дисперсии $\hat{\sigma}^2$. Поэтому сл. в.

$$d^{-1} \hat{\sigma}^{-2} \|\hat{\eta} - \eta\|^2 = d^{-1} \hat{\sigma}^{-2} (\hat{\theta} - \theta)' X' X (\hat{\theta} - \theta), \quad (6)$$

имеет распределение $F_{d, n-d}$, так что с вероятностью $1-\alpha$ выполняется неравенство

$$d^{-1} \hat{\sigma}^{-2} (\hat{\theta} - \theta)' X' X (\hat{\theta} - \theta) \leq x_{1-\alpha}(F_{d, n-d}). \quad (7)$$

Допустим, что матрица X имеет полный ранг: $d=r$. Тогда область переменных $(\theta_1, \dots, \theta_r)$, удовлетворяющая (7), представляет собой эллипсоид, являющийся α -доверительным множеством в пространстве параметров R^r : неизвестная точка $(\theta_1, \dots, \theta_r)$ накрывается эллипсоидом (7) с вероятностью $1-\alpha$. Заметим, что доверительный эллипсоид (7) (при $d=r$) получается из следующего не зависящего от выборки эллипса:

$$r^{-1} \theta' X' X \theta \leq x_{1-\alpha}(F_{d, n-d})$$

растяжением с коэффициентом $\hat{\sigma}$ и переносом на вектор $\hat{\theta}$.

Линейная модель (1) фактически определена подпространством V , в котором лежит вектор среднего η ; α -доверительным множеством для вектора η является d -мерный шар ($d = \dim V$):

$$d^{-1}\hat{\sigma}^{-2}\|\hat{\eta} - \eta\|^2 \leq x_{1-\alpha}(F_{d,n-d}), \quad \eta \in V. \quad (8)$$

Выбирая различные базисы в пространстве V , будем получать для координат вектора η в этих базисах описанные выше α -доверительные эллипсоиды; в частном случае, когда базис ортонормированный, эллипсоид превращается в шар.

Если V^s — подпространство V размерности s , то аналогично (8) можно записать α -доверительный эллипсоид для $\eta_s = \text{пр}_{V^s}\eta$. Полагая $\hat{\eta}_s = \text{пр}_{V^s}\hat{\eta}$, имеем

$$s^{-1}\hat{\sigma}^{-2}\|\hat{\eta}_s - \eta_s\|^2 \leq x_{1-\alpha}(F_{s,n-d}), \quad \eta \in V. \quad (9)$$

В развернутой форме (9) имеет вид

$$\frac{1}{s} \|\text{пр}_{V^s}(\hat{\eta} - \eta_s)\|^2 / \left(\frac{1}{n-d} \|\text{пр}_{R^{n-d}}\hat{\eta}\|^2 \right) \leq x_{1-\alpha}(F_{s,n-d}), \quad \eta \in V. \quad (10)$$

3. Проверка гипотез с помощью доверительных эллипсоидов.

Рассмотрим линейную модель

$$\hat{\eta} = \eta + \sigma e, \quad \eta \in V^d, \quad (11)$$

где неизвестный вектор среднего η лежит в заданном линейном подпространстве V^d размерности d . До тех пор пока мы работаем с данной моделью, предположение $\eta \in V^d$ не подвергается сомнению. Естественной гипотезой в линейной модели оказывается предположение о том, что η лежит в некотором подпространстве размерности $k < d$: $V^k \subseteq V^d$. Назовем это предположение гипотезой H . Иначе можно записать

$$H : \text{пр}_{V^d} \Theta_{V^k} \eta = 0.$$

Рассмотрим α -доверительный эллипсоид (10) для вектора $\eta_s = \text{пр}_{V^s}\eta$, $V^s = V^d \ominus V^k$, $s = d - k$. В соответствии с методологией, которую мы уже неоднократно применяли в гл. I, если α -доверительный эллипсоид содержит вектор 0, то гипотезу $H : \eta_s = 0$ следует принять на уровне значимости α . Подставляя $\eta_s = 0$ в (10), получаем статистику критерия для проверки гипотезы H в виде

$$F = \frac{1}{d-k} \|\text{пр}_{V^d} \Theta_{V^k} \hat{\eta}\|^2 / \left(\frac{1}{n-d} \|\text{пр}_{R^{n-d}} \hat{\eta}\|^2 \right); \quad (12)$$

в предположении H она имеет $F_{d-k, n-d}$ -распределение.

Оптимальные свойства сформулированного F -критерия будут рассмотрены в других главах. Пока мы приведем некоторые дополнительные соображения в пользу этого критерия. Заметим

прежде всего, что знаменатель статистики F , с точностью до постоянного множителя, равен

$$\|\text{pr}_{R^n \ominus V^d} \mathbf{Y}\|^2 = \|\mathbf{Y} - \text{pr}_{V^d} \mathbf{Y}\|^2 = \|\mathbf{Y} - \hat{\eta}\|^2 \quad (13)$$

и представляет собой меру отклонения вектора наблюдений \mathbf{Y} от оценки его вектора среднего $\hat{\eta}$. Если гипотеза H верна, то за оценку вектора среднего следует принять вектор проекции \mathbf{Y} на подпространство V^k : $\hat{\eta}_k = \text{pr}_{V^k} \mathbf{Y}$, и тогда мера отклонения \mathbf{Y} от $\hat{\eta}_k$ равна $\|\mathbf{Y} - \hat{\eta}_k\|^2$. Но

$$\mathbf{Y} = \text{pr}_{V^k} \mathbf{Y} + \text{pr}_{V^d \ominus V^k} \mathbf{Y} + \text{pr}_{R^n \ominus V^d} \mathbf{Y}$$

и, следовательно,

$$\begin{aligned} \|\mathbf{Y} - \hat{\eta}_k\|^2 &= \|\text{pr}_{V^d \ominus V^k} \mathbf{Y} + \text{pr}_{R^n \ominus V^d} \mathbf{Y}\|^2 = \\ &= \|\text{pr}_{V^d \ominus V^k} \mathbf{Y}\|^2 + \|\text{pr}_{R^n \ominus V^d} \mathbf{Y}\|^2. \end{aligned} \quad (14)$$

Отношение (14) к (13) равно

$$1 + \|\text{pr}_{V^d \ominus V^k} \mathbf{Y}\|^2 / \|\text{pr}_{R^n \ominus V^d} \mathbf{Y}\|^2 = 1 + \frac{d-k}{n-d} F. \quad (15)$$

Если гипотеза H верна, то

$$\|\text{pr}_{V^d \ominus V^k} \mathbf{Y}\|^2 = \sigma^2 \|\text{pr}_{V^d \ominus V^k} \mathbf{e}\|^2, \quad (16)$$

т. е. увеличение (14) по сравнению с (13) вызвано только случайными причинами. Если же гипотеза H нарушена, то

$$\begin{aligned} \|\text{pr}_{V^d \ominus V^k} \mathbf{Y}\|^2 &= \|\text{pr}_{V^d \ominus V^k} (\eta + \sigma \mathbf{e})\|^2 = \\ &= \|\text{pr}_{V^d \ominus V^k} \eta + \sigma \text{pr}_{V^d \ominus V^k} \mathbf{e}\|^2. \end{aligned} \quad (17)$$

Величина (17) стохастически больше (16) (см. п. 4 § 12), и, следовательно, большие значения статистики (15) могут рассматриваться как свидетельство против гипотезы H . С точки зрения задачи проверки гипотезы статистика (15) равнозначна статистике F . Подводя итог, отметим, что распределение статистики F свободно от мешающего параметра σ , критическая область с доверительным уровнем α имеет вид

$$F > x_{1-\alpha}(F_{d-k, n-d}). \quad (18)$$

F -критерий (18) называют также *дисперсионным критерием*, так как статистика F представляет собой отношение двух независимых оценок дисперсий: в знаменателе — несмещенная оценка $\hat{\sigma}^2$, в числителе — независящая от знаменателя статистика, являющаяся несмещенной оценкой σ^2 в предположении гипотезы H .

4. Пример: сравнение средних в нескольких нормальных выборках.

Пусть имеется r независимых выборок:

$$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i}), \quad i = 1, \dots, r, \quad n_1 + \dots + n_r = n,$$

причем наблюдения внутри выборки независимые и $N(\mu_i, \sigma^2)$ -распределенные. Положив $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_r)$ и обозначив через $\mathbf{1}_l$, $\mathbf{0}_l$ вектора размерности l , целиком состоящие из 1 и 0 соответственно, образуем линейную модель

$$\mathbf{Y}' = \begin{bmatrix} \mathbf{1}_{n_1}^\top \mathbf{0}_{n_1}^\top \dots \mathbf{0}_{n_r}^\top \\ \mathbf{0}_{n_1}^\top \mathbf{1}_{n_2}^\top \dots \mathbf{0}_{n_r}^\top \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n_r}^\top \mathbf{0}_{n_r}^\top \dots \mathbf{1}_{n_r}^\top \end{bmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \end{pmatrix} + \sigma \mathbf{e}', \quad (19)$$

где $\mathbf{e} = (e_1, \dots, e_n)$ — стандартный нормальный вектор. Матрица X линейной модели (19) — ортогональная, поэтому легко находятся оценки

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}, \quad i = 1, \dots, r.$$

Оценка дисперсии имеет вид

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-r} \| \mathbf{Y} - (\hat{\mu}_1 \mathbf{1}_{n_1}, \hat{\mu}_2 \mathbf{1}_{n_2}, \dots, \hat{\mu}_r \mathbf{1}_{n_r}) \|^2 = \\ &= \frac{1}{n-r} \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2 = \frac{1}{n-r} \sum_{i=1}^r \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{1}{n-r} \sum_{i=1}^r n_i \hat{\mu}_i^2. \end{aligned}$$

Рассмотрим гипотезу $H: \mu_1 = \mu_2 = \dots = \mu_r$. Подпространство V^k , где в соответствии с гипотезой H лежит вектор среднего η , в данном случае одномерно: $k=1$ и порождается вектором $\mathbf{1}_n$. Вычисляем

$$\| \text{pr}_{V^k \ominus V^1} \mathbf{Y} \|^2 = \| \text{pr}_{V^1} \mathbf{Y} \|^2 - \| \text{pr}_{V^k} \mathbf{Y} \|^2 = \sum_{i=1}^r n_i \hat{\mu}_i^2 - n \bar{Y}^2.$$

Таким образом, статистика критерия равняется

$$\frac{1}{r-1} \left(\sum_{i=1}^r n_i \hat{\mu}_i^2 - n \bar{Y}^2 \right) / \frac{1}{n-r} \left(\sum_{i=1}^r \sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^r n_i \hat{\mu}_i^2 \right). \quad (20)$$

Приведем численный пример, заимствованный из [7]. Производились измерения окружностей головы $n=142$ черепов, принадлежащих к $r=3$ сериям, при этом $n_1=83$, $n_2=51$, $n_3=8$. Результаты измерений приведены ниже в виде суммарных характеристик:

$$\sum_{j=1}^{n_1} y_{1j} = 11277; \quad \sum_{j=1}^{n_2} y_{2j} = 7049; \quad \sum_{j=1}^{n_3} y_{3j} = 1102;$$

$$\sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij}^2 - \bar{y}^2) = 4616,64,$$

откуда

$$\hat{\mu}_1 = 135,87; \quad \hat{\mu}_2 = 138,22; \quad \hat{\mu}_3 = 137,75;$$

$$\sum_{i=1}^r \sum_{j=1}^{n_i} y_{ij} = 19428; \quad \bar{y} = 136,817.$$

Дальнейшие вычисления удобно провести по формулам

$$\begin{aligned} \sum_{i=1}^r n_i \hat{\mu}_i^2 - n \bar{y}^2 &= \sum_{i=1}^r (\hat{\mu}_i - \bar{y}) \sum_{j=1}^{n_i} y_{ij}, \\ \sum_{i=1}^r \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^r n_i \hat{\mu}_i^2 &= \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{ij}^2 - \bar{y}^2) - \left(\sum_{i=1}^r n_i \hat{\mu}_i^2 - n \bar{y}^2 \right), \end{aligned}$$

которые приводят соответственно к значениям $238,59$ и $4616,4 - 238,59 \approx 4378$;

$$F_{\text{набл}} = \frac{1}{2} 238,59 / \left(\frac{1}{139} 4378 \right) = 119,29 / 31,50 = 3,79.$$

Найдем p -квантиль $F_{2,139}$ -распределения для $p_1=0,95$ и $p_2=0,975$ по табл. 3.5 [1]. Имеем по таблице

$$x_{p_1}(F_{2,120}) = 3,0718; \quad x_{p_1}(F_{2,\infty}) = 2,9957;$$

$$x_{p_2}(F_{2,120}) = 3,8046; \quad x_{p_2}(F_{2,\infty}) = 3,6889.$$

Значения $x_p(F_{m,n})$ для $n=139$ вычислим линейной интерполяцией по аргументу $1/n$:

$$\begin{aligned} x_{p_1}(F_{2,139}) &= x_{p_1}(F_{2,120}) + (x_{p_1}(F_{2,\infty}) - x_{p_1}(F_{2,120})) \times \\ &\times \left(\frac{1}{139} - \frac{1}{120} \right) \left(-\frac{1}{120} \right)^{-1} = 3,0718 - (3,0718 - 2,9957) 0,1367 = \\ &= 3,0614; \end{aligned}$$

$$x_{p_2}(F_{2,139}) = 3,8046 - (3,8046 - 3,6889) 0,1367 = 3,7888.$$

Итак, наблюденный уровень значимости

$$\alpha_{\text{набл}} = P(F \geq F_{\text{набл}}) \approx 0,025,$$

что можно рассматривать как неудовлетворительное согласие с гипотезой. Во всяком случае для уровня значимости $\alpha=0,05$ $F_{\text{набл}}=3,79$ значительно превосходит границу $x_{0,95}=3,0614$, и на этом уровне гипотеза должна быть отвергнута.

ГЛАВА III

ДОСТАТОЧНЫЕ СТАТИСТИКИ

§ 14. СТАТИСТИЧЕСКАЯ МОДЕЛЬ, ПОДОБНЫЕ СТАТИСТИКИ

1. Статистическая модель.

В предыдущих главах были приведены примеры реальных статистических данных и предложены вероятностные модели для их описания. Объектом исследования являлась последовательность сл. в. $X_i = X_i(\omega)$, $i = 1, 2, \dots$, определенных на основном вероятностном пространстве $(\Omega, \mathcal{A}, \{\mathcal{P}\})$ (см. п. 4 § 1). Вектор наблюдений $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ определяет вероятностную меру P :

$$P(B) = \mathcal{P}(\{\omega : X_n(\omega) \in B\}), \quad (1)$$

заданную на системе \mathcal{B}_n так называемых *борелевских множеств* R^n : \mathcal{B}_n — наименьшая система подмножеств R^n , включающая множества вида $\{x_n = (x_1, \dots, x_n) : a_i < x_i < b_i, i = 1, \dots, n\}$, $a_i < b_i$ — произвольные числа, и замкнутая относительно перехода к дополнению и образования счетного объединения. Меру $P(B)$ мы будем называть *распределением вероятностей* вектора \mathbf{X}_n . В последующем мы встретимся с моделями, в которых число наблюдений n заранее не фиксировано и результатами наблюдений могут оказаться числовые векторы $x_n = (x_1, \dots, x_n)$ разной размерности n . Имея в виду это обстоятельство, введем отдельное обозначение $\mathcal{X} = \{x\}$ для пространства выборок x , не накладывая ограничений на структуру этого множества, хотя, как правило, речь будет идти о $\mathcal{X} = R^n$, $x = x_n = (x_1, \dots, x_n)$. В случае выборочного множества \mathcal{X} произвольной структуры будем считать, что в \mathcal{X} выделена система \mathcal{B} так называемых *измеримых* множеств. Система \mathcal{B} должна содержать \mathcal{X} как свой элемент, должна быть замкнута относительно перехода к дополнению и образованию счетного объединения. Отображение $X(\omega)$ множества Ω в \mathcal{X} , такое, что для любого $B \subset \mathcal{B}$

$$\{\omega : X(\omega) \in B\} \in \mathcal{A}$$

называется измеримым; в случае, когда фиксирована мера \mathcal{P} на (Ω, \mathcal{A}) , функцию $X(\omega)$ называют *случайным элементом на* $(\mathcal{X}, \mathcal{B})$. Пару из множества и системы измеримых подмножеств называют *измеримым пространством*; таковы пары (Ω, \mathcal{A}) , $(\mathcal{X}, \mathcal{B})$. Название связано с возможностью задать меру на систе-

ме измеримых подмножеств — это мера \mathcal{P} в случае (Ω, \mathcal{A}) и мера \mathbf{P} , определяемая формулой

$$\mathbf{P}(B) = \mathcal{P}(\omega : X(\omega) \in B), \quad (2)$$

в случае пространства $(\mathcal{X}, \mathcal{B})$.

Предметом изучения теории вероятностей является отдельное *вероятностное пространство* $(\Omega, \mathcal{A}, \mathcal{P})$. Как мы могли убедиться в предшествующих главах, в математической статистике вероятностная модель, которая предлагается для описания опыта, определена не полностью. И как раз задача статистики — уточнить эту модель по результатам испытаний. Другими словами, модель математической статистики определяется целым семейством вероятностных мер $\{\mathcal{P}\}$ на (Ω, \mathcal{A}) , и тем самым объектом изучения в математической статистике оказывается семейство вероятностных пространств $\{(\Omega, \mathcal{A}, \mathcal{P})\}$, где \mathcal{P} принадлежит некоторому заданному множеству $\{\mathcal{P}\}$. Формула (2) порождает семейство вероятностных мер $\{\mathbf{P}\}$ на измеримом пространстве $(\mathcal{X}, \mathcal{B})$. Тройку $(\mathcal{X}, \mathcal{B}, \{\mathbf{P}\})$ называют *статистической моделью*. При фиксированном \mathbf{P} вероятностное пространство $(\mathcal{X}, \mathcal{B}, \mathbf{P})$ назовем *выборочным*. Если семейство $\{\mathbf{P}\}$ параметризовано и имеет вид $\{\mathbf{P}_\theta : \theta \in \Theta\}$, где θ — скалярный или векторный параметр, то модель $(\mathcal{X}, \mathcal{B}, \{\mathbf{P}_\theta : \theta \in \Theta\})$ называют *параметрической*.

Приведем некоторые примеры уже известных нам статистических моделей.

(I) $\mathcal{X} = R^n$; $\mathcal{B} = \mathcal{B}_n$ — система борелевских множеств в R^n ; мера \mathbf{P}_F , где F — одномерная непрерывная ф. р., задается n -мерной ф. р. вида

$$F(x_1) \cdot \dots \cdot F(x_n),$$

что соответствует независимой выборке с ф. р. отдельного наблюдения $F(x)$. Семейство $\{\mathbf{P}_F\}$ образуется из всех \mathbf{P}_F с непрерывной ф. р. F . К этой модели мы по существу обращались при доверительном оценивании теоретической ф. р. через э. ф. р. с помощью статистики Колмогорова (§ 2), а также при построении доверительных интервалов для квантилей (§ 3).

(II) Измеримое пространство $(\mathcal{X}, \mathcal{B})$ и мера \mathbf{P}_F те же, что и в (I), но F пробегает множество ф. р., отличающихся сдвигом и масштабом от некоторой заданной ф. р. $F_0(x)$. (Например, $F_0(x) = 1 - e^{-x}$, $x > 0$ или $F_0(x) = \Phi(x)$.) Семейство $\{\mathbf{P}\}$ в данном случае параметризовано: каждое $\mathbf{P}_{\mu, \sigma}$ задается парой чисел (μ, σ) , $-\infty < \mu < \infty$, $\sigma > 0$.

(III) \mathcal{X}, \mathcal{B} — те же, что и выше, семейство $\{\mathbf{P}\}$ состоит из мер \mathbf{P} следующего вида. Пусть $F(x_1, \dots, x_n)$ — соответствующая \mathbf{P} n -мерная ф. р., $F_i(x_i)$, $F_{ij}(x_i, x_j)$ — наборы одномерных и двумерных ф. р.:

$$F_i(x_i) = F(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty),$$

$$F_{ij}(x_i, x_j) = F(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty, x_j, +\infty, \dots, +\infty).$$

Предполагается, что выполнены следующие соотношения:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j dF_{ij}(x_i, x_j) = \int_{-\infty}^{\infty} x_i dF_i(x_i) \int_{-\infty}^{\infty} x_j dF_j(x_j), \quad i < j,$$

$$\int_{-\infty}^{\infty} x_i dF_i(x_i) = \theta_1 x_{i1} + \dots + \theta_r x_{ir} \equiv \eta_i, \quad \int_{-\infty}^{\infty} (x_i - \eta_i)^2 dF_i(x_i) = \sigma^2,$$

$$i = 1, \dots, n,$$

где x_{i1}, \dots, x_{ir} , $i=1, \dots, n$, — некоторые заданные числа (одни и те же для всех \mathbf{P}), $\theta_1, \dots, \theta_r$, σ^2 — числовые параметры, свои для каждого \mathbf{P} . Множество всех распределений \mathbf{P} с указанными ограничениями на первые два момента и составляет семейство $\{\mathbf{P}\}$. Легко понять, что в данном случае речь идет о линейной статистической модели с некоррелированными наблюдениями, с равными дисперсиями и матрицей модели $X = [x_{ij}, i=1, \dots, n; j=1, \dots, r]$. Следует подчеркнуть, что каждому элементу \mathbf{P} сопоставлен набор параметров $(\theta_1, \dots, \theta_r, \sigma^2)$, однако это соответствие не является взаимно-однозначным. Поэтому семейство $\{\mathbf{P}\}$ не является параметризованным.

(IV) В условиях предыдущего примера предположим дополнительно, что \mathbf{P} — n -мерное нормальное распределение. В результате получаем параметрическую модель (см. § 13), где $\mathbf{P}_{\theta,n}$ есть $N_n(\theta X', \sigma^2 I)$, $\theta = (\theta_1, \dots, \theta_r)$, $-\infty < \theta_i < \infty$, $i=1, \dots, r$, $\sigma > 0$.

Проведенное нами различие между вероятностной и статистической моделями приводит к различию в понятиях и обозначениях, которое раньше не всегда явно отмечалось. Это относится, в частности, к понятию статистики.

Определение 1. Статистикой назовем любую функцию $T(x)$, задающую измеримое отображение пространства $(\mathcal{X}, \mathcal{B})$ в (R^1, \mathcal{B}_1) , т. е. такую, что для любого $B \in \mathcal{B}_1$

$$\{x : T(x) \in B\} \in \mathcal{B}.$$

Векторной статистикой назовем упорядоченный набор статистик:

$$(T_1(x), \dots, T_k(x)) = \mathbf{T}_k(x).$$

Таким образом, понятие статистики оказывается не связанным с семейством мер $\{\mathbf{P}\}$. Переход от статистики к случайной величине — функции от выборки — производится подстановкой в $T(x)$ вместо x случайного элемента X . К сожалению, уступая традиции, не всегда удается отказаться от употребления термина статистика по отношению к сл. в. $T(X)$.

Заметим, что обозначение $T(X)$ не отражает зависимости от распределения \mathcal{P} (или \mathbf{P}). Это не приводит к путанице, когда речь идет о вероятностях $\mathcal{P}(T(X) \in B)$, но для математического ожидания мы введем явное указание на распределение: M_P или M_n , если семейство параметрическое.

2. Подобные статистики.

Существуют статистики $T(x)$, для которых распределение вероятностей $T(X)$ одно и то же для всех P из заданного семейства $\{P\}$. Такие статистики называются *подобными* (они подобны случайному величинам). Рассмотрим несколько примеров.

(I) Пусть R^{n-m} — выборочное пространство, P_F — мера в R^{n-m} , задаваемая непрерывной ф. р., имеющей вид

$$F(x_1) \cdot \dots \cdot F(x_n) \cdot F(x_{n+1}) \cdot \dots \cdot F(x_{n+m}),$$

$\{P_F\}$ — семейство всех таких мер. Положим $x^{(1)} = (x_1, \dots, x_n)$, $x^{(2)} = (x_{n+1}, \dots, x_{n+m})$, $x = (x^{(1)}, x^{(2)})$ и введем э. ф. р. $F_n(x; x^{(1)})$, $F_m(x; x^{(2)})$. При фиксированном x каждая из этих э. ф. р. является статистикой. Образуем статистики Смирнова:

$$T^+(x) = \sup_x (F_n(x; x^{(1)}) - F_m(x; x^{(2)})),$$

$$T(x) = \sup_x |F_n(x; x^{(1)}) - F_m(x; x^{(2)})|.$$

Совершенно аналогично статистике Колмогорова (см. п. 3 § 2) проверяется, что распределение каждой из статистик T^+ , T одно и то же для всех $P \in \{P_F\}$, и, следовательно, обе эти статистики подобные.

(II) Пусть $\{P_\mu : -\infty < \mu < \infty\}$ — семейство распределений в R^n с параметром сдвига μ : P_μ задается ф. р. $F_0(x_1 - \mu) \cdot F_0(x_2 - \mu) \cdot \dots \cdot F_0(x_n - \mu)$. Статистика $T(x) = \sum_{i=1}^n (x_i - \bar{x})^2$ — подобная.

(III) Для семейства $\{P_\sigma : 0 < \sigma < +\infty\}$ в R^n , где P_σ задается ф. р. $F_0(x_1/\sigma) F_0(x_2/\sigma) \cdot \dots \cdot F_0(x_n/\sigma)$, подобной статистикой будет $T(x_n) = \bar{x}^2 / \sum_{i=1}^n (x_i - \bar{x})^2$.

(IV) В линейной статистической модели с нормальными наблюдениями (см. п. 3 § 13) статистика

$$\frac{1}{r} \| \text{pr}_{V^k} \Theta_{V^k} Y \| ^2 / \left(\frac{1}{n-k} \| \text{pr}_{R^n \Theta_{V^r}} Y \| ^2 \right)$$

является подобной в предположении, что верна гипотеза H : вектор среднего лежит в V^k .

(V) В модели сдвига-масштаба статистики

$$(\bar{x} - \mu_0)^2 / \sum_{i=1}^n (x_i - \bar{x})^2 \text{ и } \frac{1}{\sigma_0} \sum_{i=1}^n (x_i - \bar{x})^2$$

будут подобными при выполнении гипотезы $\mu = \mu_0$ для первой статистики и $\sigma = \sigma_0$ — для второй.

Поскольку распределение подобной статистики одно и то же для всех мер P в пределах рассматриваемого семейства $\{P\}$, по-

добная статистика не помогает уточнить вопрос о том, какое именно из распределений P лучше всего соответствует наблюденным данным. Зато подобную статистику можно использовать для проверки соответствия модели наблюденным значениям (хотя бы в отношении таких отклонений от модели, которые улавливает данная подобная статистика). Например, в задаче о сравнении двух независимых выборок гипотеза о том, что наблюдения в обеих выборках $x^{(1)}$ и $x^{(2)}$ имеют общую ф. р., помещают нас в условия примера (I). Понятно, что умеренные значения статистик $T^+(x)$, $\bar{T}(x)$ будут говорить в пользу такого предположения. Благодаря подобию, критерии с критическими областями вида $T^+(x) > c$, $T^+(x) < -c$ имеют при любом c определенный уровень значимости, не зависящий от выбора меры P в пределах семейства $\{P\}$.

Задача проверки данных на согласие с моделью, как мы могли убедиться в предыдущих главах, близка к вопросу о построении доверительного интервала для параметра. Как видно из примера (V), этот вопрос также можно решать, имея подходящую подобную статистику.

§ 15. ДОСТАТОЧНЫЕ СТАТИСТИКИ В ДИСКРЕТНОЙ МОДЕЛИ

1. Введение.

Почему, наблюдая выборочную точку x в статистической модели $(\mathcal{X}, \mathcal{B}, \{P\})$, можно рассчитывать на уточнение наших сведений об истинном распределении вероятностей P ? Конечно, потому, что точка x , полученная в результате эксперимента с распределением вероятностей P , несет информацию о своем распределении вероятностей. Если статистическая модель уже выбрана, то интерес представляет лишь та информация в x , которая служит задаче различия элементов P из $\{P\}$. По многим причинам статистические выводы приходится основывать на одной или нескольких статистиках-функциях от x . Как выбрать эти статистики, чтобы не потерять информации по интересующему нас вопросу? Например, рассмотренные в предыдущем параграфе подобные статистики несут информацию лишь о таких чертах распределения P , которые являются общими для всех элементов из семейства $\{P\}$ и, следовательно, ничего не дают для решения задачи о различии элементов P . Нас интересует такой, по возможности минимальный, набор статистик $(T_1(x), \dots, T_k(x)) = T(x)$, который содержал бы всю имеющуюся в x различающую информацию об элементах P и ничего лишнего. Такая статистика $T(x)$ будет называться *достаточной статистикой* для нашей модели.

2. Примеры.

(I) Пусть $\mathcal{X} = \{x_n = (x_1, \dots, x_n) : x_i = 0 \text{ или } 1, i = 1, \dots, n\}$,

$$P_\theta(x_n) = \theta^{s_n} (1-\theta)^{n-s_n}, \quad s_n = x_1 + \dots + x_n,$$

где $0 < \theta < 1$. Это хорошо известная схема испытаний Бернулли.

Для оценки параметра θ — неизвестной вероятности — используется статистика $s_n: \hat{\theta} = s_n/n$. Допустим, статистика s_n приняла некоторое значение s . Что мы получим нового, если узнаем, какая именно из C_n^s выборок, содержащих ровно s единиц, осуществилась в результате эксперимента? Если мы рассматриваем случай $s_n = s$, то естественно перейти к модели с выборочным пространством $\mathcal{X}_s = \{x_n : x_1 + \dots + x_n = s\}$ и семейством условных вероятностных мер, получаемых из P_θ сужением на \mathcal{X}_s . Эти меры:

$$P_e(x_n | \mathcal{X}_s) = P_\theta(x_n) / P_\theta(\mathcal{X}_s) = \theta^s (1-\theta)^{n-s} / (C_n^s \theta^s (1-\theta)^{n-s}) = \\ = (C_n^s)^{-1}, \quad x_n \in \mathcal{X}_s,$$

как видно, не зависят от θ , а все семейство состоит из единственного элемента — равномерного распределения на \mathcal{X}_s . Итак, дополнительная вероятностная информация, содержащаяся в x_n , помимо той, что $x_n \in \mathcal{X}_s$, никак не связана с параметром θ , и потому статистику s_n следует признать содержащей всю информацию о параметре θ .

Соображения, лежащие в основе приведенных рассуждений, можно пояснить с иной точки зрения. Реализуем испытания Бернулли с помощью извлечения шаров из урны (либо по таблице случайных чисел) по следующей схеме. Возьмем рациональное θ , $0 < \theta < 1$, и урну с N шарами $n+1$ цветов, причем доля шаров i -го цвета равна биномиальной вероятности $p_i = C_n^i \theta^i (1-\theta)^{n-i}$, $i = 0, 1, \dots, n$ (в качестве N можно взять наименьшее число, при котором все N_{p_i} , $i = 0, 1, \dots, n$, целые). Кроме того, возьмем еще $n+1$ урн, причем урна с номером i , $i = 0, 1, \dots, n$, содержит C_n^i ($C_n^0 = 1$) шаров, занумерованных C_n^i наборами (x_1, \dots, x_n) , $x_j = 0$ или 1 , $j = 1, \dots, n$, содержащими ровно i единиц. Организуем теперь двухступенчатый выбор. На первом этапе выбирается шар из урны с шарами $n+1$ цветов и отмечается цвет i вытащенного шара. На втором этапе из урны с номером i извлекается шар и номер извлеченного шара — последовательность (x_1, \dots, x_n) — объявляется результатом полного эксперимента. Элементарный вероятностный расчет показывает, что результат опыта — (x_1, \dots, x_n) — представляет собой реализацию испытаний Бернулли с вероятностью θ выпадения 1. С другой стороны, поскольку вторая ступень нашего сложного опыта связана с параметром θ лишь через номер i — результат первого этапа, совершенно ясно, что к информации о θ , содержащейся в исходе i первой ступени опыта, вторая ступень ничего нового добавить не может.

(II) Пусть $\mathcal{X} = \{x_n = (x_1, \dots, x_n) : x_i = 0, 1, 2, \dots, i = 1, 2, \dots, n\}$, а мера P_θ соответствует независимой выборке из пуассонского распределения с параметром $\theta > 0$:

$$P_\theta(x_n) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = e^{-n\theta} \theta^{\sum_{i=1}^n x_i} / \prod_{i=1}^n x_i!, \quad x_n \in \mathcal{X}. \quad (1)$$

Оценкой параметра θ — теоретического среднего выборки — служит выборочное среднее $\bar{x} = s_n/n$, $s_n = \sum_{i=1}^n x_i$. Рассмотрим условную модель в предположении $s_n = s$ с выборочным пространством $\mathcal{X}_s = \{x : x_1 + \dots + x_n = s\}$ и условной мерой

$$P_\theta(x_n | \mathcal{X}_s) = P_\theta(x_n) / P_\theta(\mathcal{X}_s), \quad x_n \in \mathcal{X}_s. \quad (2)$$

Знаменатель в (2) представляет собой вероятность того, что сумма независимых пуассоновских сл. в., с параметром θ каждая, примет значение s . Легко проверить, что при сложении двух независимых сл. в., распределенных по Пуассону с параметрами λ и μ соответственно, получаемая сл. в. $X + Y$ также имеет распределение Пуассона с параметром $\lambda + \mu$:

$$\begin{aligned} P(X + Y = u) &= \sum_{v=0}^u P(X = v) P(Y = u - v) = \\ &= \sum_{v=0}^u \frac{\lambda^v}{v!} e^{-\lambda} \frac{\mu^{u-v}}{(u-v)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{u!} \sum_{v=0}^u \frac{u!}{v! (u-v)!} \lambda^v \mu^{u-v} = \\ &= \frac{e^{-(\lambda+\mu)}}{u!} (\lambda + \mu)^u. \end{aligned}$$

Таким образом,

$$P_\theta(\mathcal{X}_s) = P_\theta\left(\sum_{i=1}^n X_i = s\right) = e^{-n\theta} (n\theta)^s / s!, \quad (2')$$

и для условной меры получаем выражение

$$\begin{aligned} P_\theta(x_n | \mathcal{X}_s) &= e^{-n\theta} \theta^{\sum_{i=1}^n x_i} / \left(\prod_{i=1}^n x_i! \right) (e^{-n\theta} (n\theta)^s / s!)^{-1} = \\ &= (s! / \prod_{i=1}^n x_i!) n^{-s}, \quad s = \sum_{i=1}^n x_i, \end{aligned} \quad (3)$$

которое не зависит от θ . Отметим, что (3) представляет собой частный случай так называемого полиномиального распределения (при $n=2$ — биномиального с вероятностью 1/2). Отсюда следует, что искусственную выборку объема n из пуассоновского распределения с параметром θ можно построить, моделируя сначала пуассоновскую сл. в. с параметром $n\theta$: если она приняла значение s , то моделируется опыт по случайному размещению s шаров по n ящикам. Результат последнего опыта записываем в виде набора (x_1, x_2, \dots, x_n) , где x_i — число шаров, попавших в ящик с номером i . Легко видеть, что этот результат имеет поли-

номиальную вероятность (3), а опыт в целом приводит к исходу (x_1, x_2, \dots, x_n) с вероятностью (1). Итак, с точки зрения информации о параметре θ интерес представляет лишь первый этап указанного двухступенчатого опыта, результат которого выражает статистика s_n .

(III) Пусть $\mathcal{X} = \{x_n = (x_1, x_2, \dots, x_n) : x_i = 1, 2, \dots, N, i = 1, 2, \dots, n\}$, $P_N(x_n) = N^{-n}$, $x_n \in \mathcal{X}$, т. е. речь идет о выборке с возвращением из урны с N запущерованными шарами. Параметр N принимает натуральные значения. Проверим, что условное распределение выборки X_n при условии, что $X_{(n)} = m$, не зависит от параметра N . Имеем

$$\begin{aligned} P_N(X_{(n)} = m) &= P_N(X_{(n)} \leq m) - P_N(X_{(n)} \leq m-1) = \\ &= P(X_i \leq m, i = 1, \dots, n) - P(X_i \leq m-1, i = 1, \dots, n) = \\ &= (m/N)^n - (m-1)/N((= N^{-n}(m^n - (m-1)^n), m \leq N, \\ P_N(X_n = x_n | X_{(n)} = m) &= P_N(X_n = x_n) / P(X_{(n)} = m) = \\ &= (m^n - (m-1)^n)^{-1}, m \leq N. \end{aligned}$$

Рассуждения, аналогичные проведенным в предыдущих примерах, показывают, что статистика $x_{(n)}$ содержит всю информацию о параметре N , которая имеется в выборке.

3. Достаточные статистики: определение.

Пусть $(\mathcal{X}, \mathcal{B}, \{P\})$ — дискретная статистическая модель: множество \mathcal{X} конечно, или счетно, а \mathcal{B} — система всех подмножеств \mathcal{X} . В этом случае любая числовая функция $T(x)$ является, очевидно, измеримой относительно \mathcal{B} , т. е. будет статистикой. Для любого P из семейства $\{P\}$ и для данной (вообще говоря, векторной) статистики $T(x)$ определим семейство условных мер P^t , положив для любого $B \in \mathcal{B}$

$$P^t(B) = P(B \cap \{x : T(x) = t\}) / P(\{x : T(x) = t\}), \quad (4)$$

где $P \in \{P\}$, а индекс t при данном P пробегает множество всех значений, для которых знаменатель в (4) отличен от нуля.

Определение 1. Статистика $T(x)$ называется *достаточной* для дискретной статистической модели $(\mathcal{X}, \mathcal{B}, \{P\})$, если при каждом t условные вероятностные меры P^t не зависят от $P \in \{P\}$.

Замечание 1. Полезно иметь в виду, что при каждом t имеется, вообще говоря, свое подмножество множества $\{P\}$ мер P таких, что знаменатель в формуле (4) отличен от нуля, и именно для этих P условные меры P^t должны совпадать между собой. Более наглядным является случай параметризованного семейства $\{P_\theta, \theta \in \Theta\}$. Здесь определение достаточности $T(x)$ удобно

сформулировать так: существует семейство мер P^t , зависящих от индекса t , таких, что равенство

$$P_\theta(B \cap \{x : T(x) = t\}) / P_\theta(\{x : T(x) = t\}) = P^t(B) \quad (4a)$$

выполняется для всех пар (θ, t) , при которых левая часть в (4a) определена.

Замечание 2. Пользуясь тем, что

$$P(B) = \mathcal{P}(X(\omega) \in B),$$

вместо условных мер (4) в выборочном подпространстве $(\mathcal{X}, \mathcal{B})$ можно ввести условные вероятности

$$\mathcal{P}(X \in B | T(X) = t) \quad (46)$$

и дать определение достаточной статистики $T(x)$ как статистики, для которой условные вероятности (46) не зависят от \mathcal{P} в пределах семейства $\{\mathcal{P}\}$ (при данном t рассматриваются лишь те \mathcal{P} , для которых (46) определено).

Из определения достаточной статистики $T(x)$ вытекает, что, как и в разобранных выше примерах, статистика $T(x)$ содержит всю информацию выборки x , полезную для различия элементов P из $\{P\}$. Понятие информации при этом не формализовано, однако содержательный смысл сказанного хорошо ясен, если искусственно моделировать выборку X^M с распределением P . Повторим это рассуждение в общем случае. Сначала реализуем опыт, который по элементу $P \in \{P\}$ моделирует (вообще говоря, векторную) сл. в. T^M , для которой

$$P(T^M = t) = P(T(X) = t). \quad (5)$$

Затем по значению t , которое принимала сл. в. T^M , и заданному семейству мер P^t строим сл. в. X^M с множеством значений $\{x : T(x) = t\}$ и вероятностями

$$P(X^M = x | T^M = t) = P^t(x). \quad (6)$$

Перемножая (5) и (6), находим с учетом (4)

$$P(X^M = x) = P(T(X) = t) P^t(x) = P(X = x),$$

т. е. X^M имеет то же распределение, что и X , но при этом только первый этап моделирования зависит от $P \in \{P\}$ и, следовательно, несет различающую элементы P из $\{P\}$ информацию. Результат T^M первого этапа, очевидно, может быть представлен в виде $T^M = T(X^M)$, где $T(x)$ — рассматриваемая достаточная статистика. Итак, в искусственном эксперименте достаточная статистика $T(x)$ является носителем всей полезной различающей элементы P из $\{P\}$ информации.

Определение 2. Достаточная статистика $S(x)$ называется *минимальной*, если она представима как (вектор-)функция от любой достаточной статистики.

Понятно, что интерес к достаточным статистикам связан с возможностью сократить ненужную информацию, содержащуюся в выборке, и тем самым облегчить возможность последующего анализа. Поэтому если от достаточной статистики $T(x)$ можно перейти с сохранением свойства достаточности к статистике $\psi(T(x))$, где функция $\psi(t)$ принимает одинаковые значения при некоторых $t_1 \neq t_2$, то статистику $\psi(T(x))$ следует предпочесть $T(x)$. Из определения минимальной достаточной статистики следует, что дальнейшее сокращение данных без потери нужной информации невозможно. Мы покажем, что статистики, рассмотренные в примерах (I)–(III), являются минимальными достаточными. Проведем подготовительные рассуждения.

Пусть $T(x)$ — достаточная статистика для дискретной модели $(\mathcal{X}, \mathcal{B}, \{P_\theta\})$ и $t_1 \neq t_2$ — два ее значения. Перейдем к статистике $\psi(T(x))$, где $\psi(t) = t$ при $t_1 \neq t_2$ и $\psi(t_2) = t_1$, т. е. статистика $\psi(T(x))$ совпадает с $T(x)$, если последняя принимает любое значение, кроме t_2 , а выборки, для которых $T(x) = t_2$, она объединяет с теми, для которых $T(x) = t_1$. Рассмотрим условия, при которых такое сокращение данных сохраняет свойство достаточности. Чтобы это имело место, условная вероятность

$$\begin{aligned} & P_\theta(X=x | T(X)=t_1) \cup | T(X)=t_2) = \\ & = \frac{P_\theta(X=x, T(X)=t_1) + P_\theta(X=x, T(X)=t_2)}{P_\theta(T(X)=t_1) + P_\theta(T(X)=t_2)} \end{aligned} \quad (7)$$

не должна зависеть от θ (при тех θ , для которых выражение (7) имеет смысл). Преобразуем числитель в (7) по формуле условной вероятности:

$$P_\theta(X=x, T(X)=t_i) = P_\theta(X=x | T(X)=t_i) P_\theta(T(X)=t_i), \quad i=1, 2. \quad (8)$$

Ввиду достаточности статистики $T(x)$ условная вероятность в (8) от параметра θ не зависит (см. (4а)). Положим для краткости $P_\theta(X=x | T(X)=t_i) = p_i(x)$, $i=1, 2$, и перепишем условную вероятность (7) в виде

$$p_1(x) + (p_2(x) - p_1(x)) \frac{P_\theta(T(X)=t_2)}{P_\theta(T(X)=t_1) + P_\theta(T(X)=t_2)}. \quad (9)$$

Выражение (9) не зависит от θ (для θ таких, что знаменатель в (9) отличен от нуля), если либо $p_1(x) = p_2(x)$, т. е. условное распределение выборки при значениях t_1 и t_2 достаточной статистики $T(x)$ одно и то же, либо множитель при $p_2(x) - p_1(x)$ в (9) не зависит от θ . Это будет так, если отношение

$$\frac{P_\theta(T(X)=t_2)}{P_\theta(T(X)=t_1)} = \frac{P_\theta(\{x : T(x)=t_2\})}{P_\theta(\{x : T(x)=t_1\})} \quad (10)$$

не зависит от θ . Обратимся к примерам (I)–(III).

В примере (I) $T(x_n) = x_1 + \dots + x_n$, $p_i(x_n) = C_n^{t_i}$ при $T(x_n) = t_i$, так что $p_1(x_n) \neq p_2(x_n)$ для $t_1 \neq t_2$. Отношение (10) равно

$$C_n^{t_1} \theta^{t_1} (1-\theta)^{n-t_1} / (C_n^{t_2} \theta^{t_2} (1-\theta)^{n-t_2}), \quad 0 < \theta < 1,$$

и, очевидно, зависит от θ при $t_1 \neq t_2$. Следовательно, если от статистики $T(x_n)$ перейти к $\psi(T(x_n))$, где $\psi(t) \equiv t$ при $t \neq t_2$ и $\psi(t_2) = t_1$, т. е. объединить значения t_1 и t_2 статистики $T(x_n)$, то свойство достаточности нарушается, и потому статистика $T(x_n) = x_1 + \dots + x_n$ минимальна.

В примере (II) ситуация аналогичная: $p_i(x_n)$ определяется формулой (3) и $p_1(x_n) \neq p_2(x_n)$. Отношение (10) выражений (2') при $s=t_2$ и $s=t_1$ равно

$$(n\theta)^{s_2-s_1} t_1! / t_2!$$

и, очевидно, зависит от θ при $s_1 \neq s_2$.

В примере (III) $T(x_n) = x_{(n)}$, $p_i(x_n) = (m_i^n - (m_i-1)^n)^{-1}$ при $x_{(n)} = m_i$ и, следовательно, $p_1(x_n) \neq p_2(x_n)$. Положим для определенности $m_1 < m_2$. Тогда отношение (10) равно $(m_2^n - (m_2-1)^n) / (m_1^n - (m_1-1)^n)$, если $N > m_2$, и принимает значение нуль, если $m_1 < N < m_2$, т. е. отношение (10) зависит от параметра N , следовательно, статистика $x_{(n)}$ минимальна.

4. Критерий достаточности.

Для наглядности рассмотрим параметрическую модель $(\mathcal{X}, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$. Пусть $T(x)$ — достаточная статистика; положим

$$g(t; \theta) = P_\theta(\{x : T(x) = t\}).$$

Основное соотношение (4а) можно записать в виде

$$P_\theta(x) = g(t; \theta) P^t(x), \quad t = T(x).$$

Введя функцию

$$h(x) = P^{T(x)}(x),$$

получаем

$$P_\theta(x) = g(T(x); \theta) h(x), \quad x \in \mathcal{X}. \quad (11)$$

Разложение распределения вероятностей $P_\theta(x)$ в модели $(\mathcal{X}, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ на множители в форме (11), где один множитель $h(x)$ не зависит от $\theta \in \Theta$, а другой — зависит от x через некоторую статистику $T(x)$, назовем *факторизацией* этого распределения. Факторизация $P_\theta(x)$ неединственна. Например, взяв производящую не обращающуюся в нуль функцию $\psi(x)$ и положив

$$g'(T(x); \theta) = g(T(x); \theta) \psi(x), \quad h'(x) = h(x) / \psi(x),$$

получим из (11)

$$P_{\theta}(x) = g'(T(x), \theta) h'(x).$$

Всегда возможна тривиальная факторизация

$$P_{\theta}(x) = P_{\theta}(x) \cdot 1.$$

Итак, если статистика $T(x)$ достаточна, то имеет место факторизация (11).

Верно и обратное. Пусть имеется факторизация (11), где $T(x)$ — некоторая статистика, $g(t; \theta)$ — функция t , $h(x)$ — функция, не зависящая от θ . Покажем, что статистика $T(x)$ является в таком случае достаточной. Так как все меры P_{θ} обращаются в нуль на множестве $\{x : h(x) = 0\}$, то исключим эти x из дальнейшего рассмотрения. Из (11) получаем

$$\begin{aligned} P_{\theta}(B \cap \{x : T(x) = t\}) &= \sum_{x \in B, T(x)=t} g(T(x); \theta) h(x) = \\ &= g(t; \theta) \sum_{x \in B, T(x)=t} h(x). \end{aligned} \quad (12)$$

При $B = \mathcal{X}$ из (12) получаем

$$P_{\theta}(\{x : T(x) = t\}) = g(t; \theta) \sum_{T(x)=t} h(x). \quad (13)$$

Положим

$$P^t(B) = \sum_{x \in B, T(x)=t} h(x) / \sum_{T(x)=t} h(x).$$

Для θ таких, что $P_{\theta}(\{x : T(x) = t\}) > 0$, поделив (12) на (13), получаем, что условная мера совпадает с $P^t(B)$ и, следовательно, статистика $T(x)$ — достаточная. Таким образом, получена

Теорема факторизации. Для того чтобы статистика $T(x)$ была достаточной для дискретной модели $(\mathcal{X}, \mathcal{A}, \{P_{\theta} : \theta \in \Theta\})$, необходимо и достаточно, чтобы имела место факторизация (11).

При соответствующих изменениях в обозначениях этот же результат, очевидно, справедлив для непараметрических моделей.

Отыскание достаточных статистик с помощью теоремы факторизации сводится к анализу распределения выборки. Достаточность статистик в примерах (I)–(III) с помощью теоремы факторизации устанавливается немедленно. Рассмотрим еще пример.

(IV) Пусть в схеме испытаний Бернулли число наблюдений заранее не фиксировано: испытания продолжают до случайного момента появления r -го успеха, где r — заданное число. Множество \mathcal{X} состоит из всех последовательностей $x_n = (x_1, \dots, x_n)$, $x_i = 0$ или 1 , $i = 1, \dots, n$; $n = 1, 2, \dots$, содержащих ровно r единиц,

причем $x_n=1$. Вероятностная мера P_θ , $0 < \theta < 1$, определяется формулой

$$P_\theta(x) = \theta^r (1-\theta)^{N(x)-r},$$

где $N(x)$ — число координат вектора x . В силу теоремы факторизации $N(x)$ — достаточная статистика.

5. Достаточные и минимальные достаточные разбиения.

Статистики $T_1(x)$ и $T_2(x)$, связанные взаимно-однозначным преобразованием $T_1(x) = \psi(T_2(x))$, назовем эквивалентными. Эквивалентные статистики несут одну и ту же информацию о статистической модели. Если одна из них достаточная, то достаточной будет и другая. Сокращение данных, которое дает каждая из этих статистик, одно и то же: системы подмножеств

$$\{x : T_1(x) = t_1\} \text{ и } \{x : T_2(x) = t_2\},$$

где t_1 и t_2 пробегают соответственно множество значений статистик T_1 и T_2 , совпадают.

Определение 3. Разбиение множества \mathcal{X} на подмножества \mathcal{X}_t , где t пробегает некоторое множество индексов:

$$\mathcal{X} = \bigcup_t \mathcal{X}_t, \quad \mathcal{X}_t \cap \mathcal{X}_s = \emptyset, \quad t \neq s,$$

называется *достаточным* для дискретной модели $(\mathcal{X}, \mathcal{B}, \{P\})$, если условные меры

$$P'(B) = P(B \cap \mathcal{X}_t) / P(\mathcal{X}_t), \quad B \in \mathcal{B},$$

не зависят от $P \in \{P\}$ (для тех P при данном t , для которых $P(\mathcal{X}_t) > 0$).

Если имеется достаточное разбиение, то разными способами можно определить числовую функцию, постоянную на элементах разбиения и принимающую разные значения на разных элементах. Эта функция будет достаточной статистикой. Полезно иметь в виду, что если элементы разбиения $\{\mathcal{Y}_a\}$ множества \mathcal{X} целиком содержатся в элементах достаточного разбиения $\{\mathcal{X}_t\}$ (т. е. для любого a найдется $t=t_a$, что $\mathcal{Y}_a \subseteq \mathcal{X}_{t_a}$), то разбиение $\{\mathcal{Y}_a\}$ также является достаточным. Действительно, для $x \in \mathcal{Y}_a$ отношение

$$\begin{aligned} P(x)/P(\mathcal{Y}_a) &= (P(x)/P(\mathcal{X}_{t_a})) (P(\mathcal{Y}_a)/P(\mathcal{X}_{t_a}))^{-1} = \\ &= (P(x)/P(\mathcal{X}_{t_a})) \mid \sum_{x \in \mathcal{Y}_a} P(x)/P(\mathcal{X}_{t_a}) \mid^{-1} \end{aligned}$$

не зависит от $P \in \{P\}$, поскольку разбиение $\{\mathcal{X}_t\}$ — достаточное.

Определим *минимальное достаточное разбиение* как разбиение, элементы которого являются объединением элементов любого достаточного разбиения (или совпадают с таким элементом). Пока-

жем, как строится минимальное достаточное разбиение в дискретной модели. Для наглядности рассмотрим параметрический случай.

В п. 3 (видели, что если достаточная статистика $T(x)$ такова, что отношение (см. (10))

$$P_\theta(\{x : T(x) = t_1\}) / P_\theta(\{x : T(x) = t_2\})$$

не зависит от $\theta \in \Theta$, то, сохраняя свойство достаточности, можно объединить элементы разбиения, соответствующие значениям t_1 и t_2 статистики T . Это же соображение лежит в основе приема минимального сокращения данных, о котором идет речь ниже. Именем введем отношение xRy на множестве \mathcal{X} , полагая, что x связано с y отношением R тогда и только тогда, когда для каждого $\theta \in \Theta$, такого, что $P_\theta(x), P_\theta(y) > 0$, отношение $P_\theta(x)/P_\theta(y)$ не зависит от θ , а для всех θ , для которых одно из $P_\theta(x), P_\theta(y)$ равно нулю, также равно нулю и другое. Отношение xRy , очевидно, рефлексивно: xRx имеет место для любого x ; симметрично: если xRy , то и yRx , и транзитивно: $xRy, yRz \Rightarrow xRz$. Таким образом, xRy есть отношение эквивалентности, и, как известно, множество \mathcal{X} разбивается на непересекающиеся классы \mathcal{X}_t :

$$\mathcal{X} = \bigcup_t \mathcal{X}_t, \quad \mathcal{X}_t \cap \mathcal{X}_s = \emptyset, \quad s \neq t, \quad (14)$$

где t пробегает некоторое множество индексов, причем каждый класс \mathcal{X}_t состоит из множества всех элементов, которые эквивалентны друг другу по отношению R . [Проверим сделанное утверждение. Пусть $x \in \mathcal{X}$ произвольно. Обозначим через \mathcal{Y}_x множество всех элементов y , таких, что xRy . Тогда xRx , и, следовательно, $x \in \mathcal{Y}_x$. Пусть $z \in \mathcal{X}$ произвольно; допустим, что $\mathcal{Y}_x \cap \mathcal{Y}_z \neq \emptyset$, тогда существует элемент u из их пересечения. Поскольку xRu и zRu , то по симметрии uRz , а по транзитивности xRz . Пусть $v \in \mathcal{Y}_z$, тогда zRv , и так как xRz , то по транзитивности xRv , т. е. $v \in \mathcal{Y}_x$ или $\mathcal{Y}_z \subseteq \mathcal{Y}_x$. Аналогично получаем $\mathcal{Y}_z \subseteq \mathcal{Y}_x$, т. е. $\mathcal{Y}_x = \mathcal{Y}_z$. Итак, система подмножеств $\mathcal{Y}_x, x \in \mathcal{X}$, такова, что любые два из них либо не пересекаются, либо совпадают. Определим функцию $t = t(x)$ так, что $t(x_1) = t(x_2)$, если $\mathcal{Y}_{x_1} = \mathcal{Y}_{x_2}$, и $t(x_1) \neq t(x_2)$, если $\mathcal{Y}_{x_1} \neq \mathcal{Y}_{x_2}$. Полагая $\mathcal{X}_t = \mathcal{Y}_t, t = t(x), x \in \mathcal{X}$, получаем, как нетрудно понять, требуемое утверждение.]

Докажем, что разбиение (14) достаточное. Определим функцию

$$\Phi(t, x, y) = P_\theta(y)/P_\theta(x), \quad x, y \in \mathcal{X}_t, \quad (15)$$

где $\theta \in \Theta$ — любое такое, что $P_\theta(x) > 0$; если такого θ не существует, то положим $\Phi(t, x, y) = \infty$. Проверим, что определение (15) корректно. Если при всех $\theta \in \Theta$ $P_\theta(x) = 0$, то все в порядке: функция Φ однозначно определена и принимает значение ∞ . Пусть θ_1 такое, что $P_{\theta_1}(x) > 0$. Тогда по определению класса \mathcal{X}_t также

и $P_{\theta_i}(y) > 0$, а для всех θ таких, что $P_{\theta}(x) > 0$, верно равенство

$$P_{\theta}(y)/P_{\theta}(x) = P_{\theta_i}(y)/P_{\theta_i}(x),$$

что и требовалось установить. Пусть $\theta \in \Theta$ — любое фиксированное. Допустим, что $P_{\theta}(\mathcal{X}_t) > 0$ и $P_{\theta}(x) > 0$. По определению класса \mathcal{X}_t в этом случае $P_{\theta}(y) > 0$ для всех $y \in \mathcal{X}_t$. Суммируя (15), получаем

$$\psi(t, x) = \sum_{y \in \mathcal{X}_t} \psi(t, x, y) = P_{\theta}(\mathcal{X}_t)/P_{\theta}(x), \quad x \in \mathcal{X}_t,$$

откуда

$$P_{\theta}(x)/P_{\theta}(\mathcal{X}_t) = (\psi(t, x))^{-1}, \quad x \in \mathcal{X}_t. \quad (16)$$

Рассуждения, приводящие к (16), справедливы при всех $\theta \in \Theta$, таких, что $P_{\theta}(x) > 0$. Если при некотором θ $P_{\theta}(x) = 0$, то по определению класса \mathcal{X}_t также и $P_{\theta}(\mathcal{X}_t) = 0$. Следовательно, (16) справедливо при всех θ , таких, что $P_{\theta}(\mathcal{X}_t) > 0$, и разбиение (14) — достаточное.

Для доказательства минимальности разбиения (14) возьмем произвольное достаточное разбиение $\{\mathcal{Y}_a\}$, и пусть $V(x)$ — порождающая разбиение $\{\mathcal{Y}_a\}$ достаточная статистика ($V(x)$ — любая функция, принимающая различные значения на различных элементах разбиения). По теореме факторизации

$$P_{\theta}(x) = g(V(x); \theta) h(x).$$

Ограничимся далее подмножеством $\mathcal{X}' = \{x : h(x) > 0\}$, которое содержит носители всех мер $P_{\theta}(x)$. Пусть $x, y \in \mathcal{X}'$ таковы, что $V(x) = V(y) = v$. Тогда

$$P_{\theta}(x) = g(v; \theta) h(x), \quad P_{\theta}(y) = g(v; \theta) h(y).$$

Если $g(v; \theta) = 0$, то $P_{\theta}(x) = P_{\theta}(y) = 0$. Если же $g(v; \theta) > 0$, то

$$P_{\theta}(x)/P_{\theta}(y) = h(x)/h(y),$$

и, следовательно, x, y находятся в отношении R и принадлежат одному элементу достаточного разбиения (14). Итак, элементы разбиения $\{\mathcal{Y}_a\}$ содержатся в элементах разбиения $\{\mathcal{X}_t\}$, что и требовалось доказать.

В примере (IV)

$$P_{\theta}(x)/P_{\theta}(y) = (1 - \theta)^{N(x) - N(y)}.$$

и это отношение не зависит от θ , $0 < \theta < 1$, только если $N(x) = N(y)$. Следовательно, разбиение с элементами

$$\mathcal{X}_n = \{x \in \mathcal{X} : N(x) = n\}$$

является минимальным достаточным, а $L(x)$ — минимальная достаточная статистика.

Отметим, что если $S(x)$ — минимальная достаточная статистика, то, например, $(S(x), S(x))$ также будет минимальна, так как, очевидно, порождает то же разбиение. Поэтому в теоретическом плане остается вопрос о выборе такой минимальной достаточной статистики, которая имела бы наименьшую размерность.

К построению минимального достаточного разбиения имеется еще и другой подход. Рассмотрим сначала случай, когда семейство $\{P\}$ состоит из конечного числа элементов P_0, P_1, \dots, P_k . Введем статистику

$$L(x) = (P_0(x), P_1(x), \dots, P_k(x)).$$

Покажем, что статистика $L(x)$ — достаточная. Пусть x и y принадлежат одному элементу разбиения, задаваемого статистикой $L(x)$:

$$L(x) = L(y),$$

т. е.

$$P_i(x) = P_i(y), \quad i=0, 1, \dots, k.$$

Очевидно, в таком случае точки x и y находятся в отношении R и, следовательно, принадлежат одному элементу минимального достаточного разбиения, (14). Итак, элементы разбиения, задаваемого статистикой $L(x)$, содержатся в элементах достаточного разбиения. Как было отмечено в начале раздела, отсюда следует, что статистика $L(x)$ — достаточная.

Теперь мы преобразуем статистику $L(x)$ в минимально достаточную. Предположим вначале, что носитель распределения $P_0(x)$ содержит носители остальных $P_i(x)$:

$$\{x : P_i(x) > 0\} \subseteq \{x : P_0(x) > 0\}. \quad (17)$$

Дальнейшие построения проведем на подмножестве $\mathcal{X}' = \{x : P_0(x) > 0\} \subseteq \mathcal{X}$. Покажем, что статистика

$$T(x) = (P_1(x)/P_0(x), \dots, P_k(x)/P_0(x))$$

является минимальной достаточной. Пусть $x, y \in \mathcal{X}'$ принадлежат одному элементу минимального достаточного разбиения (14). В таком случае отношение $P_i(x)/P_i(y)$ не зависит от i при всех i , таких, что $P_i(x), P_i(y) > 0$, а для остальных i $P_i(x) = P_i(y) = 0$. Отсюда следует, что либо

$$P_i(x)/P_i(y) = P_0(x)/P_0(y),$$

либо $P_i(x) = P_i(y) = 0$. Во всех случаях получается, что

$$P_i(x)/P_0(x) = P_i(y)/P_0(y)$$

для всех $i=1, \dots, k$, т. е. $T(x) = T(y)$. Таким образом, элементы разбиения, порождаемого статистикой $T(x)$, содержат в себе эле-

менты минимального достаточного разбиения (суженного на \mathcal{X}') а потому $T(x)$ — минимальная достаточная статистика.

Если условие (17) не выполнено, то, положив

$$P_0^*(x) = \sum_{i=0}^k P_i(x),$$

легко обнаружить, что статистика $T(x)$ с заменой $P_0(x)$ на $P_0^*(x)$ оказывается минимальной достаточной.

В случае произвольного семейства $\{P_\theta : \theta \in \Theta\}$ рассуждения сохраняются без изменений, если выполнено условие, аналогичное (17). Достаточное разбиение порождается системой статистик, зависящих от θ как параметра:

$$L(x; \theta) = P_\theta(x), \quad \theta \in \Theta. \quad (18)$$

Если множество Θ конечно, то (18) можно рассматривать как векторную статистику. В общем случае, ради краткости, систему (18) назовем *функциональной статистикой*. При каждом фиксированном $\theta = \theta_0$ «координата» $L(x; \theta_0)$ является обычной скалярной статистикой, совокупность всех таких координат $L(x; \theta)$ составляет функциональную статистику (18).

Минимальное достаточное разбиение порождается функциональной статистикой

$$T(x; \theta) = L(x; \theta)/L(x; \theta_0), \quad \theta \in \Theta,$$

где $\theta_0 \in \Theta$ таково, что для всех $\theta \in \Theta$

$$\{x : P_\theta(x) > 0\} \subseteq \{x : P_{\theta_0}(x) > 0\}.$$

В примере (II) выборки из пуассоновского распределения

$$L(x; \theta)/L(x; \theta_0) = e^{-n(\theta-\theta_0)} (\theta/\theta_0)^{\sum_{i=1}^n x_i}$$

Прологарифмировав, получаем

$$\ln(L(x; \theta)/L(x; \theta_0)) = -n(\theta - \theta_0) + \sum_{i=1}^n x_i \ln \left(\frac{\theta}{\theta_0} \right). \quad (19)$$

Поскольку взаимно-однозначное преобразование приводит к эквивалентной статистике, то (19) также является минимальной достаточной. Как видно из (19), любая «координата» функциональной статистики (19) выражается через любую другую с $\theta \neq \theta_0$. Поэтому «избыточную» функциональную статистику (19) можно заменить одной скалярной, например статистикой $\sum_{i=1}^n x_i$, и, следовательно, эта статистика минимальна.

§ 16. ДОСТАТОЧНЫЕ СТАТИСТИКИ В НЕПРЕРЫВНОЙ МОДЕЛИ

1. Примеры достаточных статистик.

Мы будем рассматривать здесь статистические модели $(\mathcal{X}, \mathcal{B}, \{P\})$, где \mathcal{X} — область в R^n , $\mathcal{B} = \mathcal{X} \cap \mathcal{B}_n$, а мера P задается некоторой n -мерной плотностью. Эта модель, которую мы назовем *непрерывной*, и дискретная модель, изучавшаяся в предыдущем параграфе, охватывают все наиболее важные области применения математической статистики к выборкам фиксированного объема. К сожалению, полное изложение теории достаточных статистик в непрерывной модели требует привлечения понятий и утверждений, выходящих за рамки данной книги, хотя в идейном отношении нет разницы между дискретным и непрерывным случаем. Математически подготовленный читатель может ознакомиться с логически полноценным изложением общей теории этого и некоторых других параграфов данной главы по многочисленным отечественным и переводным монографиям по математической статистике, список которых можно найти в конце книги.

Определение достаточной статистики будет дано в следующем разделе, а пока займемся примерами достаточных (и на самом деле минимальных) статистик, обращаясь к неформальному понятию информации, различающей элементы P из $\{P\}$. Отметим, что, как и в случае дискретной модели, статистики $T(x_n)$ и $S(x_n)$, связанные измеримым взаимно-однозначным преобразованием ψ , являются эквивалентными носителями вероятностной информации и будут называться *эквивалентными*. Полезнее, однако, несколько более слабая форма эквивалентности:

$$T(x_n) = \psi(S(x_n))$$

на некотором множестве P -меры 1 при каждом $P \in \{P\}$.

(I) Пусть мера P_f задается многомерной плотностью вида $f(x_1)f(x_2)\dots f(x_n)$, где $f(x)$ — произвольная одномерная плотность. Таким образом, статистическая модель $(R^n, \mathcal{B}_n, \{P_f\})$ соответствует последовательности n независимых одинаково распределенных испытаний с неизвестным законом распределения, обладающим плотностью. Это непараметрическая модель, так как индекс f здесь функциональный.

Покажем, что вариационный ряд $T_n = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ содержит всю информацию о плотности f . Для этого заметим, что

$$\mathcal{P}_f(X_{(1)} \in B_1, \dots, X_{(n)} \in B_n) = n! \mathcal{P}_f(X_1 \in B_1, \dots, X_n \in B_n, X_1 < \dots < X_n),$$

и поэтому плотность $T_n(X_n)$ равна при $x_1 < x_2 < \dots < x_n$

$$n! f(x_1) \dots f(x_n).$$

Нетрудно понять, что условное распределение X_n при условии

$$T_n(X_n) = (t_1, \dots, t_n), \quad t_1 < t_2 < \dots < t_n,$$

— равномерное и сосредоточено в $n!$ точках $(t_{i_1}, \dots, t_{i_n})$, где (i_1, \dots, i_n) — произвольная перестановка из индексов $(1, \dots, n)$. Мы не будем проводить формального доказательства этого факта, а вместо этого сразу обратимся к моделированию, как это делалось в дискретном случае. На первом этапе образуем последовательность сл. в. $Y_n = (Y_1, Y_2, \dots, Y_n)$ с плотностью

$$f_{Y_n}(y_n) = n! f(y_1) f(y_2) \dots f(y_n), \quad y_1 < y_2 < \dots < y_n$$

и $f_{Y_n}(y_n) = 0$ при остальных значениях аргумента. Затем построим независимую от Y_n последовательность сл. в. (v_1, v_2, \dots, v_n) , которая с равными вероятностями принимает $n!$ значений (i_1, i_2, \dots, i_n) — перестановок из чисел $(1, 2, \dots, n)$. Определим результат составного эксперимента как последовательность сл. в.:

$$X_n = (X_1, X_2, \dots, X_n), \quad X_i = Y_{v_i}, \quad i = 1, 2, \dots, n.$$

Найдем плотности распределения случайного вектора X_n . Выберем произвольные попарно различные числа x_1, x_2, \dots, x_n и $\delta > 0$, такое, что интервалы $(x_i, x_i + \delta]$, $i = 1, \dots, n$, попарно не пересекаются. По данным x_n найдем такую перестановку (k_1, k_2, \dots, k_n) индексов $(1, 2, \dots, n)$, что

$$x_{k_1} < x_{k_2} < \dots < x_{k_n}.$$

Тогда, учитывая, что $Y_1 < Y_2 < \dots < Y_n$ с вероятностью 1, получаем $P(x_i < X_i < x_i + \delta, i = 1, \dots, n) = P(x_i < Y_{v_i} < x_i + \delta, i = 1, \dots, n) =$

$$\begin{aligned} &= P(x_{k_j} < Y_{v_{k_j}} < x_{k_j} + \delta, j = 1, \dots, n) = \\ &= P(x_{k_j} < Y_{v_{k_j}} < x_{k_j} + \delta, v_{k_j} = j, j = 1, \dots, n) = \\ &= P(x_{k_j} < Y_j < x_{k_j} + \delta, v_{k_j} = j, j = 1, \dots, n) = \\ &= P(x_{k_j} < Y_j < x_{k_j} + \delta, j = 1, \dots, n) P(v_{k_j} = j, j = 1, \dots, n) = \\ &= \int_{x_{k_1}}^{x_{k_1} + \delta} \dots \int_{x_{k_n}}^{x_{k_n} + \delta} n! f(y_1) f(y_2) \dots f(y_n) dy_1 \dots dy_n (n!)^{-1}. \end{aligned}$$

Производя сокращение на $n!$, получаем отсюда, что случайный вектор X_n имеет плотность $f(x_1) f(x_2) \dots f(x_n)$. Итак, вектор X_n получен в два этапа: на первом — образован вектор $T_n = (X_{(1)}, \dots, X_{(n)})$, на втором — с помощью случайного механизма, не зависящего от неизвестной плотности $f(x)$, получен уже сам вектор X_n . Проведенная конструкция показывает, что T_n содержит всю информацию о плотности $f(x)$.

(II) Пусть $\mathcal{X} = R_+^n = \{x_n \in R^n : x_i > 0, i = 1, \dots, n\}$, $\mathcal{B} = \mathcal{B}^+ = \mathcal{B}_n \cap \cap R_+^n$, а мера P_θ , $0 < \theta < +\infty$, задается плотностью

$$f(x_n; \theta) = \theta^{-n} \text{ при } x_n \in R_+^n, \quad x_i < \theta, \quad i = 1, \dots, n$$

и $f(x_n; \theta) = 0$ для остальных x_n . Таким образом, рассматривается случайная выборка объема n из равномерного на $(0, \theta)$ распределения. Мы покажем, что вся информация о параметре θ , содержащаяся в выборке x_n , заключена в статистике $x_{(n)} = \max(x_1, \dots, x_n)$. Используя предыдущий пример, перейдем от выборки x_n к ее вариационному ряду $T_n = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$, не потеряв при этом информации о параметре θ . Плотность распределения $T_n(X_n)$ запишем в виде

$$f_{T_n}(t_n) = n! \theta^{-n}, \quad t_n \in R_+, \quad t_1 < t_2 < \dots < t_n < \theta \quad (1)$$

и положим ее равной нулю при остальных значениях t_n . Так как

$$\mathcal{P}_\theta(X_{(n)} \leq t) = t^n \theta^{-n}, \quad 0 < t < \theta,$$

то

$$f_{X_{(n)}}^\theta(t) = n! t^{n-1} \theta^{-n}, \quad 0 < t < \theta.$$

Отсюда условная плотность $X_{(1)}, \dots, X_{(n-1)}$ при условии $X_{(n)}$ равна при $0 < t_1 < \dots < t_n < \theta$

$$f_{X_{(1)}, \dots, X_{(n-1)} | X_{(n)}}^\theta(t_1, \dots, t_{n-1} | t_n) = f_{T_n}^\theta(t_n) / f_{X_{(n)}}^\theta(t_n) = (n-1)! t_n^{n-1}. \quad (2)$$

Функция

$$g(t_1, \dots, t_{n-1}; t) = (n-1)! t^{-(n-1)}, \quad 0 < t_1 < \dots < t_{n-1} < t,$$

представляет собой совместную плотность порядковых статистик случайной выборки объема $n-1$ из равномерного на $(0, t)$ распределения (ср. (1)). Соотношение (2) устанавливает, что

$$f_{X_{(1)}, \dots, X_{(n-1)} | X_{(n)}}^\theta(t_1, \dots, t_{n-1} | t_n) = g(t_1, \dots, t_{n-1}; t_n)$$

для всех t, θ , таких, что $f_{X_{(n)}}^\theta(t) > 0$. Это означает отсутствие дополнительной информации о θ в значениях $X_{(1)}, \dots, X_{(n-1)}$, если известна $X_{(n)}$.

(III) Пусть, как и в предыдущем примере, $\mathcal{X} = R_+^n$, $\mathcal{B} = \mathcal{B}_+^+$, а мера $P_{\mu, \sigma}$, $-\infty < \mu < +\infty$, $\sigma > 0$, задается n -мерной плотностью

$$f(x_n; \mu, \sigma) = \prod_{i=1}^n \sigma^{-1} f((x_i - \mu)/\sigma), \quad f(x) = e^{-x}, \quad x > 0.$$

Покажем, что двумерная статистика

$$S_2(x_n) = \left(x_{(1)}, \sum_{i=2}^n (x_{(i)} - x_{(1)}) \right)$$

содержит всю информацию о двумерном параметре (μ, σ) . Перейдем от X_n к $T_n(X_n) = (X_{(1)}, \dots, X_{(n)})$, сохранив всю информацию

о (μ, σ) , а затем взаимно-однозначным преобразованием переходим к вектору

$$Y_n = (Y_1, \dots, Y_n), \quad Y_j = (n-j+1)(X_{(j)} - X_{(j-1)}), \quad j=1, \dots, n,$$

$$X_{(0)} = 0.$$

Ввиду леммы 1 § 6 сл. в. Y_j , $j=1, \dots, n$, независимы и экспоненциально распределены: Y_1 — с плотностью $\sigma^{-1} \exp(-(y_1 - n\mu)/\sigma)$, $y_1 > n\mu$, а Y_j , $j \geq 2$, — с плотностью $\sigma^{-1} \exp(-y_j/\sigma)$, $y_j > 0$, т. е. плотность распределения вектора Y_n равна

$$f_{Y_n}^{\mu, \sigma}(y_n) = \sigma^{-n} \exp \left[-\sigma^{-1} \left(\sum_{j=1}^n y_j - n\mu \right) \right],$$

$$y_j > 0, \quad j=2, \dots, n, \quad y_1 > n\mu. \quad (3)$$

По лемме 1 § 7

$$Y_2 + \dots + Y_n = \sum_{j=2}^n (n-j+1)(X_{(j)} - X_{(j-1)}) =$$

$$= \sum_{j=2}^n X_{(j)} - (n-1)X_{(1)} = \sum_{j=2}^n (X_{(j)} - X_{(1)})$$

имеет распределение $G(\sigma^{-1}, n-1)$ с плотностью

$$f_{Y_2 + \dots + Y_n}^{\mu, \sigma}(t) = \frac{t^{n-2}}{(n-2)! \sigma^{n-1}} \exp(-t/\sigma), \quad t > 0.$$

Совместная плотность пары $(Y_1, Y_2 + \dots + Y_n) = (X_1, \sum_{j=2}^n (X_{(j)} - X_{(1)}))$ равна

$$f_{Y_1, Y_2 + \dots + Y_n}^{\mu, \sigma}(s, t) = f_{Y_1}^{\mu, \sigma}(s) f_{Y_2 + \dots + Y_n}^{\mu, \sigma}(t) =$$

$$= \frac{t^{n-2}}{(n-2)! \sigma^n} \exp(-\sigma^{-1}(t+s-n\mu)), \quad t > 0, \quad s > n\mu. \quad (4)$$

Сделаем взаимно-однозначный переход от вектора Y_n к вектору Z_n по формулам

$$Z_1 = Y_1, \quad Z_j = Y_2 + \dots + Y_j, \quad j=2, \dots, n.$$

Якобиан перехода от Y_n к Z_n равен 1, так что ввиду (3) и (4)

$$f_{Z_n}^{\mu, \sigma}(z_n) = \sigma^{-n} \exp(-\sigma^{-1}(z_1 + z_n - n\mu)), \quad 0 < z_2 < \dots < z_n, \quad z_1 > n\mu,$$

$$f_{Z_1, Z_n}^{\mu, \sigma}(z_1, z_n) = \frac{z_n^{n-2}}{(n-2)! \sigma^n} \exp(-\sigma^{-1}(z_1 + z_n - n\mu)), \quad z_n > 0, \quad z_1 > n\mu.$$

Отсюда получаем при $z_1 > n\mu$, $z_n > z_{n-1} > \dots > z_2 > 0$:

$$\begin{aligned} f_{z_1, \dots, z_{n-1} | z_1, z_n}(z_1, \dots, z_{n-1} | z_1, z_n) = \\ = f_{z_n}^{\mu, \sigma}(z_n) / f_{z_1, z_n}^{\mu, \sigma}(z_1, z_n) = (n-2)! / z_n^{n-2}. \end{aligned}$$

Таким образом, условная плотность Z_2, \dots, Z_{n-1} при условии Z_1, Z_n не зависит от параметров μ, σ , и, следовательно, вся информация о параметрах содержится в статистике

$$(z_1, z_n) = (y_1, y_2 + \dots + y_n) = (x_{(1)}, \sum_{j=2}^n (x_{(j)} - x_{(1)})),$$

или в эквивалентной ей статистике $(x_{(1)}, \bar{x})$. Отметим, что оценки параметров μ, σ , с которыми мы встречались в § 6 и 11, являются функциями статистик $(x_{(1)}, \bar{x})$.

(IV) Пусть $\mathcal{X}=R^n$, $\mathcal{B}=\mathcal{B}^n$, а мера $P_{\mu, \sigma}$, $-\infty < \mu < +\infty$, $\sigma > 0$, задается n -мерной нормальной плотностью

$$(2\pi\sigma)^{-n} \exp \left(-\frac{1}{2\sigma} \sum_{i=1}^n (x_i - \mu)^2 \right).$$

От случайного вектора X_n с плотностью $N(\mu \cdot 1, \sigma^2 I)$, где $1 = (1, \dots, 1)$, I — единичная $n \times n$ -матрица, с помощью ортогональной матрицы C , у которой последний столбец имеет вид $n^{-1/2}1'$, перейдем к вектору $Y_n = X_n C$, имеющему распределение $N(\mu 1C, \sigma^2 I)$. При этом (см. п. 3 § 8)

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i, \quad \sum_{i=1}^{n-1} Y_i^2 = \sum_{i=1}^n (X_i - \bar{X})^2, \quad (5)$$

с. в. Y_i , $i = 1, \dots, n$, независимы, Y_n имеет распределение $N(\mu \sqrt{n}, \sigma)$, а вектор $\sigma^{-1}(Y_1, \dots, Y_{n-1})$ — стандартный нормальный. Вектор Y_n получен из X_n взаимно-однозначным преобразованием и, следовательно, как носитель информации полностью эквивалентен X_n . Поскольку плотность распределения вектора (Y_1, \dots, Y_{n-1})

$$f_{Y_1, \dots, Y_{n-1}}^{\mu, \sigma}(y_1, \dots, y_{n-1}) = (\sqrt{2\pi}\sigma)^{-n+1} \exp \left(-\frac{1}{2\sigma} \sum_{i=1}^{n-1} y_i^2 \right) \quad (6)$$

сферически симметрична, то нетрудно понять, что условное распределение точки (Y_1, \dots, Y_{n-1}) на сфере $y_1^2 + \dots + y_{n-1}^2 = r^2$ — равномерное. Нам важно лишь, что это распределение не зависит

от μ и σ . Проверим это, переходя к полярным координатам r , $\varphi_1, \dots, \varphi_{n-2}$ по формулам

$$\begin{aligned} y_1 &= r \cos \varphi_1 & (0 \leq \varphi_1 \leq \pi), \\ y_2 &= r \sin \varphi_1 \cos \varphi_2 & (0 \leq \varphi_2 \leq \pi), \end{aligned} \quad (7)$$

$$\begin{aligned} y_{n-2} &= r \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{n-3} \cos \varphi_{n-2} & (0 \leq \varphi_{n-2} \leq \pi), \\ y_{n-1} &= r \sin \varphi_1 \sin \varphi_2 \dots \sin \varphi_{n-3} \sin \varphi_{n-2} & (0 < r). \end{aligned}$$

При этом, очевидно, $y_1^2 + \dots + y_{n-1}^2 = r^2$, а индукцией по n легко установить, что якобиан перехода равен

$$\left| \frac{\partial(y_1, \dots, y_{n-1})}{\partial(r, \varphi_1, \dots, \varphi_{n-2})} \right| = r^{n-2}\xi,$$

$$\xi = \sin^{n-3} \varphi_1 \sin^{n-2} \varphi_2 \dots \cdot \sin^2 \varphi_{n-4} \sin \varphi_{n-3}. \quad (8)$$

Введем сл. в. $R, \Phi_1, \dots, \Phi_{n-2}$, связанные с Y_1, \dots, Y_{n-1} теми же соотношениями (7), что координаты $r, \varphi_1, \dots, \varphi_{n-2}$ связаны с y_1, \dots, y_{n-1} . Вероятность попадания точки $(R, \Phi_1, \dots, \Phi_{n-2})$ в борелевское множество B вычисляется как интеграл

$$\int_{B'} \dots \int f_{Y_1, \dots, Y_{n-1}}^{\mu, \sigma}(y_1, \dots, y_{n-1}) dy_1 \dots dy_{n-1} \quad (9)$$

по области B' , являющейся образом B при отображении (7). Совершая в интеграле (9) замену переменных (7) и учитывая (6) и (8), приведем его к виду

$$\int_B \dots \int (\sqrt{2\pi}\sigma)^{-n+1} \exp\left(-\frac{1}{2\sigma^2} r^2\right) r^{n-2}\xi d\varphi_1 \dots d\varphi_{n-2} dr, \quad (10)$$

где ξ , определенное в (8), является функцией только $\varphi_1, \dots, \varphi_{n-2}$. Из (10) вытекает, что совместная плотность сл. в. $R, \Phi_1, \dots, \Phi_{n-2}$ равна

$$f_{R, \Phi_1, \dots, \Phi_{n-2}}(r, \varphi_1, \dots, \varphi_{n-2}) = (\sqrt{2\pi}\sigma)^{-n+1} \exp\left(-\frac{1}{2\sigma^2} r^2\right) r^{n-2}\xi. \quad (11)$$

(Отметим, что полярные координаты находятся во взаимно-однозначном соответствии с декартовыми при условии $r \sin \varphi_1 \dots \sin \varphi_{n-3} \neq 0$, но так как $R \sin \Phi_1 \dots \sin \Phi_{n-3} \neq 0$ с вероятностью 1 по любой мере $P_{\mu, \sigma}$, то никаких осложнений не возникает.) Поскольку $\sigma^{-2}R^2 = \sigma^{-2}(Y_1^2 + \dots + Y_{n-1}^2)$ имеет распределение $\chi_{n-1}^2 = G(1/2, (n-1)/2)$, то R^2 имеет распределение $G(1/(2\sigma^2), (n-1)/2)$ с плотностью (см. (2), § 7)

$$f_{R^2}^{\mu, \sigma}(x) = \frac{x^{(n-3)/2}}{(2\sigma^2)^{(n-1)/2} \Gamma\left(\frac{n-1}{2}\right)} e^{-x/(2\sigma^2)}.$$

откуда

$$f_R^{\mu, \sigma}(r) = 2r f_R^{\mu, \sigma}(r^2) = -\frac{r^{n-2}}{2^{(n-3)/2} \sigma^{n-1} \Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{r^2}{2\sigma^2}}. \quad (12)$$

Интегрируя (11) по множеству $\{r > 0\}$ и учитывая, что интеграл от плотности (12) равен 1, найдем плотность

$$f_{\Phi_1, \dots, \Phi_{n-2}}(\varphi_1, \dots, \varphi_{n-2}) = c_n \sin^{n-3} \varphi_1 \sin^{n-4} \varphi_2 \cdots \sin^2 \varphi_{n-4} \sin \varphi_{n-3}, \quad (13)$$

где c_n — постоянная, которая не зависит от μ и σ . Совместная плотность (11) равна произведению плотностей (12) и (13). Таким образом, случайный вектор $(Y_n, R, \Phi_1, \dots, \Phi_{n-2})$ находится во взаимно-однозначном соответствии с вектором X_n , за исключением множества значений $P_{\mu, \sigma}$ -меры нуль. Его плотность равна

$$\begin{aligned} & f_{Y_n, R, \Phi_1, \dots, \Phi_{n-2}}^{\mu, \sigma}(y_n, r, \varphi_1, \dots, \varphi_{n-2}) = \\ & = f_{Y_n}^{\mu, \sigma}(y_n) f_R^{\mu, \sigma}(r) f_{\Phi_1, \dots, \Phi_{n-2}}(\varphi_1, \dots, \varphi_{n-2}). \end{aligned}$$

Отсюда условная плотность величин $\Phi_1, \dots, \Phi_{n-2}$ при условии сл. в. Y_n, R равна их безусловной плотности (13) и не зависит от (μ, σ) . Следовательно, двумерная статистика

$$(y_n, r) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i, \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \quad (14)$$

содержит всю информацию о параметрах μ, σ . Вместо (14) можно рассматривать эквивалентные статистики

$$\left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2 \right) \text{ или } (\bar{x}, s^2), \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (15)$$

связанные с (14) взаимно-однозначным преобразованием. Таким образом, выборочные среднее \bar{x} и дисперсия s_0^2 образуют достаточную статистику.

2. Определение достаточной статистики, теорема факторизации.

Как мы уже отмечали, понятие достаточной статистики в непрерывной модели вполне аналогично дискретному случаю. Однако математические тонкости, связанные с введением условных вероятностей, не позволяют нам дать определение достаточной статистики в его общей форме. Мы будем рассматривать только параметрические модели и, если не оговорено противное, только такие статистики $T_r(x_n) = (T_1(x_n), \dots, T_r(x_n))$, которые можно до-

полинить статистикой $\mathbf{Y}_{n-r}(\mathbf{x}_n) = (Y_1(\mathbf{x}_n), \dots, Y_{n-r}(\mathbf{x}_n))$ такой, что замена переменных

$$t_1 = T_1(\mathbf{x}_n), \dots, t_r = T_r(\mathbf{x}_n), \quad y_1 = Y_1(\mathbf{x}_n), \dots, y_{n-r} = Y_{n-r}(\mathbf{x}_n) \quad (16)$$

взаимно-однозначна на множестве P_θ -меры единица при любом $\theta \in \Theta$, а плотности меры P_θ в координатах \mathbf{x}_n и (t_r, y_{n-r}) связаны на множестве P_θ -меры единица соотношением

$$f_{\mathbf{x}_n}^\theta(\mathbf{x}_n) = f_{T_r(\mathbf{x}_n), Y_{n-r}(\mathbf{x}_n)}^\theta(t_r, y_{n-r}) |J|, \quad (17)$$

где

$$|J| = \left| \frac{\partial(t_1, \dots, t_r, y_1, \dots, y_{n-r})}{\partial(x_1, x_2, \dots, x_n)} \right|$$

— якобиан перехода от координат (t_r, y_{n-r}) к (\mathbf{x}_n) .

Иначе говоря, предположение (16), (17) означает, что, во-первых, существует плотность случайного вектора $(T_r(\mathbf{X}_n), Y_{n-r}(\mathbf{X}_n))$, и, во-вторых, при использовании этой плотности для вычисления интеграла

$$\int_B \dots \int f_{T_r(\mathbf{x}_n), Y_{n-r}(\mathbf{x}_n)}^\theta(t_r, y_{n-r}) dt_1 \dots dt_r dy_1 \dots dy_{n-r}$$

можно воспользоваться заменой переменных (16) и свести этот интеграл к следующему

$$\int_{B'} \dots \int f_{\mathbf{x}_n}^\theta(\mathbf{x}_n) dx_1 \dots dx_n,$$

где B' является прообразом B при отображении (16).

Определение. Пусть статистика $T_r(\mathbf{x}_n)$ такова, что выполняются предположения (16), (17). Если условная плотность, определенная при t_r, θ , таких, что $f_{T_r(\mathbf{x}_n)}^\theta(t_r) > 0$, формулой

$$f_{Y_{n-r}(\mathbf{x}_n) | T_r(\mathbf{x}_n)}^\theta(y_{n-r} | t_r) = f_{T_r(\mathbf{x}_n), Y_{n-r}(\mathbf{x}_n)}^\theta(t_r, y_{n-r}) / f_{T_r(\mathbf{x}_n)}^\theta(t_r) \quad (18)$$

не зависит от $\theta \in \Theta$, то статистика $T_r(\mathbf{x}_n)$ называется *достаточной* для данной модели.

Отметим, что

$$f_{T_r(\mathbf{x}_n)}^\theta(t_r) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{T_r(\mathbf{x}_n), Y_{n-r}(\mathbf{x}_n)}^\theta(t_r, y_{n-r}) dy_1 \dots dy_{n-r}.$$

Теорема факторизации. В предположении (16), (17) статистика $T_r(\mathbf{x}_n)$ является достаточной тогда и только тогда, когда для любого $\theta \in \Theta$ на некотором множестве P_θ -меры единица выполняется равенство

$$f_{\mathbf{x}_n}(\mathbf{x}_n; \theta) = g(T_r(\mathbf{x}_n); \theta) h(\mathbf{x}_n), \quad (19)$$

здесь $g(t_r; \theta)$ и $h(x_n)$ — неотрицательные измеримые функции аргументов t_r и x_n соответственно, $h(x_n)$ не зависит от θ .

Доказательство. Пусть статистика $T_r(x_n)$ — достаточная. Положим

$$g(t_r; \theta) = f_{T_r(x_n)}^{\theta}(t_r),$$

и для каждой пары (t_r, y_{n-r}) , для которой существует $\theta \in \Theta$, что $f_{T_r(x_n)}^{\theta}(t_r) > 0$, положим

$$h_1(t_r, y_{n-r}) = f_{Y_{n-r}(x_n) | T_r(x_n)}^{\theta}(y_{n-r} | t_r). \quad (20)$$

Ввиду достаточности статистики $T_r(x_n)$ функция $h_1(t_r, y_{n-r})$ формулой (20) задается корректно (и является функцией только от t_r, y_{n-r}), а равенство (20) при каждом θ, t_r , для которых правая часть определена, выполняется на множестве P_θ -меры единицы. Учитывая введенные обозначения, из (18) получаем равенство

$$f_{T_r(x_n), Y_{n-r}(x_n)}^{\theta}(t_r, y_{n-r}) = g(t_r; \theta) h_1(t_r, y_{n-r}), \quad (21)$$

справедливое на множестве P_θ -меры единица, откуда, используя (17), получаем

$$f_{X_n}^{\theta}(x_n) = g(T_r(x_n); \theta) h_1(T_r(x_n), Y_{n-r}(x_n)) |J|. \quad (22)$$

Поскольку якобиан $|J|$ не зависит от θ , то, полагая $h = h_1 |J|$, получаем представление (19).

Обратно, пусть имеет место факторизация (19). Определим функцию $h_1(t_r, y_{n-r})$ условием

$$h(x_n) = h_1(t_r, y_{n-r}), \quad t_r = T_r(x_n), \quad y_{n-r} = Y_{n-r}(x_n)$$

и перейдем в (19) к координатам (t_r, y_{n-r}) . Все дальнейшие рассуждения проведем на подмножестве \mathcal{X} , на котором $h_1 > 0$, $h > 0$. Воспользовавшись формулой преобразования (17), перепишем (19) в виде

$$f_{T_r(x_n), Y_{n-r}(x_n)}^{\theta}(t_r, y_{n-r}) |J| = g(t_r; \theta) h_1(t_r, y_{n-r}). \quad (23)$$

Найдем отсюда плотность статистики $T_r(X_n)$:

$$\begin{aligned} f_{T_r(x_n)}^{\theta}(t_r) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{T_r(x_n), Y_{n-r}(x_n)}^{\theta}(t_r, y_{n-r}) dy_1 \dots dy_{n-r} = \\ &= g(t_r; \theta) \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_1(t_r, y_{n-r}) |J|^{-1} dy_1 \dots dy_{n-r}. \end{aligned} \quad (24)$$

Если $g(t_r; \theta) > 0$, то, поделив (23) на (24), получаем, что условная плотность равна

$$f_{Y_{n-r}(x_n) | T_r(x_n)}^{\theta} (y_{n-r} | t_r) = \\ = h_1(t_r, y_{n-r}) |J|^{-1} / \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_1(t_r, z_{n-r}) |J|^{-1} dz_1 \dots [dz_{n-r}]$$

и не зависит от θ , что и требовалось установить. ■

Достаточность рассматривавшихся в примерах (II)–(IV) статистик с помощью теоремы факторизации устанавливается совсем просто. Например, расписав плотность случайной выборки из нормального распределения $N(\mu, \sigma)$ в виде

$$(V 2\pi \sigma)^{-n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{\mu^2}{2\sigma^2} \right), \quad (25)$$

обнаруживаем, что пара $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ является достаточной статистикой, что было непосредственно проверено выше (см. (15)). К примеру (I) теорема факторизации в предложенной выше форме неприменима, так как условное распределение выборки при условии достаточной статистики — вариационного ряда — является дискретным. Это один из редких в приложениях случаев, когда не выполняются ограничивающие предположения (16), (17). Обратимся к другим примерам.

(V) Рассмотрим линейную статистическую модель с нормальными наблюдениями в канонической форме (см. (7) § 10):

$$U_i = \psi_i + \sigma \delta_i, \quad i = 1, \dots, r; \quad U_i = \sigma \delta_i, \quad i = r+1, \dots, n, \quad (26)$$

где $\psi = (\psi_1, \dots, \psi_r)$, $\sigma > 0$ — неизвестные параметры, $(\delta_1, \delta_2, \dots, \delta_n)$ — стандартный нормальный вектор. Плотность распределения U равна

$$f_U^{\psi, \sigma}(\mathbf{u}) = (V 2\pi \sigma)^{-n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^r (u_i - \psi_i)^2 - \frac{1}{2\sigma^2} \sum_{i=r+1}^n u_i^2 \right) = \\ = (V 2\pi \sigma)^{-n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n u_i^2 + \frac{1}{\sigma^2} \sum_{i=1}^r u_i \psi_i - \frac{1}{2\sigma^2} \sum_{i=1}^r \psi_i^2 \right)$$

По теореме факторизации, статистики

$$T_i(\mathbf{u}) = u_i, \quad i = 1, \dots, r, \quad S_0^2(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n u_i^2$$

образуют $r+1$ -мерную достаточную статистику. Эквивалентная ей статистика

$$(T_1(\mathbf{u}), \dots, T_r(\mathbf{u}), S_1^2(\mathbf{u})), S_1^2(\mathbf{u}) = \frac{1}{n-r} \sum_{i=r+1}^n u_i^2$$

также является достаточной. Заметим, что оценки $\hat{\psi}_i = T_i(\mathbf{u})$, $\hat{\sigma}^2 = S_1^2(\mathbf{u})$ параметров ψ , σ линейной модели были введены в § 10.

Аналогично обстоит дело и для нормальной линейной модели в общей форме (1) § 13 с матрицей X полного ранга. Плотность распределения наблюдений в этом случае равна

$$\begin{aligned} f_V^{\theta, \sigma}(\mathbf{y}) &= (\sqrt{2\pi}\sigma)^{-n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^r \theta_j x_{ij} \right)^2 \right) = \\ &= (\sqrt{2\pi}\sigma)^{-n} \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \boldsymbol{\theta}X'\|^2 \right), \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_r), \end{aligned} \quad (27)$$

$$X = [\mathbf{x}_1, \dots, \mathbf{x}_r], \quad \mathbf{x}_j = (x_{1j}, \dots, x_{nj}), \quad j = 1, \dots, r.$$

Введем статистики $T_i(\mathbf{y})$, $i = 1, \dots, r$, как коэффициенты разложения проекции вектора \mathbf{y} на подпространство V , порожденное векторами \mathbf{x}_j , $j = 1, \dots, r$:

$$\text{pr}_V \mathbf{y} = T_1(\mathbf{y}) \mathbf{x}_1 + \dots + T_r(\mathbf{y}) \mathbf{x}_r.$$

Так как вектор

$$\boldsymbol{\theta}X' = \theta_1 \mathbf{x}_1 + \dots + \theta_r \mathbf{x}_r \in V,$$

то, применяя теорему Пифагора, найдем

$$\begin{aligned} \|\mathbf{y} - \boldsymbol{\theta}X'\|^2 &= \|(\mathbf{y} - \text{pr}_V \mathbf{y} + \text{pr}_V \mathbf{y} - \boldsymbol{\theta}X')\|^2 = \|\mathbf{y} - \text{pr}_V \mathbf{y}\|^2 + \|\text{pr}_V \mathbf{y} - \boldsymbol{\theta}X'\|^2 = \\ &= \left\| \mathbf{y} - \sum_{i=1}^r T_i(\mathbf{y}) \mathbf{x}_i \right\|^2 + \left\| \sum_{i=1}^r (T_i(\mathbf{y}) - \theta_i) \mathbf{x}_i \right\|^2. \end{aligned} \quad (28)$$

Подставляя (28) в (27) и применяя теорему факторизации, получаем достаточную статистику в виде

$$(T_1(\mathbf{y}), \dots, T_r(\mathbf{y}), S_1^2(\mathbf{y})), S_1^2(\mathbf{y}) = \frac{1}{n-r} \left\| \sum_{i=1}^r (T_i(\mathbf{y}) - \theta_i) \mathbf{x}_i \right\|^2.$$

Остается заметить, что известные нам оценки параметров $\boldsymbol{\theta}$, σ линейной модели имеют вид

$$\hat{\theta}_i = T_i(\mathbf{y}), \quad i = 1, \dots, r, \quad \hat{\sigma}^2 = S_1^2(\mathbf{y}).$$

(VI) Рассмотрим цензурируемую выборку (см. п. 7 § 11)

$$X_{(r_1+1)} \leq X_{(r_1+2)} \leq \dots \leq X_{(n-r_2)}, \quad r_1, r_2 \geq 0, \quad (29)$$

считая наблюдения X_1, \dots, X_n независимыми $N(\mu, \sigma)$ -распределенными, а r_1, r_2 — фиксированными числами. Найдем плотность распределения вектора

$$\mathbf{Y} = (Y_1, \dots, Y_N) = (X_{(r_1+1)}, X_{(r_1+2)}, \dots, X_{(n-r_2)}), \quad N = n - r_1 - r_2,$$

в точке $\mathbf{y} = (y_1, \dots, y_N)$, $y_1 < y_2 < \dots < y_N$. Проведем вначале нестрогие рассуждения. Плотность вероятности того, что какие-то N из сл. в. X_1, \dots, X_n примут значения $y_1 < y_2 < \dots < y_N$, равна

$$n(n-1)\dots(n-N+1) \prod_{i=1}^N \frac{1}{\sigma} \varphi((y_i - \mu)/\sigma),$$

где $\varphi(x) = N(0, 1)$ -плотность. Вероятность, что слева от y_1 и справа от y_N окажется ровно r_1 и r_2 соответственно каких-то из $n-N$ оставшихся наблюдений, равна

$$C_{n-N}^{r_1} \Phi((y_1 - \mu)/\sigma)^{r_1} (1 - \Phi((y_N - \mu)/\sigma))^{r_2},$$

где $\Phi(x)$ — ф.р. закона $N(0, 1)$. Используя независимость, перемножим полученные выражения и получим

$$f_Y(\mathbf{y}) = \frac{n!}{r_1! r_2!} \Phi((y_1 - \mu)/\sigma)^{r_1} (1 - \Phi((y_N - \mu)/\sigma))^{r_2} \times \\ \times \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right). \quad (30)$$

[Проведем формальную выкладку. Возьмем $\delta > 0$ настолько малым, что интервалы $(y_i, y_i + \delta)$, $i = 1, \dots, N$, не пересекаются и запишем

$$\mathcal{P}_{\mu, \sigma}(Y_j \in (y_j, y_j + \delta], j = 1, \dots, N) = \\ + \sum_{k_1, k_2, \dots, k_N} \mathcal{P}_{\mu, \sigma}(Y_j = X_{k_j}, Y_j \in (y_j, y_j + \delta], j + 1, \dots, N), \quad (31)$$

где суммирование ведется по всем $n(n-1)\dots(n-N+1)$ упорядоченным наборам индексов (k_1, k_2, \dots, k_N) без повторения из множества $1, 2, \dots, n$. Учитывая, что все слагаемые суммы (31) равны между собой, перепишем (31) в виде

$$\frac{n!}{(n-N)!} \mathcal{P}_{\mu, \sigma}(Y_j = X_{r_1+j}, Y_j \in (y_j, y_j + \delta], j = 1, \dots, N) = \\ = \sum_{i_1, \dots, i_{r_1}} \frac{n!}{(n-N)!} \mathcal{P}_{\mu, \sigma}(\max(X_{i_1}, \dots, X_{i_{r_1}}) < X_{r_1+1}, \\ Y_j = X_{r_1+j}, Y_j \in (y_j, y_j + \delta], j = 1, \dots, N, \\ X_{n-r_1} < \min(X_{i_1}, \dots, X_{i_{r_1}})),$$

где суммирование ведется по всем $C_{n-N}^{r_1}$ неупорядоченным наборам индексов i_1, i_2, \dots, i_{r_1} без повторения из множества $\{1, \dots, r_1, n-r_2+1, \dots, n\}$, а l_1, l_2, \dots, l_{r_2} — остающиеся элементы этого множества. Снова учитывая, что слагаемые суммы (32) равны между собой, получим для (32) выражение

$$C_{n-N}^{r_1} \frac{n!}{(n-N)!} \mathcal{P}_{\mu, \sigma}(\max(X_1, \dots, X_{r_1}) < X_{r_1+1}, X_{r_1+j} \in (y_j, y_j + \delta]),$$

$$j = 1, \dots, N, X_{n-r_2} < \min(X_{n-r_2+1}, \dots, X_n)). \quad (33)$$

Запишем очевидную оценку сверху для (33):

$$\frac{n!}{r_1! r_2!} \mathcal{P}_{\mu, \sigma}(\max(X_1, \dots, X_{r_1}) < y_1 + \delta, X_{r_1+j} \in (y_j, y_j + \delta],$$

$$j = 1, \dots, N, y_N < \min(X_{n-r_2+1}, \dots, X_n)) =$$

$$= \frac{n!}{r_1! r_2!} \Phi\left(\frac{y_1 + \delta - \mu}{\sigma}\right)^{r_1} \prod_{j=1}^N \left(\Phi\left(\frac{y_j + \delta - \mu}{\sigma}\right) - \Phi\left(\frac{y_j - \mu}{\sigma}\right)\right) \times$$

$$\times \left(1 - \Phi\left(\frac{y_N - \mu}{\sigma}\right)\right)^{r_2}. \quad (34)$$

Аналогично получаем оценку снизу:

$$\frac{n!}{r_1! r_2!} \Phi\left(\frac{y_1 - \mu}{\sigma}\right)^{r_1} \prod_{j=1}^N \left(\Phi\left(\frac{y_j + \delta - \mu}{\sigma}\right) - \Phi\left(\frac{y_j - \mu}{\sigma}\right)\right) \times$$

$$\times \left(1 - \Phi\left(\frac{y_N + \delta - \mu}{\sigma}\right)\right)^{r_2}. \quad (35)$$

Разделив выражения (34), (35) на δ^N и устремляя δ к нулю, приходим к (30).]

Из формулы (30) и теоремы факторизации нетрудно вывести, что достаточной является четырехмерная статистика (при $r_1 > 0, r_2 > 0$)

$$(y_1, y_N, \sum_{i=1}^N y_i, \sum_{i=1}^N y_i^2).$$

Если цензурирование одностороннее: $r_1 = 0$ или $r_2 = 0$, то y_1 или y_N соответственно надо отбросить, и достаточная статистика оказывается трехмерной.

3. Экспоненциальные семейства, минимальная достаточность.

Рассмотрим непрерывную параметрическую модель

$(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$, где $\theta = (\theta_1, \dots, \theta_k)$, $k \geq 1$, $\Theta \subseteq R^k$,

а плотности $f_\theta(x)$ мер P_θ заданы формулой

$$f_{\theta}(x) = h(x) \exp \left(\sum_{i=1}^k \theta_i T_i(x) + b(\theta_1, \dots, \theta_k) \right). \quad (36)$$

Семейство \mathbf{P}_0 в этом случае назовем *k-параметрическим распределением экспоненциального типа* (в естественной параметризации).

Ряд важных семейств распределений повторной выборки после подходящей параметризации сводится к (36). Например, если отдельное наблюдение в последовательности n независимых испытаний имеет распределение $G(\lambda; p)$, то совместная плотность равна при $x_1, \dots, x_n > 0$ (см. (2) § 7)

$$\begin{aligned} f_{\lambda, p}(x_n) &= \frac{\lambda^{np}}{\Gamma(p)^n} x_1^{p-1} \dots x_n^{p-1} e^{-\lambda(x_1+\dots+x_n)} = \\ &= (x_1, \dots, x_n)^{-1} \exp \left(p \sum_{i=1}^n \ln x_i - \lambda \sum_{i=1}^n x_i + \ln (\lambda^{np}/\Gamma(p)^n) \right), \end{aligned} \quad (37)$$

так что, полагая $\theta_1 = p$, $\theta_2 = -\lambda$, $T_1 = \sum_{i=1}^n \ln x_i$, $T_2 = \sum_{i=1}^n x_i$, приходим к (36). Другой пример: независимая выборка из распределения $N(\mu, \sigma)$ имеет совместную плотность

$$\begin{aligned} f_{\mu, \sigma}(x_n) &= (\sqrt{2\pi}\sigma)^{-n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right) = \\ &= (\sqrt{2\pi})^{-n} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{\mu^2}{2\sigma^2} - n \ln \sigma \right) \end{aligned} \quad (38)$$

экспоненциального типа (36), если положить $\theta_1 = -1/\sigma^2$, $\theta_2 = \mu/\sigma^2$. ■

Более общая экспоненциальная модель вида

$$f_{\theta}(x) = h(x) \exp \left(\sum_{i=1}^k a_i(\theta) T_i(x) + b(\theta) \right), \quad (39)$$

где $\theta = (\theta_1, \dots, \theta_k)$, $\theta \in \Theta \subseteq R^k$, фактически может быть сведена к (36), если ввести естественную параметризацию $\psi_i = a_i(\theta)$. Чтобы убедиться в этом, достаточно заметить, что $b(\theta)$ на самом деле является функцией от $a_i(\theta)$, $i = 1, \dots, k$, поскольку $\exp(b(\theta))$ выполняет в (39) роль нормирующего множителя.

Отметим, что повторная выборка из распределения экспоненциального типа снова принадлежит к этому типу:

$$f_{\theta}(x^{(1)}, x^{(2)}, \dots, x^{(m)}) = h(x^{(1)}) \dots h(x^{(m)}) \exp \left(\sum_{i=1}^k a_i(\theta) \times \right. \\ \left. \times \sum_{j=1}^m T_i(x^{(j)}) + mb(\theta) \right).$$

Поэтому в примерах (37), (38) можно было рассматривать только случай $n=1$. ■

Из теоремы факторизации при дополнительных ограничениях (16), (17) вытекает, что векторная статистика $T(x) = (T_1(x), \dots, T_k(x))$ является достаточной в модели (36). Отметим, что ограничения (16), (17) являются излишними при более широком понимании достаточности. Как и в дискретной модели, интерес представляют достаточные статистики, сокращающие данные в наибольшей степени. Достаточную статистику $S(x)$ назовем *минимальной достаточной статистикой*, если она представима как (вектор-) функция от любой достаточной статистики.

С достаточной статистикой $T(x)$ связем *достаточное разбиение* множества \mathcal{X} :

$$\mathcal{X} = \bigcup \{x : T(x) = t\}. \quad (40)$$

С минимальной статистикой $S(x)$ связано *минимальное достаточное разбиение* — такое, что элементы любого достаточного разбиения содержатся в элементах минимального:

$$\{x : T(x) = t\} \subseteq \{x : S(x) = st\}.$$

Как и в дискретной модели, минимальное достаточное разбиение, правда при некоторых ограничениях, образуется из классов эквивалентности по отношению R :

$$xRy \Leftrightarrow f_{\theta}(x)/f_{\theta}(y) \text{ не зависит от } \theta. \quad (41)$$

Проверим это утверждение на экспоненциальной модели (39). Заметим, во-первых, что плотности (39) обращаются в нуль на общем множестве $\{x : h(x) = 0\}$ и потому можно исключить из рассмотрения все такие x . Далее, отношение

$$f_{\theta}(x)/f_{\theta}(y) \exp \left(\sum_{i=1}^k a_i(\theta) (T_i(x) - T_i(y)) \right)$$

для данных x, y не зависит от θ , если либо $T(x) = T(y)$, либо функции $1, a_i(\theta), \theta \in \Theta, i=1, \dots, k$, линейно зависимы и коэффициенты $c_i = T_i(x) - T_i(y)$ таковы, что

$$c_1 a_1(\theta) + \dots + c_k a_k(\theta) \equiv \text{const}, \quad \theta \in \Theta. \quad (42)$$

Если при некоторых $c_1, \dots, c_k, c_1^2 + \dots + c_k^2 > 0$, имеет место равенство (42), то модель можно редуцировать, взяв из системы $1, a_i(\theta), i=1, \dots, k$, максимальную линейно независимую подсистему

и заменив в (39) остальные функции $a_i(\theta)$ на линейное представление через эту систему. Поэтому мы предположим, что 1, $a_i(\theta)$, $i=1, \dots, k$, — линейно независимые. В таком случае разбиение на классы эквивалентности по отношению к (41) совпадает с достаточным разбиением (40). Покажем, что это разбиение минимально. Пусть достаточная статистика $S(x)$ такова,

$$S(x) = \psi(T(x)). \quad (43)$$

По теореме факторизации, если $S(x) = S(y) = s$, то

$$f_{\theta}(x) = g_1(s; \theta) h_1(x), \quad f_{\theta}(y) = g_1(s; \theta) h_1(y)$$

и для всех $\theta \in \Theta$, таких, что $g_1(s; \theta) > 0$, отношение

$$f_{\theta}(x)/f_{\theta}(y) = h_1(x)/h_1(y)$$

не зависит от θ , т. е. $T(x) = T(y)$, что и требовалось установить. ■

Подчеркнем, что в (43) рассматриваются лишь такие минимальные статистики $S(x)$, которые удовлетворяют ограничениям (16), (17), и тем самым утверждение о минимальности $T(x)$ доказано лишь в этом классе статистик. В примерах (37), (38) требования (16), (17) удовлетворяются, и достаточные статистики $(\sum_{i=1}^n \ln x_i, \sum_{i=1}^n x_i)$ и $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ являются минимальными соответственно для выборок из распределений $G(\lambda; p)$ и $N(\mu, \sigma)$. В линейной статистической модели с нормальными наблюдениями (27) статистики (29), служащие оценками коэффициентов θ и σ^2 , также являются минимальными достаточными.

Равномерное распределение на $(0; \theta)$ (см. пример (II)), экспоненциальное распределение с параметрами сдвига-масштаба (пример (III)) не принадлежат к семейству экспоненциального типа. Характерная особенность этих распределений заключается в том, что множество, где плотность распределения отлична от нуля (носитель распределения), зависит от параметра, в то время как в (36) носитель у всех распределений один и тот же — $\{x : h(x) > 0\}$. Однако минимальность рассматриваемых в примерах (II), (III) достаточных статистик доказывается теми же рассуждениями, что и для семейств экспоненциального типа. Такое же заключение справедливо и для цензурированной выборки примера (IV). ■

К построению минимального достаточного разбиения имеется и другой подход. Оказывается, что, как и в дискретной модели, достаточное разбиение порождается следующей системой статистик:

$$L(z; \theta) = f_{\theta}(x), \quad \theta \in \Theta,$$

зависящих от θ как от параметра, а минимальное достаточное разбиение порождается функциональной статистикой

$$L(x; \theta)/L(x; \theta_0); \quad \theta \in \Theta, \quad (44)$$

в предположении, что существует такое $\theta_0 \in \Theta$, что носители всех мер P_θ содержатся в носителе P_{θ_0} .

Если носители всех мер P_θ совпадают, то в (44) удобнее перейти к логарифмам.

$$\ln L(x; \theta) - \ln L(x; \theta_0) \equiv g(x; \theta), \quad \theta \in \Theta. \quad (45)$$

Рассмотрим, к чему приводит этот подход в экспоненциальной модели (39). Подставляя $f_\theta(x)$ из (39) в (45), получаем

$$g(x; \theta) = \sum_{i=1}^k (a_i(\theta) - a_i(\theta_0)) T_i(x) + (b(\theta) - b(\theta_0)), \quad \theta \in \Theta. \quad (46)$$

Следовательно, разбиение X , порожденное совокупностью статистик (45), совпадает в этом случае с разбиением, порожденным векторной статистикой $T(x) = (T_1(x), \dots, T_k(x))$, которая, как мы уже проверяли, является минимальной достаточной. Таким образом, в экспоненциальной модели (39) удается сократить минимальную достаточную функциональную статистику (45) до k -мерной векторной $T(x)$. Полезно отметить, что если в экспоненциальной модели (39) система функций 1, $a_i(\theta)$, $i=1, \dots, k$, $\theta \in \Theta$ взята линейно независимой, то, как нетрудно понять, можно подобрать значения $\theta^{(i)} \in \Theta$, $i=1, \dots, k$, так, что матрица $[a_i(\theta^{(j)}) - a_i(\theta_0)]$, $i, j = 1, \dots, k$ невырождена, и из соотношений (46) $T_i(x)$, $i=1, \dots, k$, могут быть линейно выражены через функции 1, $g(x; \theta^{(i)})$, $i=1, \dots, k$:

$$T_i(x) = c_0^{(i)} + \sum_{j=1}^n c_j^{(i)} g(x; \theta^{(j)}).$$

Перейдем к произвольной непрерывной модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$, у которой носители всех мер P_θ совпадают. Рассмотрим пространство \mathcal{L} всевозможных конечных линейных комбинаций:

$$\Phi(x) = c_0 + \sum_{j=1}^l c_j g(x; \theta_j),$$

где $\theta_j \in \Theta$, c_j — произвольные постоянные $j=1, \dots, l$, l — любое натуральное. \mathcal{L} является линейным пространством, вообще говоря, бесконечномерным. В случае экспоненциальной модели, как мы только что видели, \mathcal{L} конечномерно: 1, $T_i(x)$, $i=1, \dots, k$, порождает \mathcal{L} . Допустим, что в \mathcal{L} имеется не более чем счетный базис 1, $\varphi_i(x)$, $i=1, \dots, k$ (k может быть и ∞), так что любой

элемент $\varphi(x) \in \mathcal{L}$ представим в виде конечной линейной комбинации элементов базиса:

$$\varphi(x) = b_0 + \sum_{i=1}^m b_i \varphi_i(x).$$

В частности, это верно и для каждой функции $g(x; \theta)$, $\theta \in \Theta$. В таком случае, k -мерная статистика

$$(\varphi_1(x), \varphi_2(x), \dots, \varphi_k(x)), \quad k \leq \infty, \quad (47)$$

эквивалентна функциональной статистике $g(x; \theta)$, $\theta \in \Theta$, так как $g(x; \theta) \in \mathcal{L}$ при любом $\theta \in \Theta$, и поэтому линейно выражается через систему. Отсюда совпадение разбиений, порождаемых системой статистик $g(x; \theta)$, $\theta \in \Theta$, и k -мерной статистикой (47), и, следовательно, статистика (47) минимально достаточна.

Интересно заметить, что если размерность k статистики (47) конечна, то, записывая при каждом $\theta \in \Theta$ разложение элемента $g(x; \theta) \in \mathcal{L}$ по базису $1, \varphi_1(x), \dots, \varphi_k(x)$, получаем

$$\ln f_\theta(x) - \ln f_{\theta_0}(x) = g(x; \theta) = \sum_{i=1}^k a_i(\theta) \varphi_i(x) + b(\theta),$$

откуда

$$f_\theta(x) = f_{\theta_0}(x) \exp \left(\sum_{i=1}^k a_i(\theta) \varphi_i(x) + b(\theta) \right), \quad (48)$$

т. е. статистическая модель является экспоненциальной. Таким образом, семейство распределений с общим носителем является экспоненциальным тогда и только тогда, когда пространство \mathcal{L} конечномерно. ■

Рассмотрим, что дает описанный подход в модели с независимой повторной выборкой, когда плотность $f_\theta(x)$, $\theta \in \Theta$, отдельного наблюдения сосредоточена на одном и том же множестве. Пусть

$$g(x; \theta) = \ln f_\theta(x) - \ln f_{\theta_0}(x), \quad \theta \in \Theta, \quad (49)$$

и \mathcal{L} — линейное пространство функций $\varphi(x)$, порожденное линейными комбинациями функций системы (49). Допустим, что $1, \varphi_i(x)$, $i = 1, \dots, k$ (k может равняться и ∞) — базис \mathcal{L} , так что при любом $\theta \in \Theta$

$$g(x; \theta) = \sum_{i=1}^{m(\theta)} c_i(\theta) \varphi_i(x) + c_0(\theta). \quad (50)$$

Из предыдущего вытекает, что k -мерная статистика

$$(\varphi_1(x), \varphi_2(x), \dots, \varphi_k(x))$$

является минимальной достаточной. Подчеркнем, что здесь всюду

x — скалярная переменная. Перейдем теперь к повторной независимой выборке объема n . Очевидно,

$$g(x_n; \theta) = \sum_{j=1}^n g(x_j; \theta), \theta \in \Theta,$$

а из (50) вытекает, что

$$\begin{aligned} g(x_n; \theta) &= \sum_{j=1}^n \left(\sum_{i=1}^{m(\theta)} c_i(\theta) \varphi_i(x_j) + c_0(\theta) \right) = \\ &= \sum_{i=1}^{m(\theta)} c_i(\theta) \sum_{j=1}^n \varphi_i(x_j) + nc_0(\theta). \end{aligned}$$

В таком случае ясно, что k -мерная статистика

$$(\psi_1(x_n), \dots, \psi_k(x_n)), \psi_i(x_n) = \sum_{j=1}^n \varphi_i(x_j), i = 1, \dots, k, \quad (51)$$

является минимальной достаточной.

Если размерность k статистики (51) конечна, то при $n > k$ она дает сокращение данных по сравнению с тривиальной достаточной статистикой — вариационным рядом. Заметим, что при этом, как вытекает из (48), распределения отдельного наблюдения и всей выборки принадлежат экспоненциальному типу. Однако при условии $k < n$ остается открытый вопрос о сокращении размерности статистики (51) с сохранением достаточности.

При дополнительном предположении, что все плотности $f_\theta(x)$, $\theta \in \Theta$, непрерывно дифференцируемы на некотором общем для всех интервале Δ , мы докажем следующее утверждение: любые $s, s \leq \min(k, n)$, компонент статистики (51) функционально независимы. При $k \leq n$ отсюда, очевидно, следует, что размерность статистики (51) не может быть понижена с сохранением свойства достаточности. Действительно, пусть некоторая статистика $(\chi_1(x_n), \dots, \chi_l(x_n))$, $l < k$, является достаточной; тогда ввиду минимальности статистики (51) все ее компоненты $\psi_i(x_n)$, $i = 1, \dots, k$, должны функционально выражаться через $\chi_1(x_n), \dots, \chi_l(x_n)$. Но поскольку $l < k$, это приводит к функциональной зависимости $\psi_1(x_n), \dots, \psi_k(x_n)$.

При $k > n$ из сформулированного утверждения вытекает, что размерность достаточной статистики не может быть сделана меньше n . В самом деле, в этом случае статистики $\psi_1(x_n), \dots, \psi_n(x_n)$ функционально независимы, и потому якобиан $\partial(\psi_1, \dots, \psi_n)/\partial(x_1, \dots, x_n)$ отличен от нуля на множестве $\Delta'' = \{x_n : x_i \in \Delta, i = 1, \dots, n\}$. Следовательно, найдется интервал $\tilde{\Delta} \subseteq \Delta$, что на подмножестве $\tilde{\Delta}'' \subset \Delta''$ зависимости $\psi_i = \psi_i(x_n)$, $i = 1, \dots, n$, можно однозначно разрешить относительно x_n . Учитывая, что

$$\psi_i(\mathbf{x}_n) = \sum_{j=1}^n \varphi_i(x_j) = \sum_{j=1}^n \varphi_i(x_{(j)}) := \psi_i(x_{(1)}, \dots, x_{(n)}),$$

получаем, что соотношения $\psi_i = \psi_i(x_{(1)}, \dots, x_{(n)})$ также однозначно разрешимы относительно $x_{(1)}, \dots, x_{(n)}$. Итак, в некоторой области $\tilde{\Delta}^n$ переменных x_1, \dots, x_n (имеющей положительную меру при каждом $\theta \in \Theta$) статистика $(x_{(1)}, \dots, x_{(n)})$ эквивалентна статистике $(\psi_1(x_{(1)}, \dots, x_{(n)}), \dots, \psi_n(x_{(1)}, \dots, x_{(n)}))$. Поэтому достаточное разбиение, которое порождает вариационный ряд в области $\tilde{\Delta}^n$, является минимально достаточным, т. е. $(x_{(1)}, \dots, x_{(n)})$ — минимально достаточная статистика по крайней мере в области $\tilde{\Delta}^n$.

Докажем теперь сформулированное выше утверждение о функциональной независимости статистик $\psi_1(\mathbf{x}_n), \dots, \psi_s(\mathbf{x}_n)$. Предположим противное. Тогда якобиан

$$\begin{aligned} & \partial(\psi_1, \dots, \psi_s)/\partial(x_1, \dots, x_s) = \\ & = \det[\partial\psi_i(\mathbf{x}_n)/\partial x_j, i, j = 1, \dots, s] = \det[\varphi'_i(x_j), i, j = 1, \dots, s] \quad (52) \end{aligned}$$

тождественно равен нулю в $\Delta^n = \{\mathbf{x}_n : x_i \in \Delta, i = 1, \dots, n\}$. Приведем это к противоречию с линейной независимостью системы $1, \varphi_1(\mathbf{x}), \dots, \varphi_s(\mathbf{x})$. Будем действовать по индукции. При $s = 1$ получаем $\varphi'_1(x_1) \equiv 0, \varphi_1(x_1) = \text{const}$, и противоречие получено. Допустим, что $\partial(\psi_1, \dots, \psi_{s-1})/\partial(x_1, \dots, x_{s-1}) \neq 0$. Тогда, разложив определитель в (52) по последнему столбцу, получаем для $\mathbf{x}_n \in \Delta^n$

$$A_{1s}\varphi'_1(x_s) + \dots + A_{ss}\varphi'_s(x_s) \equiv 0, \quad (53)$$

где $A_{js} = A_{js}(x_1, \dots, x_{s-1})$ — алгебраические дополнения соответствующих элементов определителя (52). Так как по предположению индукции

$$A_{ss} = \partial(\psi_1, \dots, \psi_{s-1})/\partial(x_1, \dots, x_{s-1}) \neq 0,$$

то найдутся такие значения x_1^0, \dots, x_{s-1}^0 , что $A_{ss}^0 = A_{ss}(x_1^0, \dots, x_{s-1}^0) \neq 0$. В таком случае из (53) получаем $A_{1s}^0\varphi_1(x_s) + \dots + A_{ss}^0\varphi_s(x_s) = \text{const}, A_{ss}^0 \neq 0$, что приводит к противоречию с линейной независимостью функций

$$1, \varphi_1(\mathbf{x}), \dots, \varphi_s(\mathbf{x}).$$

§ 17. ДОСТАТОЧНОСТЬ И НЕСМЕЩЕННОЕ ОЦЕНИВАНИЕ

1. Полные достаточные статистики.

Рассмотрим статистическую модель $(\mathcal{X}, \mathcal{B}, \{\mathcal{P}\})$. Мы уже встречались с различными примерами числовых функций $\varphi(\mathbf{P})$ неизвестного распределения вероятностей выборки \mathbf{P} , которые требуется оценить по результатам наблюдений. Так, в модели независимых повторных испытаний обычно

представляют интерес математическое ожидание и дисперсия отдельного наблюдения. В параметрических моделях $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ возникает задача оценивания различных числовых функций $\varphi(\theta)$ от неизвестного параметра θ .

Различают оценки *точечные*, когда по выборке x вычисляется значение t некоторой статистики $T(x)$ и это значение t предлагается в качестве приближения к неизвестному $\varphi(\theta)$ ($\varphi(P)$), и оценки *интервальные*, когда по выборке x строится интервал $(T_1(x), T_2(x))$, любая точка t которого рассматривается как возможное значение неизвестного $\varphi(\theta)$ ($\varphi(P)$).

Сделаем замечание о терминологии. Занимаясь задачей точечного оценивания, приходится рассматривать различные статистики $T(x)$ как возможные оценки $\varphi(\theta)$ ($\varphi(P)$). Принято в таком случае, вне зависимости от того, насколько хорошо $T(x)$ приближает неизвестную функцию, называть статистику $T(x)$ *оценкой*.

Определение 1. Оценка $T(x)$ функции $\varphi(\theta)$ ($\varphi(P)$) называется *несмешенной*, если

$$M_\theta T(X) = \varphi(\theta), \quad \theta \in \Theta \quad (M_P T(X) = \varphi(P), \quad P \in \{P\}),$$

где X — случайная выборка.

Несмешенные оценки не всегда существуют, но в случае, когда они имеются, естественной мерой качества оценки служит ее дисперсия $D_\theta T(X), \theta \in \Theta$ ($D_P T(X), P \in \{P\}$).

Будем рассматривать для определенности параметрический случай. Допустим, что $S(x)$ — достаточная статистика в модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$. Так как, по нашим представлениям, $S(x)$ содержит всю информацию о θ , содержащуюся в выборке x , то, решая задачу оценивания $\varphi(\theta)$, по-видимому, целесообразно рассматривать лишь статистики вида $g(S(x))$ — функции достаточной статистики $S(x)$. Представляется, что при этом сужении класса рассматриваемых оценок мы не должны потерять в качестве: если $T(x)$ — оценка с определенным значением какого-то критерия качества, то должна найтись оценка вида $g(S(x))$ с таким же или лучшим, чем у $T(x)$, значением этого критерия. Мы убедимся в справедливости этого утверждения при решении задачи несмешенного оценивания, когда за критерий качества принята величина дисперсии оценки. А пока заметим, что многие оценки, встречавшиеся нам в первых двух главах, являются функциями минимальной достаточной статистики (см. § 15, 16).

Переход от произвольных статистик $T(x)$ к функциям от достаточной $g(S(x))$ облегчает решение задачи оценивания, если статистика $S(x)$ существенно сокращает данные, например, когда размерность $S(x)$ невелика и не зависит от размерности выборки. В этом случае можно надеяться, что условие несмешенности

$$M_\theta g(S(X)) = \varphi(\theta), \quad \theta \in \Theta, \quad (1)$$

рассматриваемое как уравнение относительно неизвестной функции

ции $g(s)$, можно решить, а в множестве его решений отыскать оценку с минимальной дисперсией.

Рассмотрим пример.

(I) Пусть

$$\mathcal{X} = \mathbb{R}_+^n = \{x_n \in \mathbb{R}^n : x_i > 0; i = 1, \dots, n\}, \quad \mathcal{B} = \mathcal{X} \cap \mathbb{R}_+^n,$$

семейство P_θ , $\theta > 0$, задается n -мерной плотностью

$$f_\theta(x_n) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right), \quad x_n \in \mathbb{R}_+^n,$$

так что случайная выборка $X_n = (X_1, \dots, X_n)$ состоит из независимых сл. в. с плотностью распределения $\theta e^{-\theta x_i}$, $x_i > 0$, каждая. Достаточная статистика: $S(x_n) = \sum_{i=1}^n x_i$. Запишем уравнение (1) для $\varphi(\theta) = \theta$:

$$M_\theta g\left(\sum_{i=1}^n X_i\right) = \int_0^\infty g(s) f_{S(X_n)}^\theta(s) ds = \theta, \quad \theta > 0. \quad (2)$$

Так как $S(X_n)$ имеет распределение $G(\theta, n)$ с плотностью (см. пп. 1, 2 § 7)

$$f_{S(X_n)}^\theta = \frac{\theta^n}{\Gamma(n)} s^{n-1} e^{-\theta s}, \quad s > 0, \quad (3)$$

то уравнение (2) принимает вид

$$\int_0^\infty g(s) s^{n-1} e^{-\theta s} ds = \theta^{-n+1} \Gamma(n), \quad \theta > 0. \quad (4)$$

Если $g(s)$ — решение (4), то, изменив его произвольным образом на множестве меры Лебега нуль, мы снова получим решение уравнения (4). Оказывается, что с точностью до такого рода неоднозначности уравнение (4) имеет единственное решение, которое нетрудно найти. Действительно, заменив в (3) n на $n-1$, запишем

$$\int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} s^{n-2} e^{-\theta s} ds = 1, \quad \theta > 0,$$

или

$$\int_0^\infty \frac{1}{s} \frac{\Gamma(n)}{\Gamma(n-1)} s^{n-1} e^{-\theta s} ds = \theta^{-n+1} \Gamma(n), \quad \theta > 0. \quad (5)$$

Сравнивая (5) с (4) и замечая, что $\Gamma(n) = (n-1)!$, находим решение уравнения (4):

$$g(s) = (n-1)/s.$$

Таким образом, несмешенная оценка параметра θ , являющаяся функцией достаточной статистики, имеет вид

$$\hat{\theta}(x_n) = (n-1)/\sum_{i=1}^n x_i. \quad (6)$$

Стоит отметить, что θ^{-1} является средним значением отдельного наблюдения и известной оценкой θ^{-1} является выборочное среднее $\bar{x} = \sum_{i=1}^n x_i/n$, которое есть, кстати, функция достаточной статистики. Однако попытка оценить θ с помощью $(\bar{x})^{-1}$, как видно, приводит к смещенной оценке.

Полагая $h(s) = g(s)s^{n-1}$, $\psi(\theta) = \theta^{-n+1}\Gamma(n)$, перепишем уравнение (4) в виде

$$\int_0^\infty h(s) e^{-\theta s} ds = \psi(\theta), \quad \theta > 0. \quad (7)$$

Проверим, что уравнение (7) имеет не более одного решения $h(s)$ (с точностью до его изменения на множестве меры нуль). Функция $\psi(\theta)$, определенная формулой (7), называется преобразованием Лапласа функции $h(s)$. Существуют различные способы обращать преобразование Лапласа — находить $h(s)$ по $\psi(\theta)$. Мы используем вероятностный подход (см.: Феллер [2], т. 2). Дифференцируя (7) k раз по θ , находим

$$(-1)^k \psi^{(k)}(\theta) = \int_0^\infty h(s) s^k e^{-\theta s} ds. \quad (8)$$

Домножим (8) на $\theta^k/k!$ и просуммируем по $k \leq \theta x$:

$$\sum_{k \leq \theta x} \frac{(-1)^k}{k!} \theta^k \psi^{(k)}(\theta) = \int_0^\infty h(s) \sum_{k \leq \theta x} \frac{(\theta s)^k}{k!} e^{-\theta s} ds. \quad (9)$$

Мы покажем, что при $\theta \rightarrow +\infty$

$$\sum_{k \leq \theta x} \frac{(\theta s)^k}{k!} e^{-\theta s} \rightarrow \begin{cases} 0, & \text{если } s > x, \\ 1, & \text{если } s < x, \end{cases} \quad (10)$$

откуда будет следовать, что при $\theta \rightarrow \infty$

$$\sum_{k \leq \theta x} \frac{(-1)^k}{k!} \theta^k \psi^{(k)}(\theta) \rightarrow \int_0^x h(s) ds = H(x), \quad (11)$$

т. е. первообразная $H(x)$ функции $h(x)$ однозначно восстанавливается по преобразованию Лапласа $\psi(\theta)$. Как известно, функция $h(x)$ восстанавливается по $H(x)$ с точностью до ее значений на множестве меры Лебега нуль. Остается проверить соотношение

(10) Левая часть (10) равна вероятности того, что пуассоновская сл. в. X с параметром θs приняла значение, не превосходящее θx . По неравенству Чебышева

$$\mathcal{P}(|X-\theta s|>a) \leq a^{-2}\theta s,$$

откуда при $x < s$

$$\mathcal{P}(X < \theta x) = \mathcal{P}(X - \theta s < \theta(x-s)) \leq s/(\theta(x-s))^2 \rightarrow 0, \quad \theta \rightarrow \infty.$$

а при $x > s$

$$\begin{aligned} \mathcal{P}(X < \theta x) &= 1 - \mathcal{P}(X > \theta x) = 1 - \mathcal{P}(X - \theta s > 0(x-s)) \geq \\ &\geq 1 - \theta s/(\theta(x-s)^2) \rightarrow 1, \quad \theta \rightarrow \infty. \end{aligned}$$

Единственность решения уравнения (7) можно выразить в следующей форме: если $h(s)$ такова, что

$$\int_0^{\infty} h(s) e^{-\theta s} ds = 0, \quad \theta > 0,$$

то $h(s) \equiv 0$ всюду, за исключением множества меры нуль.

Определение 2. Семейство распределений Q_{θ} , $\theta \in \Theta$, на измеримом пространстве $(\mathcal{Y}, \mathcal{A})$ называется *полным*, если для любой измеримой функции $h(y)$ уравнение

$$M_{\theta}h(Y) = 0, \quad \theta \in \Theta, \tag{12}$$

где Y — случайный элемент в $(\mathcal{Y}, \mathcal{A}, Q_{\theta})$, имеет только нулевое решение $h(y) = 0$ с точностью до изменения $h(y)$ на множестве Q_{θ} -меры нуль при любом $\theta \in \Theta$.

Будем говорить, что уравнение

$$M_{\theta}g(Y) = \varphi(\theta), \quad \theta \in \Theta,$$

имеет *существенно единственное* решение, если для любых двух решений $g_1(y), g_2(y)$ этого уравнения множество

$$\{y : g_1(y) \neq g_2(y)\}$$

имеет Q_{θ} -меру нуль при любом $\theta \in \Theta$.

Если семейство распределений Q_{θ} полно, то решение уравнения несмещено

$$M_{\theta}g(Y) = \varphi(\theta), \quad \theta \in \Theta,$$

существенно единственно. В самом деле, пусть $g_1(y), g_2(y)$ — два таких решения. Тогда $h(y) = g_1(y) - g_2(y)$ является решением уравнения (12), и, следовательно, множество $\{y : h(y) = 0\}$ имеет Q_{θ} -меру нуль при каждом $\theta \in \Theta$. ■

Как мы видели в примере (I), семейство распределений $G(\theta, n)$ полно на (R^1, \mathcal{B}_1) . Семейство распределений P_{θ} в том же при-

мере не является полным на (R^n, \mathcal{B}_n) , $n > 1$, так как, скажем, $M_\theta(X_1 - X_2) = 0$, $\theta \in \Theta$.

Определение 3. Статистика $T(x)$ на $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ со значениями в измеримом пространстве $(\mathcal{Y}, \mathcal{A})$ называется *полной*, если семейство ее распределений вероятностей

$$Q_\theta(A) = P_\theta(x : T(x) \in A), \quad A \in \mathcal{A},$$

является полным.

Если достаточная статистика $T(x)$ полна и уравнение (1) разрешимо, то его решение существенно единствено. Мы увидим, что соответствующая несмещенная оценка имеет минимально возможную дисперсию. Рассмотрим примеры полных достаточных статистик.

(II) Статистика $x_{(n)}$ в модели повторных независимых испытаний с равномерным на $(0, \theta)$, $\theta > 0$, распределением полна (см. пример (II) § 16): из уравнения

$$M_\theta h(X_{(n)}) = \int_0^\theta h(t) n t^{n-1} \theta^{-n} dt = 0, \quad \theta > 0,$$

сразу получаем, что первообразная функции $h(t)t^{n-1}$ равна нулю, а следовательно, $h(t) = 0$ всюду, кроме множества меры Лебега нуль.

(III) Рассмотрим статистику $(nx_{(1)}, \sum_{i=2}^n (x_{(i)} - x_{(1)}))$ в выборке из экспоненциального распределения с плотностью

$$\sigma^{-1} \exp(-\sigma^{-1}(x-\mu)), \quad x > \mu, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

(см. пример (III) § 16). Так как $X = nX_{(1)}$ и $Y = \sum_{i=2}^n (X_{(i)} - X_{(1)})$ независимы и имеют соответственно экспоненциальную и гамма плотности:

$$f_X^{\mu, \sigma}(x) = \sigma^{-1} \exp(-(x-\mu)\sigma^{-1}), \quad x > \mu,$$

$$f_Y(y) = \frac{y^{n-2}}{(n-2)! \sigma^{n-1}} \exp(-y\sigma^{-1}), \quad y > 0,$$

то уравнение (12) преобразуем следующим образом:

$$\begin{aligned} 0 &= \int_{n\mu}^{\infty} \int_0^{\infty} h(x, y) f_X^{\mu, \sigma}(x) f_Y(y) dx dy = \int_{n\mu}^{\infty} f_X^{\mu, \sigma}(x) dx \int_0^{\infty} h(x, y) f_Y(y) dy = \\ &= \int_{n\mu}^{\infty} f_X^{\mu, \sigma}(x) h^\sigma(x) dx = \sigma^{-1} \exp(n\mu\sigma^{-1}) \int_{n\mu}^{\infty} \exp(-x\sigma^{-1}) h^\sigma(x) dx. \end{aligned} \quad (13)$$

Рассматривая (13) как уравнение относительно μ при фиксированном σ , легко видеть, что, за исключением множества меры Лебега нуль, $h^\sigma(x) = 0$, так что для почти всех x

$$\int_0^\infty h(x, y) f_Y^{\mu, \sigma}(y) dy = 0. \quad (14)$$

Используя полноту семейства $G(\sigma^{-1}, n-1)$, получаем $h(x, y) = 0$ для почти всех y при всех x , для которых выполнено (14), что и доказывает полноту статистики $(nx_{(1)}, \sum_{i=2}^n (x_{(i)} - x_{(1)}))$.

(IV) В модели повторной выборки из распределения $N(\mu, \sigma^2)$ полной является статистика $(\sum_{i=1}^n x_i, \sum_{i=1}^n (x_i - \bar{x})^2)$. Проверим сначала, что распределение $N(\mu, \sigma_0^2)$ на (R^1, \mathcal{B}_1) , где $-\infty < \mu < +\infty, \sigma_0^2 > 0$ — фиксировано, является полным. Воспользуемся для этого тем, что двустороннее преобразование Лапласа

$$\Phi(\theta) = \int_{-\infty}^\infty e^{-\theta x} g(x) dx, \quad -\infty < \theta < +\infty, \quad (15)$$

однозначно определяет функцию $g(x)$ для почти всех x .

Уравнение (12) для семейства $N(\mu, \sigma_0^2)$ после элементарных преобразований принимает вид

$$\int_{-\infty}^\infty \exp\left(\mu \frac{x}{\sigma_0^2}\right) h(x) \exp\left(-\frac{x}{2\sigma_0^2}\right) dx = 0, \quad -\infty < \mu < +\infty,$$

которое приводится к (15) заменой переменного $y = x/\sigma_0^2$, так что $h(x) \equiv 0$ почти всюду.

Положим $X = \sum_{i=1}^n X_i, Y = \sum_{i=1}^n (X_i - \bar{X})^2$. Так как X и Y независимы и имеют соответственно распределения $N(n\mu, \sigma^2)$ и $G(1/(2\sigma^2), (n-1)/2)$, то уравнение (12) для пары статистик $(\sum_{i=1}^n x_i, \sum_{i=1}^n (x_i - \bar{x})^2)$ записывается в виде

$$0 = \int_{-\infty}^\infty \int_0^\infty h(x, y) f_X^{\mu, \sigma}(x) f_Y^\sigma(y) dxdy = \int_{-\infty}^\infty f_X^{\mu, \sigma}(x) dx \int_0^\infty h(x, y) f_Y^\sigma(y) dy.$$

Фиксируем σ ; пользуясь полнотой семейства $f_X^{\mu, \sigma}(x)$, $-\infty < \mu < \infty$, находим, что для почти всех x

$$\int_0^\infty h(x, y) f_Y^\sigma(y) dy = 0, \quad y > 0.$$

Теперь, используя полноту семейства $G(1/(2\sigma^2), (n-1)/2)$ (см. пример (1)), получаем, что для почти всех y и почти всех x $h(x, y) = 0$. ■

Имеет место следующая

Теорема 1. Пусть плотность $f_\theta(x)$ принадлежит k -параметрическому экспоненциальному семейству:

$$f_\theta(x) = h(x) \exp \left(\sum_{i=1}^k a_i(\theta) T_i(x) + b(\theta) \right), \quad \theta \in \Theta.$$

Если множество значений функции $a(\theta) = (a_1(\theta), \dots, a_k(\theta))$, $\theta \in \Theta$, содержит k -мерный параллелипипед, то $T(x) = (T_1(x), \dots, \dots, T_k(x))$ — полная достаточная статистика.

Доказательство теоремы можно найти в [11, 12].

2. Наилучшие несмешенные оценки в дискретной модели.

Рассмотрим дискретную статистическую модель $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$, и пусть $T(x)$ — полная достаточная статистика. Допустим, что существует несмешенная оценка $S(x)$ параметрической функции $\varphi(\theta)$:

$$M_\theta S(X) = \varphi(\theta), \quad \theta \in \Theta. \quad (16)$$

Будем искать функцию $g(t)$ такую, что

$$M_\theta g(T(X)) = \varphi(\theta), \quad \theta \in \Theta. \quad (17)$$

Введем разбиение $\{\mathcal{X}_t\}$ множества \mathcal{X} , порождаемое достаточной статистикой $T(x)$:

$$\mathcal{X}_t = \{x : T(x) = t\}. \quad (18)$$

Запишем уравнение (17) в виде

$$\sum_x g(T(x)) P_\theta(x) \equiv \sum_t P_\theta(\mathcal{X}_t) g(t) = \varphi(\theta), \quad \theta \in \Theta, \quad (19)$$

где суммирование при каждом $\theta \in \Theta$ ведется по тем t , для которых $P_\theta(\mathcal{X}_t) > 0$. Из (16) имеем, что для всех $\theta \in \Theta$

$$\varphi(\theta) = \sum_x S(x) P_\theta(x) \equiv \sum_t P_\theta(\mathcal{X}_t) \sum_{x \in \mathcal{X}_t} S(x) P_\theta(x) / P_\theta(\mathcal{X}_t), \quad (20)$$

с теми же ограничениями на область суммирования по t , что и в (19). Сравнивая (19) и (20) и учитывая, что ввиду полноты $T(x)$ решение уравнения (19) существенно единственno, обнаруживаем, что

$$g(t) = \sum_{x \in \mathcal{X}_t} S(x) P_\theta(x) / P_\theta(\mathcal{X}_t). \quad (21)$$

где семейство условных мер

$$P^t(x) = P_\theta(x)/P_\theta(\mathcal{X}_t), \quad x \in \mathcal{X}_t, \quad (22)$$

не зависит от θ ввиду достаточности статистики $T(x)$. (Напомним, что формула (22) при данном t определяет меру $P^t(x)$, если существует хотя бы одно $\theta \in \Theta$, для которого $P^t(\mathcal{X}_t) > 0$ и для всех таких θ значение $P^t(x)$, $x \in \mathcal{X}_t$, одно и то же; см. § 15.) Перешием формулу (21), задающую функцию $g(t)$, с учетом (22) в обозначениях, не содержащих θ :

$$g(t) = \sum_{x \in \mathcal{X}_t} S(x) P^t(x). \quad (23)$$

Итак, получено следующее утверждение.

Лемма 1. Если $T(x)$ — полная достаточная статистика в дискретной модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ и существует статистика $S(x)$, несмешенно оценивающая параметрическую функцию $\varphi(\theta)$, то статистика $g(T(x))$, где $g(t)$ задается формулой (23), также несмешенно оценивает $\varphi(\theta)$. Статистика $g(T(x))$ существенно единственна. ■

Формула (23) определяет $g(t)$ вне зависимости от того, является ли достаточная статистика $T(x)$ полной или нет — оценка $g(T(x))$ остается все равно несмешенной:

$$\begin{aligned} M_\theta g(T(x)) &= \sum_t g(t) P_\theta(\mathcal{X}_t) = \\ &= \sum_t P_\theta(\mathcal{X}_t) \sum_{x \in \mathcal{X}_t} S(x) P^t(x) = \sum_x S(x) P_\theta(x) = \varphi(\theta). \end{aligned}$$

Таким образом, справедлива

Лемма 2. Если $T(x)$ — достаточная статистика в дискретной модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ и существует несмешенная оценка $S(x)$ параметрической функции $\varphi(\theta)$, то статистика $g(T(x))$, где $g(t)$ определена формулой (23), также несмешенно оценивает $\varphi(\theta)$.

Лемма 3. В условиях леммы 2 имеет место соотношение

$$D_\theta S(X) = D_\theta g(T(X)) + M_\theta(S(X) - g(T(X)))^2. \quad (24)$$

Доказательство. Для дисперсии $S(X)$ имеет место следующее представление:

$$\begin{aligned} D_\theta S(X) &= \sum_x (S(x) - \varphi(\theta))^2 P_\theta(x) = \\ &= \sum_t P_\theta(\mathcal{X}_t) \sum_{x \in \mathcal{X}_t} (S(x) - \varphi(\theta))^2 P_\theta(x)/P_\theta(\mathcal{X}_t), \end{aligned} \quad (25)$$

где \mathcal{X}_t определено в (18). Преобразуем внутреннюю сумму в (25), используя обозначение (22):

$$\begin{aligned} & \sum_{x \in \mathcal{X}_t} (S(x) - g(t) + g(t) - \varphi(\theta)^2 P^t(x)) = \\ & = \sum_{x \in \mathcal{X}_t} (S(x) - g(t))^2 P^t(x) + 2(g(t) - \varphi(\theta)) \times \\ & \times \sum_{x \in \mathcal{X}_t} (S(x) - g(t)) P^t(x) + (g(t) - \varphi(\theta))^2. \end{aligned} \quad (26)$$

Ввиду (23) среднее слагаемое в (26) равно нулю. Подставляя (26) в (25), получим

$$\begin{aligned} D_\theta S(X) &= \sum_t P_\theta(\mathcal{X}_t) \sum_{x \in \mathcal{X}_t} (S(x) - g(t))^2 P_\theta(x)/P_\theta(\mathcal{X}_t) + \\ &+ \sum_t P_\theta(\mathcal{X}_t) (g(t) - \varphi(\theta))^2 = \\ &= \sum_x (S(x) - g(t))^2 P_\theta(\mathcal{X}_t) + D_\theta g(T(X)), \end{aligned}$$

откуда и вытекает равенство (24).

Из формулы (24) следует, что $D_\theta S(X) \geq D_\theta g(T(X))$, причем равенство выполняется, лишь когда $S(X)$ и $g(T(X))$ совпадают с P_θ -вероятностью единица. Если достаточная статистика $T(x)$ полна, то имеется существенно единственная функция $g(t)$, такая, что $g(T(x))$ несмешенно оценивает $\varphi(\theta)$. Формула (23) при этом определяет существенно одну и ту же функцию $g(t)$, вне зависимости от того, какая там взята несмешенная оценка $S(x)$. В таком случае, используя лемму 2, получаем, что $g(T(x))$ имеет наименьшую возможную дисперсию в классе всех несмешенных оценок. Подведем итог всему сказанному.

Теорема 2*). *Если $T(x)$ — достаточная статистика в дискретной модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ и $S(x)$ — какая-либо несмешенная оценка параметрической функции $\varphi(\theta)$, то статистика $g(T(x))$, где $g(t)$ определена формулой (23), также несмешенно оценивает $\varphi(\theta)$ и*

$$D_\theta g(T(X)) \leq D_\theta S(X),$$

причем равенство соблюдается лишь при условии совпадения $g(T(X))$ и $S(X)$ с P_θ -вероятностью единица. Если к тому же достаточная статистика $T(x)$ полна, то $g(T(x))$ — существенно

* Теоремы 2 и 3 являются вариантами теоремы Блекуэла — Колмогорова — Рао, а также включают утверждение теоремы Лемана — Шеффе о полных достаточных статистиках.

единственная статистика, являющаяся функцией $T(x)$, которая несмешенно оценивает $\varphi(\theta)$, и она имеет наименьшую возможную дисперсию в классе всех несмешенных оценок. Функция $g(t)$ может быть найдена как решение уравнения

$$M_\theta g(T(X)) = \sum g(t) P_\theta(\mathcal{X}_t) = \varphi(\theta), \quad \theta \in \Theta,$$

либо получена из формулы (23), где $S(x)$ — произвольная несмешенная оценка $\varphi(\theta)$.

Рассмотрим примеры.

(V) Пусть

$$\mathcal{X} = \{x_n = (x_1, \dots, x_n) : x_i = 0 \text{ или } 1, i = 1, \dots, n\},$$

$$P_\theta(x_n) = \theta^{T(x_n)} (1-\theta)^{n-T(x_n)},$$

$$T(x_n) = \sum_{i=1}^n x_i, \quad 0 < \theta < 1.$$

Статистика $T(x_n)$ — полная достаточная. Действительно, приравнивая математическое ожидание к нулю:

$$\begin{aligned} M_\theta h(T(X_n)) &= \sum_{t=0}^n h(t) C_n^t \theta^t (1-\theta)^{n-t} = \\ &= \sum_{t=0}^n h(t) C_n^t \left(\frac{\theta}{1-\theta}\right)^t (1-\theta)^n = 0, \end{aligned} \quad (27)$$

получаем, что многочлен от переменного $x=\theta/(1-\theta)$, $x>0$, тождественно равен нулю. Следовательно, его коэффициенты равны нулю, т. е. $h(t)=0$, $t=0, 1, \dots, n$.

Возьмем статистику $S_1(x_n) = x_1$. Поскольку

$$P_\theta(X_1=1) = \theta, \quad P_\theta(X_1=0) = 1-\theta, \quad M_\theta X_1 = \theta,$$

то оценка $S_1(x_n)$ — несмешенная. Далее, при $T(x_n)=t$ имеем по формуле (22)

$$P^t(x_n) = \theta^t (1-\theta)^{n-t} / (C_n^t \theta^t (1-\theta)^{n-t}) = (C_n^t)^{-1}$$

Вычисляя по формуле (23), находим

$$g_1(t) = \sum_{x_n: T(x_n)=t} x_1 (C_n^t)^{-1} = (C_n^t)^{-1} \sum_{x_1+\dots+x_n=t-1} 1 = (C_n^t)^{-1} C_{n-1}^{t-1} = t/n.$$

Таким образом, наилучшая несмешенная оценка θ есть

$$g_1(T(x_n)) = \sum_{i=1}^n x_i/n = \bar{x}.$$

Возьмем статистику $S_2(x_n) = (x_2 - x_1)^2/2$. Имеем

$$M_\theta(X_2 - X_1)^2/2 = M_\theta(X_2^2 - 2X_1X_2 + X_1^2)/2 = \theta(1-\theta),$$

т. е. статистика $S_2(x_n)$ несмешенно оценивает $\theta(1-\theta)$ — дисперсию отдельного наблюдения. По формуле (23) находим

$$\begin{aligned} g_2(t) &= \frac{1}{2} \sum_{x_n: T(x_n)=t} (\dot{x}_2 - x_1)^2 (C_n^t)^{-1} = \\ &= \frac{1}{2} \left(2 \sum_{x_2 + \dots + x_n = t-1} 1 (C_n^t)^{-1} \right) = (C_n^t)^{-1} C_{n-2}^{t-1} = \frac{t(n-t)}{n(n-1)}, \end{aligned}$$

откуда получаем наилучшую несмешенную оценку для $\theta(1-\theta)$:

$$g_2(T(x_n)) = \frac{1}{n(n-1)} \sum_{i=1}^n x_i \left(n - \sum_{i=1}^n x_i \right) = \frac{n}{n-1} \bar{x}(1-\bar{x}).$$

Она не совпадает с результатом подстановки в формулу $\theta(1-\theta)$ наилучшей несмешенной оценки $\hat{\theta} = \bar{x}$, но она совпадает с несмешенной выборочной оценкой дисперсии s^2 :

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

где мы учли, что $x_i^2 = x_i$.

(VI) Пусть

$$\mathcal{X} = \{x_n = (x_1, \dots, x_n) : x_i = 1, 2, \dots, N; i = 1, \dots, n\},$$

$$P_N(x_n) = N^{-n}, \quad x_n \in \mathcal{X}, \quad N = 1, 2, \dots$$

Проверим, что достаточная статистика $T(x_n) = x_{(n)}$ является полной. Имеем (см. пример (III) § 15)

$$P_N\{x_n : T(x_n) = t\} = N^{-n} (t^n - (t-1)^n), \quad t = 1, 2, \dots, N,$$

откуда уравнение (12) приобретает вид

$$0 = M_N h(T(x_n)) = \sum_{t=1}^N h(t) N^{-n} (t^n - (t-1)^n), \quad N = 1, 2, \dots,$$

или

$$\sum_{t=1}^N h(t) (t^n - (t-1)^n) = 0, \quad N = 1, 2, \dots,$$

и потому $h(t) = 0$ при $t = 1, 2, \dots$. Наилучшая несмешенная оценка N равняется $g(x_{(n)})$, где $g(t)$ находится из условия

$$\sum_{t=1}^N g(t) N^{-n} (t^n - (t-1)^n) = N, \quad N = 1, 2, \dots \quad (28)$$

Из (28) получаем

$$g(N) (N^n - (N-1)^n) = N^{n+1} - (N-1)^{n+1},$$

или

$$g(t) = t(1 - (1 - 1/t)^{n+1}) / 1 - (1 - 1/t)^n.$$

При больших t $g(t) \approx t(n+1)/n$.

Если $T(x)$ — достаточная статистика для дискретной модели $(\mathcal{X}, \mathcal{B}, \{\mathbf{P}_\theta, \theta \in \Theta\})$, то для любой статистики $S(x)$ можно рассмотреть статистику $g(T(x))$, определенную формулой (23):

$$g(t) = \sum_{x \in \mathcal{X}_t} S(x) \mathbf{P}^t(x), \quad \mathcal{X}_t = \{x : T(x) = t\}, \quad (29)$$

$$\mathbf{P}^t(x) = \mathbf{P}_\theta(x) / \mathbf{P}_\theta(\mathcal{X}_t), \quad \theta \in \Theta.$$

Статистика $g(T(x))$ принимает постоянное значение на элементах разбиения $\{\mathcal{X}_t\}$, и это значение равно среднему значению статистики $S(x)$ по условной мере $\mathbf{P}^t(x)$. В теории вероятностей так определенную $g(T(x))$ называют *условным математическим ожиданием* $S(x)$ при условии $T(x)$, или, иначе, при условии разбиения $\{\mathcal{X}_t\}$, и обозначают

$$g(T(x)) = M_\theta(S(x) | T(x)) = M_\theta(S(x) | \{\mathcal{X}_t\}). \quad (30)$$

Индекс θ в обозначениях (30) имеет чисто формальный смысл: объект, определяемый формулой (30), от θ не зависит. Однако если статистика $T(x)$ не является достаточной, то условное математическое ожидание (30), вообще говоря, зависит от θ и уже не представляет такого интереса для математической статистики.

3. Наилучшие несмешенные оценки в непрерывной модели.

В случае непрерывной модели $(R^n, \mathcal{B}_n, \{\mathbf{P}_\theta, \theta \in \Theta\})$ мы рассматриваем лишь такие достаточные статистики $T_r(x_n) = (T_1(x_n), \dots, T_r(x_n))$, что отображение $t_i = T_i(x_n)$, $i = 1, \dots, r$, можно дополнить некоторыми статистиками $Y_{n-r}(x_n) = (Y_1(x_n), \dots, Y_{n-r}(x_n))$ до взаимно-однозначного преобразования всего пространства R^n (кроме, быть может, множества \mathbf{P}_θ -меры нуль), и это преобразование регулярно, так что можно делать обычную замену переменных в интегралах (см. § 16).

Рассматривая различные несмешанные оценки $S(x_n)$ параметрической функции $\varphi(\theta)$, мы будем предполагать, что отображение $(t_r, s) = (T_r(x_n), S(x_n))$ также можно дополнить до регулярного преобразования всего R^n . Из определения достаточности $T_r(x_n)$ вытекает, что условная плотность

$$f_{S(x_n)|T_r(x_n)}^{\theta}(s|t_r) = f(s|t_r) \quad (31)$$

не зависит от θ .

Теорема 3. Пусть $T_r(x_n)$ — достаточная статистика в непрерывной модели $(R^n, \mathcal{B}_n, \{P_\theta, \theta \in \Theta\})$, $S(x_n)$ — несмешенная оценка $\varphi(\theta)$ и пара $T_r(x_n), S(x_n)$ удовлетворяет сформулированным выше условиям регулярности. Положим

$$g(t_r) = \int_{-\infty}^{\infty} sf(s|t_r) ds, \quad (32)$$

где условная плотность $f(s|t_r)$ определена формулой (31). Тогда статистика $g(T_r(X_n))$ несмешенно оценивает $\varphi(\theta)$:

$$D_\theta g(T_r(X_n)) \leq D_\theta(S(X_n)), \quad (33)$$

и равенство достигается лишь при условии совпадения $g(T_r(X_n))$ и $S(X_n)$ с P_θ -вероятностью единицы. Если к тому же статистика $T_r(x_n)$ полна, то $g(T_r(x_n))$ — существенно единственная оценка $\varphi(\theta)$, являющаяся функцией $T_r(x_n)$, и она имеет наименьшую дисперсию в классе несмешенных оценок. Функция $g(t_r)$ может быть найдена по формуле (32) либо как существенно единственное решение уравнения

$$\varphi(\theta) = M_\theta g(T_r(X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(t_r) f_{T_r(x_n)}^{\theta}(t_r) dt_1 \dots dt_r, \quad \theta \in \Theta. \quad (34)$$

Доказательство проводится аналогично дискретному случаю с заменой сумм на интегралы. Имеем с учетом (31)

$$\begin{aligned} M_\theta g(T_r(X_n)) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{T_r(x_n)}^{\theta}(t_r) dt_1 \dots dt_r \int_{-\infty}^{\infty} sf(s|t_r) ds = \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} sf_{S(x_n), T_r(x_n)}^{\theta}(s, t_r) ds dt_1 \dots dt_r = \\ &= \int_{-\infty}^{\infty} s ds \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{S(x_n), T_r(x_n)}^{\theta}(s, t_r) dt_1 \dots dt_r = \\ &= \int_{-\infty}^{\infty} sf_{S(x_n)}^{\theta}(s) ds = \varphi(\theta). \end{aligned}$$

Для дисперсии оценки $S(\mathbf{x}_n)$ получаем

$$\begin{aligned} \mathbf{D}_{\theta} S(\mathbf{X}_n) &= \int_{-\infty}^{\infty} (s - \varphi(\theta))^2 f_{S(\mathbf{X}_n)}(s) ds = \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (s - \varphi(\theta))^2 f_{S(\mathbf{X}_n), T_r(\mathbf{X}_n)}^{\theta}(s, t_r) ds dt_1 \dots dt_r = \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{T_r(\mathbf{X}_n)} dt_1 \dots dt_r \int_{-\infty}^{\infty} (s - \varphi(\theta))^2 f(s | t_r) ds. \end{aligned} \quad (35)$$

Преобразуем интеграл по s в (35) к виду

$$\begin{aligned} &\int_{-\infty}^{\infty} [(s - g(t_r))^2 + 2(s - g(t_r))(g(t_r) - \varphi(\theta)) + \\ &+ (g(t_r) - \varphi(\theta))^2] f(s | t_r) ds. \end{aligned} \quad (36)$$

Ввиду (32) интеграл от среднего слагаемого в (36) обращается в нуль. Подставляя (36) в (35), получаем

$$\mathbf{D}_{\theta}(S(\mathbf{X}_n)) = \mathbf{D}_{\theta}g(T_r(\mathbf{X}_n)) + M_{\theta}(S(\mathbf{X}_n) - g(T_r(\mathbf{X}_n)))^2, \quad (37)$$

откуда следует неравенство (33). Наконец, если статистика $T_r(\mathbf{X}_n)$ — полная, то решение уравнения (34) существенно единственно, так что при любой несмещенной оценке $S(\mathbf{x}_n)$ формулы (31), (32) определяют существенно одну и ту же функцию $g(t_r)$, и тем самым дисперсия оценки $g(T_r(\mathbf{X}_n))$ минимальная.

З а м е ч а н и е 1. Пусть $T_r(\mathbf{x}_n)$ — достаточная статистика, а $Y(\mathbf{x}_n)$ — такая статистика, что отображение $(t_r, y) = (T_r(\mathbf{x}_n), Y(\mathbf{x}_n))$ можно дополнить до регулярного преобразования всего R^n . Пусть несмещенная оценка $h(Y(\mathbf{x}_n))$ не обязательно имеет плотность, например принимает лишь дискретные значения. Просматривая доказательство теоремы 3, легко убедиться в том, что оно сохраняется при замене $S(\mathbf{x}_n)$ на $h(Y(\mathbf{x}_n))$, если положить

$$g(t_r) = \int_{-\infty}^{\infty} h(y) f_{Y(\mathbf{x}_n) | T_r(\mathbf{x}_n)}(y | t_r) dy. \quad (38)$$

Таким образом, теорема 3 справедлива и для всех статистик указанного вида $h(Y(\mathbf{X}_n))$.

З а м е ч а н и е 2. Отметим, что, как и в дискретном случае, статистика $g(T_r(\mathbf{x}_n))$, определенная в (32) или (38), называется *условным математическим ожиданием* соответственно $S(\mathbf{x}_n)$ или $h(Y(\mathbf{x}_n))$ при условии статистики $T_r(\mathbf{x}_n)$. Обозначение:

$$M_{\theta}(S(\mathbf{x}_n) | T_r(\mathbf{x}_n)) \text{ или } M_{\theta}(h(Y(\mathbf{x}_n)) | T_r(\mathbf{x}_n)). \blacksquare$$

Используя теорему 3, получаем, что в примере (I) оценка (6) — наилучшая. В примере (III) повторной выборки из рас-

пределения с плотностью $\sigma^{-1} \exp(\sigma^{-1}(x-\mu))$, $x > \mu$, наилучшие линейные несмешанные оценки σ и μ , полученные в § 11, являются наилучшими в классе всех несмешанных оценок. То же верно для повторной выборки из распределения $N(\mu, \sigma^2)$: выборочные среднее и дисперсия — наилучшие несмешанные оценки. ■

Рассмотрим дополнительные примеры.

(VII) Нормальная линейная модель в общей форме:

$$f_{Y|\Theta, \sigma}^{\theta, \sigma}(y) = (\sqrt{2\pi}\sigma)^{-n} \exp\left(-\frac{1}{2\sigma^2} \|y - \theta X'\|^2\right) \quad (39)$$

с матрицей X полного ранга приводится с помощью невырожденного линейного преобразования переменного y и невырожденного линейного преобразования вектора параметров θ к канонической форме (см. п. 5 § 10, пример (V) § 16):

$$f_U^{\psi, \sigma}(u) = (\sqrt{2\pi}\sigma)^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^r (u_i - \psi_i)^2 - \frac{1}{2\sigma^2} \sum_{i=r+1}^n u_i^2\right). \quad (40)$$

Оценки наименьших квадратов $\hat{\theta}_i$, $i=1, \dots, r$, $\hat{\sigma}^2$ образуют $(r+1)$ -мерную достаточную статистику в модели (39) (см. пример (V) § 16). Ее полнота вытекает из полноты соответствующих оценок в преобразованной модели (40), которые имеют вид

$$\hat{\psi}_i(u) = u_i, \quad i = 1, \dots, r, \quad \hat{\sigma}^2(u) = \frac{1}{n-r} \sum_{i=r+1}^n u_i^2. \quad (41)$$

Для $r=1$ полнота статистики (41) доказана в примере (IV); для $r>1$ доказательство полностью аналогично случаю $r=1$. Таким образом, найденные в § 9 линейные несмешанные оценки $\hat{\theta}$, $\hat{\sigma}^2$ для нормальной модели (39) являются оценками с минимально возможной дисперсией в классе всех несмешанных оценок, а не только оценок, линейных по наблюдениям.

(VIII) Рассмотрим, как в примере (I), повторную выборку из распределения с плотностью $\theta \exp(-\theta x)$, $x > 0$. Статистика $S(x_n) = n\bar{x}$ как было показано, — полная достаточная. Параметрическая функция

$$\varphi_t(\theta) = \exp(-\theta t),$$

где $t > 0$ — любое фиксированное число, допускает несмешанную оценку. Действительно, положим

$$I_{\{x_1 > t\}} = \begin{cases} 1, & \text{если } x_1 > t, \\ 0, & \text{если } x_1 \leq t, \end{cases}$$

тогда

$$M_\theta I_{\{x_1 > t\}} = P(X_1 > t) = \varphi_t(0), \quad -\infty < 0 < \infty.$$

Следовательно, наилучшая несмешенная оценка $\varphi_t(\theta)$ может быть получена в виде (см. замечания 1, 2 к теореме 3)

$$M_\theta(I_{\{x_1 > t\}} | S(x_n)). \quad (42)$$

Как показано в лемме 2 § 5, условная плотность

$$f_{X_1|S(x_n)}(x_1 | s) \quad (43)$$

совпадает с плотностью $U_{(1)}$, где U_1, U_2, \dots, U_{n-1} независимы и равномерно распределены на $(0, s)$. Но

$$\begin{aligned} \mathcal{P}(U_{(1)} > u) &= \mathcal{P}(U_1 > u, \dots, U_{n-1} > u) = \\ &= ((s-u)/s)^{n-1}, \quad 0 < u < s, \end{aligned} \quad (44)$$

так что плотность (43) равна

$$(n-1)s^{-n+1}(s-x_1)^{n-2}, \quad 0 < x_1 < s. \quad (45)$$

По формуле (38) получаем

$$\begin{aligned} g(s) &= \int_0^\infty I_{\{x_1 > t\}} f_{X_1|S(x_n)}(x_1 | s) dx_1 = \\ &= \int_0^s I_{\{x_1 > t\}} (n-1)s^{-n+1}(s-x_1)^{n-2} dx_1. \end{aligned} \quad (46)$$

Если $s \leq t$, то $I_{\{x_1 > t\}} = 0$ при $0 < x_1 < s$ и, следовательно, $g(s) = 0$. При $0 < t < s$ имеем $I_{\{x_1 > t\}} = 1$ при $t < x_1 < s$ и $I_{\{x_1 > t\}} = 0$ при $0 < x_1 < t$, так что интеграл (46) равняется в этом случае интегралу от плотности (45) по области $t < x_1 < s$, т. е. равен вероятности (44) при $u = t$. Таким образом,

$$g(s) = \begin{cases} 0 & \text{при } s \leq t, \\ \left(1 - \frac{t}{s}\right)^{n-1} & \text{при } s > t. \end{cases}$$

Наилучшая несмешенная оценка $\varphi_t(\theta)$ равна

$$g(S(x_n)) = \begin{cases} 0 & \text{при } S(x_n) \leq t, \\ (1 - t/S(x_n))^{n-1} & \text{при } S(x_n) > t. \end{cases}$$

Наилучшая несмешенная оценка неизвестной функции распределения $1 - \varphi_t(\theta) = 1 - \exp(-\theta t)$ в точке t равна $1 - g(S(x_n))$.

§ 18. ИНФОРМАЦИЯ В СТАТИСТИКЕ

1. Байесовский подход в статистике.

Неопределенность в исходе опыта при его описании статистической моделью $(\mathcal{X}, \mathcal{A}, \{P_\theta, \theta \in \Theta\})$ складывается из двух

составляющих: неопределенности вероятностного характера, обусловленной случайностью механизма, реализующего x по распределению вероятностей P_θ , и неопределенности, связанной с незнанием истинного значения θ_0 параметра θ , в соответствии с которым реализуется x . Байесовский метод в статистике рассматривает вторую неопределенность также с позиций теории вероятностей.

Рассмотрим для примера задачу выборочного контроля производства, когда параметр θ описывает некоторые свойства партии изделий и может рассматриваться как случайный, меняющийся от партии к партии. Опыт прошлых наблюдений может быть обобщен в форме задания распределения вероятностей Q параметра θ . Обозначим через Θ соответствующую случайную величину. В результате мы имеем пару случайных элементов X и Θ . Пусть $f_\theta(x)$, $f_\Theta(\theta)$ — плотности распределений P_θ и Q соответственно. Тогда пара (X, Θ) имеет плотность $f_\theta(x) f_\Theta(\theta)$. Зависимость случайной выборки X от сл. в. Θ характеризуется условной плотностью

$$f_{x|\Theta}(x|\theta) = f_\theta(x). \quad (1)$$

Можно представить себе, что вначале реализуется значение сл. в. Θ с плотностью $f_\Theta(\theta)$, затем в соответствии с условной плотностью (1) образуется случайная выборка X . Значение θ , принятое сл. в. Θ , неизвестно, значение x случайной выборки X наблюдается. По результатам наблюдения x надо сделать выводы о значении θ .

Назовем байесовской статистической моделью структуру, состоящую из статистической модели $(\mathcal{X}, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, измеримого пространства (Θ, \mathcal{T}) и вероятностной меры Q на нем, называемой априорным распределением.

Выбор априорного распределения является принципиальным моментом в байесовском методе. Примеры, подобные приведенному выше, когда распределение Q устанавливается на основе частотных соображений по результатам прошлых наблюдений за явлением, весьма немногочисленны. В теории разработаны различные подходы к этой проблеме. Суть проблемы заключается в том, каким способом можно перейти от априорных сведений или представлений неформального характера к заданию вероятностного распределения на (Θ, \mathcal{T}) . Мы не будем здесь останавливаться на этом вопросе, отметим лишь, что, несмотря на различие методологий частотного и байесовского методов в статистике, в конечном счете они приводят к сопоставимым статистическим процедурам.

Если априорное распределение Q каким-то образом выбрано, то задачу оценки θ можно отнести к области чистой теории вероятностей. Случайный элемент Θ измеримого пространства (Θ, \mathcal{T}) с распределением Q полностью определен своим распределением вероятностей. Информация о Θ , содержащаяся в слу-

чайной выборке X , полностью заключается в условной плотности (рассматриваемой на множестве $\{x : f_X(x) > 0\}$):

$$\begin{aligned} f_{\Theta|x}(\theta|x) &= f_{X|\Theta}(x, \theta)/f_X(x) = \\ &= f_\theta(x) f_{\Theta}(\theta)/f_X(x) = f_{X|\Theta}(x|\theta) f_\theta(\theta)/f_X(x), \end{aligned} \quad (2)$$

где мы предполагаем, что меры P_θ и Q задаются плотностями и

$$f_X(x) = \int \dots \int f_\theta(x) f_\Theta(\theta) d\theta_1 \dots d\theta_k. \quad (3)$$

Условную плотность (2) параметра θ называют *апостериорной*.

Формула (2) представляет собой непрерывный аналог так называемой теоремы Байеса. Легко понять, как трансформируются выражения (2), (3), если одна из величин X , Θ или обе дискретны. Чтобы не вводить лишних обозначений, примем соглашение, что в случае дискретных X , Θ под плотностями $f_\Theta(\theta)$, $f_\theta(x)$ и т. д. будут пониматься соответствующие распределения вероятностей: $P_\theta(x)$, $Q(\theta)$ и т. д. В таком случае формула (2) сохраняет свой прежний вид, если X или Θ дискретны, в формуле (3) интегралы заменяются на суммы, если Θ дискретно. ■

Апостериорная плотность (2) служит источником всех дальнейших статистических выводов о θ . Скажем, в качестве точечной оценки берут обычно среднее значение по этой плотности:

$$\hat{\theta}_i = \int \theta_i f_{\Theta_i|x}(\theta_i|x) d\theta_i, \quad i = 1, \dots, k, \quad (4)$$

где

$$f_{\Theta|x}(\theta_i|x) = \int \dots \int f_{\Theta|x}(\theta|x) d\theta_1 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_k$$

— апостериорная плотность компоненты θ_i вектора $\Theta = (\theta_1, \dots, \theta_k)$. ■

Рассмотрим примеры байесовских моделей.

(1) Пусть $(X_1, \dots, X_n) = X_n$ испытания Бернулли с неизвестной вероятностью $0 < \theta < 1$, априорное распределение есть бета-распределение с плотностью $B(a, b)^{-1}\theta^{a-1}(1-\theta)^{b-1}$, $0 < \theta < 1$ (см. п. 5 § 7). Тогда апостериорная плотность (2) пропорциональна

$$\theta^{S(x_n)} (1-\theta)^{n-S(x_n)} \theta^{a-1} (1-\theta)^{b-1}.$$

$$S(x_n) = x_1 + \dots + x_n, \quad x_i = 0 \text{ или } 1$$

и является бета-плотностью с параметрами $a + S(x_n)$, $b + n - S(x_n)$, т. е. параметры априорного распределения (a, b) с учетом результатов испытания изменились на

$$(a + S(x_n), b + n - S(x_n)).$$

Априорное среднее параметра θ равно (см. § 7)

$$\int_0^1 \theta B(a, b)^{-1} \theta^{a-1} (1-\theta)^{b-1} d\theta = B(a+1, b)/B(a, b) = a/(a+b),$$

а апостериорное — $(a+S(x_n))/(a+b+n)$. Если число наблюдений n велико, то байесовская оценка $(a+S(x_n))/(a+b+n)$, каковы бы ни были $a>0$ и $b>0$, приближается $S(x_n)/n$ — частотой положительного исхода в испытаниях Бернулли.

(II) Пусть X_1, \dots, X_n — независимые сл. в. с экспоненциальной плотностью $G(\theta, 1)$, априорное распределение параметра θ — $G(\lambda, p)$ с некоторыми заданными λ, p . Тогда апостериорная плотность пропорциональна

$$\theta^n e^{-S(x_n)\theta} \theta^{p-1} e^{-\lambda\theta}, \quad \theta > 0, \quad S(x_n) = x_1 + \dots + x_n,$$

т. е. представляет собой плотность распределения $G(\lambda+S(x_n), p+n)$. Найдем байесовскую оценку для функции надежности

$$P_\theta(X_1 > t) = e^{-\theta t}.$$

Априорное среднее от $\exp(-\theta t)$ равно

$$\begin{aligned} \int_0^\infty e^{-\theta t} \frac{\lambda^p}{\Gamma(p)} \theta^{p-1} e^{-\lambda\theta} d\theta &= \left(\frac{\lambda}{\lambda+t} \right)^p \int_0^\infty \frac{(\lambda+t)^p}{\Gamma(p)} \theta^{p-1} e^{-(\lambda+t)\theta} d\theta = \\ &= (\lambda/(\lambda+t))^p. \end{aligned}$$

Априорное среднее от $\exp(-\Theta t)$ равно

$$((\lambda+S(x_n))/(\lambda+t+S(x_n)))^{p+n} = (1+t/(\lambda+S(x_n)))^{-(n+p)} \quad (5)$$

Если n велико, то байесовская оценка (5) близка к наилучшей несмешенной оценке (см. пример (VIII) § 17):

$$g(S(x_n)) = \begin{cases} (1-t/S(x_n))^{n-1} & \text{при } S(x_n) > t, \\ 0 & \text{при } S(x_n) \leq t; \end{cases}$$

однако байесовская оценка дает нетривиальный результат при всех t , в том числе и $t > S(x_n)$. ■■

Лемма 1. Пусть $T(x)$ — достаточная статистика для статистической модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$. Тогда для любого априорного распределения Q на (Θ, \mathcal{T}) имеет место равенство (на подмножестве X , имеющем P_θ -меру единица при любом θ)

$$f_{\Theta|X}(\theta|x) = f_{\Theta|T(x)}(\theta|t), \quad t = T(x), \quad (6)$$

где участвующие в (6) плотности в дискретном случае понимаются как распределения вероятностей.

Доказательство проведем для непрерывной модели. По критерию факторизации (см. § 16)

$$f_{\theta}(x) \equiv f_x^{\theta}(x) = g(T(x); \theta) h(x) \quad (7)$$

на подмножестве X , имеющем P_{θ} -меру единица, $\theta \in \Theta$, так что апостериорная плотность пропорциональна (см. (2))

$$g(T(x); \theta) f_{\theta}(\theta). \quad (8)$$

Покажем, что плотность $f_{\Theta|T(x)}(\theta|t)$, $t = T(x)$, также пропорциональна правой части (8). По формуле (24) § 16

$$f_{\Theta|T(x)}^{\theta}(t) = g(t; \theta) h_2(t),$$

откуда по формуле (2), примененной к паре $T(X), \Theta$, находим $f_{\Theta|T(x)}(\theta|t)$ пропорциональна

$$g(t; \theta) f_{\theta}(\theta). \quad (9)$$

Сравнивая (8) и (9), получаем утверждение леммы. ■

Из леммы 1 вытекает, что при байесовском подходе понятие достаточности играет ту же роль, что и в классической статистике. Именно, как следует из формулы (6), по достаточной статистике восстанавливается апостериорное распределение θ при условии выборки x , а это и все, что нужно знать при байесовском методе.

2. Информация по Шенону.

Достаточная статистика выступает носителем всей полезной информации о неизвестном параметре θ как при классическом, так и байесовском подходе. Но само понятие информации остается пока предметом наших представлений, не имеющим точного математического выражения. В байесовском подходе и выборка и параметр являются случайными элементами, и потому к ним применимо хорошо известное и оказавшееся чрезвычайно полезным понятие информации (и энтропии), введенное Шенном.

Пусть, для начала, X и Θ — дискретные случайные элементы, определенные на некотором общем вероятностном пространстве $(\Omega, \mathcal{A}, \mathcal{P})$ и отображающие его в измеримые пространства $(\mathcal{X}, \mathcal{B})$ и (Θ, \mathcal{T}) соответственно;

$$Q(\theta) = \mathcal{P}(\Theta = \theta), \quad P_{\theta}(x) = \mathcal{P}(X = x | \Theta = \theta).$$

До проведения каких бы то ни было наблюдений все наши сведения о значении, которое может в результате опыта принять Θ , заключены в распределении вероятностей $Q(\theta)$, которое в статистике называют *априорным*. Если стало известно значение x ,

принятое случайной выборкой X , то вся полнота знаний о θ воплощается в условном распределении:

$$\mathcal{P}(\Theta = \theta | X = x) = \mathcal{P}(\Theta = \theta, X = x) / \mathcal{P}(X = x),$$

называемом в статистике *апостериорным*. Таким образом, дополнительные сведения, содержащиеся в результате опыта, изменили неопределенность наших представлений о возможном значении θ : до опыта она выражалась априорным распределением вероятностей, после опыта — апостериорным. Представляется естественным определить какую-либо меру расхождения между этими распределениями и с ее помощью измерять информацию, получающую в результате опыта.

Первое, что может прийти в голову, это рассмотреть разность

$$\mathcal{P}(\Theta = \theta) - \mathcal{P}(\Theta = \theta | X = x)$$

в качестве такой меры. Однако мы уже имели возможность убедиться в § 15, 16, что отношение вероятностей или логарифм этого отношения (т. е. разность логарифмов вероятностей) оказались полезными при изучении информативных свойств выборки. Также и здесь оказывается полезной логарифмическая мера расхождения:

$$I_{\theta, x}(\theta; x) = \log(\mathcal{P}(\Theta = \theta | X = x) / \mathcal{P}(\Theta = \theta)), \quad (10)$$

называемая *информацией*, содержащейся в событии $A = \{X = x\}$ относительно события $B = \{\Theta = \theta\}$. Поскольку

$$\frac{\mathcal{P}(B|A)}{\mathcal{P}(B)} = \frac{\mathcal{P}(AB)}{\mathcal{P}(A)\mathcal{P}(B)} = \frac{\mathcal{P}(A|B)}{\mathcal{P}(A)},$$

то информация, содержащаяся в B относительно A , равна информации, содержащейся в A относительно B , и потому меру (10) называют *взаимной информацией* между A и B . (Конечно, все делается в предположении $\mathcal{P}(A), \mathcal{P}(B) > 0$.) Перепишем (10) в симметричной форме:

$$I_{\theta, x}(\theta, x) = \log \frac{\mathcal{P}(\Theta = \theta, X = x)}{\mathcal{P}(\Theta = \theta)\mathcal{P}(X = x)}. \quad (10a)$$

Подчеркнем, что равенство информаций не означает равнозначности выводов, которые можно сделать об A из знания B и, наоборот, о B из знания A . Пусть, скажем, $A \subseteq B$, тогда $\mathcal{P}(B|A) = 1$ и информация в A относительно B , равная

$$\log(1/\mathcal{P}(B)) = \log(1/\mathcal{P}(\Theta = \theta)) = I_{\theta}(\theta). \quad (11)$$

полностью определяет событие B . Та же информация (11) недостаточна для определения A по B , если только A и B не совпадают с вероятностью единица. Очевидно, дело заключается в том, что собственная неопределенность событий A и B , вообще говоря, различна. Примем выражение (11) за собственную меру не-

определенности события B , или, напр., *свойственную информацию*, содержащуюся в событии B о самом себе. Ее можно интерпретировать как информацию, необходимую для полного разрешения неопределенности относительно события B . С учетом сказанного величину

$$\log(1/\mathcal{P}(\Theta = \theta | X = x)) \equiv I_{\Theta|X}(\theta | x) \quad (12)$$

можно назвать *условной собственной информацией*, содержащейся в событии $\{\Theta = \theta\}$ при условии события $\{X = x\}$, и интерпретировать как информацию, необходимую для определения события $\{\Theta = \theta\}$, после того как стало известно, что $X = x$. Поэтому разность

$$I_\Theta(\theta) - I_{\Theta|X}(\theta | x) = I_{\Theta,X}(\theta, x) \quad (13)$$

может рассматриваться как информация, содержащаяся в событии $\{X = x\}$ относительно события $\{\Theta = \theta\}$ и, как видно, совпадает с взаимной информацией (10).

Формулы (10)–(13) определяют понятие информации для отдельных событий $\{\Theta = \theta\}$, $\{X = x\}$, тогда как желательно определить информацию о Θ в X , как случайных элементах. Это достигается усреднением введенных количеств информации по всем значениям участвующих там случайных величин.

Энтропия Θ определяется как среднее значение собственной информации и задается формулой

$$H(\Theta) = \sum_{\theta} \mathcal{P}(\Theta = \theta) I_\Theta(\theta) = - \sum_{\theta} \mathcal{P}(\Theta = \theta) \log \mathcal{P}(\Theta = \theta). \quad (14)$$

Средняя условная энтропия Θ при условии X определяется равенством

$$\begin{aligned} H(\Theta | X) &= \sum_{\theta,x} \mathcal{P}(\Theta = \theta, X = x) I_{\Theta|X}(\theta | x) = \\ &= - \sum_{\theta,x} \mathcal{P}(\Theta = \theta, X = x) \log \mathcal{P}(\Theta = \theta | X = x). \end{aligned} \quad (15)$$

Средняя взаимная энтропия между Θ и X получается усреднением формулы (13) и равна

$$I(\Theta, X) = H(\Theta) - H(\Theta | X). \quad (16)$$

Определенная формулой (16) *информация по Шеннону* обладает рядом свойств, которые было бы естественно потребовать от любой формализации понятия информации. Перечислим эти свойства.

1. Информация всегда неотрицательна: $I(\Theta, X) \geq 0$.
2. Если статистика $T(x)$ подобная, то $I(\Theta, T(X)) = 0$.

3. Информация, содержащаяся в статистике, не превосходит информации, содержащейся во всей выборке:

$$I(\Theta, T(X)) \leq I(\Theta, X).$$

4. Если $T(x)$ — достаточная статистика для

$$(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\}), \text{ то } I(\Theta, T(X)) = I(\Theta, X).$$

Проверим свойства 2 и 4. В случае подобной статистики $T(X)$ получаем

$$I_{\Theta, T(X)}(\Theta, x) \log \frac{\mathcal{P}(T(X)=x | \Theta=\theta)}{\mathcal{P}(T(X)=x)} = \log \frac{\mathcal{P}(T(X)=x)}{\mathcal{P}(T(X)=x)} = 0,$$

так что равна нулю и средняя взаимная информация $I(\Theta, T(X))$.

Пусть $T(X)$ — достаточная статистика. Рассмотрим

$$I(\Theta, X) = \sum_{\theta, x} \mathcal{P}(\Theta=\theta, X=x) \log \frac{\mathcal{P}(X=x | \Theta=\theta)}{\mathcal{P}(X=x)} \quad (17)$$

По формуле полной вероятности

$$\mathcal{P}(X=x) = \sum_{\tau} \mathcal{P}(X=x | \Theta=\tau) \mathcal{P}(\Theta=\tau). \quad (18)$$

Используя достаточность $T(X)$, запишем факторизацию (11) из § 15:

$$\mathcal{P}(X=x | \Theta=\theta) = g(T(X); \theta) h(x). \quad (19)$$

Логарифмический множитель в (17) после подстановки (18), (19) приобретает вид

$$\log(g(t; \theta) / \sum_{\tau} g(t; \tau) \mathcal{P}(\Theta=\tau)), \quad t = T(X). \quad (20)$$

Учитывая, что (см. (13) § 15)

$$\mathcal{P}(T(X)=t | \Theta=\theta) = g(t; \theta) \sum_{x: T(x)=t} h(x), \quad (21)$$

и домножая числитель и знаменатель дроби в (20) на второй множитель в (21), получаем для выражения (20) значение

$$\begin{aligned} & \log(\mathcal{P}(T(X)=t | \Theta=\theta) / \sum_{\tau} \mathcal{P}(T(X)=t | \Theta=\tau) \mathcal{P}(\Theta=\tau)) = \\ & = \log(\mathcal{P}(T(X)=t | \Theta=\theta) / \mathcal{P}(T(X)=t)) = I_{\Theta, T(X)}(\theta, t), \end{aligned} \quad (22)$$

где $t = T(x)$. Подставляя (22) на место логарифмического множителя в (17) и снова воспользовавшись формулой полной вероятности, находим

$$\begin{aligned}
I(\Theta, X) &= \sum_{\theta} P(\Theta = \theta) \sum_x P(X = x | \Theta = \theta) I_{\Theta, T(X)}(\theta, T(x)) = \\
&= \sum_{\theta} P(\Theta = \theta) \sum_t I_{\Theta, T(X)}(\theta, t) \sum_{x: T(X)=t} P(X = x | \Theta = \theta) = \\
&= \sum_{\theta, t} P(\Theta = \theta) P(T(X) = t | \Theta = \theta) I_{\Theta, T(X)}(\theta, t) = I(\Theta, T(X)). \quad \blacksquare
\end{aligned}$$

Для непрерывных моделей и непрерывного априорного распределения информационные характеристики вводятся путем замены в формулах (10) — (16) дискретных распределений вероятностей на соответствующие плотности, а сумм — на интегралы. Запишем, для примера, выражения для взаимной и средней взаимной информации между Θ и X :

$$I_{\Theta, X}(\theta, x) = \log(I_{\Theta, X}(\theta, x)/f_{\Theta}(\theta) f_X(x)), \quad (23)$$

$$I(\Theta, X) = \int \dots \int I_{\Theta, X}(\theta, x) T_{\Theta, X}(\theta, x) d\theta_1 \dots d\theta_k dx_1 \dots dx_n. \quad (24)$$

Аналогично дискретному случаю доказываются сформулированные выше информационные свойства подобных и достаточных статистик.

Рассмотрим один пример.

(III) Допустим вначале, что параметрическое множество $\Theta = \{\theta_1, \theta_2, \dots, \theta_r\}$ конечно и выборка также дискретна. Запишем взаимную информацию:

$$I_{\Theta, X}(\theta_i, x) = \log(P(\Theta = \theta_i | X = x)/P(\Theta = \theta_i)). \quad (25)$$

Задачу различия, какое именно значение приняла сл. в. Θ , будем решать следующим образом: при данном наблюдении x в качестве предполагаемого значения неизвестного параметра выберем θ_j такое, что выборка содержит о нем наибольшую информацию:

$$I_{\Theta, X}(\theta_j, x) \geq I_{\Theta, X}(\theta_i, x), \quad i = 1, \dots, r. \quad (26)$$

Иначе говоря, выбирается такое значение θ_j , что отношение ее апостериорной вероятности к априорной наибольшее:

$$\mathcal{P}(\Theta = \theta_j | X = x) / \mathcal{P}(\Theta = \theta_i) \geq \mathcal{P}(\Theta = \theta_i | X = x) / \mathcal{P}(\Theta = \theta_i), \quad i = 1, \dots, r. \quad (27)$$

Перепишем (27) в виде

$$\mathcal{P}(X = x | \Theta = \theta_j) / \mathcal{P}(X = x) \geq \mathcal{P}(X = x | \Theta = \theta_i) / \mathcal{P}(X = x), \quad i = 1, \dots, r,$$

или

$$\mathcal{P}(X = x | \Theta = \theta_j) \geq \mathcal{P}(X = x | \Theta = \theta_i), \quad i = 1, \dots, r. \quad (28)$$

Из (28) видно, что предложенное правило различия на самом деле не зависит от априорного распределения. Если вообще отка-

ваться от байесовского метода и записать неравенство (28) для классического подхода:

$$P_{\theta_i}(x) \geq P_{\theta_j}(x), \quad i = 1, \dots, r, \quad (29)$$

то правило выбора (29) неизвестного значения параметра выглядит так: в качестве оценки выбирается такое θ_i , при котором вероятность наблюденной выборки x максимальна. Этот метод называется методом максимума правдоподобия, и о нем будет идти речь в следующей главе.

Отметим, что в непрерывном случае также можно рассмотреть правило (26), заменив дискретные распределения на плотности. Аналогично (29) имеем оценку $\theta^* = \theta^*(x)$ максимального правдоподобия:

$$f_{\theta^*}(x) \geq f_{\theta}(x), \quad \theta \in \Theta. \quad (30)$$

3. Информация по Кульбаку.

В байесовской модели информация по Шеннону служила мерой расхождения между апостериорным и априорным распределениями параметра. При классическом подходе также играют важную роль меры расхождения, но уже между распределениями $P_{\theta}, \theta \in \Theta$. Фактически любая форма статистического вывода о неизвестном параметре θ модели $(\mathcal{X}, \mathcal{B}, \{P_{\theta}, \theta \in \Theta\})$ представляет собой некоторый способ различия мер P_{θ} по наблюденному значению выборки. Если $P_{\theta_1} = P_{\theta_2}$ для некоторых $\theta_1 \neq \theta_2$, то понятно, что различить по выборке значения θ_1 и θ_2 невозможно. С другой стороны, если носители мер P_{θ_1} и P_{θ_2} не пересекаются, то задача различия θ_1 и θ_2 становится тривиальной. Эти крайние случаи не интересны для теории, но наводят на мысль, что возможности различия θ_1 и θ_2 зависят от того, насколько расходятся между собой распределения P_{θ_1} и P_{θ_2} .

Предположим для определенности, что статистическая модель непрерывна и меры P_{θ} задаются плотностями $f_{\theta}(x)$. Информацией по Кульбаку в точке x для различия в пользу θ_1 против θ_2 называется величина

$$I(\theta_1 : \theta_2, x) = \log(f_{\theta_1}(x)/f_{\theta_2}(x)). \quad (31)$$

Выражение (31) равно разности логарифмов плотностей $f_{\theta_1}(x)$ и $f_{\theta_2}(x)$ и как мера расхождения представляется довольно естественной. Возвратимся на время к байесовскому подходу и рассмотрим взаимную информацию по Шеннону между x и θ_i :

$$I_{\theta, x}(\theta_i, x) = \log(f_{\theta_i}(x)/f_X(x)), \quad i = 1, 2. \quad (32)$$

Составляя разность выражений (32), получаем

$$I_{\Theta_1, X}(\theta_1, x) - I_{\Theta_2, X}(\theta_2, x) = \log(f_{\theta_1}(x)/f_{\theta_2}(x)) = I(\theta_1 : \theta_2, x). \quad (33)$$

Итак, различающая информация в пользу f_{θ_1} против f_{θ_2} в точке x по Кульбаку равна разности информаций Шеннона о θ_1 и θ_2 , содержащихся в x , вне зависимости от априорного распределения (при условии, что априорные плотности или априорные вероятности в точках θ_1 и θ_2 положительны).

Определим среднюю различающую информацию в пользу f_{θ_1} против f_{θ_2} относительно меры P_{θ_1} как среднее значение (31) по этой мере:

$$\begin{aligned} I(\theta_1 : \theta_2) &= \int \dots \int I_x(\theta_1 : \theta_2; x) f_{\theta_1}(x) dx_1 \dots dx_n = \\ &= \int \dots \int f_{\theta_1}(x) \log(f_{\theta_1}(x)/f_{\theta_2}(x)) dx_1 \dots dx_n, \end{aligned} \quad (34)$$

где мы предполагаем, что носитель меры P_{θ_1} содержится в носителе P_{θ_2} :

$$\{x : f_{\theta_1}(x) > 0\} \subseteq \{x : f_{\theta_2}(x) > 0\},$$

и интегрирование в (34) распространяется на все x , для которых $f_{\theta_1}(x) > 0$. Иногда удобно вместо $I(\theta_1 : \theta_2)$ писать $I_X(\theta_1 : \theta_2)$, чтобы подчеркнуть, что речь идет о различающей информации по всей выборке. С другой стороны, если $T = T(x)$ — некоторая статистика, то аналогично (34) определяется различающая информация по статистике:

$$I_T(\theta_1 : \theta_2) = \int \dots \int f_T^{\theta_1}(t) \log(f_T^{\theta_1}(t)/f_T^{\theta_2}(t)) dt, dt = dt_1 \dots dt_k. \quad (35)$$

Рассмотрим пример.

(IV) Пусть $f_1(x_1, x_2)$ есть двумерная нормальная плотность $N_2(0, Q)$, где (см. пример п. 1 § 12)

$$Q = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

а $f_2(x_1, x_2) = \sigma_1^{-1}\varphi(x_1/\sigma_1)\sigma_2^{-1}\varphi(x_2/\sigma_2)$, где $\varphi(x)$ — плотность $N(0, 1)$. Тогда

$$\begin{aligned} \log(f_1(X_1, X_2)/f_2(X_1, X_2)) &= 2^{-1} \log(1 - \rho^2) - \sigma_1^{-2} 2^{-1} (1 - \rho^2)^{-1} X_1^2 + \\ &+ \rho(1 - \rho^2)^{-1} \sigma_1^{-1} \sigma_2^{-1} X_1 X_2 - \sigma_2^2 2^{-1} (1 - \rho^2)^{-1} X_2^2 - 2^{-1} \sigma_1^{-2} X_1^2 + 2^{-1} \sigma_2^2 X_2^2. \end{aligned}$$

Взяв математическое ожидание M_1 по мере с плотностью $f_1(x_1, x_2)$ от обеих частей равенства и учитывая, что

$$M_1 X_1^2 = \sigma_1^2, \quad M_1 X_2^2 = \sigma_2^2, \quad M_1 X_1 X_2 = \rho\sigma_1\sigma_2,$$

получаем

$$I(1:2) = 2^{-1} \log(1-\rho^2) - 2^{-1} (1-\rho^2)^{-1} + \rho^2 (1-\rho^2)^{-1} - \\ - 2^{-1} (1-\rho^2)^{-1} + 2^{-1} + 2^{-1} = 2^{-1} \log(1-\rho^2).$$

Таким образом, $I(1:2)$ является функцией только коэффициента корреляции ρ . ■

Перечислим некоторые свойства информации (34).

1. Информация всегда неотрицательна: $I(\theta_1 : \theta_2) \geq 0$, причем равенство имеет место тогда и только тогда, когда множество $\{x : f_{\theta_1}(x) \neq f_{\theta_2}(x)\}$ имеет P_{θ_i} -меру нуль, $i=1, 2$.

2. Если $T(x)$ — некоторая статистика, то

$$I_{T(X)}(\theta_1 : \theta_2) \leq I_X(\theta_1 : \theta_2)$$

с равенством тогда и только тогда, когда $T(X)$ — достаточная статистика.

3. Пусть $X = (X^{(1)}, X^{(2)})$ и $X^{(1)}, X^{(2)}$ — независимы при каждом из распределений P_{θ_1} и P_{θ_2} . Тогда

$$I_{X^{(1)}, X^{(2)}}(\theta_1 : \theta_2) = I_{X^{(1)}}(\theta_1 : \theta_2) + I_{X^{(2)}}(\theta_1 : \theta_2).$$

Коротко: для независимых наблюдений информация складывается.

Проверим последнее свойство, предполагая, что меры P_{θ_i} , $i=1, 2$, определяются плотностями. Имеем

$$I_{X^{(1)}, X^{(2)}}(\theta_1 : \theta_2) = \int \int f_{\theta_1}(u, v) \log(f_{\theta_1}(u, v)/f_{\theta_2}(u, v)) du dv, \quad (36)$$

где $du = du_1 \dots du_k$, $dv = dv_1 \dots dv_{n-k}$, а интегралы в (36) соответственно k -мерный по u и $(n-k)$ -мерный по v . Ввиду независимости

$$f_{\theta_i}(u, v) \equiv f_{X^{(1)}, X^{(2)}}^{\theta_i}(u, v) = f_{X^{(1)}}^{\theta_i}(u) f_{X^{(2)}}^{\theta_i}(v), \quad i=1, 2. \quad (37)$$

Подставляя (37) в (36) и записывая

$$\log \frac{f_{\theta_1}(u, v)}{f_{\theta_2}(u, v)} = \log \frac{f_{X^{(1)}}^{\theta_1}(u)}{f_{X^{(1)}}^{\theta_2}(u)} + \log \frac{f_{X^{(2)}}^{\theta_1}(v)}{f_{X^{(2)}}^{\theta_2}(v)}, \quad (38)$$

разложим интеграл (36) в сумму интеграла

$$J_1 = \int \int f_{X^{(1)}}^{\theta_1}(u) f_{X^{(2)}}^{\theta_2}(v) \log(f_{X^{(1)}}^{\theta_1}(u)/f_{X^{(1)}}^{\theta_2}(u)) du dv \quad (39)$$

и интеграла J_2 , получаемого из J_1 заменой логарифмического члена на второе слагаемое из (38). Заменяя интеграл (39) повторным и интегрируя сначала по dv , найдем, что (39) равен

$$\int f_{X^{(1)}}^{\theta_1}(u) \log(f_{X^{(1)}}^{\theta_1}(u)/f_{X^{(1)}}^{\theta_2}(u)) du = I_{X^{(1)}}(\theta_1 : \theta_2).$$

Аналогично $J_2 = I_{X^{(2)}}(\theta_1 : \theta_2)$, что и дает требуемый результат.

Проверим равенство информаций

$$I_{T(x)}(\theta_1 : \theta_2) = I_x(\theta_1 : \theta_2),$$

когда статистика $T = T(x)$ — достаточная. В соответствии с предположениями § 16 найдется статистика $Y = Y(x)$ такая, что преобразование $(t, y) = (T(x), Y(x))$ регулярно и с вероятностью единица взаимно-однозначно, так что

$$\begin{aligned} I_x(\theta_1 : \theta_2) &= I_{T(x), Y(x)}(\theta_1 : \theta_2) = \\ &= \iint f_{T,Y}^{\theta_1}(t, y) \log(f_{T,Y}^{\theta_1}(t, y) / f_{T,Y}^{\theta_2}(t, y)) dt dy, \end{aligned} \quad (40)$$

где $dt = dt_1 \dots dt_r$, $dy = dy_1 \dots dy_{n-r}$ и интегралы имеют соответствующую размерность. Подставляя в (40) факторизацию (23) § 16, приведем (40) к виду

$$\iint g(t; \theta_1) h_1(t, y) |J|^{-1} \log(g(t; \theta_1) / g(t; \theta_2)) dt dy. \quad (41)$$

Заменим интеграл (41) повторным и проинтегрируем сначала по y . Учитывая формулу (24) § 16

$$f_{T,x}^{\theta_1}(t) = g(t; \theta_1) \int h_1(t, y) |J|^{-1} dy,$$

получаем для (41) выражение

$$\int f_T^{\theta_1}(t) \log(f_T^{\theta_1}(t) / f_T^{\theta_2}(t)) dt = I_{T(x)}(\theta_1 : \theta_2),$$

что и требовалось установить.

4. Информация по Фишеру.

Предположим вначале, что θ — скалярный параметр, и рассмотрим информацию по Кульбаку:

$$I(\theta : \theta + \Delta\theta) = \iint f(x; \theta) \log(f(x; \theta) / f(x; \theta + \Delta\theta)) dx, \quad (42)$$

где $dx = dx_1 \dots dx_n$, а интеграл понимается как n -мерный. Будем предполагать, что носитель у всех мер P_θ , $\theta \in \Theta$, общий, а от плотностей $f(x; \theta)$ потребуем выполнение свойств регулярности, которые мы не будем перечислять, но которые обеспечивают обоснованность проводимых ниже выкладок. Разложим логарифмический член в (42) по формуле Тейлора:

$$\log f(x; \theta) - \log f(x; \theta + \Delta\theta) = - \frac{\partial \log f(x; \theta)}{\partial \theta} \Delta\theta - \frac{1}{2} \frac{\partial^2 \log f(x; \theta'_x)}{\partial \theta^2} \Delta\theta^2, \quad (43)$$

где θ'_x лежит между θ и $\theta + \Delta\theta$. Поскольку (мы считаем логарифм натуральным)

$$\int f(x; \theta) \frac{\partial \log f(x; \theta)}{\partial \theta} dx = \int f(x; \theta) \frac{f'_\theta(x; \theta)}{f(x; \theta)} dx = \\ = f'_\theta(x; \theta) dx = \frac{\partial}{\partial \theta} \int f(x; \theta) dx = \frac{\partial}{\partial \theta} 1 = 0, \quad (44)$$

то, подставляя (43) в (42), имеем

$$I(\theta : \theta + \Delta\theta) = -\frac{1}{2} \int f(x; \theta) \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \Delta\theta^2 dx. \quad (45)$$

Положим

$$I(\theta) = -\int f(x; \theta) \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} dx = -M_\theta \frac{\partial^2 \log f(X; \theta)}{\partial \theta^2}.$$

Тогда (45) можно переписать в виде

$$I(\theta : \theta + \Delta\theta) = \frac{1}{2} I(\theta) \Delta\theta^2 + o(\Delta\theta^2). \quad (46)$$

Дифференцируя, получаем

$$\frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} = \frac{1}{f(x; \theta)} \frac{\partial^2 f(x; \theta)}{\partial \theta^2} - \frac{1}{f^2(x; \theta)} \left(\frac{\partial f(x; \theta)}{\partial \theta} \right)^2. \quad (47)$$

Но (ср. (44))

$$M_\theta \frac{1}{f(X; \theta)} \frac{\partial^2 f(X; \theta)}{\partial \theta^2} = \int \frac{\partial^2 f(x; \theta)}{\partial \theta^2} dx = 0,$$

поэтому, взяв математическое ожидание от обеих частей равенства (47), где полагаем $x = X$, имеем

$$I(\theta) = M_\theta \frac{1}{f^2(X; \theta)} \left(\frac{\partial f(X; \theta)}{\partial \theta} \right)^2 = M_\theta \left(\frac{\partial \log f(X; \theta)}{\partial \theta} \right)^2. \quad (48)$$

Формула (48) определяет информацию по Фишеру в выборке X о параметре θ (мы полагаем при этом, что Θ — интервал в R^1). При условиях регулярности, для которых оправданы приведенные выше выкладки, информация по Фишеру определяется по формуле (46) информацию по Кульбаку для бесконечно близких значений параметра θ и $\theta + \Delta\theta$. Используя равенство (см. (44))

$$M_\theta (\partial \log f(X; \theta) / \partial \theta) = 0, \quad (49)$$

запишем для $I(\theta)$ эквивалентную формулу

$$I(\theta) = D_\theta (\partial \log f(X; \theta) / \partial \theta). \quad (50)$$

Работая с информацией по Фишеру, мы будем всегда предполагать, что выполнено равенство (49), что обеспечивается требованием совпадения носителей мер P_θ и возможностью дифференцирования под знаком интеграла (44).

Если статистическая модель дискретна, то формула (50) заменяется на следующую:

$$I(\theta) = \mathbf{D}_{\theta} (\partial \log P_{\theta}(\mathbf{X}) / \partial \theta), \quad (51)$$

при этом требуется выполнение равенства, аналогичного (49).

Перейдем к векторному параметру $\theta = (\theta_1, \dots, \theta_k)$, предполагая, что Θ — область в R^k . Информационной матрицей Фишера назовем матрицу ковариаций $I(\theta)$ случайного вектора

$$\text{grad}_{\theta} \log f(\mathbf{X}; \theta) = (\partial \log f(\mathbf{X}; \theta) / \partial \theta_1, \dots, \partial \log f(\mathbf{X}; \theta) / \partial \theta_k), \quad (52)$$

где \mathbf{X} имеет распределение P_{θ} . При этом предполагается, что вектор (52) имеет нулевое математическое ожидание:

$$\mathbf{M}_{\theta} \text{grad}_{\theta} \log f(\mathbf{X}; \theta) = 0, \quad \theta \in \Theta, \quad (53)$$

и, таким образом,

$$I(\theta) = \mathbf{M}_{\theta} [(\text{grad}_{\theta} \log f(\mathbf{X}; \theta))' \text{grad}_{\theta} \log f(\mathbf{X}; \theta)]. \quad (54)$$

Аналогично определяется информационная матрица Фишера для статистики $T(\mathbf{x})$:

$$I_{T(\mathbf{x})}(\theta) = \mathbf{M} [(\text{grad}_{\theta} \log f_{T(\mathbf{x})}(T(\mathbf{x}); \theta))' \text{grad}_{\theta} \log f_{T(\mathbf{x})}(T(\mathbf{x}); \theta)], \quad (55)$$

где $f_{T(\mathbf{x})}(t; \theta)$ — плотность распределения $T(\mathbf{x})$ в предположении, что выборка имеет распределение P_{θ} . В частности, $I(\theta) = I_{\mathbf{x}}(\theta)$.

Если выборка или статистика дискретны, то плотности в (53), (54), (55) заменяются на распределения вероятностей.

Перечислим некоторые свойства информации по Фишеру.

1. Если $T(\mathbf{x})$ — подобная статистика, то $I_{T(\mathbf{x})}(\theta) = 0, \theta \in \Theta$.
2. Для любой статистики $T(\mathbf{x})$ имеет место неравенство^{*}

$$I_{T(\mathbf{x})}(\theta) \leq I_{\mathbf{x}}(\theta), \quad \theta \in \Theta.$$

3. Если $T(\mathbf{x})$ — достаточная статистика, то

$$I_{T(\mathbf{x})}(\theta) = I_{\mathbf{x}}(\theta), \quad \theta \in \Theta.$$

4. Пусть $\mathbf{X} = (\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$, $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ — независимы при каждом $P_{\theta}, \theta \in \Theta$, и для них определена информация по Фишеру. Тогда

$$I_{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}}(\theta) = I_{\mathbf{X}^{(1)}}(\theta) + J_{\mathbf{X}^{(2)}}(\theta),$$

Проверим свойство 4. Ввиду независимости $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$

$$\log f_{\mathbf{X}^{(1)} \mathbf{X}^{(2)}}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}; \theta) = \log f_{\mathbf{X}^{(1)}}(\mathbf{x}^{(1)}; \theta) + \log f_{\mathbf{X}^{(2)}}(\mathbf{x}^{(2)}; \theta)$$

* Напомним, что знак неравенства $A \leq B$ между матрицами означает, что разность $B - A$ неотрицательно определена.

и

$$\begin{aligned} \text{grad}_{\theta} \log f_{X^{(1)}, X^{(2)}}(X^{(1)}, X^{(2)}; \theta) &= \\ = \text{grad}_{\theta} \log f_{X^{(1)}}(X^{(1)}; \theta) + \text{grad}_{\theta} \log f_{X^{(2)}}(X^{(2)}; \theta), \end{aligned} \quad (56)$$

причем векторы в правой части (56) независимы, имеют нулевое математическое ожидание и у них существует матрица ковариаций. В таком случае математическое ожидание вектора в левой части (56) равно нулю, а его матрица ковариаций равна сумме матриц ковариаций векторов в правой части, что и требовалось установить.

Свойство 1 очевидно. Свойства 2, 3 ради сокращения выкладок докажем для скалярного параметра и в дискретном варианте. Положим

$$Q_{\theta}(t) = \sum_{x:T(x)=t} P_{\theta}(x).$$

Рассмотрим

$$\begin{aligned} M_{\theta} \left(\frac{\partial \log P_{\theta}(X)}{\partial \theta} - \frac{\partial \log Q_{\theta}(T(X))}{\partial \theta} \right)^2 &= \\ = I_X(\theta) + I_{T(X)}(\theta) - 2M_{\theta} \left(\frac{\partial \log P_{\theta}(X)}{\partial \theta} \frac{\partial \log Q_{\theta}(T(X))}{\partial \theta} \right). \end{aligned} \quad (57)$$

Распишем математическое ожидание в правой части (57):

$$\begin{aligned} \sum_x \frac{P'_{\theta}(x)}{P_{\theta}(x)} \frac{Q'_{\theta}(T(x))}{Q_{\theta}(T(x))} P_{\theta}(x) &= \sum_t \frac{Q'_{\theta}(t)}{Q_{\theta}(t)} \sum_{x:T(x)=t} P'_{\theta}(x) = \\ = \sum_t \frac{Q'_{\theta}(t)}{Q_{\theta}(t)} \frac{\partial}{\partial \theta} \sum_{x:T(x)=t} P_{\theta}(x) &= \sum_t \frac{(Q'_{\theta}(t))^2}{Q_{\theta}(t)} = \sum_t \left(\frac{Q'_{\theta}(t)}{Q_{\theta}(t)} \right)^2 Q_{\theta}(t) = I_{T(X)}(\theta). \end{aligned}$$

Таким образом, левая часть (57) равна

$$I_X(\theta) - I_{T(X)}(\theta), \quad (58)$$

а так как правая часть (57) всегда неотрицательна, то свойство 2 установлено.

Допустим, что $T(x)$ — достаточная статистика. По теореме факторизации (см. п. 5 § 15) равенство

$$P_{\theta}(x) = g(T(x); \theta) h(x), \quad (59)$$

выполняется при каждом θ на некотором подмножестве \mathcal{X}_{θ} , имеющем P_{θ} -меру единица. Изменим при каждом θ меру P_{θ} на подмножестве $\mathcal{X} \setminus \mathcal{X}_{\theta}$. P_{θ} -меры нуль так, чтобы равенство (59) выполнялось при всех $x \in \mathcal{X}$. Формально статистическая модель при этом изменилась, но с содержательной точки зрения

такое видоизменение несущественно. Суммируя (59), находим

$$Q_\theta(t) = g(t; \theta) \sum_{x: T(x)=t} h(x). \quad (60)$$

Подставляя (59), (60) в левую часть (57), обращаем ее в нуль, что с учетом (58) доказывает свойство 3.

Отметим, что свойства 1, 3 допускают обращение, если наложить некоторые ограничения, исключающие «патологические» случаи. Предположим, для краткости, что Θ — скаляр, а Θ — некоторый интервал, и потребуем, например, чтобы производная

$$\partial \log f_{T(X)}(t; \theta) / \partial \theta \quad (61)$$

была непрерывна по θ . Допустим, что $f_{T(X)}(\theta) = 0$, $\theta \in \Theta$, т. е.

$$D_\theta (\partial \log f_{T(X)}(T(X); \theta) / \partial \theta) = 0, \quad \theta \in \Theta,$$

и, следовательно, (см. (53))

$$\partial \log f_{T(X)}(t; \theta) / \partial \theta = 0, \quad \theta \in \Theta, \quad (62)$$

на подмножестве $\mathcal{T}_\theta \subseteq \mathcal{T}$ переменного t Q_θ -вероятности единица, где Q_θ — распределение статистики $T(X)$. В таком случае для подмножества $\Theta' \subseteq \Theta$, состоящего из всех рациональных точек из Θ , соотношение (62) выполнено на некотором подмножестве $\mathcal{T}' \subseteq \mathcal{T}$, Q_θ -мера которого равна единице для всех $\theta \in \Theta'$. По непрерывности (62) выполнено при всех $\theta \in \Theta$ на \mathcal{T}' :

$$\frac{\partial}{\partial \theta} \log f_{T(X)}(t; \theta) = 0, \quad t \in \mathcal{T}', \quad \theta \in \Theta,$$

и, следовательно,

$$f_{T(X)}(t; \theta) = f(t), \quad t \in \mathcal{T}', \quad \theta \in \Theta, \quad (63)$$

где $f(t)$ — плотность, не зависящая от θ . Но из этого вытекает, что при всех $\theta \in \Theta$

$$\int_{\mathcal{T}'} f_{T(X)}(t, \theta) dt = \int_{\mathcal{T}'} f(t) dt = 1,$$

т. е. $Q_\theta(\mathcal{T}') = 1$ при всех $\theta \in \Theta$, что доказывает подобие статистики $T(X)$.

Чтобы доказать обращение свойства 3, воспользуемся следующим утверждением.

Лемма 2 [13, с. 62]. Пусть $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ — непрерывная или дискретная модель. Тогда существует счетное подмножество $\Theta' \subseteq \Theta$, что для любого подмножества $X' \subseteq X$, для которого при всех $\theta \in \Theta'$ верно $P_\theta(X') = 0$, это равенство справедливо и для всех $\theta \in \Theta$.

Предположим, что Θ — интервал прямой, статистика $T(x)$ такова, что разность информаций (58) обращается в нуль. Потребуем от модели, чтобы производные

$$\frac{\partial}{\partial \theta} \log P_{\theta}(x), \quad \frac{\partial}{\partial \theta} \log Q_{\theta}(T(x)), \quad \theta \in \Theta,$$

были непрерывны по θ (при x , для которых они определены). Пусть Θ' — подмножество, определенное в лемме 2, Θ'' — подмножество рациональных точек из Θ . Поскольку левая часть (57), по предположению, обращается в нуль, то при каждом $\theta \in \Theta$

$$\frac{\partial}{\partial \theta} \log P_{\theta}(x) - \frac{\partial}{\partial \theta} \log Q_{\theta}(T(x)) = 0 \quad (64)$$

на подмножестве $\mathcal{X}_0 \subseteq \mathcal{X}$, имеющем P_{θ} -меру единица. В таком случае найдется подмножество $\mathcal{X}' \subseteq \mathcal{X}$, что равенство (64) справедливо при всех $\theta \in \Theta' \cup \Theta''$ и $x \in \mathcal{X}'$, и при этом для $\theta \in \Theta' \cup \Theta''$ верно

$$P_{\theta}(\mathcal{X}') = 1. \quad (65)$$

По непрерывности равенство (64) выполнено при всех $\theta \in \Theta$, $x \in \mathcal{X}'$, а ввиду леммы 2 равенство (65) выполнено при всех $\theta \in \Theta$. Итак, с P_{θ} -вероятностью 1 при каждом $\theta \in \Theta$

$$\frac{\partial}{\partial \theta} \log P_{\theta}(x) - \frac{\partial}{\partial \theta} \log Q_{\theta}(T(x)) = 0, \quad \theta \in \Theta,$$

откуда

$$P_{\theta}(x) = Q_{\theta}(T(x))k(x), \quad \theta \in \Theta,$$

где $k(x)$ не зависит от θ , т. е. статистика $T(x)$ — достаточная. ■

Основание логарифмов в рассмотренных нами информационных мерах не является существенным и определяет лишь единицу измерения. Расчеты в статистике удобнее проводить с натуральными логарифмами.

§ 19. НЕРАВЕНСТВО ФРЕШЕ—РАО—КРАМЕРА

1. Скалярный параметр.

Обратимся к задаче несмешенного оценивания параметрической функции $\varphi(\theta)$ в модели $(\mathcal{X}, \mathcal{B}, \{P_{\theta}, \theta \in \Theta\})$. Пусть $T(x)$ — достаточная статистика. Как было установлено в § 17, если за эффективность оценки принять величину ее дисперсии, то можно ограничиться оценками вида $S(x) = g(T(x))$, не проиграв в эффективности. Если же оценка ищется в виде $h(V(x))$, где статистика $V(x)$ не является достаточной, то происходит проигрыш в эффективности, если только $h(V(X))$ не совпадает с $g(T(X))$ с P_{θ} -вероятностью единица, $\theta \in \Theta$. Интересно, что проигрыш в эффективности несмешенного оценивания может быть оценен непосредственно через количество информации по Фишеру. Напомним, что информация по Фишеру не теряется при переходе от выборки x к достаточной статистике $T(x)$ и, напротив, уменьшается при переходе к статистике $V(x)$, не являющейся достаточной.

В качестве основания логарифмов в определении $I_X(0)$ примем число e . Вкладом выборки называют величину

$$U(x; \theta) = \frac{\partial}{\partial \theta} \ln f(x; \theta) \text{ или } \frac{\partial}{\partial \theta} \ln P_\theta(x), \quad (1)$$

соответственно для непрерывной или дискретной моделей. В случае повторных независимых наблюдений

$$U(x_n; \theta) = \sum_{i=1}^n U(x_i; \theta),$$

где $U(x_i; \theta)$ называется вкладом наблюдения x_i и определяется формулой, аналогичной (1). Мы предполагаем выполненным условие (49) § 18:

$$M_\theta U(X_i; \theta) = 0, \quad (2)$$

так что из (2) получаем

$$I_{X_n}(\theta) = D_\theta U(X_n; \theta) = \sum_{i=1}^n D_\theta U(X_i; \theta) = n I_{X_1}(\theta).$$

Ниже мы докажем, что при некоторых условиях регулярности дисперсия несмещенной оценки $S(x) = g(T(x))$, где $T(x)$ — произвольная статистика, удовлетворяет неравенству Фреше—Рао—Крамера:

$$D_\theta(S(X)) \geq \varphi'(\theta)^2 / I_{T(X)}(\theta). \quad (3)$$

Если статистика $T(x)$ — достаточная, то $I_{T(X)}(0) = I_X(0)$ и неравенство (3) приобретает вид

$$D_\theta S(X) \geq \varphi'(\theta)^2 / I_X(\theta). \quad (4)$$

Если статистика $T(X)$ не является достаточной, то (см. п. 4 § 18)

$$I_{T(X)}(0) < I_X(0),$$

и нижняя граница (3) дисперсии любой несмещенной оценки, являющейся функцией $T(X)$, больше, чем нижняя граница несмещенной оценки — функции всей выборки X .

В случае повторных независимых наблюдений неравенство (4) имеет вид

$$D_\theta S(X_n) \geq \varphi'(\theta)^2 / (n I_{X_1}(0)). \quad (5)$$

Стонит подчеркнуть, что правая часть (5) имеет порядок $1/n$, так что стандартное отклонение имеет порядок $1/\sqrt{n}$. Если оценка $S(x)$ смещенная:

$$M_\theta S(X) = \varphi(\theta) + b(\theta) = \psi(\theta),$$

то она является несмещенной оценкой $\psi(\theta)$, и, следовательно, неравенства (4), (5) сохраняются с заменой $\varphi(\theta)$ на $\psi(\theta)$.

Рассмотрим примеры.

(I) $\mathcal{X} = \{x_n = (x_1, \dots, x_n) : x_i = 0 \text{ или } 1, i = 1, \dots, n\}$,

$$P_\theta(x_n) = \theta^{T(x_n)} (1-\theta)^{n-T(x_n)}, \quad T(x_n) = \sum_{i=1}^n x_i, \quad 0 < \theta < 1.$$

Наилучшая несмещенная оценка параметра θ имеет дисперсию

$$D(T(X_n)/n) = \theta(1-\theta)/n.$$

Вычислим информацию по Фишеру, пользуясь независимостью и одинаковой распределенностью сл. в. X_1, \dots, X_n :

$$I_{X_n}(\theta) = \sum_{i=1}^n I_{X_i}(\theta) = n I_{X_1}(\theta).$$

Так как

$$\begin{aligned} U(x_1; \theta) &= \frac{\partial}{\partial \theta} \ln P_\theta(X_1 = x_1) = \frac{\partial}{\partial \theta} \ln (\theta^{x_1} (1-\theta)^{1-x_1}) = \\ &= \frac{\partial}{\partial \theta} (x_1 \ln \theta + (1-x_1) \ln (1-\theta)) = x_1/\theta - (1-x_1)/(1-\theta) = \\ &= (x_1 - \theta)/(\theta(1-\theta)), \end{aligned}$$

то

$$I_{X_1}(\theta) = D_\theta U(X_1; \theta) = (\theta(1-\theta))^{-1},$$

так что в данном случае неравенство (5) обращается в равенство.

(II) Для выборки из экспоненциального распределения мера P_θ , $\theta > 0$, задается плотностью

$$f_\theta(x_n) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right), \quad x_i > 0, \quad i = 1, \dots, n.$$

Наилучшая несмещенная оценка θ равна (см. пример (I) § 17)

$$T(x_n) = (n-1) / \sum_{i=1}^n x_i.$$

Вычислим дисперсию $T(X_n)$:

$$\begin{aligned} M_\theta T(X_n)^2 &= \int_0^\infty \left(\frac{n-1}{t}\right)^2 \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} dt = \\ &= (n-1)^2 \theta^2 \frac{\Gamma(n-2)}{\Gamma(n)} \int_0^\infty \frac{\theta^{n-2}}{\Gamma(n-2)} t^{(n-2)-1} e^{-\theta t} dt = \end{aligned}$$

$$= (n-1)^2 \theta^2 \frac{\Gamma(n-2)}{\Gamma(n)} = \frac{n-1}{n-2} \theta^2,$$

$$\mathbf{D}_\theta T(\mathbf{X}_n) = \frac{n-1}{n-2} \theta^2 - \theta^2 = \frac{\theta^2}{n-2}. \quad (6)$$

Подсчитаем информацию Фишера:

$$U(x_1; \theta) = \frac{\partial}{\partial \theta} \ln(\theta \exp(-\theta x_1)) = \theta^{-1} - x_1,$$

$$I_{X_1}(\theta) = \mathbf{D}_\theta(\theta^{-1} - x_1) = \theta^{-2}.$$

Неравенство (5) принимает вид

$$\mathbf{D}_\theta S(X_n) \geq \theta^2/n. \quad (7)$$

Таким образом, минимальная достижимая дисперсия (6) несмешенной оценки параметра θ больше, чем граница (7) Фреше—Рао—Крамера. При больших n расхождение между (6) и (7) пренебрежимо мало.

Для параметрической функции $\varphi(\theta) = \theta^{-1}$, совпадающей со средним значением отдельного наблюдения, оценка $\bar{x} = \sum_{i=1}^n x_i/n$ — несмешенная и является функцией полной достаточной статистики $\sum_{i=1}^n x_i$. Ее дисперсия равна $1/(n\theta)^2$, и, с учетом $\varphi'(\theta) = -\theta^{-2}$, получаем для нижней границы (5) значение

$$\varphi'(\theta)^2/(n I_{X_1}(\theta)) = \theta^{-4}/(n\theta^{-2}) = 1/(n\theta^2),$$

т. е. для наилучшей несмешенной оценки \bar{x} неравенство (5) обращается в равенство. ■

Условия, в которых будет доказано неравенство Фреше—Рао—Крамера (4), включают в себя прежде всего требование (49) § 18, сформулированное при определении информации по Фишеру:

$$\mathbf{M}_\theta U(\mathbf{X}; \theta) = 0, \quad \theta \in \Theta. \quad (8)$$

Так как

$$\int \dots \int f_{\mathbf{X}}(\mathbf{x}; \theta) dx_1 \dots dx_n = 1, \quad (9)$$

то в предположении, что операцию дифференцирования по θ можно перенести под знак интеграла, из (9) получаем

$$\int \dots \int \frac{\partial}{\partial \theta} f_{\mathbf{X}}(\mathbf{x}; \theta) dx_1 \dots dx_n = 0. \quad (10)$$

Сравним левую часть (10) с выражением

$$\int \dots \int \left(\frac{\partial}{\partial \theta} f_X(x; \theta) / f_X(x; \theta) \right) f_X(x; \theta) dx_1 \dots dx_n = \\ = \int \dots \int U(x; \theta) f_X(x; \theta) dx_1 \dots dx_n = M_\theta U(X; \theta). \quad (11)$$

Различие между интегралами (10) и (11) заключается в том, что интегрирование в (11) фактически производится лишь по носителю меры P_θ :

$$X_\theta = \{x : f_X(x; \theta) > 0\}. \quad (12)$$

Очевидно, если верно соотношение (10), то (8) выполнено тогда и только тогда, когда

$$\int_{x \in X_\theta} \dots \int \frac{\partial}{\partial \theta} f_X(x; \theta) dx_1 \dots dx_n = 0. \quad (13)$$

Соотношение (13) выполнено, если

$$\frac{\partial}{\partial \theta} f_X(x; \theta) = 0 \text{ при } x \in X \setminus X_\theta. \quad (14)$$

Для повторной независимой выборки равенство (9) приобретает вид

$$(\int f_{X_1}(x_1; \theta) dx_1)^n = 1,$$

а требование (13) сводится к следующему:

$$\int \frac{\partial}{\partial \theta} f_{X_1}(x_1; \theta) dx_1 = 0. \quad (15)$$

Еще одно требование, необходимое для доказательства неравенства (4), — это возможность перенести операцию дифференцирования под знак интеграла в следующем выражении:

$$\varphi'(\theta) = \frac{d}{d\theta} \int \dots \int S(x) f_X(x; \theta) dx_1 \dots dx_n = \\ = \int \dots \int S(x) \frac{\partial}{\partial \theta} f_X(x; \theta) dx_1 \dots dx_n. \quad (16)$$

С учетом условия (14) соотношение (16) можно переписать в виде (ср. (11))

$$\varphi'(\theta) = M_\theta(S(X) U(X; \theta)). \quad (17)$$

Теорема 1. Допустим, что плотность выборки $f_X(x, \theta)$ дифференцируема по $\theta \in \Theta$, Θ — интервал прямой, операция дифференцирования по θ перестановочна с интегрированием по x в интегралах (10), (16) и выполнено условие (14). Если S — несме-

щенная оценка параметрической функции $\phi(\theta)$ и существуют дисперсии

$$\mathbf{D}_{\theta}S(\mathbf{X}), \quad \mathbf{D}_{\theta}U(\mathbf{X}; \theta) = I_{\mathbf{X}}(\theta),$$

то имеет место неравенство (4); в равенство оно обращается тогда и только тогда, когда $S(x)$ является линейной функцией вклада выборки:

$$S(x) = \varphi(\theta) + a(\theta) U(x; \theta) \quad (18)$$

на подмножестве \mathbb{P}_{θ} -вероятности единицы.

Доказательство. Из условия теоремы вытекают соотношения (8) и (17). Используя (8), перепишем (17) в виде

$$\phi'(\theta) = M(S(\mathbf{X}) - \varphi(\theta)) U(\mathbf{X}; \theta).$$

По неравенству Коши—Шварца—Буняковского

$$\phi'(\theta)^2 \leq \mathbf{D}_{\theta}S(\mathbf{X}) \mathbf{D}_{\theta}U(\mathbf{X}; \theta),$$

причем равенство имеет место лишь при выполнении условия (18), что и требовалось доказать. ■

Замечание 1. Для дискретной модели сохраняются все предшествующие рассуждения с заменой плотностей на распределения вероятностей и интегралов на суммы.

Замечание 2. Если $T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_r(\mathbf{X}))$ — статистика и $f_T(t; \theta)$ удовлетворяет условиям теоремы 1, то для несмещенных оценок — функций $T(\mathbf{X})$ — справедливо неравенство (3).

Следствие. Если неравенство (4) обращается в равенство при всех $\theta \in \Theta$, то из (18) находим

$$\frac{\partial}{\partial \theta} \ln f_{\mathbf{X}}(\mathbf{x}; \theta) = a(\theta)^{-1} S(\mathbf{x}) - \varphi(\theta) a(\theta)^{-1}, \quad (19)$$

$$f_{\mathbf{X}}(\mathbf{x}, \theta) = h(x) \exp(A(\theta) S(\mathbf{x}) + B(\theta)), \quad (20)$$

и при этом

$$B'(\theta)/A'(\theta) = -\varphi(\theta). \quad (21)$$

Наоборот, если $f_{\mathbf{X}}(\mathbf{x}; \theta)$ имеет вид (20), то верно (19), а следовательно, (18), т. е. неравенство Фреше—Рао—Крамера (4) обращается в равенство. Таким образом, обращение неравенства (4) в равенство имеет место только в случае экспоненциальной модели (20) и только для оценок, имеющих вид $\alpha S(\mathbf{x}) + \beta$ (α, β — любые постоянные), или, другими словами, для параметрических функций вида

$$\varphi(\theta) = \gamma B'(\theta)/A'(\theta) + \delta$$

(γ, δ — любые постоянные).

2. Векторный параметр.

Теорема 2. Допустим, что плотность выборки $f_{\mathbf{X}}(\mathbf{x}; \theta)$, $\theta = (\theta_1, \dots, \theta_k) \in \Theta$ (Θ — область в R^k) дифференцируема по каж-

дому 0_i , $i=1, \dots, k$, векторная статистика $\mathbf{S}(\mathbf{x}) = (S_1(\mathbf{x}), \dots, S_r(\mathbf{x}))$ несмешенно оценивает вектор-функцию $\Phi(\theta) = (\varphi_1(\theta), \dots, \varphi_r(\theta))$ и имеют место соотношения

$$\frac{\partial}{\partial \theta_i} \int f_{\mathbf{x}}(\mathbf{x}; \theta) d\mathbf{x} = \int \frac{\partial}{\partial \theta_i} f_{\mathbf{x}}(\mathbf{x}; \theta) d\mathbf{x},$$

$$\frac{\partial}{\partial \theta_i} \int \mathbf{S}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}; \theta) d\mathbf{x} = \int \mathbf{S}(\mathbf{x}) \frac{\partial}{\partial \theta_i} f_{\mathbf{x}}(\mathbf{x}; \theta) d\mathbf{x},$$

$i=1, \dots, k$, $d\mathbf{x}=dx_1 \dots dx_n$, а интегралы понимаются как n -мерные. Предположим также, что

$$\partial f_{\mathbf{x}}(\mathbf{x}; \theta) / \partial \theta_i = 0, \quad i = 1, \dots, k \text{ при } \mathbf{x}: f_{\mathbf{x}}(\mathbf{x}; \theta) = 0.$$

Если существует матрица ковариаций

$$R_{\mathbf{S}(\mathbf{x})}(\theta) = \mathbf{M}_{\theta}(\mathbf{S}(\mathbf{X}) - \Phi(\theta))' (\mathbf{S}(\mathbf{X}) - \Phi(\theta)),$$

и информационная матрица Фишера $I_{\mathbf{x}}(\theta)$ положительно определена, то ^{*}

$$R_{\mathbf{S}(\mathbf{x})}(\theta) \geq \Delta(\theta) I_{\mathbf{x}}^{-1}(\theta) \Delta(\theta)', \quad (22)$$

где

$$\Delta(\theta) = \left[\frac{\partial}{\partial \theta_j} \varphi_i(\theta), \quad i = 1, \dots, r; \quad j = 1, \dots, k \right]$$

и штрих обозначает операцию транспонирования. Равенство в (22) имеет место тогда и только тогда, когда

$$\mathbf{S}(\mathbf{x}) = \Phi(\theta) + \mathbf{U}(\mathbf{x}; \theta) A(\theta) \quad (23)$$

на подмножестве \mathbf{P}_{θ} -вероятности единицы; $A(\theta)$ — $k \times r$ -матрица полного ранга, зависящая лишь от θ .

Доказательство. Аналогично тому, как это сделано в случае скалярного параметра, легко показать, что

$$\mathbf{M}_{\theta} \mathbf{U}(\mathbf{X}; \theta) = 0, \quad \mathbf{U}(\mathbf{x}; \theta) = \operatorname{grad} \ln f_{\mathbf{x}}(\mathbf{x}; \theta),$$

$$\frac{\partial}{\partial \theta_j} \varphi_i(\theta) = \mathbf{M}_{\theta} S_i(\mathbf{X}) \frac{\partial}{\partial \theta_j} \ln f_{\mathbf{x}}(\mathbf{X}; \theta).$$

так что

$$I_{\mathbf{x}}(\theta) = R_{\mathbf{U}(\mathbf{x}; \theta)} = \mathbf{M}_{\theta} \mathbf{U}'(\mathbf{X}; \theta) \mathbf{U}(\mathbf{X}; \theta), \quad (24)$$

$$\Delta(\theta) = \mathbf{M}_{\theta} (\mathbf{S}(\mathbf{X}) - \mathbf{M}_{\theta} \mathbf{S}(\mathbf{X}))' \mathbf{U}(\mathbf{X}; \theta).$$

Найдем ковариационную матрицу вектора

$$\mathbf{W}(\mathbf{X}; \theta) = \mathbf{S}(\mathbf{X}) - \mathbf{M}_{\theta} \mathbf{S}(\mathbf{X}) - \mathbf{U}(\mathbf{X}; \theta) I_{\mathbf{x}}^{-1}(\theta) \Delta'(\theta).$$

* См. сноску на с. 182.

Проведя элементарные преобразования, получаем с учетом (24)

$$\begin{aligned} R_w &= M_0(S - M_0 S - U I^{-1} \Delta')'(S - M_0 S - U I^{-1} \Delta') = \\ &= R_s + \Delta I^{-1} R_u I^{-1} \Delta' - \Delta I^{-1} M_0 U' (S - M_0 S) - \\ &\quad - M_0 (S - M_0 S)' U I^{-1} \Delta' = R_s - \Delta I^{-1} \Delta', \end{aligned}$$

что и дает утверждение (22) теоремы 2. Равенство в (22) возможно, лишь когда $R_w = [0]$, т. е. с P_θ -вероятностью единица

$$S(X) - M_0 S(X) - U(X; \theta) I^{-1}(\theta) \Delta'(\theta) = 0, \quad (25)$$

откуда и следует (23). ■■■

Следствие 1. Пусть неравенство (22) обращается в равенство, $\Delta(\theta)$ обратима, $I_X^{-1}(\theta)$, $\varphi(\theta)$ непрерывны по θ . Тогда найдутся такие функции $H_1(\theta), \dots, H_r(\theta)$, $B(\theta)$, $K(x)$, что

$$f_X(x; \theta) = \exp \left[\sum_{i=1}^r H_i(\theta) S_i(x) + B(\theta) + K(x) \right]. \quad (26)$$

т. е. статистическая модель является экспоненциальной.

Действительно, из (23) получаем, что вектор градиента по θ

$$\operatorname{grad}_\theta \ln f(x; \theta) = S(x) A^{-1}(\theta) - \varphi(\theta) A^{-1}(\theta)$$

непрерывно зависит от θ и, следовательно, функция $\ln f(x; \theta)$ — дифференцируемая функция переменных $\theta = (\theta_1, \dots, \theta_n)$ в области Θ . В курсах математического анализа доказывается, что в этом случае

$$\ln f(x; \theta) - \ln f(x; \theta_0) = \int_{\theta_0}^\theta (S(x) A^{-1}(\theta) - \varphi(\theta) A^{-1}(\theta)) d\Gamma,$$

где криволинейный интеграл берется вдоль любого гладкого пути, лежащего в области Θ и соединяющего точки θ_0 и θ , $d\Gamma$ — векторный элемент пути, штрих обозначает транспонирование. Производя интегрирование, получаем представление (26).

Следствие 2. Если X_n — повторная независимая случайная выборка, то

$$\begin{aligned} \ln f_{X_n}(x_n; \theta) &= \sum_{i=1}^n \ln f_{X_i}(x_i; \theta), \\ U(x_n; \theta) &= \operatorname{grad}_\theta \ln f_{X_n}(x_n; \theta) = \\ &= \sum_{i=1}^n \operatorname{grad}_\theta \ln f_{X_i}(x_i; \theta) = \sum_{i=1}^n U(x_i; \theta), \\ I_{X_n}(\theta) &= R_{U(X_n; \theta)} = \sum_{i=1}^n R_{U(X_i; \theta)} = n I_{X_1}(\theta), \end{aligned}$$

и неравенство (22) приобретает вид

$$R_{S(x)}(\theta) \geq \frac{1}{n} \Delta(\theta) I_{X_1}^{-1}(\theta) \Delta(\theta)' . \quad (27)$$

Рассмотрим пример.

(III) Для повторной независимой выборки из распределения $N(\mu, \sigma^2)$ положим $\theta_1 = \mu$, $\theta_2 = \sigma^2$. Тогда

$$f_X(x; \theta) = (\sqrt{2\pi\theta_2})^{-n} \exp\left(-\frac{1}{\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2\right).$$

Меры P_θ имеют общий носитель — все пространство R^n , и выполняются требования дифференцируемости интегралов теоремы 2. Находим

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \ln f_{X_1}(x_1; \theta) &= \frac{\partial}{\partial \theta_1} \left(-\ln \sqrt{2\pi} - \frac{1}{2} \ln \theta_2 - \frac{1}{2\theta_2} (x_1 - \theta_1)^2 \right) = \\ &= (x_1 - \theta_1) \theta_2^{-1}, \\ \frac{\partial}{\partial \theta_2} \ln f_{X_1}(x_1; \theta) &= -\frac{1}{2\theta_2} + \frac{(x_1 - \theta_1)^2}{2\theta_2^2}, \end{aligned}$$

откуда для вектора $U(x_1; \theta)$ и симметричной матрицы $U'(x_1; \theta) \times U(x_1; \theta)$ получаем выражения

$$U(x_1; \theta) = ((x_1 - \theta_1) \theta_2^{-1}, (x_1 - \theta_1)^2 \theta_2^{-2} - \theta_2^{-1}),$$

$$U'(x_1; \theta) U(x_1; \theta) = \begin{bmatrix} \theta_2^{-2} (x_1 - \theta_1)^2 & \theta_2^{-3} (x_1 - \theta_1)^3 - (x_1 - \theta_1) \theta_2^{-2} \\ \dots & \theta_2^{-2} ((x_1 - \theta_1)^2 \theta_2^{-1} - 1)^2 \end{bmatrix}$$

Таким образом

$$I_{X_1}(\theta) = M U'(X_1; \theta) U(X_1; \theta) = \begin{bmatrix} \theta_2^{-1} & 0 \\ 0 & 2\theta_2^{-2} \end{bmatrix},$$

где мы использовали следующие соотношения для стандартной нормальной сл. в. $Y_1 = (X_1 - \theta_1)/\theta_2^{-1/2}$ (см. п. 3 § 7): $MY_1^3 = 0$, $MY_1^4 = 3$. Поскольку

$$\Phi(\theta) = M \theta S(X) = (\theta_1, \theta_2), \Delta(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

то неравенство (27) для любой несмещенной оценки $S(x)$, удовлетворяющей требованиям теоремы 2, принимает вид

$$R_{S(x)}(\theta) \geq \frac{1}{n} I_{X_1}(\theta) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}. \quad (28)$$

В частности,

$$DS_1(\mathbf{X}) \geq \sigma^2/n, \quad DS_2(\mathbf{X}) \geq 2\sigma^4/n.$$

Наилучшие несмешенные оценки

$$S_1(\mathbf{x}_n) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad S_2(\mathbf{x}_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

очевидно, удовлетворяют требованиям регулярности теоремы 2, а их матрица ковариаций равна

$$R_{S(\mathbf{x})}(\theta) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/(n-1) \end{bmatrix}. \quad (29)$$

где мы использовали, что \bar{X} и $Z = \sum_{i=1}^n (X_i - \bar{X})^2$ независимы,

$\sigma_2^{-1}Z$ имеет распределение χ_{n-1}^2 , т. е. $\sigma_2^{-1}Z$ распределена как сумма $n-1$ квадратов независимых $N(0, 1)$ сл. в. Сравнивая (28) и (29), мы видим, что по θ_2 нижняя граница не достигается, но при больших n разница пренебрежимо мала.

3. Границы дисперсии при нарушении условий регулярности

Как видно из неравенства Фреше—Рао—Крамера, в регулярном случае дисперсия несмешенных оценок в повторной выборке имеет с ростом n порядок не выше n^{-1} . Если дисперсия оценки $T(\mathbf{X}_n)$ имеет порядок $1/n$, то из неравенства Чебышева вытекает, что отклонение оценки от ее математического ожидания имеет порядок $1/\sqrt{n}$.

В то же время были примеры, в которых порядок убывания дисперсии равнялся $1/n^2$. Так было для выборки из равномерного на $(0, \theta)$ распределения и статистики $x_{(n)}$, а также для распределения с экспоненциальной плотностью $\sigma^{-1} \exp((x - \mu)/\sigma)$ и статистики $x_{(1)}$. В обоих случаях оценивается точка разрыва плотности $f(x; \theta)$ распределения отдельного наблюдения, или, другими словами, точка излома функции распределения. Можно показать, что, вообще говоря, точка x_θ , в которой ф. р. $F(x; \theta)$ не имеет производной, может быть оценена по результатам повторной выборки с точностью, превышающей $1/\sqrt{n}$; здесь мы ограничимся примером.

(IV) Пусть одномерное распределение повторной выборки сосредоточено на интервале $(0, a(\theta))$, а ф. р. $F(x; \theta)$ вблизи точки $x=a(\theta)$ ведет себя следующим образом:

$$1 - F(a(\theta) - \delta, \theta) \sim c\delta^\alpha, \quad \delta \rightarrow 0, \quad \alpha > 0, \quad c > 0.$$

В таком случае при $n \rightarrow \infty$

$$\mathcal{P}_\theta(n^{1/\alpha}(a(\theta) - X_{(n)}) > t) = \mathcal{P}_\theta(X_{(n)} < a(\theta) - tn^{-1/\alpha}) =$$

$$= F(a(0) - tn^{-1/\alpha})^n = (1 - (1 - F(a(0) - tn^{-1/\alpha})))^n \sim \\ \sim (1 - ct^\alpha n^{-1})^n \rightarrow \exp(-ct^\alpha). \quad (30)$$

Таким образом, отклонение $X_{(n)}$ от $a(\theta)$ имеет порядок $n^{-1/\alpha}$, так что при $\alpha < 2$ статистика $x_{(n)}$ является лучшей оценкой $a(\theta)$, чем это можно было бы ожидать в регулярном случае. ■

Неравенство Фреше—Рао—Крамера можно модифицировать так, что оно даст нижнюю границу для дисперсии оценок и при нарушении условий регулярности. Запишем в случае скалярного параметра θ условие несмещенностии оценки $S(x)$ в точках θ и $\theta + \Delta\theta$:

$$\int S(x) f_x(x; \theta) dx = \varphi(\theta); \quad (31)$$

$$\int S(x) f_x(x; \theta + \Delta\theta) dx = \varphi(\theta + \Delta\theta),$$

или

$$\int (S(x) - \varphi(\theta)) f_x(x; \theta) dx = 0; \quad (32)$$

$$\int (S(x) - \varphi(\theta)) f_x(x; \theta + \Delta\theta) dx = \varphi(\theta + \Delta\theta) - \varphi(\theta),$$

где $dx = dx_1 \dots dx_n$, а интегралы понимаются как n -мерные. Вычитая выражения (32) одно из другого, получаем

$$\int (S(x) - \varphi(\theta)) \Delta_\theta f_x(x; \theta) dx = \Delta\varphi(\theta), \quad (33)$$

где $\Delta_\theta f_x(x; \theta) = f_x(x; \theta + \Delta\theta) - f_x(x; \theta)$, $\Delta\varphi(\theta) = \varphi(\theta + \Delta\theta) - \varphi(\theta)$. Допустим, что для данного $\Delta\theta$

$$\{x : f_x(x; \theta + \Delta\theta) > 0\} \subseteq \{x : f_x(x; \theta) > 0\}. \quad (34)$$

Тогда, ограничивая область интегрирования в (33) правым множеством (34), перепишем (33) в виде

$$\int (S(x) - \varphi(\theta)) \frac{\Delta_\theta f_x(x; \theta)}{f_x(x; \theta)} f_x(x; \theta) dx = \Delta\varphi(\theta). \quad (35)$$

Применяя к (35) неравенство Коши—Шварца—Буняковского, получаем

$$\Delta\varphi(\theta)^2 \leq D_\theta S(X) \int \frac{(\Delta_\theta f_x(x; \theta))^2}{f_x(x; \theta)} dx, \quad (36)$$

где интеграл в (36) берется по носителю меры P_θ . В случае повторной выборки и при соблюдении условия (34) имеем

$$\frac{\Delta_\theta f_x(x; \theta)}{f_x(x; \theta)} = \prod_{i=1}^n \frac{f_{X_i}(x_i; \theta + \Delta\theta)}{f_{X_i}(x_i; \theta)} - 1,$$

и интеграл в правой части (36) можно записать в виде

$$\begin{aligned} M_{\theta} \left(\frac{\Delta_{\theta} f_X(X; \theta)}{f_X(X; \theta)} \right)^2 &= \left(M_{\theta} \left(\frac{f_{X_1}(X_1; \theta + \Delta\theta)}{f_{X_1}(X_1; \theta)} \right)^2 \right)^n - \\ &- 2M_{\theta} \prod_{i=1}^n \frac{f_{X_i}(X_i; \theta + \Delta\theta)}{f_{X_i}(X_i; \theta)} + 1. \end{aligned} \quad (37)$$

Учитывая независимость сл. в. X_1, \dots, X_n , а также, что

$$M_{\theta} \left(\frac{f_{X_1}(X_1; \theta + \Delta\theta)}{f_{X_1}(X_1; \theta)} \right) = \int f_{X_1}(x_1; \theta + \Delta\theta) dx_1 = 1.$$

получаем для (37) выражение

$$\left(M_{\theta} \left(\frac{f_{X_1}(X_1; \theta + \Delta\theta)}{f_{X_1}(X_1; \theta)} \right)^2 \right)^n - 1 = \left(M_{\theta} \left(\frac{\Delta_{\theta} f_{X_1}(X_1; \theta)}{f_{X_1}(X_1; \theta)} \right)^2 + 1 \right)^n - 1, \quad (38)$$

где также было использовано равенство, аналогичное (37), но для $n=1$. Таким образом, для повторной независимой выборки неравенство (36) приобретает вид

$$\Delta_{\Phi}(\theta)^2 \leq D_{\theta} S(X) \left(\left(M_{\theta} \left(\frac{\Delta_{\theta} f_{X_1}(X_1; \theta)}{f_{X_1}(X_1; \theta)} \right)^2 + 1 \right)^n - 1 \right) \quad (39)$$

(V) Рассмотрим частный случай примера (IV), полагая, что плотность одномерного распределения повторной выборки равна

$$f(x; \theta) = \alpha \theta^{-\alpha} (\theta - x)^{\alpha-1}, \quad 0 < x < \theta, \quad 1/2 < \alpha < 1,$$

и равна нулю при остальных значениях x . Найдем асимптотику при $\Delta\theta \rightarrow 0$ математического ожидания, входящего в правую часть (39), полагая $\Delta\theta = -\theta\delta$, $\delta > 0$, что обеспечивает выполнение одномерного варианта условия (34). Проводя элементарные преобразования, получаем

$$\begin{aligned} \int_0^{\theta} \frac{(\Delta_{\theta} f(x; \theta))^2}{f(x; \theta)} dx &= \int_0^{\theta} \frac{(f(x; \theta - \theta\delta) - f(x; \theta))^2}{f(x; \theta)} dx = \\ &= \int_0^{\theta} \frac{f(x; \theta - \theta\delta)^2}{f(x; \theta)} dx - 2 \int_0^{\theta} f(x; \theta - \theta\delta) dx + \int_0^{\theta} f(x; \theta) dx. \end{aligned} \quad (40)$$

Два последних слагаемых в правой части (40) дают в сумме -1 , и выражение (40) приводится к виду

$$-1 + \int_0^{\theta(1-\delta)} (\alpha(\theta - \theta\delta)^{-\alpha} (\theta - \theta\delta - x)^{\alpha-1})^2 \alpha^{-1} \theta^{\alpha} (\theta - x)^{-\alpha+1} dx =$$

$$= -1 + \int_0^1 \alpha (1-\delta)^{-1} (1-x)^{\alpha-1} (1+\delta x (1-x)^{-1})^{1-\alpha} dx. \quad (41)$$

Мы покажем, что при $\delta \rightarrow 0$ выражение (41) эквивалентно $c\delta^\alpha$, где c — некоторая постоянная. Имея в виду асимптотику, отбросим в интеграле (41) множитель $(1-\delta)^{-1}$ и, положив $y=1-x$, запишем остающееся выражение в виде

$$\int_0^1 a y^{\alpha-1} ((1+\delta y^{-1}-\delta)^{1-\alpha}-1) dy = \delta^\alpha c \int_0^\infty (z+\delta)^{-\alpha-1} ((1+z)^{1-\alpha}-1) dz,$$

где сделана замена переменного $z=\delta y^{-1}-\delta$, что и доказывает наше утверждение, причем

$$c = a \int_0^\infty z^{-\alpha-1} ((1+z)^{1-\alpha}-1) dz.$$

Возьмем $\delta^\alpha = (bn)^{-1}$, где b — произвольная постоянная. Тогда из (39) находим

$$D_\theta S(X) \geq \Delta\varphi(\theta)^2 \left(\left(\int_0^\theta \frac{1(\Delta\theta f(x; \theta))^2}{f(x; \theta)} dx + 1 \right)^n - 1 \right)^{-1} \underset{n \rightarrow \infty}{\sim} \quad (42)$$

$$\sim \theta^2 \varphi'(\theta)^2 (bn)^{-2/\alpha} ((c(bn)^{-1} + 1)^n - 1)^{-1} \sim \\ \sim \theta^2 \varphi'(\theta)^2 b^{-2/\alpha} (c^{cb^{-1}} - 1)^{-1} n^{-2/\alpha}. \quad (43)$$

Значение b можно выбрать так, чтобы максимизировать множитель при $n^{-2/\alpha}$ в (43). Если δ^α взять порядка $n^{-1+\epsilon}$ или $n^{-1-\epsilon}$, $\epsilon > 0$, то легко обнаружить, что правая часть (42) будет иметь порядок $o(1)n^{-2/\alpha}$, где $o(1) \rightarrow 0$, $n \rightarrow \infty$. Следовательно, граница (43) является наилучшей возможной (по порядку), которую может дать неравенство (39). Заметим, что оценка параметра θ с помощью порядковой статистики $x_{(n)}$, предложенная в примере (IV), как раз дает оптимальный порядок отклонения (см. (30)).

ГЛАВА IV

ПРАВДОПОДОБИЕ

§ 20. МЕТОД МАКСИМУМА ПРАВДОПОДОБИЯ

1. Функция правдоподобия.

Положим

$$L(x; \theta) = f(x; \theta) \text{ или } P_\theta(x), x \in \mathcal{X}, \theta \in \Theta,$$

соответственно для непрерывной или дискретной статистической модели (\mathcal{X} , Θ , $\{P_\theta, \theta \in \Theta\}$). В § 15, 16, рассматривая $L(x, \theta)$ при каждом фиксированном θ как статистику, мы видели, что совокупность статистик $\{L(x, \theta), \theta \in \Theta\}$ порождает достаточное разбиение \mathcal{X} . Это означает, что любые два наблюдения $x^{(1)}$ и $x^{(2)}$, для которых

$$L(x^{(1)}; \theta) = L(x^{(2)}; \theta), \theta \in \Theta, \quad (1)$$

несут равнозначную информацию о неизвестном параметре θ . Если какая-то статистическая процедура дает различные ответы для $x^{(1)}$, $x^{(2)}$, удовлетворяющих (1), то следует признать, что при этом используется информация, не имеющая отношения к задаче различия распределений P_θ , $\theta \in \Theta$. Хотя некоторые вопросы теории приобретают более естественную математическую форму, если допустить и такие процедуры, тем не менее основное теоретическое и прикладное значение имеют процедуры, которые для любых $x^{(1)}$ и $x^{(2)}$ с совпадающими функциями $L(x^{(1)}; \theta)$ и $L(x^{(2)}; \theta)$, $\theta \in \Theta$, приводят к одним и тем же статистическим выводам. Функция

$$L(x; \theta), \theta \in \Theta, \quad (1)'$$

у которой θ рассматривается как основное переменное, а x — как параметр, называется *функцией правдоподобия*. Значение функции $L(x; \theta)$ при данном наблюдении x называется *правдоподобием* θ (или меры P_θ) при этом x . Значительное практическое и теоретическое удобство достигается переходом к логарифмам. Функцию

$$l(x; \theta) = \ln L(x; \theta), \theta \in \Theta,$$

называют *логарифмической функцией правдоподобия*. В схеме повторной независимой выборки

$$l(\mathbf{x}; \theta) = \sum_{i=1}^n l(x_i; \theta), \quad l(x_i; \theta) = \ln f(x_i; \theta), \quad i = 1, \dots, n. \quad (2)$$

Итак, функция правдоподобия $L(\mathbf{x}; \theta)$, $\theta \in \Theta$, содержит всю информацию о неизвестном параметре θ , заключенную в выборке \mathbf{x} . Но, как мы знаем, возможно дальнейшее сокращение данных, если перейти к минимальному достаточному разбиению, при котором в один класс помещаются выборки $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, для которых отношение

$$L(\mathbf{x}^{(1)}; \theta)/L(\mathbf{x}^{(2)}; \theta) = k(\mathbf{x}^{(1)}; \mathbf{x}^{(2)})$$

не зависит от $\theta \in \Theta$ (см. подробнее п. 6 § 15, п. 3 § 16). Принцип, требующий от статистической процедуры одинакового заключения для выборок с пропорциональными функциями правдоподобия, называется *принципом правдоподобия*. Как видно, он совпадает с *принципом достаточности*, в соответствии с которым по выборкам $\mathbf{x}^{(1)}$ и $\mathbf{x}^{(2)}$ с одним и тем же значением минимальной достаточной статистики должны производиться идентичные статистические выводы. Подробное обсуждение различных принципов статистического вывода и их интерпретация содержатся в книге Кокса и Хинкли [14].

2. Оценки максимального правдоподобия.

Задача точечного оценивания, решаемая с позиций принципа достаточности, приводит к оценкам, являющимся функциями минимальной достаточной статистики. При этом, однако, остается открытым вопрос, какую именно из функций достаточной статистики следует выбрать для оценивания заданной параметрической функции $\phi(\theta)$. Подход, основанный на выборе несмешенной оценки с минимальной дисперсией, был подробно рассмотрен в § 17. Этот подход эффективен, однако, лишь когда имеется полная достаточная статистика небольшой размерности, что практически сводится к моделям экспоненциального типа. Требование несмещности также ограничивает возможности применения метода; к тому же, ничто не препятствует существованию оценки $S(\mathbf{X})$ с небольшим смещением и среднеквадратическим отклонением

$$M_\theta(S(\mathbf{X}) - \phi(\theta))^2,$$

меньшим, чем дисперсия наилучшей несмешенной оценки. Реализацией принципа правдоподобия в задаче точечного оценивания можно считать оценивание по методу максимального правдоподобия: оценка $\hat{\theta}(\mathbf{x})$, максимизирующая функцию правдоподобия

$$L(\mathbf{x}; \hat{\theta}(\mathbf{x})) \geq L(\mathbf{x}; \theta), \quad \theta \in \Theta,$$

или, что то же, логарифмическую функцию правдоподобия, называется оценкой максимального правдоподобия (о.м.п.) параметра θ . Если статистическая модель дискретна, то $\hat{\theta}(\mathbf{x})$ есть такое

значение параметра θ , при котором вероятность $P_\theta(x)$ наблюдаемой выборки x наибольшая; в непрерывной модели плотность $f(x, \theta)$ наблюденного значения выборки x становится максимальной при значении $\theta = \hat{\theta}(x)$.

Допустим, что $\phi(\theta) = (\phi_1(\theta), \dots, \phi_l(\theta))$ есть некоторая векторная параметрическая функция. Что следует понимать под оценкой максимального правдоподобия функции $\phi(\theta)$? Не ограничивая общности, будем предполагать, что система $\phi_1(\theta), \dots, \phi_l(\theta)$ функционально независима и ее можно дополнить до взаимно-однозначного преобразования всего параметрического пространства. Пусть, для краткости, $l=k$ и отображение $\phi=\phi(\theta)$ уже взаимно-однозначно. Функция правдоподобия $L(x; \phi)$ в новой параметризации и $L(x; \theta)$ связаны формулой

$$L(x; \phi(\theta)) = L(x; \theta).$$

Максимизация по ϕ функции $L(x; \phi)$, очевидно, эквивалентна максимизации $L(x; \theta)$ по θ , и точки максимума связаны соотношением

$$\hat{\phi} = \phi(\hat{\theta}(x)). \quad (3)$$

Естественно определить о. м. п. любой параметрической функции $\phi(\theta)$ как результат подстановки $\theta = \hat{\theta}(x_n)$ в оцениваемую параметрическую функцию. Отметим, что в задаче наилучшего несмешенного оценивания оценка для $\phi(\theta)$ ищется независимо для каждой $\phi(\theta)$. ■

Отметим, что если $T(x)$ — достаточная статистика, то по теореме факторизации

$$L(x; \theta) = h(x)g(T(x); \theta),$$

и задача построения о. м. п. сводится к поиску максимума по θ функции $g(T(x); \theta)$. Если точка максимума по θ функции $g(t; \theta)$ единственная, то о. м. п. $\hat{\theta}(x)$ определяется однозначно и является функцией достаточной статистики $T(x)$. В частности, при соблюдении указанного условия получаем, что $\hat{\theta}(x)$ является функцией минимальной достаточной статистики. ■

В случае, когда функция правдоподобия дифференцируема, нахождение о. м. п. сводится к решению системы

$$\frac{\partial l(x; \theta)}{\partial \theta_i} = 0, \quad i=1, \dots, k, \quad \theta = (\theta_1, \dots, \theta_k),$$

называемой *системой уравнений правдоподобия*. Точка максимума функции $l(x, \theta)$ находится среди решений этой системы (а также значений функции правдоподобия на границе области Θ). ■

Многие из оценок, с которыми мы до сих пор встречались, являются оценками максимума правдоподобия. Рассмотрим некоторые примеры.

(!) Для испытаний Бернулли с вероятностью положительного исхода 0 получаем

$$L(x_n; \theta) = \theta^{S(x_n)} (1-\theta)^{n-S(x_n)}, \quad S(x_n) = x_1 + \dots + x_n,$$

$$l(x_n; \theta) = S(x_n) \ln \theta + (n - S(x_n)) \ln (1-\theta).$$

Для нахождения максимума приравняем нулю производную по θ логарифмической функции правдоподобия:

$$\begin{aligned} 0 &= \partial l(x_n; \theta) / \partial \theta = S(x_n) \theta^{-1} - (n - S(x_n)) (1-\theta)^{-1} = \\ &= S(x_n) (\theta (1-\theta))^{-1} - n (1-\theta)^{-1} = (S(x_n) \theta^{-1} - n) (1-\theta)^{-1}. \end{aligned}$$

Производная $\partial l(x_n; \theta) / \partial \theta$ меняет знак с минуса на плюс при переходе через точку $\widehat{\theta}(x_n) = \frac{1}{n} S(x_n) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$, и потому \bar{x} является о. м. п.

(II) Для повторной выборки из нормального распределения $N(\mu, \sigma^2)$, полагая $\theta_1 = \mu$, $\theta_2 = \sigma^2$, получаем

$$\begin{aligned} l(x_n; \theta) &= -n \ln \sqrt{2\pi\theta_2} - \frac{1}{2\theta_2} \sum_{i=1}^n (x_i - \theta_1)^2 = \\ &= -n \ln \sqrt{2\pi\theta_2} - \sum_{i=1}^n (x_i - \bar{x})^2 / (2\theta_2) - n(\bar{x} - \theta_1)^2 / (2\theta_2), \\ \partial l(x_n; \theta) / \partial \theta_1 &= n(\bar{x} - \theta_1) / \theta_2, \\ \partial l(x_n; \theta) / \partial \theta_2 &= -n/2\theta_2 + \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta_1)^2 \right) / (2\theta_2^2). \end{aligned}$$

Приравнивая производные к нулю, находим

$$\widehat{\theta}_1(x_n) = \bar{x}, \quad \widehat{\theta}_2(x_n) = \sum_{i=1}^n (x_i - \bar{x})^2 / n.$$

Так как (исключая случай $x_i = \bar{x}$, $i = 1, \dots, n$)

$$\frac{\partial^2 l(x_n; \theta)}{\partial \theta_1^2} \Big|_{\theta=\widehat{\theta}} = -\theta_2^{-1} \Big|_{\theta=\widehat{\theta}} = -\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{-1} < 0,$$

$$\begin{aligned} \frac{\partial^2 l(x_n; \theta)}{\partial \theta_2^2} \Big|_{\theta=\widehat{\theta}} &= \left(\frac{n}{2\theta_2^2} - \frac{1}{\theta_2^3} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta_1)^2 \right) \right) \Big|_{\theta=\widehat{\theta}} = \\ &= n/2\widehat{\theta}_2^2 - n\widehat{\theta}_2^2 < 0, \end{aligned}$$

$$\frac{\partial^2 l(x_n; \theta)}{\partial \theta_1 \partial \theta_2} \Big|_{\theta=\widehat{\theta}} = -\frac{n(\bar{x} - \theta_1)}{\theta_2^2} \Big|_{\theta=\widehat{\theta}} = 0,$$

то дискриминант

$$(\partial^2 l / \partial \theta_1^2)(\partial^2 l / \partial \theta_2^2) - \partial^2 l / (\partial \theta_1 \partial \theta_2) > 0$$

и, следовательно, в точке $\hat{\theta}_1, \hat{\theta}_2$ функция правдоподобия достигает максимального значения.

(III) Для повторной выборки из равномерного распределения на $(0, \theta)$

$$L(x_n; \theta) = \begin{cases} \theta^{-n}, & 0 < x_{(1)} \leq x_{(n)} < \theta, \\ 0 & \text{в противном случае.} \end{cases}$$

Функция $L(x_n; \theta)$ равна нулю при $\theta \in (0; x_{(n)})$, делает в точке $\theta = x_{(n)}$ скачок величины $x_{(n)}^{-n}$, а затем монотонно убывает. Формально о. м. п. не существует, однако если переопределить плотность распределения в исходной модели, положив $f(x; \theta) = \theta^{-1}$ при $0 < x \leq \theta$, то фактически, не изменив модели, получим о. м. п. $\hat{\theta}(x_n) = x_{(n)}$.

(IV) Для повторной выборки из экспоненциального распределения с плотностью

$$\theta_2^{-1} \exp(-(x - \theta_1)\theta_2^{-1}), \quad x \geq \theta_1, \quad -\infty < \theta_1 < +\infty, \quad \theta_2 > 0,$$

получаем

$$l(x_n; \theta_1, \theta_2) = -n \ln \theta_2 - \sum_{i=1}^n (x_i - \theta_1) \theta_2^{-1}.$$

Очевидно, при любом θ_2 максимум $l(x_n; \theta_1, \theta_2)$ по θ_1 достигается при

$$\hat{\theta}_1 = \min_{1 \leq i \leq n} x_i = x_{(1)}.$$

Далее, из уравнения

$$0 = \partial l(x_n; \hat{\theta}_1, \theta_2) / \partial \theta_2 = -n/\theta_2 + \sum_{i=1}^n (x_i - \hat{\theta}_1) \theta_2^{-2}$$

находим

$$\hat{\theta}_2 = \sum_{i=1}^n (x_i - x_{(1)})/n.$$

Напомним, что полной достаточной статистикой в рассматриваемой модели является $(x_{(1)}, \sum_{i=1}^n (x_i - x_{(1)}))$ (см. пример (III) § 17), и потому несмешанные оценки (см. п. 3 § 6)

$$\theta_2^* = \frac{n}{n-1} \hat{\theta}_2, \quad \theta_1^* = \hat{\theta}_1 - \theta_2^*/n$$

имеют минимальную дисперсию. Как видно, отличие между $(\hat{\theta}_1, \hat{\theta}_2)$ и (θ_1^*, θ_2^*) незначительно при больших n . ■

Рассмотрим повторную выборку из семейства экспоненциального типа с плотностью

$$f(x; \theta) = h(x) \exp\left(\sum_{i=1}^n \theta_i T_i(x) + b(\theta)\right), \quad \theta = (\theta_1, \dots, \theta_k).$$

Логарифмическая функция правдоподобия выборки равна

$$l(x_n; \theta) = \sum_{i=1}^n \ln h(x_i) + \sum_{i=1}^k \theta_i \sum_{j=1}^n T_i(x_j) + nb(\theta).$$

Дифференцируя по θ_i , $i = 1, \dots, k$, приходим к системе уравнений

$$0 = \partial l(x_n; \theta) / \partial \theta_i = \sum_{j=1}^n T_i(x_j) + n \partial b(\theta) / \partial \theta_i, \quad i = 1, \dots, k. \quad (4)$$

Условием однозначной разрешимости системы (4) является невырожденность матрицы

$$\left[\frac{\partial^2 l(x_n; \theta)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, k \right] = \left[n \frac{\partial^2 b(\theta)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, k \right]. \quad (5)$$

Чтобы решение системы (4) было точкой максимума функции правдоподобия, достаточно, как известно, положительной определенности матрицы

$$[-\partial^2 l(x_n; \theta) / \partial \theta_i \partial \theta_j], \quad i, j = 1, \dots, k. \quad (6)$$

При условиях регулярности, которые здесь выполнены, имеют место соотношения (ср. (46)–(48) § 18)

$$-M_\theta \frac{\partial^2 l(x_n; \theta)}{\partial \theta_i \partial \theta_j} = M_\theta \frac{\partial l(x_n; \theta)}{\partial \theta_i} \frac{\partial l(x_n; \theta)}{\partial \theta_j}, \quad i, j = 1, \dots, k.$$

Так как матрица (6) не зависит от x_n (см. (5)), то отсюда следует, что (6) есть матрица Фишера $I_{x_n}(\theta)$ и потому неотрицательно определена. Итак, при условии невырожденности матрицы $[\partial^2 b(\theta) / \partial \theta_i \partial \theta_j, i, j = 1, \dots, k]$ о. м. п. находится как единственное решение системы (4). Отметим, что поскольку статистика

$$S(x_n) = (S_1(x_n), \dots, S_k(x_n)), \quad S_i(x_n) = \frac{1}{n} \sum_{j=1}^n T_i(x_j), \quad (7)$$

является достаточной, то статистика $\hat{\theta}(x_n)$, эквивалентная $S(x_n)$, является достаточной.

Напомним, что неравенство Фреше–Рао–Крамера обращается в равенство для оценки (7). При этом из соотношения

$$M_\theta \operatorname{grad}_\theta \ln f(x_n; \theta) = 0$$

и из (4) вытекает (см. также следствие 1 теоремы 2 § 19)

$$M_{\theta} S_i(X_n) = -\partial b(\theta)/\partial \theta_i, \quad i = 1, \dots, k.$$

С другой стороны, подставляя в (4) $\theta = \hat{\theta}(x_n)$, находим

$$M_{\theta} S_i(X_n) = -M_{\theta} \left(\frac{\partial b(\theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}(x_n)} \right), \quad i = 1, \dots, k,$$

так что

$$M_{\theta} \partial b(\hat{\theta}(X_n))/\partial \theta_i = \partial b(\theta)/\partial \theta_i, \quad i = 1, \dots, k.$$

Таким образом, о. м. п. $\text{grad}_{\theta} b(\hat{\theta})$ параметрической функции $\text{grad}_{\theta} b(\theta)$ является несмешеною, а ее ковариационная матрица принимает минимально возможное значение:

$$(nI_X(\theta))^{-1} = [-n\partial^2 b/\partial \theta_i \partial \theta_j, \quad i, j = 1, \dots, n]^{-1}.$$

Из теоремы 2 § 19 (см. следствие) и проведенных выше рассуждений вытекает следующее утверждение: если соблюдаются условия регулярности и существует несмешенная оценка, обращающая неравенство Фреше—Рао—Крамера в равенство, то эта оценка является о. м. п.

(V) Пусть (X_i, Y_i) , $i = 1, \dots, n$ — повторная выборка из двумерного нормального распределения $N_2(0, Q)$ (см. пример п. 1 § 12):

$$Q = \begin{pmatrix} \sigma^2 & \sigma^2 \rho \\ \sigma^2 \rho & \sigma^2 \end{pmatrix}.$$

Вводя параметризацию

$$\varphi_1 = -\frac{1}{2\sigma^2(1-\rho^2)}, \quad \varphi_2 = \frac{\rho}{\sigma^2(1-\rho^2)}. \quad (8)$$

приведем плотность распределения отдельного наблюдения к виду

$$f(x, y) = (2\pi)^{-1} \exp(\varphi_1 T_1(x, y) + \varphi_2 T_2(x, y) + b(\varphi_1, \varphi_2)),$$

где

$$T_1(x, y) = x^2 + y^2, \quad T_2(x, y) = xy, \quad b(\varphi_1, \varphi_2) = 1/2 \ln(4\varphi_1^2 - \varphi_2^2).$$

Запишем систему уравнений (4):

$$\frac{1}{n} \sum_{i=1}^n T_1(x_i) = -\frac{\partial b(\varphi_1, \varphi_2)}{\partial \varphi_1} \equiv \frac{4\varphi_1}{4\varphi_1^2 - \varphi_2^2}, \quad (9)$$

$$\frac{1}{n} \sum_{i=1}^n T_2(x_i) = -\frac{\partial b(\varphi_1, \varphi_2)}{\partial \varphi_2} \equiv \frac{\varphi_2}{4\varphi_1^2 - \varphi_2^2}.$$

Разрешая систему (8), находим

$$\rho = -\varphi_2 / 2\varphi_1, \quad \sigma^2 = -2\varphi_1(4\varphi_1^2 - \varphi_2^2). \quad (10)$$

Из (9) и (10) получаем о. м. п.:

$$\hat{\rho} = 2 \sum_{i=1}^n x_i y_i / \sum_{i=1}^n (x_i^2 + y_i^2), \quad \hat{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^n (x_i^2 + y_i^2). \quad \blacksquare$$

В гл. V будет показано, что оценки максимума правдоподобия обладают рядом оптимальных свойств для выборок большого объема ($n \rightarrow \infty$). В случае умеренных и небольших n свойства о. м. п. изучались лишь для частных моделей. Затруднение заключается в том, что о. м. п. в общем случае не выписываются в явном виде и для их нахождения приходится применять численные методы.

(VI) Рассмотрим цензурированную выборку из распределения $N(\mu, \sigma^2)$, полагая, что наблюдению доступны порядковые статистики $x_{(1)}, \dots, x_{(N)}$, $N = n - r$, $r > 0$. Логарифмическая функция правдоподобия при $x_{(1)} \leq \dots \leq x_{(N)}$ имеет вид (см. пример (VI) § 16)

$$l(x; \mu, \sigma) = c + r \ln(1 - \Phi((x_{(N)} - \mu)/\sigma)) - N \ln \sigma - \sum_{i=1}^N (x_{(i)} - \mu)^2 / (2\sigma^2), \quad x = (x_{(1)}, \dots, x_{(N)}).$$

Здесь $\Phi(x)$ — ф.р. нормального закона $N(0, 1)$, $\varphi(x) = \Phi'(x)$. Положим

$$\xi = (x_{(N)} - \mu)/\sigma, \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_{(i)}, \quad s_0^2 = \frac{1}{N} \sum_{i=1}^N (x_{(i)} - \bar{x})^2$$

и выпишем систему уравнений правдоподобия

$$0 = \frac{\partial l(x; \mu, \sigma)}{\partial \mu} = -\frac{r\varphi(\xi)}{\sigma(1 - \Phi(\xi))} + \frac{1}{\sigma^2} \sum_{i=1}^N (x_{(i)} - \mu), \quad (11)$$

$$0 = \frac{\partial l(x; \mu, \sigma)}{\partial \sigma} = -\frac{r\xi\varphi(\xi)}{\sigma(1 - \Phi(\xi))} - \frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^N (x_{(i)} - \mu)^2.$$

Введем функцию

$$B(y) = B(h, y) = \frac{h}{1-h} \frac{\varphi(y)}{1 - \Phi(y)}, \quad h = r/n,$$

и перепишем систему (11) в виде

$$\bar{x} - \mu = \sigma B(\xi), \quad (12)$$

$$s_0^2 + (\bar{x} - \mu)^2 = \sigma^2(1 + \xi B(\xi)). \quad (13)$$

Подставляя уравнение (12) в (13), получим

$$s_0^2 + \sigma^2 B(\xi)^2 = \sigma^2 (1 + \xi B(\xi)). \quad (14)$$

С другой стороны, воспользовавшись уравнением (12), имеем

$$\sigma \xi = x_{(N)} - \mu = x_{(N)} - \bar{x} + (\bar{x} - \mu) = x_{(N)} - \bar{x} + \sigma B(\xi),$$

откуда

$$\sigma = (x_{(N)} - \bar{x}) / (\xi - B(\xi)). \quad (15)$$

Подставляя (15) в (14), получим уравнение для нахождения ξ :

$$s_0^2 / (x_{(N)} - \bar{x})^2 = (1 - B(\xi)^2 - \xi B(\xi)) / (\xi - B(\xi))^2. \quad (16)$$

Уравнение (16) решается численно. Подставляя его решение ξ в (15), найдем σ , а подставляя $\widehat{\xi}$ и $\widehat{\sigma}$ в (12), найдем $\widehat{\mu}$. Получим для μ , σ более удобные для применения формулы. Переставляя члены выражения (14), используя (15) и полагая

$$\lambda(\xi) = B(\xi) / (B(\xi) - \xi), \quad (17)$$

получаем

$$\sigma^2 = s_0^2 + \sigma^2 B(\xi) (B(\xi) - \xi) = s_0^2 + \lambda(\xi) (x_{(N)} - \bar{x})^2.$$

Подставляя (15) в (12), находим

$$\mu = \bar{x} + \sigma B(\xi) = \bar{x} + \lambda(\xi) (x_{(N)} - \bar{x}).$$

Таким образом, о. м. п. представляются формулами

$$\widehat{\mu} = \bar{x} + \lambda(\widehat{\xi}) (x_{(N)} - \bar{x}), \quad \widehat{\sigma}^2 = s_0^2 + \lambda(\widehat{\xi}) (x_{(N)} - \bar{x})^2, \quad (18)$$

где $\widehat{\xi}$ определяется из уравнения (16).

Чтобы иметь практическую возможность пользоваться оценками (18), достаточно рассчитать для разных $h=r/n$ таблицу решений уравнения (16):

$$(1 - B(h, \xi)^2 - \xi B(h, \xi)) / (\xi - B(h, \xi)) = \gamma \quad (19)$$

для значений γ , взятых с некоторым шагом. Подставляя решение $\xi = \xi(h, \gamma)$ в формулу (17), получим таблицу значений $\lambda = \lambda(h, \gamma)$. Такая таблица, рассчитанная Коэном, приведена в [9, с. 147—148].

Проиллюстрируем изложенную методику на примере (I) § 11, где $n=10$; $r=3$; $N=7$; $h=0,3$; $\bar{x}=1,70471$; $s_0^2=0,003514$; $x_{(N)}=1,778$. Тогда $\gamma=s_0^2/(x_{(N)}-\bar{x})^2=0,654$. Из таблицы, приведенной в [9, с. 148], интерполяцией по γ находим $\lambda=0,512$. Подставляя в (18), получаем

$$\widehat{\mu}=1,742; \quad \widehat{\sigma}=0,079.$$

Всестороннее обсуждение этого примера содержится на с. 150—151 книги [9].

3. Метод Монте-Карло в модели сдвига-масштаба.

Рассмотрим статистическую модель, задаваемую n -мерной плотностью вида

$$f(x; \mu, \sigma) = \sigma^{-n} f(\sigma^{-1}(x - \mu \cdot 1)), \quad 1 = (1, \dots, 1), \quad (20)$$

где $f(x) = f(x; 0, 1)$ — некоторая известная плотность, μ, σ — неизвестные параметры. Модель (20) охватывает важный для приложений случай цензурированной выборки, когда наблюдению доступны порядковые статистики

$$(x_{(r_1+1)}, x_{(r_1+2)}, \dots, x_{(n-r_1)}) \equiv (y_1, \dots, y_N) \equiv y$$

и плотность наблюдений равна (ср. пример (VI) § 16)

$$\begin{aligned} f_Y(y; \mu, \sigma) &= \\ &= \frac{n!}{r_1! r_2!} F((y_1 + \mu)/\sigma)^{r_1} (1 - F((y_N - \mu)/\sigma))^{r_2} \prod_{i=1}^N \sigma^{-1} f((y_i - \mu)/\sigma), \end{aligned}$$

где $F(y)$ — ф. р. плотности $f(y)$. Это так называемый II тип цензурирования; модель (20) включает и I тип цензурирования, когда наблюдаются значения $x_{(i)}$, попавшие в заранее известный интервал (t, s) (см. [9, § 6.3]).

Покажем, как, используя искусственное моделирование выборки с плотностью $f(x)$, можно строить доверительные оценки для параметров μ и σ . Мы будем исходить из о. м. п. $\hat{\mu}(x)$, $\hat{\sigma}(x)$, хотя метод применим к любым оценкам $\hat{\mu}, \hat{\sigma}$, для которых распределение сл. в.

$$\sigma^{-1}(\hat{\mu}(X) - \mu), \quad \sigma^{-1}\hat{\sigma}(X), \quad (21)$$

где X имеет плотность распределения (20), не зависит от параметров μ и σ . Покажем, что совместное распределение сл. в. (21) не зависит от параметров μ, σ , когда $\hat{\mu}(x), \hat{\sigma}(x)$ — о. м. п. Фиксируем значения μ_0, σ_0 параметров μ, σ , и пусть $X^{(0)}$ — случайная выборка с плотностью $f(x; \mu_0, \sigma_0)$, $Y = \sigma_0^{-1}(X^{(0)} - \mu_0 \cdot 1)$ — стандартный случайный вектор с плотностью $f(y; 0, 1) = f(y)$. Очевидно,

$$\begin{aligned} L(X^{(0)}; \mu, \sigma) &= \\ &= L(\sigma_0 Y + \mu_0 \cdot 1; \mu, \sigma) = \sigma^{-n} f(\sigma^{-1}(\sigma_0 Y + \mu_0 \cdot 1 - \mu \cdot 1)) = \\ &= \sigma_0^{-n} (\sigma/\sigma_0)^{-n} f((\sigma/\sigma_0)^{-1}(Y - (\mu - \mu_0)/\sigma_0 \cdot 1)) = \\ &= \sigma_0^{-n} L(Y; (\mu - \mu_0)/\sigma_0, \sigma/\sigma_0). \end{aligned} \quad (22)$$

Пусть $\hat{\mu}(X^{(0)})$ и $\hat{\sigma}(X^{(0)})$ — о. м. п. Как видно из (22), точку максимума $\hat{\mu}(X^{(0)}), \hat{\sigma}(X^{(0)})$ левой части выражения (22) можно связать

с точкой максимума функции $L(Y; \mu', \sigma')$, которая равна $\mu' = \hat{\mu}(Y)$, $\sigma' = \hat{\sigma}(Y)$. Именно

$$\sigma_0^{-1}(\hat{\mu}(X^{(0)} - \mu_0)) = \hat{\mu}(Y), \quad \sigma_0^{-1}\hat{\sigma}(X^{(0)}) = \hat{\sigma}(Y). \quad (23)$$

Итак, совместное распределение сл. в. (21) совпадает с распределением сл. в. $\hat{\mu}(Y)$, $\hat{\sigma}(Y)$, что и требовалось установить.

Для приближенного численного расчета совместного распределения вероятностей

$$H(u, v) = \mathcal{P}(\hat{\mu}(Y) < u, \hat{\sigma}(Y) < v)$$

с помощью метода статистических испытаний необходимо иметь, во-первых, алгоритм вычисления функций $\hat{\mu}(x)$, $\hat{\sigma}(x)$:

$$L(x; \hat{\mu}(x), \hat{\sigma}(x)) = \max_{\mu, \sigma} L(x; \mu, \sigma)$$

и, во-вторых, устройство, моделирующее последовательность независимых испытаний:

$$Y^{(1)}, Y^{(2)}, \dots, Y^{(N)}, \quad (24)$$

с плотностью $f_{Y^{(i)}}(y) = f(y; 0, 1)$, $i = 1, \dots, N$. По (24) строим последовательность N независимых испытаний пар сл. в.:

$$(\hat{\mu}(Y^{(i)}), \hat{\sigma}(Y^{(i)})), \quad i = 1, \dots, N, \quad (25)$$

с ф. р. $H(u, v)$, откуда рассчитываем э. ф. р. $H_N(u, v)$ и полагаем $H(u, v) \equiv H_N(u, v)$.

В результате получаем, используя (23):

$$P_{\mu, \sigma}(x : \sigma^{-1}\hat{\sigma}(x) < u) = H_1(u) \equiv H(u, +\infty),$$

$$\begin{aligned} P_{\mu, \sigma}(x : \sigma(x)^{-1}(\hat{\mu}(x) - \mu) < v) &= P_{\mu, \sigma}(x : (\hat{\mu}(x) - \mu)\sigma^{-1}\hat{\sigma}(x)^{-1}\sigma < v) = \\ &= P_{0, 1}(x : \hat{\mu}(x)\hat{\sigma}(x)^{-1} < v) = G(v), \end{aligned}$$

где

$$G(v) = \iint_{\hat{\sigma}(x)^{-1} < v} h(x, y) dx dy; \quad h(x, y) = \partial^2 H(x, y) / \partial x \partial y.$$

Таким образом, α -доверительные интервалы для σ и μ имеют вид

$$\hat{\sigma}(x)/(x_{1-\alpha/2}(H_1))^{-1} \leq \sigma \leq \hat{\sigma}(x)/(x_{\alpha/2}(H_1))^{-1}, \quad (26)$$

$$\hat{\mu}(x) - \hat{\sigma}(x)x_{1-\alpha/2}(G) \leq \mu \leq \hat{\mu}(x) - \hat{\sigma}(x)x_{\alpha/2}(G), \quad (27)$$

где $x_\alpha(H_1)$ и $x_\alpha(G)$ обозначают α -квантили соответствующих распределений.

Отметим, что знание совместного распределения оценок $\hat{\mu}$, $\hat{\sigma}$ позволяет вычислять доверительные интервалы для любой параметрической функции $\psi(\mu, \sigma)$. В частности, предположим, что

речь идет о повторной (цензурированной) выборке с одномерной ф. р. $F(\sigma^{-1}(x-\mu))$, где $F(x)$ — известная ф. р. Покажем, как строится доверительный интервал для

$$\psi_t(\mu, \sigma) = F(\sigma^{-1}(t-\mu)),$$

где t — любое фиксированное. Ввиду монотонности функции F достаточно построить α -доверительную границу для $\psi_t(\mu, \sigma) = \sigma^{-1}(t-\mu)$, так как

$$a \leq \psi_t(\mu, \sigma) \leq b \Leftrightarrow F(a) \leq \psi_t(\mu, \sigma) \leq F(b).$$

Для краткости рассмотрим только верхнюю α -доверительную границу. Имеем, используя (23):

$$\begin{aligned} P_{\mu, \sigma}(x : \hat{\sigma}(x)^{-1}(t - \hat{\mu}(x)) \leq s) &= \\ = P_{\mu, \sigma}\left(x : \frac{t - \mu}{\sigma} \frac{\sigma}{\hat{\sigma}(x)} - \frac{\hat{\mu}(x) - \mu}{\sigma} \frac{\sigma}{\hat{\sigma}(x)} \leq s\right) &= \\ = P_{0,1}(x : \sigma^{-1}(t - \mu) \hat{\sigma}(x)^{-1} - \hat{\mu}(x) \hat{\sigma}(x)^{-1} \leq s). \end{aligned} \quad (28)$$

Положим

$$P_{0,1}(x : \tau \hat{\sigma}(x)^{-1} - \hat{\mu}(x) \hat{\sigma}(x)^{-1} \leq s) = Q_\tau(s). \quad (29)$$

По результатам моделирования (25) строим э. ф. р. $Q_{t,N}(s)$, так что $Q_\tau(s) \approx Q_{t,N}(s)$.

С помощью ф. р. $Q_\tau(s)$ находятся доверительные границы для $\psi_t(\mu, \sigma) = \sigma^{-1}(t - \mu)$ при любом t . Отступая от принятых обозначений, α -квантиль распределения Q_τ обозначим через $x_{1-\alpha}(\tau)$:

$$P_{0,1}(x : \tau \hat{\sigma}(x)^{-1} - \hat{\mu}(x) \hat{\sigma}(x)^{-1} \leq x_{1-\alpha}(\tau)) = 1 - \alpha. \quad (30)$$

Из (29) вытекает, что функция $Q_\tau(s)$ убывает по τ и, следовательно, функция $y = x_{1-\alpha}(\tau)$ возрастает по τ . Предположим для краткости, что $Q_\tau(s)$ непрерывна и строго монотонна по τ , так что функция $y = x_{1-\alpha}(\tau)$ имеет обратную $\tau = x_{1-\alpha}^{-1}(y)$. Из соотношений (28) и (30) получаем

$$\begin{aligned} 1 - \alpha &= P_{\mu, \sigma}(x : \hat{\sigma}(x)^{-1}(t - \hat{\mu}(x)) \leq x_{1-\alpha}(\sigma^{-1}(t - \mu))) = \\ &= P_{\mu, \sigma}(x : x_{1-\alpha}^{-1}(\hat{\sigma}(x)^{-1}(t - \hat{\mu}(x))) \leq \sigma^{-1}(t - \mu)). \end{aligned}$$

Таким образом, α -доверительная нижняя граница для $\sigma^{-1}(t - \mu)$ равна

$$x_{1-\alpha}^{-1}(\hat{\sigma}(x)^{-1}(t - \hat{\mu}(x))),$$

а для $F(\sigma^{-1}(t - \mu))$ она равна

$$F(x_{1-\alpha}^{-1}(\hat{\sigma}(x)^{-1}(t - \hat{\mu}(x)))).$$

Проиллюстрируем эту методику на примере повторной выборки из распределения Вейбулла с ф. р.:

$$F_{\beta, \gamma}(x) = 1 - \exp(-x/\beta)^{\gamma}, \quad x > 0.$$

Преобразованием $Y_i = \ln X_i$ выборка X_1, \dots, X_n из распределения $F_{\beta, \gamma}(x)$ переводится в выборку Y_1, \dots, Y_n из распределения

$$F(\sigma^{-1}(y - \mu)), \quad \mu = \ln \beta, \quad \sigma = \gamma^{-1}, \quad F(y) = 1 - \exp(-\exp y)$$

(см. п. 1 § 4). Томан, Бэйн, Антл^{*)} рассчитали таблицы для построения доверительных интервалов для различных n и нескольких вариантов цензурирования. Воспользуемся их примером для

числовых данных по испытанию до разрушения шарикоподшипников. Ниже приводятся результаты $x_{(i)}$ испытаний в миллионах оборотов $n=23$ шарикоподшипников (взятых ими из работы Либлейна, Целена):

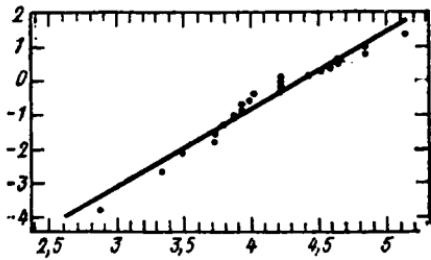


Рис. 12. Точки $(y_{(i)}, F^{-1}((i-0.5)/23))$, $F(y) = 1 - \exp(-\exp y)$ и прилегающая к ним прямая; $y_{(i)}$ — логарифм числа оборотов (в миллионах) до разрушения выборки из 23 шарикоподшипников

17,88	28,92	33,00
48,48	51,84	51,96
68,64	68,64	68,88
105,12	105,84	127,92
41,52	42,12	45,60
54,12	55,56	67,80
84,12	93,12	98,64
128,04	173,40	

На рис. 12 нанесены точки $(y_{(i)}, F^{-1}((i-0.5)/23))$, $y_{(i)} = \ln x_{(i)}$, $i=1, \dots, 23$, $F(y) = 1 - \exp(-\exp y)$ и проведена на глаз прямая

$$x = \tilde{\sigma}y + \tilde{\mu}, \quad (31)$$

где $\tilde{\mu} = 4,38$; $\tilde{\sigma} = 0,43$.

Согласие точек $(y_{(i)}, F^{-1}((i-0.5)/23))$ с линейной зависимостью можно считать удовлетворительным. Соответствующие оценки β и γ равны:

$$\tilde{\beta} = \exp(\tilde{\mu}) = 79,8; \quad \tilde{\gamma} = \tilde{\sigma}^{-1} = 2,3.$$

Систему уравнений правдоподобия после некоторых преобразований можно записать в виде

$$n/\hat{\gamma} - n \sum_{i=1}^n x_i^{\hat{\gamma}} \ln x_i / \sum_{i=1}^n x_i^{\hat{\gamma}} + \sum_{i=1}^n \ln x_i = 0,$$

$$\hat{\beta} = \left(\sum_{i=1}^n x_i^{\hat{\gamma}} / n \right)^{1/\hat{\gamma}}$$

^{*)} Technometrics, 1969, т. 11, с. 445—460; 1970, т. 12, с. 363—371. Авторы не использовали описанного нами общего подхода, а опирались на частные свойства распределения Вейбулла.

Численное решение приводит к значениям

$$\hat{\gamma} = 2,102; \hat{\beta} = 81,99.$$

Выпишем несколько значений квантилей x_α распределения сл. в. $\hat{\gamma}/\gamma = (\hat{\sigma}/\sigma)^{-1}$, рассчитанных Томаном и др. методом Монте-Карло:

$n \backslash \alpha$	0,02	0,05	0,10	0,90	0,95	0,98
22	0,752	0,798	0,843	1,320	1,418	1,583
24	0,759	0,805	0,848	1,301	1,392	1,504

Линейной интерполяцией по n находим α -доверительные границы для γ при $\alpha=0,1$:

$$1,50 = \hat{\gamma}/x_{1-\alpha/2} < \gamma < \hat{\gamma}/x_{\alpha/2} = 2,62.$$

Выпишем несколько значений квантилей x_α распределения сл. в.

$$\hat{\gamma} \ln(\hat{\beta}/\beta) = (\hat{\mu} - \mu)/\hat{\sigma},$$

рассчитанных Томаном и др.:

$n \backslash \alpha$	0,02	0,05	0,10	0,90	0,95	0,98
22	-0,509	-0,404	-0,314	0,302	0,398	0,519
24	-0,483	-0,384	-0,299	0,288	0,379	0,494

Линейной интерполяцией по n находим α -доверительные границы для β при $\alpha=0,1$:

$$68,04 = \hat{\beta} \exp(-x_{1-\alpha/2}/\hat{\gamma}) < \beta < \hat{\beta} \exp(-x_{\alpha/2}/\hat{\gamma}) = 98,76.$$

§ 21. КРИТЕРИЙ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ

i. Проверка статистических гипотез.

В гл. I, II нам неоднократно встречалась форма статистического вывода, называемая *проверкой гипотез*. Напомним некоторые примеры. При сравнении средних двух независимых нормальных выборок из $N(\mu_1, \sigma^2)$ и $N(\mu_2, \sigma^2)$ рассматривалась гипотеза $H_0: \mu_1 = \mu_2$. В линейной статистической модели по выборке из $N_n(\eta, \sigma^2 I)$, где вектор η принадлежит некоторому линейному подпространству V , проверялась гипотеза $H_0: \eta \in V'$, где V' — некоторое заданное линейное подпространство V . Наконец, когда имеется α -доверительный интервал для некоторой параметрической функции $\varphi = \varphi(0)$ статистической модели $(\mathcal{X}, \mathfrak{B}, \{P_\theta, \theta \in \Theta\})$,

возникает возможность проверить гипотезу $H_0: \phi = \phi_0$, где ϕ_0 — любое фиксированное значение параметрической функции. Во всех этих примерах имеется статистика $T(x)$, называемая *статистикой критерия*, такая, что область отверждения гипотезы принимает вид

$$W_\alpha = \{x : T(x) > c_\alpha\}$$

или

$$W_\alpha = \{x : T(x) < c'_\alpha, \text{ либо } T(x) > c''_\alpha\}.$$

Сама гипотеза H_0 могла быть представлена в форме

$$H_0: P \in \{P\}_0 \subseteq \{P\}, \quad (1)$$

где $\{P\}$ — семейство распределений вероятностей рассматриваемой статистической модели, а $\{P\}_0$ — некоторое его подмножество. Во всех встречающихся нам задачах оказывалось, что вероятность

$$P(W_\alpha) = \alpha, \quad P \in \{P\}_0, \quad (2)$$

не зависит от P при условии справедливости гипотезы H_0 , т. е. для всех $P \in \{P\}_0$. Вообще говоря, для проверки одной и той же гипотезы H_0 можно использовать различные статистики критерия. В приложениях бывает иногда ясно, какие возможные отклонения от гипотезы H_0 необходимо проверить и, следовательно, на какой статистике критерия остановить свой выбор. Но даже и в этом случае весьма интересно оценить свойства выбранного критерия по сравнению с другими.

В общем случае нет необходимости непременно связывать понятие критерия со статистикой. Статистическим критерием для проверки гипотезы (1) назовем любую систему подмножеств $\{W_\alpha\}$ множества \mathcal{X} , индексированную числами α из интервала $(0, 1)$. Если при этом выполнено соотношение (2), то α называется *размером критической области* W_α . Потребуем дополнительно, чтобы

$$W_{\alpha_1} \subseteq W_{\alpha_2}, \quad \alpha_1 < \alpha_2.$$

Будем говорить, что данный критерий *отвергает гипотезу H_0 на уровне значимости α* , если результат наблюдений x принадлежит критической области W_α . Степень согласия данных с гипотезой удобно характеризовать значением

$$\alpha(x) = \inf\{\alpha : x \in W_\alpha\},$$

так что x отвергается на уровне значимости $\alpha(x) + \epsilon$ и принимается на уровне $\alpha(x) - \epsilon$, каково бы ни было $0 < \epsilon < \alpha(x)$. Чаще, однако, используют набор стандартных значений $\alpha: 0, 1; 0, 05; 0, 01$ и др., и степень согласия наблюдения x с гипотезой H_0 -характеризуют наименьшим α из этого набора, для которого $x \in W_\alpha$. Для одного и того же наблюдения x при одних уровнях значимости критерий отвергает гипотезу H_0 , а при других — принимает. Поэтому следует подчеркнуть, что теория проверки гипотез не рассчитана на такое ее применение, когда по результатам наблюдения

ний гипотеза H_0 будет на самом деле принята или отвергнута. Применение статистического критерия к результатам наблюдений есть форма извлечения из выборки полезной информации, дальнейшее использование которой зависит от наблюдателя.

Если сравниваются два критерия $\{W_\alpha\}$ и $\{W'_{\alpha'}\}$, то естественно сопоставить критические области одного и того же размера. Применение любой из критических областей W_α и $W'_{\alpha'}$ в предположении, что для каждой из них имеет место (2), приводит к одной и той же вероятности так называемой ошибки первого рода — отвергнуть гипотезу H_0 , если на самом деле она верна. Говорят, что совершена ошибка второго рода, если гипотеза H_0 принята, когда она неверна. Чтобы уточнить математическую постановку вопроса, введем непересекающееся с $\{P\}_0$ семейство распределений $\{P\}_1 \subseteq \{P\}$ и назовем гипотезу $H_1 : P \in \{P\}_1$ альтернативной к гипотезе H_0 из (1) (гипотезу H_0 назовем нулевой). Будем оценивать качество критерия $\{W_\alpha\}$ по проверке гипотезы H_0 против альтернативы H_1 набором вероятностей

$$\beta = P(W_\alpha), \quad P \in \{P\}_1. \quad (3)$$

При фиксированном α вероятность β рассматривают как функцию от P и называют функцией мощности критерия. Значение $1 - \beta$ при любом фиксированном P_1 равно вероятности ошибки следующего вида: гипотеза H_0 принята, хотя она неверна, а выборка на самом деле извлечена из распределения P_1 . Если

$$P(W'_\alpha) \leq P(W_\alpha), \quad P \in \{P\}_1 \quad (4)$$

и хотя бы при одном P неравенство (4) является строгим, то критическую область W_α размера α следует предпочесть области того же размера $W'_{\alpha'}$, поскольку критическая область W_α обеспечивает меньшую вероятность ошибки второго рода. Неравенство (4) задает отношение (частичного) порядка в множестве критериев, и тем самым можно поставить вопрос о существовании оптимального критерия, мощность которого превосходит мощности всех других критериев.

Поскольку функция мощности (3) зависит еще от параметра α , то сравнение критериев (4) было бы целесообразно проводить при всех α . В теории это, однако, привело бы к излишним усложнениям, и потому принято изучать и сравнивать критические области, а не критерии, тем более что в приложениях критерии задаются обычно статистиками и наличие неравенства (4) при каком-либо одном α влечет, как правило, такое же неравенство и при других уровнях значимости.

В связи со сказанным в теории принято критическую область размера α называть также критерием размера α или даже просто критерием. Индикаторную функцию области W_α :

$$\varphi(x) = 1, \text{ если } x \in W_\alpha, \text{ и } \varphi(x) = 0, \text{ если } x \notin W_\alpha. \quad (5)$$

называют критической функцией критерия W_α . Из контекста всег-

да будет ясно, идет ли речь о критерии как системе вложенных критических областей либо как об одной такой области.

Параметрические модели, как и в задаче оценивания, играют первостепенную роль в теории проверки статистических гипотез. Пусть

$$\{P_\theta\}_0 = \{P_\theta : \theta \in \Theta_0\}, \quad \{P_\theta\}_1 = \{P_\theta : \theta \in \Theta_1\}, \quad \Theta_0 \cap \Theta_1 = \emptyset,$$

тогда гипотезы о распределениях становятся гипотезами о параметрах

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_1,$$

а функция мощности $\beta(\theta) = P_\theta(W_\alpha)$ рассматривается как функция переменного θ . Рассмотрим примеры.

(I) Для повторной выборки из экспоненциального распределения с плотностью $f(x) = \exp(-(x-\theta))$, $x > \theta$, а-доверительный интервал для параметра θ из п. 4 § 3 имеет вид

$$x_{(1)} + n^{-1} \ln(1-a) \leq \theta \leq x_{(1)}.$$

Критерий W_α размера a для проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta \neq \theta_0$ выбираем в виде

$$W_\alpha = \{x_n : \theta_0 > x_{(1)}, \text{ либо } \theta_0 < x_{(1)} + n^{-1} \ln(1-a)\}.$$

Отметим, что включение в W_α точек x_n , для которых $x_{(1)} < \theta_0$, не влияет на размер критерия, так как подмножество $\{x_n : x_{(1)} \leq \theta_0\}$ имеет P_{θ_0} -меру нуль. Функция мощности критерия W_α :

$$\begin{aligned} \beta(\theta) &= P_\theta(x_n : x_{(1)} < \theta_0, \text{ либо } x_{(1)} > \theta_0 - n^{-1} \ln(1-a)) = \\ &= \mathcal{P}_\theta(X_{(1)} < \theta_0) + \mathcal{P}_\theta(X_{(1)} > \theta_0 - n^{-1} \ln(1-a)) \end{aligned}$$

принимает значения

$$\beta(\theta) = \begin{cases} 1 - a e^{n(\theta - \theta_0)} & \text{при } \theta - \theta_0 < 0, \\ (1-a) e^{n(\theta - \theta_0)} & \text{при } 0 < \theta - \theta_0 < -n^{-1} \ln(1-a), \\ 1 & \text{при } \theta - \theta_0 > -n^{-1} \ln(1-a). \end{cases} \quad (6)$$

Функция мощности $\beta(\theta)$ принимает минимальное значение в точке θ_0 : $\beta(\theta_0) = 1-a$ и монотонно стремится к 1, если $\theta \rightarrow \pm\infty$. Если $n \rightarrow \infty$, то при любом $\theta \neq \theta_0$ $\beta(\theta) = \beta^{(n)}(0) \rightarrow 1$.

Рассмотрим другую альтернативу $H_1^+ : \theta > \theta_0$. В этом случае представляется естественным взять критерий исходя из нижней доверительной границы для θ :

$$W_{+\alpha}^+ = \{x_n : \theta_0 < x_{(1)} + n^{-1} \ln(1-a)\}.$$

Легко видеть, что функция мощности этого критерия дается последними двумя строками в (6), и, следовательно, мощность одностороннего критерия $W_{+\alpha}^+$ в области $\theta > \theta_0$ такая же, как и двустороннего W_α .

(II) Для повторной выборки из нормального распределения $N(\mu, 1)$ при проверке гипотезы $H_0: \mu = \mu_0$ против альтернативы $H_1: \mu > \mu_0$ будем исходить из нижней доверительной границы для μ (см. п. 1 § 8):

$$W_{\alpha}^+ = \{x_n : \mu_0 < \bar{x} - x_{1-\alpha}/\sqrt{n}\},$$

где x_{α} — α -квантиль распределения $N(0, 1)$. Мощность критерия W_{α}^+ равна

$$\begin{aligned} \beta^+(\mu) &= \beta^+(\mu; \alpha) = \\ &= \mathcal{P}_{\mu} \left(\frac{1}{n} \sum_{i=1}^n X_i > \mu_0 + \frac{x_{1-\alpha}}{\sqrt{n}} \right) = \\ &= \mathcal{P}_{\mu} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) > \right. \\ &\quad \left. > \sqrt{n} (\mu_0 - \mu) + x_{1-\alpha} \right) = \\ &= \mathcal{P}_0 \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i > \sqrt{n} (\mu_0 - \mu) + \right. \\ &\quad \left. + x_{1-\alpha} \right) = 1 - \Phi \left(\sqrt{n} (\mu_0 - \mu) + \right. \\ &\quad \left. + x_{1-\alpha} \right) = \Phi \left(\sqrt{n} (\mu - \mu_0 - x_{1-\alpha}/\sqrt{n}) \right) \end{aligned}$$

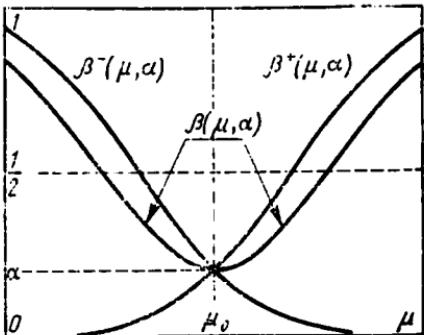


Рис. 13. Функции мощности одностороннего и двустороннего критериев отношения правдоподобий выборки из нормального распределения $N(\mu, 1)$

и представляет собой по μ ф. р. закона $N(\mu_0 + x_{1-\alpha}/\sqrt{n}, n^{-1})$ (рис. 13). Отметим, что $\beta^+(\mu) < \alpha$ при $\mu < \mu_0$ и стремится к 0 при $\mu \rightarrow -\infty$, так что критерий W_{α}^+ не годится против альтернативы $H_1^-: \mu < \mu_0$. Критическая область для проверки гипотезы H_0 против альтернативы H_1^- строится по верхней α -доверительной границе для μ :

$$W_{-\alpha}^- = \{x_n : \mu_0 > \bar{x} - x_{\alpha}/\sqrt{n}\}.$$

Функция мощности критерия $W_{-\alpha}^-$ получается из $\beta^+(\mu)$ симметричным отражением относительно точки μ_0 (см. рис. 13):

$$\beta^-(\mu) = \beta^-(\mu; \alpha) = \Phi \left(\sqrt{n} (-\mu + \mu_0 - x_{1-\alpha}/\sqrt{n}) \right).$$

Против двусторонней альтернативы $H_1: \mu \neq \mu_0$ критерий и функция мощности, основанные на симметричном α -доверительном интервале, имеют вид

$$\begin{aligned} W_{\alpha} &= W_{\alpha/2}^+ + W_{\alpha/2}^-, \quad \beta(\mu) = \beta(\mu; \alpha) = \mathcal{P}_{\mu}(W_{\alpha}) = \\ &= \mathcal{P}_{\mu}(W_{\alpha/2}^+) + \mathcal{P}_{\mu}(W_{\alpha/2}^-) = \beta^-(\mu; \alpha/2) + \beta^+(\mu; \alpha/2). \end{aligned}$$

Нетрудно проверить, что

$$\begin{aligned}\beta(\mu; \alpha) &< \beta^+(\mu; \alpha) \text{ при } \mu > \mu_0, \\ \beta(\mu; \alpha) &< \beta^-(\mu; \alpha) \text{ при } \mu < \mu_0,\end{aligned}\tag{7}$$

так что двусторонний критерий W_α менее мощен, чем $W_{-\alpha}^+$ или $W_{-\alpha}^-$, если его применять против соответствующих односторонних альтернатив (см. рис. 13). Зато критерии $W_{-\alpha}^+$, $W_{-\alpha}^-$, имея большую мощность при односторонних альтернативах, дают неудовлетворительное решение против альтернативы $H_1: \mu \neq \mu_0$. ■

Отметим, что в обоих примерах критические функции рассматриваемых критериев являются функциями минимальной достаточной статистики $T(\mathbf{x})$. В результате задача фактически сводится к проверке гипотезы о параметре распределения статистики $T(\mathbf{x})$ по результату единичного испытания. Например, критерий W_α из (II) можно преобразовать в критерий

$$\tilde{W}_\alpha = \{t : t < \mu_0 + x_\alpha / \sqrt{n}, \text{ либо } t > \mu_0 + x_{1-\alpha} / \sqrt{n}\}$$

для проверки гипотезы $H_0: \mu = \mu_0$ по результатам наблюдений t в модели $(R^1, \mathcal{B}, N(\mu, n^{-1}))$ против альтернативы $H_1: \mu \neq \mu_0$.

2. Лемма Неймана—Пирсона.

Рассмотрим статистическую модель вида $(\mathcal{X}, \mathcal{B}, \{P_0, P_1\})$, где семейство вероятностных мер состоит всего из двух элементов. Допустим, что

$$H_0: P = P_0, H_1: P = P_1, \tag{8}$$

т. е. при каждой из гипотез H_0 и H_1 распределение выборки полностью определено. Такие гипотезы называются *простыми* (в разобранном выше примерах (I), (II) нулевая гипотеза была простой, а альтернативная — не простой). Оказывается, что критерий максимальной мощности в случае (8) всегда существует. Поскольку функция мощности любого критерия W_α в постановке (8) сводится к двум вероятностям:

$$P_0(W_\alpha), \quad P_1(W_\alpha),$$

то задача поиска оптимальной критической области ставится следующим образом: среди всех подмножеств $W_\alpha \in \mathcal{B}$, для которых выполнено равенство

$$\alpha = P_0(W_\alpha), \quad 0 < \alpha < 1, \tag{9}$$

найти подмножество $W^{*\alpha}$, такое, что при всех $W_\alpha \in \mathcal{B}$

$$P_1(W^{*\alpha}) \geq P_1(W_\alpha). \tag{10}$$

Поясним, как строится это множество, а затем проведем формальное доказательство. Рассмотрим вначале следующую простую ситуацию: $\mathcal{B} = \{x_1, x_2, \dots, x_N\}$, $P_0(x_i) = N^{-1}$, $i = 1, \dots, N$. Заметим

прежде всего, что условие (9) выполнимо лишь при $a=a_k=k/N$; $k=1, 2, \dots, N-1$. Равенство

$$a_k = P_0(W_{a_k})$$

выполняется для любого подмножества W_{a_k} , состоящего ровно из k точек. Так как

$$P_1(W_{a_k}) = \sum_{x_i \in W_{a_k}} P_1(x_i),$$

то ясно, что, упорядочив вероятности

$$P_1(x_{i_1}) \geq P_1(x_{i_2}) \geq \dots \geq P_1(x_{i_N})$$

и положив

$$W_{a_k}^* = \{i_1, i_2, \dots, i_k\}. \quad (11)$$

мы придем к свойству (10).

Рассмотрим другой частный случай: множество \mathcal{X} то же самое, но предположим, что $P_0(x_i) = m_i \delta$, где $\delta > 0$, m_i — некоторые целые числа, т. е. набор вероятностей $P_0(x_i)$, $i=1, \dots, N$, состоит из соизмеримых чисел. Трансформируем эту модель, заменив каждый элемент $x_i \in \mathcal{X}$ на m_i элементов $x_i(1), \dots, x_i(m_i)$ и положив

$$\tilde{\mathcal{X}} = \{x_i(j), j=1, \dots, m_i, i=1, \dots, N\}, \quad P_0(x_i(j)) = \delta,$$

$$P_1(x_i(j)) = P_1(x_i)/m_i, \quad j=1, \dots, m_i, \quad i=1, \dots, N.$$

Ясно, что исходная статистическая модель получается из трансформированной объединением групп элементов и сложением соответствующих вероятностей. Оптимальное множество в новой модели получается аналогично (11) упорядочиванием пар (i, j) , при котором

$$P_1(x_{i_1}(j_1)) \geq P_1(x_{i_2}(j_2)) \geq \dots \quad (12)$$

Поскольку вероятности $P_1(x_i(j))$ равны между собой при всех j и любом фиксированном i , то упорядочивание (12) можно провести с соблюдением следующего свойства: вероятности $P_1(x_i(j))$, $j=1, 2, \dots, m_i$, расположены в (12) подряд. В таком случае всякое подмножество $W_a \subseteq \tilde{\mathcal{X}}$, включающее все пары $x_i(j)$ из ряда (12) вплоть до пары вида $x_i(m_i)$ при некотором i , с одной стороны, является оптимальным в новой модели, а с другой стороны, после объединения групп элементов, переводится в множество $W_a \subseteq \mathcal{X}$, являющееся оптимальным в исходной модели. Запишем отрезок из ряда (12):

$$\dots \geq P_1(x_i(1)) = \dots = P_1(x_i(m_i)) \geq P_1(x_k(1)) = P_1(x_k(m_k)) \geq \dots \quad (13)$$

Так как

$$P_0(x_i) = P_0(x_i(1))m_i = \delta m_i,$$

го

$$P_1(x_i(1)) = \frac{P_1(x_i)}{m_i} = \frac{P_1(x_i)}{\delta m_i} \delta = \frac{P_1(x_i)}{P_0(x_i)} \delta,$$

и, следовательно, с учетом (13) получаем, что порядок в $\tilde{\mathcal{X}}$, порожденный соотношением (12), индуцирует в \mathcal{X} следующий порядок:

$$P_1(x_{i_1})/P_0(x_{i_1}) \geq P_1(x_{i_2})/P_0(x_{i_2}) \geq \dots, \quad (14)$$

и потому образованное выше оптимальное множество W_a может быть описано так:

$$W_a = \{x_{i_1}, x_{i_2}, \dots, x_{i_k}\},$$

где x_{i_j} определены в (14), а k таково, что соблюдается равенство (9). Итак, оптимальное множество W_a включает точки x с наибольшим отношением правдоподобий $P_1(x)/P_0(x)$.

Поясним, что для непрерывной модели имеет место то же самое. Полагая носитель плотности $f_0(x)$ конечным, разобьем его на конечное число областей Δ_i , $i=1, \dots, N$, с объемами

$$|\Delta_i| = \delta / f_0(x_i), \quad (15)$$

где x_i — некоторые точки из Δ_i , $i=1, \dots, N$, а δ выбрано настолько малым, что можно пренебречь погрешностью в равенствах

$$P_0(\Delta_i) \approx f_0(x_i) |\Delta_i|, \quad P_1(\Delta_i) \approx f_1(x_i) |\Delta_i|, \quad i=1, \dots, N. \quad (16)$$

Из (15) и (16) имеем

$$P_0(\Delta_i) \approx \delta, \quad P_1(\Delta_i) \approx \delta f_1(x_i) / f_0(x_i). \quad (17)$$

Используя разобранную выше дискретную модель с равновероятными исходами по мере P_0 , получаем, что оптимальное множество W_a строится из областей Δ_i с наибольшим отношением правдоподобий $f_1(x_i)/f_0(x_i)$, $x_i \in \Delta_i$.

Лемма 1 (Неймана—Пирсона). Для статистической модели $(\mathcal{X}, \mathcal{B}, (P_0, P_1))$ и гипотез $H_0: P=P_0$, $H_1: P=P_1$ при любом k область

$$W^* = \{x : L_1(x)/L_0(x) > k\} \quad (18)$$

(где $L_i(x)$ — правдоподобие меры P_i , $i=0, 1$) имеет наибольшую мощность среди всех областей W того же или меньшего размера:

$$P_1(W^*) \geq P_1(W), \quad (19)$$

где

$$P_0(W) \leq P_0(W^*). \quad (20)$$

Доказательство проведем для непрерывной модели. Пусть (см. (5)) $\varphi(x)$, $\varphi^*(x)$ — критические функции W и W^* соответственно. Рассмотрим функцию

$$(\varphi^*(x) - \varphi(x))(f_1(x) - k f_0(x)). \quad (21)$$

При $x \in W^*$ оба сомножителя в (21) неотрицательны: первый — так как $\varphi^*(x) = 1$, а второй — по определению (18). При $x \notin W^*$ первый сомножитель в (21) равен — $\varphi(x)$ и, следовательно, неположителен, а ввиду (18) и второй сомножитель неположителен. В таком случае, полагая $dx = dx_1 \dots dx_n$ и рассматривая интеграл как n -мерный, получаем

$$\begin{aligned} 0 &\leq \int (\varphi^*(x) - \varphi(x)) (f_1(x) - kf_0(x)) dx = \\ &= \int \varphi^*(x) f_1(x) dx - \int \varphi(x) f_1(x) dx - \\ &- k \left(\int \varphi^*(x) f_0(x) dx - \int \varphi(x) f_0(x) dx \right) = \\ &= P_1(W^*) - P_1(W) - k(P_0(W^*) - P_0(W)). \end{aligned}$$

Отсюда ввиду (20) следует (19). ■

Лемма 2. *Если множество $V = \{x : f_1(x) - kf_0(x) = 0\}$ имеет меру Лебега нуль, то соотношение (18) определяет единственную с точностью до множества меры Лебега нуль наиболее мощную критическую область заданного размера.*

Доказательство. Из равенства $P_1(W^*) = P_1(W)$ вытекает

$$\begin{aligned} 0 &= \int (\varphi^*(x) - \varphi(x)) (f_1(x) - kf_0(x)) dx = \\ &= \int_{x \in V} (\varphi^*(x) - \varphi(x)) (f_1(x) - kf_0(x)) dx. \end{aligned}$$

Но при $x \in \mathcal{X} \setminus V$ и таких, что $f_1(x) - kf_0(x) > 0$, имеем $\varphi^*(x) = 1$ и, следовательно, $\varphi^*(x) - \varphi(x) \geq 0$, а при $x \in \mathcal{X} \setminus V$ и таких, что $f_1(x) - kf_0(x) < 0$, имеем $\varphi^*(x) = 0$ и, следовательно, $\varphi^*(x) - \varphi(x) \leq 0$. Таким образом, при $x \in \mathcal{X} \setminus V$ подынтегральная функция

$$(\varphi^*(x) - \varphi(x)) (f_1(x) - kf_0(x)) \geq 0,$$

и, следовательно, на самом деле эта функция равна нулю для почти всех $x \in \mathcal{X} \setminus V$ по мере Лебега. Но $f_1(x) - kf_0(x) \neq 0$ на $\mathcal{X} \setminus V$, и потому $\varphi^*(x) = \varphi(x)$ для почти всех $x \in \mathcal{X} \setminus V$, а следовательно, и для почти всех $x \in \mathcal{X}$. ■

Критерий W^* , определенный в (18), называют *критерием отношения правдоподобий*. Напомним, что в рассматриваемой статистической модели статистика $L_1(x)/L_0(x)$ порождает минимальное достаточное разбиение (см. подробнее п. 6 § 15, п. 3 § 16), так что статистика критерия W^* является функцией минимальной достаточной статистики.

3. Примеры. Равномерно наиболее мощные критерии.

(III) В условиях примера (I), обозначая через $I_{\{x_{(1)} > \theta\}}$ индикатор множества $\{x_n : x_{(1)} > \theta\}$, получаем

$$\begin{aligned} \frac{L(x_n; \theta_1)}{L(x_n; \theta_0)} &= \frac{\exp \left(-\sum_{i=1}^n (x_i - \theta_1) \right) / I_{\{x_{(1)} > \theta_1\}}}{\exp \left(-\sum_{i=1}^n (x_i - \theta_0) \right) / I_{\{x_{(1)} > \theta_0\}}} = \\ &= e^{n(\theta_0 - \theta_1)} / I_{\{x_{(1)} > \theta_1\}} / I_{\{x_{(1)} > \theta_0\}}. \end{aligned} \quad (22)$$

Если $\theta_1 > \theta_0$, то для x_n таких, что $\theta_0 < x_{(1)} < \theta_1$, отношение правдоподобий (22) равно нулю, а при $x_{(1)} > \theta_1$ оно постоянно и равно $e^{n(\theta_0 - \theta_1)}$. Положим для любого $\tilde{W} \subseteq R^1$

$$W = \{x_n : x_{(1)} \in \tilde{W}\}. \quad (23)$$

Тогда наилучшая критическая область W^*_{α} размера α для гипотезы $\theta = \theta_0$ против простой альтернативы $\theta = \theta_1 > \theta_0$ получается с помощью (23) из одномерной области \tilde{W}^*_{α} :

$$\tilde{W}^*_{\alpha} = \tilde{W}'_{\alpha} \cup \tilde{W}''_{\alpha}, \quad \tilde{W}'_{\alpha} \subseteq (\theta_0, \theta_1), \quad \tilde{W}''_{\alpha} \subseteq (\theta_1, +\infty),$$

где $\tilde{W}''_{\alpha} = \emptyset$, если $P_{\theta_0}(\tilde{W}''_{\alpha}) \geq \alpha$, и в таком случае в качестве \tilde{W}''_{α} берется любое подмножество из $(\theta_1, +\infty)$ такое, что $P_{\theta_0}(\tilde{W}''_{\alpha}) = \alpha$; если же $P_{\theta_0}(\tilde{W}''_{\alpha}) < \alpha$, то $\tilde{W}''_{\alpha} = (\theta_1, +\infty)$, а в качестве \tilde{W}''_{α} берется любое подмножество интервала (θ_0, θ_1) , так чтобы $P_{\theta_0}(W_{\alpha}) + P_{\theta_0}(\tilde{W}''_{\alpha}) = \alpha$. В частности, критерий W^*_{α} из примера (I) удовлетворяет всем перечисленным требованиям и, следовательно, является наиболее мощным.

Для $\theta_1 < \theta_0$ отношение правдоподобий (22) равно $+\infty$ при $x_n : \theta_1 < x_{(1)} < \theta_0$ и равно $e^{n(\theta_0 - \theta_1)}$ при $x_{(1)} > \theta_0$. Следовательно, оптимальная область получается с помощью (23) из одномерной области

$$\tilde{W}^*_{\alpha} = (\theta_1, \theta_0) \cup \tilde{W}''_{\alpha},$$

где $\tilde{W}''_{\alpha} \subseteq (\theta_0, +\infty)$ — любое такое, что $P_{\theta_0}(\tilde{W}''_{\alpha}) = \alpha$. Обсудим полученные в этом примере результаты. Во-первых, наилучшую критическую область против альтернативы $\theta_1 > \theta_0$ можно выбрать в виде

$$W^*_{\alpha} = \{x_n : x_{(1)} > \theta_0 - n^{-1} \ln(1-\alpha)\},$$

а против альтернативы $\theta_1 < \theta_0$, принимая во внимание, что $P_{\theta_1}(x_n : x_{(1)} \leq \theta_1) = 0$, — в виде

$$W^*_{\alpha} = \{x_n : x_{(1)} \leq \theta_0\} \cup W^+_{\alpha}. \quad (24)$$

Таким образом, критическая область W^*_{α} — одна и та же против альтернативы $\theta_1 > \theta_0$ при любом θ_1 . То же самое верно и для W^-_{α} . Более того, область W^-_{α} годится как оптимальная и против любой альтернативы $\theta_1 > \theta_0$, так как левое подмножество в объединении (24) имеет меру нуль по P_{θ_0} и P_{θ_1} , $\theta_1 > \theta_0$. В результате область

(24) является наиболее мощной против любой простой альтернативы $\theta_1 \neq \theta_0$. ■

Определение. Рассмотрим статистическую модель $(\mathcal{X}, \mathcal{B}, \{P_{\theta}, \theta \in \Theta\})$. Критерий W_{α}^* размера α для проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \in \Theta_1 \subseteq \Theta$ называется *равномерно наиболее мощным*, если для любого $\theta \in \Theta_1$ и любой другой области $W_{\alpha} \in \mathcal{B}$ размера, не превосходящего α , выполнено неравенство

$$P_{\theta}(W_{\alpha}^*) \geq P_{\theta}(W_{\alpha}), \quad \theta \in \Theta_1. \quad \blacksquare \quad (25)$$

(IV) В условиях примера II имеем

$$L(\mathbf{x}_n; \mu) = (2\pi)^{-n/2} \exp \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n}{2} (\bar{x} - \mu)^2 \right).$$

$$L(\mathbf{x}_n; \mu_1)/L(\mathbf{x}_n; \mu_0) = \exp(n(\mu_1 - \mu_0)\bar{x} + (\mu_0^2 - \mu_1^2)).$$

При $\mu_1 > \mu_0$ отношение правдоподобий возрастает с ростом \bar{x} , и поэтому наилучшая критическая область для проверки гипотезы $\mu = \mu_0$ против простой альтернативы $\mu = \mu_1 > \mu_0$ имеет вид

$$W_{\alpha}^+ = \{\mathbf{x}_n : \bar{x} > k_{\alpha}\}. \quad (26)$$

Из условия

$$\begin{aligned} \alpha &= P_{\mu_0}(W_{\alpha}^+) = P_{\mu_0} \left(\mathbf{x}_n : \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i > k_{\alpha} \sqrt{n} \right) = \\ &= P_{\mu_0} \left(\mathbf{x}_n : \frac{1}{\sqrt{n}} \sum_{i=1}^n (x_i - \mu_0) > -\sqrt{n} \mu_0 + k_{\alpha} \sqrt{n} \right) = \\ &= P_{\mu_0} \left(\mathbf{x}_n : \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i > -\sqrt{n} \mu_0 + k_{\alpha} \sqrt{n} \right) \end{aligned}$$

получаем уравнение для определения k_{α} :

$$-\sqrt{n} \mu_0 + k_{\alpha} \sqrt{n} = x_{1-\alpha}(N(0, 1)),$$

откуда

$$k_{\alpha} = \mu_0 + x_{1-\alpha}(N(0, 1)) / \sqrt{n}. \quad (27)$$

Таким образом, критическая область (26), (27) равномерно наиболее мощная против сложной альтернативы $\mu_1 > \mu_0$. Аналогично область

$$W_{-\alpha} = \{\mathbf{x}_n : \bar{x} < \mu_0 + x_{\alpha}(N(0, 1)) / \sqrt{n}\}$$

является равномерно наиболее мощной против сложной альтернативы $\mu < \mu_0$. Как видно, в данном случае не существует области, равномерно наиболее мощной против сложной альтернативы

$\mu_1 \neq \mu_0$. Для проверки гипотезы $\mu_1 = \mu_0$ против сложной альтернативы $\mu_1 \neq \mu_0$ в примере (II) была предложена область $W_\alpha = W_{\alpha/2} \cup W_{-\alpha/2}$, которую можно рассматривать как некоторый компромисс. Как следует из леммы 2, область W_α строго менее мощная, чем W^+_α и W^-_α , против соответствующих односторонних альтернатив.

(V) Рассмотрим повторную выборку из однопараметрического семейства распределений экспоненциального типа:

$$f_\theta(x) = h(x) \exp(a(\theta)T(x) + b(\theta)).$$

Отношение правдоподобий:

$$L(\mathbf{x}_n; \theta_1)/L(\mathbf{x}_n; \theta_0) = \exp \left((a(\theta_1) - a(\theta_0)) \sum_{i=1}^n T(x_i) + n(b(\theta_1) - b(\theta_0)) \right)$$

зависит от выборки через достаточную статистику, и при $a(\theta_1) - a(\theta_0) > 0$ наиболее мощная критическая область для гипотезы $\theta = \theta_0$ против простой альтернативы $\theta = \theta_1 > \theta_0$ имеет вид

$$W_\alpha^+ = \left\{ \mathbf{x}_n : \sum_{i=1}^n T(x_i) > k_\alpha \right\}. \quad (28)$$

Если функция $a(\theta)$ монотонна по θ на некотором подмножестве $\Theta' \subseteq \Theta$, включающем точку θ_0 , то область $W_{\alpha/2}^+$ является равномерно наиболее мощной против сложной гипотезы $\theta > \theta_0$, $\theta \in \Theta'$. В случае, если разность $a(\theta) - a(\theta_0)$ меняет знак при прохождении точки θ_0 , равномерно наиболее мощная критическая область против гипотезы $\theta_1 < \theta_0$ будет иметь вид

$$W_\alpha^- = \left\{ \mathbf{x}_n : \sum_{i=1}^n T(x_i) < l_\alpha \right\},$$

и, следовательно, не существует равномерно наиболее мощного критерия против двусторонней альтернативы $\theta \neq \theta_0$.

4. Близкие гипотезы.

Хотя в рассмотренных выше примерах были найдены равномерно наиболее мощные критерии при сложных альтернативах, существования таких критериев в общем случае ожидать не приходится. Имеются различные варианты других постановок оптимизационных задач в теории проверки статистических гипотез. Здесь мы остановимся на критериях, которые будут в определенном смысле равномерно наиболее мощными по отношению к «близким» гипотезам $H_{+1}^+ : \theta \in (0_0, \theta_0 + \delta)$, $\delta \rightarrow 0$, либо $H_{-1}^- : \theta \in (-\theta_0 - \delta, 0_0)$, $\delta \rightarrow 0$.

Для повторной выборки наиболее мощный критерий для гипотезы $\theta = \theta_0$ против простой альтернативы $\theta_0 + \delta$, $\delta > 0$, запишем в виде

$$W^{(\delta)} = \left\{ x_n : \ln \frac{L(x_n; \theta_0 + \delta)}{L(x_n; \theta_0)} > k^{(\delta)} \right\} = \\ = \{x_n : l(x_n; \theta_0 + \delta) - l(x_n; \theta_0) > k^{(\delta)}\}.$$

Предполагая функцию правдоподобия дифференцируемой по θ , имеем при $\delta \rightarrow 0$

$$l(x_n; \theta_0 + \delta) - l(x_n; \theta_0) = \frac{\partial l(x_n; \theta_0)}{\partial \theta} \delta + o(\delta) = \\ = \sum_{i=1}^n \frac{\partial l(x_i; \theta_0)}{\partial \theta} \delta + o(\delta). \quad (29)$$

Если, к примеру, вторая производная по θ логарифмической функции правдоподобия $l(x_i; \theta)$ ограничена по x_i , то остаточный член в (29) мал равномерно по x_n и область $W^{(\delta)}$ при малых δ будет близка к области

$$W = \{x_n : \partial l(x_n; \theta_0) / \partial \theta > k\}, \quad (30)$$

где выбор k определяет размер критерия W . Будем называть критерий W локально наиболее мощным по отношению к близким альтернативам $\theta > \theta_0$.

Для повторной выборки из распределения $N(\mu, 1)$ имеем

$$\partial l(x_1; \mu_0) / \partial \mu = x_1 - \mu_0,$$

и критическая область (30) принимает вид

$$W = \{x_n : \bar{x} > \mu_0 + k\},$$

что совпадает с равномерно наиболее мощным критерием против сложной односторонней альтернативы $\theta > \theta_0$. ■

(VI) Рассмотрим повторную выборку из распределения Коши с параметром сдвига

$$f(x; \theta) = (\pi(1 + (x - \theta)^2))^{-1}.$$

Имеем

$$\partial l(x_1; \theta_0) / \partial \theta = 2(x_1 - \theta_0) / (1 + (x_1 - \theta_0)^2),$$

так что статистика локально наиболее мощного критерия имеет вид

$$2 \sum_{i=1}^n (x_i - \theta_0) / (1 + (x_i - \theta_0)^2). ■$$

Чтобы воспользоваться критерием (30), необходимо уметь вычислять его размер. Это можно сделать приближенно, используя центральную предельную теорему. В условиях регулярности, при

которых была введена информация по Фишеру (см. подробнее п. 1 § 19), имеем

$$M_{\theta}(\partial l(X_1; \theta)/\partial \theta) = 0, \quad D_{\theta}(\partial l(X_1; \theta)/\partial \theta) = I_{X_1}(\theta),$$

и нормальное приближение дает

$$P_{\theta_0}(x_n : \sum_{i=1}^n \partial l(x_i; \theta_0)/\partial \theta > \sqrt{n} I_{X_1}^{1/2}(\theta_0) x_{1-\alpha}(N(0, 1))) \approx 1 - \alpha. \quad (31)$$

Нормальное приближение (31) практически бывает достаточным уже при умеренных n , хотя следует иметь в виду, что чем меньше α , тем погрешность в (31) больше.

5. Сложные гипотезы.

Предположим, что и нулевая гипотеза, и альтернатива

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_1,$$

— сложные, т. е. параметрические множества Θ_0 и Θ_1 состоят более чем из одного элемента. Одна из возможностей строить критерии в этом случае основана на статистике

$$T_1(x) = \sup_{\theta \in \Theta_1} L(x; \theta) / \sup_{\theta \in \Theta_0} L(x; \theta). \quad (32)$$

Полагая, что оба супремума в (32) достигаются, можно сказать, что по выборке x строятся оценки максимума правдоподобия $\hat{\theta}_0(x)$ и $\hat{\theta}_1(x)$:

$$L(x; \hat{\theta}_i(x)) = \max_{\theta \in \Theta_i} L(x; \theta), \quad i = 0, 1,$$

и затем сравниваются правдоподобия выборки x при этих самых «правдоподобных» значениях параметра θ при гипотезах H_0 и H_1 , как это делается в задаче проверки простой гипотезы против простой альтернативы. Критерий

$$W^{(1)} = \{x : T_1(x) > k\}$$

называют, как и в случае простой гипотезы, критерием отношения правдоподобий. Употребляется и другой вариант критерия отношения правдоподобий, основанный на статистике

$$T_2(x) = \sup_{\theta \in \Theta} L(x; \theta) / \sup_{\theta \in \Theta_0} L(x; \theta), \quad \Theta = \Theta_0 \cup \Theta_1. \quad (33)$$

Так как

$$\sup_{\theta \in \Theta} L(x; \theta) = \max \left(\sup_{\theta \in \Theta_0} L(x; \theta), \sup_{\theta \in \Theta_1} L(x; \theta) \right),$$

то

$$T_2(x) = \max(1, T_1(x)),$$

так что критерий

$$W^{(2)} = \{x : T_2(x) > k\} \quad (34)$$

при $k \geq 1$ совпадает с $W^{(1)}$.

(VII) Рассмотрим линейную модель в канонической форме с нормальными наблюдениями:

$$f^{\Psi, \sigma}(u) = (2\pi\sigma^2)^{-n/2} \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^r (u_i - \psi_i)^2 + \sum_{i=r+1}^n u_i^2 \right) \right),$$

где $\Psi = (\psi_1, \dots, \psi_r)$, $u = (u_1, \dots, u_n)$. Проверим линейную гипотезу $H_0: \psi_1^2 = \dots = \psi_{r-k}^2 = 0$ против альтернативы $H_1: \psi_1^2 + \dots + \psi_{r-k}^2 > 0$ с помощью критерия (34). Имеем

$$l(u; \Psi, \sigma) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^r (u_i - \psi_i)^2 + \sum_{i=r+1}^n u_i^2 \right). \quad (35)$$

Безусловный максимум логарифмической функции правдоподобия по Ψ достигается, очевидно, при $\hat{\psi}_i = u_i$, $i = 1, \dots, r$. Дифференцируя (35) по σ^2 при $\psi_i = \hat{\psi}_i$ и приравнивая производные нулю, находим

$$0 = \partial l(u; \hat{\Psi}, \sigma) / \partial \sigma^2 = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} \left(\sum_{i=r+1}^n u_i^2 \right),$$

откуда

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=r+1}^n u_i^2.$$

Отметим, кстати, что о. м. п. $\hat{\Psi}$ совпадают с наилучшими линейными несмещенными оценками из § 10, а оценка $\hat{\sigma}^2$ отличается множителем.

Проводя аналогичные вычисления в предположении, что верна гипотеза H_0 , т. е. полагая в (35) $\psi_1 = \dots = \psi_{r-k} = 0$, находим

$$\hat{\Psi}_i = u_i, \quad i = r-k, \dots, r, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=r-k+1}^n u_i^2.$$

Таким образом,

$$L(u; \hat{\Psi}, \hat{\sigma}) = \left(\frac{2\pi}{n} \sum_{i=r+1}^n u_i^2 \right)^{-n/2} \exp(-n/2).$$

$$L(u; \hat{\Psi}, \hat{\sigma}) = \left(\frac{2\pi}{n} \sum_{i=r-k+1}^n u_i^2 \right)^{-n/2} \exp(-n/2).$$

и статистика T_2 из (33) принимает вид

$$T_2(u) = \left(\sum_{i=r+1}^n u_i^2 / \sum_{i=r-k+1}^n u_i^2 \right)^{-n/2} = \left(1 + \frac{r-k}{n-r} F(u) \right)^{n/2}, \quad (36)$$

где $F(u)$ — статистика F -критерия из § 13. Из (36) вытекает, что $T_2(u)$ и $F(u)$ задают один и тот же критерий. Этот же вывод, очевидно, справедлив в любой параметризации линейной модели.

§ 22. ПОСЛЕДОВАТЕЛЬНЫЙ КРИТЕРИЙ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ

1. Метод последовательного статистического анализа.

Мы будем рассматривать схему повторного выбора и задачу о проверке простой гипотезы $H_0: P = P_0$ против простой альтернативы $H_1: P = P_1$, но в отличие от предыдущего допустим, что число наблюдений заранее не фиксировано, а может зависеть от результатов уже проведенных испытаний. Определим статистическую модель, отвечающую такому опыту.

Пусть $(R^n, \mathcal{B}_n, P^{(n)})$ — последовательность вероятностных моделей ($n=1, 2, \dots$). Предположим, что последовательность мер $P^{(n)}$, $n=1, 2, \dots$, согласована в следующем смысле: для любого $B_n \in \mathcal{B}_n$:

$$P^{(n)}(B_n) = P^{(n+1)}((x_1, \dots, x_{n+1}) : (x_1, \dots, x_n) \in B_n), \quad (1)$$

что по существу означает, что мера $P^{(n+1)}$ в \mathcal{B}_{n+1} , суженная на \mathcal{B}_n , совпадает с $P^{(n)}$, т. е. последовательность $(R^n, \mathcal{B}_n, P^{(n)})$, $n=1, 2, \dots$, описывает неограниченно продолжающийся вероятностный эксперимент. Пусть $\mathcal{X} = \{x = (x_1, x_2, \dots) : x_i \in R^{(1)}, i=1, 2, \dots\}$ — множество всех бесконечных последовательностей действительных чисел. Примем обозначение $[x]_n = (x_1, \dots, x_n)$ для конечного отрезка последовательности x и $[B]_n = \{[x]_n : x \in B\}$, $B \subseteq \mathcal{X}$. Введем минимальную систему \mathcal{B} подмножеств \mathcal{X} , содержащую все множества вида $B = \{x : [x]_n \in B_n\}$, где $B_n \in \mathcal{B}_n$, $n=1, 2, \dots$, замкнутую относительно перехода к дополнению и образования счетного объединения. Известная теорема Колмогорова утверждает, что при условии (1) может быть определена единственным образом мера P на \mathcal{B} , такая, что для любого $B_n \in \mathcal{B}_n$:

$$P(x : [x]_n \in B_n) = P^{(n)}(B_n) \quad (2)$$

(см. [15, 16]). Равенство (2) означает, что P -вероятность любого события, связанного с результатом только первых n испытаний, совпадает с вероятностью соответствующего события в модели $(R^n, \mathcal{B}_n, P^{(n)})$. Определенная так вероятностная модель $(\mathcal{X}, \mathcal{B}, P)$ соответствует опыту, состоящему из бесконечной последовательности, вообще говоря, зависимых испытаний.

В случае, когда на каждом (R^n, \mathcal{B}_n) задано семейство вероятностных мер $\{P_\theta^{(n)}, \theta \in \Theta\}$ и каждая последовательность $P_\theta^{(n)}$, $n = 1, 2, \dots$, удовлетворяет условию согласованности (1), на измеримом пространстве $(\mathcal{X}, \mathcal{B})$ определено, в соответствии с теоремой Колмогорова, семейство вероятностных мер $\{P_\theta, \theta \in \Theta\}$. Статистическая модель $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$ отвечает опыту с бесконечным числом испытаний и неизвестным вероятностным распределением, принадлежащим семейству $\{P_\theta, \theta \in \Theta\}$. При такой постановке интерес представляют такие статистические процедуры, которые основываются на конечных отрезках $[x]_n$ результатов наблюдений. Введем статистику $N(x)$ с натуральными значениями, которая будет служить «моментом прекращения» наблюдений, так что статистические выводы будут функцией от $[x]_{N(x)}$. Не всякая статистика $N(x)$ годится в качестве такого рода момента, так как естественно, чтобы решение о прекращении испытаний было основано лишь на результатах уже проведенных испытаний.

Статистику $N(x)$ с натуральными значениями назовем *моментом остановки*, если

$$N(x) = n \Rightarrow N(y) = n \text{ для любого } y : [y]_n = [x]_n. \quad (3)$$

Можно сказать, что свойство (3) статистики $N(x)$ означает, что решение о прекращении наблюдений не предвосхищает будущее. Последовательный статистический анализ включает в себя любую статистическую процедуру, зависящую от $x \in \mathcal{X}$ лишь через $[x]_{N(x)}$, где $N(x)$ — некоторый момент остановки.

Рассмотрим наиболее важную для приложений модель повторных независимых испытаний, так что

$$P_\theta(x : x_1 \in B^{(1)}, \dots, x_n \in B^{(n)}) = \prod_{i=1}^n P_\theta(x : x_i \in B^{(i)})$$

для любых n , $B^{(i)} \in \mathcal{B}_1$, $i = 1, \dots, n$, и $P_\theta(x : x_i \in B^{(i)})$ не зависит от i при любом $B \in \mathcal{B}_1$, $x = (x_1, x_2, \dots)$. Ниже нам потребуются моменты остановки специального вида. Пусть $h(x)$ — некоторая \mathcal{B}_1 -измеримая функция и

$$S_n(x) = h(x_1) + \dots + h(x_n), \quad x = (x_1, \dots, x_n, \dots), \quad n = 1, 2, \dots \quad (4)$$

Введем момент $N(x)$ первого достижения последовательностью (4) интервала $(a, +\infty)$:

$$\begin{aligned} N(x) = n &\Leftrightarrow S_i(x) \leq (a, +\infty), \quad i = 1, \dots, n-1, \\ S_n(x) &\in (a, +\infty). \end{aligned} \quad (5)$$

Очевидно, что соотношения (4), (5) задают момент остановки. Покажем, что если

$$\mu_\theta = M_\theta h(X_1) > 0, \quad \theta \in \Theta,$$

то сл. в. $N(X)$ конечна с P_θ -вероятностью единица:

$$P_\theta(N(X) < \infty) = 1, \quad \theta \in \Theta. \quad (6)$$

Воспользовавшись усиленным законом больших чисел, имеем

$$P_\theta\left(x : \frac{1}{n} S_n(x) \rightarrow \mu_\theta\right) = 1. \quad (7)$$

Из (7) вытекает, что при P_θ -почти всех x найдется номер $n_\theta = n_\theta(x)$, начиная с которого $S_n(x) > n\mu_\theta/2$, и, следовательно, момент остановки $N(x)$, определенный в (5), удовлетворяет требованию (6). Если $\mu_\theta = +\infty$, то, как известно, $n^{-1}S_n(x) \rightarrow +\infty$ на множестве P_θ -вероятности единица и момент $N(x)$ также является конечным.

2. Последовательный критерий отношения правдоподобий.

Рассмотрим модель повторных (неограниченно продолжающихся) независимых испытаний. Введем отношение правдоподобий

$$R_n(x) = L([x]_n; \theta_1) / L([x]_n; \theta_0) \quad (8)$$

для отрезка наблюдений длины n , $x = (x_1, x_2, \dots)$, $[x]_n = (x_1, x_2, \dots, x_n)$. Мы видели, что для любого фиксированного n критерий, основанный на статистике (8), максимизирует мощность при простых гипотезе и альтернативе — $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$. В критическую область входят все выборки, для которых отношение (8) превосходит заданное число:

$$\{[x]_n : R_n(x) > A\}. \quad (9)$$

Наглядный смысл неравенства (9) состоит в том, что правдоподобие параметра θ_1 при выборке $[x]_n$ в A раз больше правдоподобия θ_0 , и потому неравенство (9) следует рассматривать как опровержение нулевой гипотезы. Число A выбирают для обеспечения заданного уровня значимости:

$$\alpha = P_{\theta_0}(x : R_n(x) > A),$$

так что в зависимости от α граница A может оказаться и меньше 1. Это означает, что для некоторых точек x отношение (9) меньше единицы, но мы интерпретируем их в пользу гипотезы H_1 . Такое обстоятельство, конечно, является не очень приятным, но для него есть оправдание. Нулевая и альтернативная гипотезы входят в постановку задачи несимметрично: нулевая гипотеза проверяется, альтернативная гипотеза рассматривается как направление возможного отклонения от H_0 . Поэтому и вероятности первого и второго рода не равнозначны для экспериментатора: вероятность ошибки первого рода α устанавливается заранее, что гарантирует ее определенный уровень, удовлетворяющий исследователя, а

вероятность ошибки второго рода минимизируется за счет выбора критерия. Другое замечание в отношении критерия с фиксированным объемом выборки состоит в том, что неравенство (9) могло иметь место при меньших значениях n и, следовательно, проведены «лишние» наблюдения.

Определим последовательный критерий отношения правдоподобий, выбрав два числа $A > 1 > B > 0$ и задав момент $N(x)$ остатков наблюдений следующим образом:

$$N(x) = n \Leftrightarrow R_i(x) \in (B, A), \quad i=1, \dots, n-1; \quad R_n(x) \in (B, A). \quad (10)$$

Иначе говоря, значения отношения правдоподобий из интервала (B, A) относятся к области «неуверенности» в принятии решений. Но как только R_n вышло из этого интервала, решение принимается: при $R_{N(x)}(x) \geq A$ гипотеза H_0 отвергается, при $R_{N(x)}(x) \leq B$ гипотеза H_0 принимается. Критическая область имеет, таким образом, вид

$$W = \{x : R_{N(x)}(x) \geq A\}. \quad (11)$$

Вообще говоря, остается возможность неограниченного продолжения испытаний для выборок x , для которых $B < R_n(x) < A$ при всех n . Покажем, что множество таких x при несовпадающих P_0, P_1 имеет меру нуль по каждой из этих мер, т. е. $N(x)$ удовлетворяет требованию конечности (6).

Рассмотрим для определенности случай, когда распределение вероятностей отдельного наблюдения задается плотностью $f_0(x)$ при H_0 и $f_1(x)$ при H_1 . Положим $a = \log A > 0$, $b = \log B < 0$,

$$h(x) = \log(f_1(x)/f_0(x)), \quad S_n(x) = \log R_n(x) = \sum_{i=1}^n h(x_i).$$

Мы покажем, что

$$-\infty \leq M_0 h(X_1) < 0, \quad 0 < M_1 h(X_1) \leq +\infty, \quad (12)$$

откуда будет следовать, что (см. конец п. 1)

$$P_0(x : N(x) < \infty) = P_1(x : N(x) < \infty) = 1. \quad (13)$$

Для доказательства неравенств (12) применим неравенство Йенсена для выпуклых функций: если $\varphi(x)$ такова, что для любого $0 < a < 1$ и любых $x \neq y$ $a\varphi(x) + (1-a)\varphi(y) < \varphi(ax + (1-a)y)$, то

$$M\varphi(Z) < \varphi(MZ), \quad (14)$$

в предположении, что MZ конечно и $\mathcal{P}(Z = MZ) < 1$.

[Докажем (14), предполагая дополнительно, что $\varphi(x)$ имеет непрерывную вторую производную $\varphi''(x) < 0$. По формуле Тейлора имеем

$$\varphi(x) - \varphi(y) = \varphi'(y)(x-y) + 1/2\varphi''(x+\theta(x, y))(x-y)^2. \quad (15)$$

где $0 < \theta(x, y) < 1$. Подставляя в (15) $y = MZ$, $x = Z$, запишем

$$\varphi(Z) - \varphi(MZ) = \varphi'(MZ)(Z - MZ) + g(Z), \quad (16)$$

где $g(x) < 0$ при всех $x \neq MZ$. Взяв математическое ожидание от обеих частей равенства (16), получаем в случае конечного $M\varphi(Z) : M\varphi(Z) - \varphi(MZ) = Mg(Z) < 0$, что и дает (14). Если $M\varphi(Z)$ бесконечно, то, отбрасывая в (16) неположительное слагаемое $g(Z)$, получаем, что $M\varphi(Z) = -\infty$, и неравенство (14) установлено.

Запишем неравенство (14) для $\varphi(Z) = h(X_1) = \log \frac{f_1}{f_0}(X_1)/f_0(X_1)$:

$$M_0 h(X_1) = \int \log \frac{f_1(x)}{f_0(x)} f_0(x) dx < \int \frac{f_1(x)}{f_0(x)} f_0(x) dx = 0, \quad (17)$$

аналогично

$$M_1 h(X_1) = \int \log \frac{f_1(x)}{f_0(x)} f_1(x) dx = - \int \log \frac{f_0(x)}{f_1(x)} f_1(x) dx > 0. \quad (17a)$$

Если в (17) на множестве положительной P_0 -меры $f_1(x) = 0$, то $M_0 h(X_1) = -\infty$, и аналогично $M_1 h(X_1) = +\infty$, если $f_0(x) = 0$ на множестве положительной P_1 -меры, так что неравенства (12) установлены и тем самым доказано, что $N(x)$ конечно. ■

Последовательный критерий W задается двумя параметрами A, B , $A > 1 > B > 0$; его размер a и мощность β в принципе могут быть рассчитаны по формулам

$$a = P_0(x : R_{N(x)}(x) \geq A), \quad \beta = P_1(x : R_{N(x)}(x) \geq A).$$

Такой расчет представляет в общем случае сложную задачу, решение которой в явном виде не находится. Однако возможен простой приближенный расчет. Мы покажем, что, выбрав

$$A' = \beta/a, \quad B' = (1-\beta)/(1-a), \quad (18)$$

мы получим последовательный критерий

$$W' = \{x : R_{N'(x)}(x) \geq A'\},$$

значения a' , β' которого связаны с a , β неравенствами

$$a' \leq a/\beta, \quad (1-\beta') \leq (1-\beta)/(1-a), \quad a' + (1-\beta') \leq a + (1-\beta). \quad (19)$$

Обычно при использовании последовательного критерия задаются вероятностями ошибок первого и второго рода порядка $0,1 \div 0,01$. В таком случае неравенства (19) показывают, что каждая из вероятностей a' и $(1-\beta')$ для критерия W' может лишь незначительно превысить a и $(1-\beta)$ соответственно, а их сумма не превосходит $a + (1-\beta)$.

Покажем сначала, что

$$a - A^{-1}\beta \leq 0, \quad 1 - \beta - B(1-a) \leq 0. \quad (20)$$

По формуле полной вероятности имеем

$$\begin{aligned} \alpha - A^{-1}\beta &= \sum_{n=1}^{\infty} (\mathbf{P}_0(\mathbf{x}: R_n(\mathbf{x}) \geq A, N(\mathbf{x}) = n) - \\ &\quad - \mathbf{P}_1(\mathbf{x}: R_n(\mathbf{x}) \geq A, N(\mathbf{x}) = n) A^{-1}) = \\ &= \sum_{n=1}^{\infty} \int_{E_n} \dots \int (f_0(x_1) \dots f_0(x_n) - A^{-1} f_1(x_1) \dots f_1(x_n)) dx_1 \dots dx_n. \end{aligned} \quad (21)$$

$$E_n = \{[\mathbf{x}]_n : f_1(x_1) \dots f_1(x_n) \geq A f_0(x_1) \dots f_0(x_n), N(\mathbf{x}) = n\}.$$

Поскольку подынтегральная функция в (21) неположительна на множестве E_n , то получаем отсюда первое из неравенств (20). Используя соотношения

$$1 - \alpha = \mathbf{P}_0(\mathbf{x}: R_{N(\mathbf{x})}(\mathbf{x}) \leq B), \quad 1 - \beta = \mathbf{P}_1(\mathbf{x}: R_{N(\mathbf{x})}(\mathbf{x}) \leq B),$$

аналогично получаем второе из неравенств (20):

$$(1 - \beta) - B(1 - \alpha) = \sum_{n=1}^{\infty} \int_{G_n} \dots \int (f_1(x_1) \dots f_1(x_n) - B f_0(x_1) \dots f_0(x_n)) \times \\ \times dx_1 \dots dx_n \leq 0,$$

$$G_n = \{[\mathbf{x}]_n : f_1(x_1) \dots f_1(x_n) \leq B f_0(x_1) \dots f_0(x_n), N(\mathbf{x}) = n\}.$$

Неравенства (20) перепишем в виде

$$A \leq \beta/\alpha, \quad B \geq (1 - \beta)/(1 - \alpha). \quad (22)$$

Значения вероятностей α' , β' для критерия W' подчиняются неравенствам того же вида:

$$A' \leq \beta'/\alpha', \quad B' \geq (1 - \beta')/(1 - \alpha').$$

Подставляя сюда значения A' , B' из (18), находим

$$\beta/\alpha \leq \beta'/\alpha', \quad (1 - \beta)/(1 - \alpha) \geq (1 - \beta')/(1 - \alpha'), \quad (23)$$

или

$$\alpha' \leq \alpha \beta'/\beta \leq \alpha/\beta, \quad (1 - \beta') \leq (1 - \beta)(1 - \alpha')/(1 - \alpha) \leq (1 - \beta)/(1 - \alpha),$$

что приводит к первым двум неравенствам в (19). Далее, из (23) имеем

$$\beta/\alpha - 1 \leq \beta'/\alpha' - 1, \quad (1 - \alpha)/(1 - \beta) - 1 \leq (1 - \alpha')/(1 - \beta') - 1,$$

или

$$\alpha/(\beta - \alpha) \geq \alpha'/(\beta' - \alpha'), \quad (1 - \beta)/(\beta - \alpha) \geq (1 - \beta')/(\beta' - \alpha'). \quad (24)$$

Складывая неравенства (24), получаем

$$(\alpha + 1 - \beta)/(\beta - \alpha) \geq (\alpha' + 1 - \beta')/(\beta' - \alpha'). \quad (25)$$

При $\beta - a, \beta' - a' > 0$ перепишем (25) в виде

$$1/(-1 + 1/(a+1-\beta)) \geq 1/(-1 + 1/(a'+1-\beta')),$$

или

$$-1 + 1/(a+1-\beta) \leq -1 + 1/(a'+1-\beta'),$$

откуда и получаем третье из неравенств (19).

3. Среднее число наблюдений в последовательном критерии.

Случайный момент времени $N(X)$ является моментом первого выхода траектории случайного блуждания

$$S_n(X) = \log R_n(X) = \sum_{i=1}^n h(X_i), \quad n = 1, 2, \dots,$$

из интервала (b, a) , $a = \log A$, $b = \log B$. Докажем следующее тождество Вальда:

$$M_i S_{N(X)}(X) = M_i h(X_1) M_i N(X), \quad i = 0, 1. \quad (26)$$

Проверим сначала, что $M_i N(X)$ конечно. Заметим, что

$$\begin{aligned} M_i N(X) &= \sum_{n=1}^{\infty} n \mathcal{P}_i(N(X) = n) = \sum_{n=1}^{\infty} n (\mathcal{P}_i(N(X) \geq n) - \\ &- \mathcal{P}_i(N(X) \geq n+1)) = \sum_{n=1}^{\infty} (n-1+1) \mathcal{P}_i(N(X) \geq n) - \\ &- \sum_{n=1}^{\infty} n \mathcal{P}_i(N(X) \geq n+1) = \sum_{n=1}^{\infty} \mathcal{P}_i(N(X) \geq n). \end{aligned} \quad (27)$$

Далее,

$$\begin{aligned} \mathcal{P}_i(N(X) > 2n) &= \mathcal{P}_i(S_j(X) \in (b, a), j = 1, \dots, 2n) \leq \\ &\leq \mathcal{P}_i(S_j(X) \in (b, a), S_{n+k}(X) - S_n(X) \in (b-a, a-b), \\ &\quad j, k = 1, \dots, n). \end{aligned} \quad (28)$$

Так как сл. в.

$$S_{n+k}(X) - S_n(X) = h(X_{n+1}) + \dots + h(X_{n+k}), \quad k = 1, 2, \dots, n,$$

не зависят от $S_j(X)$, $j = 1, \dots, n$, то из (28) получаем

$$\mathcal{P}_i(N(X) > 2n) \leq \mathcal{P}_i(S_j(X) \in (b-a, a-b), j = 1, \dots, n)^2. \quad (29)$$

Но вероятность в правой части (29) меньше единицы при некотором n , поскольку, как было доказано выше, момент выхода $N(X)$ из любого интервала (a, b) конечен с вероятностью единица. Таким образом, при некотором n_0

$$\mathcal{P}_i(N(X) > 2n_0) < \gamma^2, \quad 0 < \gamma < 1,$$

и, аналогично, при любом k

$$\mathcal{P}_i(N(X) \geq kn_0) \leq \gamma^k.$$

Учитывая монотонность членов ряда (27), заключаем, что этот ряд сходится.

Введем сл. в.

$$Y_j = \begin{cases} 1, & \text{если } b < S_k < a, k=0, 1, \dots, j-1, S_0=0, \\ 0, & \text{в противном случае.} \end{cases}$$

Легко видеть, что

$$N(X) = \sum_{j=1}^{\infty} Y_j, \quad S_{N(X)} = h(X_1) + \dots + h(X_{N(X)}) = \sum_{j=1}^{\infty} h(X_j) Y_j. \quad (30)$$

Из (30) получаем

$$M_i N(X) = \sum_{j=1}^{\infty} M_i Y_j,$$

а, используя независимость $h(X_j)$ и Y_j , имеем

$$M_i \sum_{j=1}^n h(X_j) Y_j = \sum_{j=1}^n M_i h(X_j) M_i Y_j = M_i h(X_1) \sum_{j=1}^n M_i Y_j. \quad (31)$$

Переходя в (31) к пределу при $n \rightarrow \infty$, приходим к (26). ■

Перепишем тождество (26) в виде

$$M_i N(X) = M_i S_{N(X)} / M_i h(X), \quad i = 0, 1. \quad (32)$$

Воспользуемся приближенными равенствами

$$S_{N(X)} \approx a \text{ либо } S_{N(X)} \approx b, \quad (33)$$

пренебрегая тем, что в момент выхода из интервала (b, a) статистика S_N выходит, вообще говоря, не на границу интервала, а за его пределы. Понятно, что для обеспечения малых вероятностей a и $1-\beta$ границы a и b по модулю должны быть велики по сравнению со средней длиной шага блуждания $M_i h(X_1)$, что и служит основанием для приближений (33). В результате получаем

$$M_i S_{N(X)} \approx a \mathcal{P}_i(S_{N(X)} \geq a) + b \mathcal{P}_i(S_{N(X)} \leq b),$$

или

$$M_0 S_{N(X)} \approx aa + b(1-a), \quad M_1 S_{N(X)} \approx a\beta + b(1-\beta). \quad (34)$$

Подставляя (34) в (32), получаем приближенные значения для среднего числа наблюдений в последовательном критерии:

$$M_0 S_{N(X)} \approx (aa + b(1-a)) / M_0 h(X_1), \quad (35)$$

$$M_1 S_{N(X)} \approx (a\beta + b(1-\beta)) / M_1 h(X_1),$$

Рассмотрим пример. Пусть $f_i(x) = \varphi(x - \mu_i)$, $i = 1, 2$, где $\varphi(x)$ -плотность закона $N(0, 1)$ и $\mu_0 < \mu_1$. Имеем

$$M_0 h(X_1) = M_0 \ln(f_1(X_1)/f_0(X_1)) = M_0 (-1/2(X_1 - \mu_1)^2 + \\ + 1/2(X_1 - \mu_0)^2) = (\mu_1 - \mu_0)^2/2, M_1 h(X_1) = -(\mu_1 - \mu_0)^2/2.$$

Границы последовательного критерия выбираются из условия (18):

$$a = \ln(\beta/\alpha), b = \ln((1-\beta)/(1-\alpha)).$$

Подставляя их в (35), получаем

$$M_0 N \approx -2(a\alpha + b(1-\alpha))/(\mu_1 - \mu_0)^2,$$

$$M_1 N \approx 2(a\beta + b(1-\beta))/(\mu_1 - \mu_0)^2.$$

Сравним среднее значение объема выборки в последовательном критерии с числом наблюдений $n_{\alpha,\beta}$ в критерии отношения правдоподобий с фиксированным объемом выборки и тем же размером α и мощностью β . Из примера (II) § 21 выпишем уравнение для определения $n = n_{\alpha,\beta}$:

$$\beta = \beta^+(\mu_1) = \Phi(\sqrt{n}(\mu_1 - \mu_0 - x_{1-\alpha}/\sqrt{n})), \quad (36)$$

где x_α — α -квантиль закона $N(0, 1)$, $\Phi(x) = \text{ф. р. } N(0, 1)$. Из (36) получаем

$$x_\beta = \sqrt{n}(\mu_1 - \mu_0) - x_{1-\alpha}, \quad n_{\alpha,\beta} = ((x_\beta + x_{1-\alpha})/(\mu_1 - \mu_0))^2.$$

В результате

$$\frac{M_0 N}{n_{\alpha,\beta}} \approx -\frac{2(a\alpha + b(1-\alpha))}{(x_\beta + x_{1-\alpha})^2}, \quad \frac{M_1 N}{n_{\alpha,\beta}} = \frac{2(a\beta + b(1-\beta))}{(x_\beta + x_{1-\alpha})^2}. \quad (37)$$

При $\alpha = 1 - \beta = 0,05$ находим $a = 2,94$; $b = -2,94$, и правые части соотношений (37) равны 0,4897, что указывает на экономию последовательного критерия в числе наблюдений.

Последовательный критерий отношения правдоподобий минимизирует средний объем выборки по отношению к любому другому последовательному критерию с теми же или меньшими вероятностями ошибок первого и второго рода (см. [11, гл. 3, разд. 12] и [7, с. 429]).

ГЛАВА V

БОЛЬШИЕ ВЫБОРКИ

§ 23. АСИМПТОТИЧЕСКИЕ СВОЙСТВА ОЦЕНОК

1. Состоятельность.

Различные методы оценивания, с которыми мы познакомились в предшествующих главах, позволяют строить оценки $T(x_n)$ при любом объеме выборки n . Свойства оценок при возрастании n систематически нами не рассматривались. В данном параграфе мы остановимся вкратце на некоторых общих свойствах оценок, проявляющихся в модели повторной независимой выборки при $n \rightarrow \infty$. Чтобы подчеркнуть зависимость статистики $T(x_n)$ от n как от параметра, будем употреблять ниже обозначение $T^{(n)}(x_n)$ вместо $T(x_n)$.

Пусть при каждом n статистика $T^{(n)}(x_n)$ определена на измеримом пространстве (R^n, \mathcal{B}_n) , $n=1, 2, \dots$. Поскольку речь идет об одном неограниченно продолжающемся эксперименте, то естественно предполагать, что меры $P^{(n)}$ на (R^n, \mathcal{B}_n) , $n=1, 2, \dots$, согласованы. В таком случае от последовательности $(R^n, B_n, \{P^{(n)}\})$, $n=1, 2, \dots$, можно перейти к единой статистической модели $(\mathcal{X}, \mathcal{B}, \{P\})$, где \mathcal{X} состоит из бесконечных последовательностей $x = (x_1, x_2, \dots)$ (см. подробнее п. 1 § 22), и рассматривать статистики $T^{(n)}([x]_n)$, зависящие от x через конечный отрезок $[x]_n = (x_1, \dots, x_n)$. Такой подход, чрезвычайно важный для теории вероятностей, на самом деле не играет столь большой роли в статистике (исключение составляет последовательный статистический анализ). Тем не менее мы будем исходить из единой статистической модели $(\mathcal{X}, \mathcal{B}, \{P\})$ ввиду ее лучшего соответствия мыслимому эксперименту и большей математической ясности.

Подчеркнем, что все последующие рассуждения сохраняют полностью свое значение, если полагать, что рассматривается последовательность моделей $(R^n, B_n, \{P^{(n)}\})$, и соответственно заменить символ вероятности P на $P^{(n)}$ (при подходящем n).

Ради упрощения выкладок мы будем употреблять обозначение x_n для отрезка (x_1, \dots, x_n) бесконечной последовательности $x = (x_1, x_2, \dots) \in \mathcal{X}$.

Определение 1. Допустим, что при каждом $n=1, 2, \dots$ статистика $T^{(n)}(x_n)$ рассматривается как оценка некоторого функционала $\phi(P)$, $P \in \{P\}$. Скажем, что последовательность оценок

$T^{(n)}(x_n)$ (или, короче, оценка $T^{(n)}(x_n)$) состоятельна, если для любого $\varepsilon > 0$ и любого $P \in \{P\}$

$$P(x_n : |T^{(n)}(x_n) - \varphi(P)| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty. \quad (1)$$

В параметрическом случае вместо (1) имеем

$$P_\theta(x_n : |T^{(n)}(x_n) - \varphi(\theta)| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty, \quad \theta \in \Theta. \quad (2)$$

В случае векторной статистики $T^{(n)}(x_n)$ и вектор-функции $\varphi(\theta)$ (или $\varphi(P)$) состоятельность означает, что соотношения (2) (или (1)) выполнены покомпонентно.

Если при фиксированном P выполнено (1), то говорят, что последовательность сл. в. $T^{(n)}(X_n)$ сходится по вероятности к $\varphi(P)$, и пишут

$$\underset{P}{T^{(n)}(X_n)} \rightarrow \varphi(P). \quad (3)$$

Можно также сказать, что разность соответствующих величин в (3) сходится по вероятности к нулю:

$$\underset{P}{T^{(n)}(X_n)} - \varphi(P) \rightarrow 0.$$

Состоятельность — это минимальное требование, которое следует предъявлять к любой разумной оценке $T^{(n)}(x_n)$, и все оценки для модели повторной выборки, с которыми мы встречались ранее, являются состоятельными. Напомним некоторые из примеров. Э. ф. р. $F_n(x; x_n)$ — состоятельная оценка теоретической ф. р. $F(x)$ в любой точке x . Состоятельность выборочных моментов (среднего, дисперсии и др.) легко вывести, опираясь на следующие два утверждения.

Теорема Хинчина (см. [2, т. 1, гл. X, § 2; т. 2, гл. VII, § 7]). Если $X_i, i=1, 2, \dots$, — последовательность независимых одинаково распределенных сл. в. и $MX_1 = \mu$ конечно, то при любом $\varepsilon > 0$

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) \rightarrow 0, \quad n \rightarrow \infty.$$

Лемма 1. Если каждая из последовательностей сл. в. $T_j(X_n)$, $j=1, \dots, k$, сходится по вероятности к некоторой постоянной a_j , и $g(t_1, \dots, t_k)$ — непрерывная в точке (a_1, \dots, a_k) функция, то

$$g(T_1^{(n)}(X_n), \dots, T_k^{(n)}(X_n)) \underset{P}{\rightarrow} g(a_1, \dots, a_k). \quad (4)$$

Доказательство леммы 1 проведем для краткости при $k=1$. Ввиду непрерывности функции $g(t)$ в точке a для любого $\varepsilon > 0$ найдется $\delta > 0$, что

$$\{t : |t - a| < \delta\} \subseteq \{t : |g(t) - g(a)| < \varepsilon\},$$

и, следовательно,

$$P\{x_n : |g(T^{(n)}(x_n)) - g(a)| \geq \varepsilon\} < P\{x_n : |T^{(n)}(x_n) - a| \geq \delta\}.$$

Но последняя вероятность стремится ввиду (1) к нулю при $n \rightarrow \infty$. ■

Выборочная квантиль $x_{[np], n}$, $0 < p < 1$ (либо $x_{[np]+1, n}$), где $x_{k, n} = x_{(k)}$ — k -я порядковая статистика вектора наблюдений x_n , является состоятельной оценкой теоретической квантили x_p в предположении, что уравнение $p = F(x)$ имеет единственное решение $x = x_p$ и в точке x_p функция F непрерывна. Докажем вначале это утверждение для повторной выборки из равномерного на $(0, 1)$ распределения. Как следует из леммы 2 § 5, если X_1, \dots, X_{n+1} — независимые экспоненциальные $G(\lambda, 1)$, $S_k = X_1 + \dots + X_k$, $k = 1, \dots, n+1$, то сл. в.

$$Z_{k,n} = S_k / S_{n+1}, \quad k = 1, \dots, n, \quad (5)$$

распределены как порядковые статистики повторной выборки из равномерного на $(0, 1)$ распределения. Применяя закон больших чисел, имеем

$$\underset{P}{S_{n+1}}/(n+1) \rightarrow \lambda, \quad \underset{P}{S_{[np]}}/[np] \rightarrow \lambda, \quad n \rightarrow \infty,$$

откуда, воспользовавшись леммой 1, получаем

$$Z_{[np], n} = \frac{S_{[np]}}{[np]} \frac{n+1}{S_{n+1}} \frac{[np]}{n+1} \underset{P}{\rightarrow} p, \quad n \rightarrow \infty. \quad (6)$$

Пусть теперь $F(x)$ -непрерывная ф. р. общего вида. Полагая $F^{-1}(z) = \sup\{x : F(x) = z\}$, легко видеть, что $X_i = F^{-1}(Z_i)$, $i = 1, \dots, \dots, n+1$, представляет собой повторную выборку из распределения $F(x)$, $F^{-1}(Z_{k,n}) = X_{k,n}$. Применяя лемму 1, получаем

$$X_{[np], n} = F^{-1}(Z_{[np], n}) \underset{P}{\rightarrow} F^{-1}(p) = x_p, \quad n \rightarrow \infty. \quad (7)$$

Если $F(x)$ имеет точки разрыва, то, применяя ту же конструкцию, но доопределив функцию $F^{-1}(z)$ по монотонности, получим утверждение (7) для любой квантили x_p , являющейся точкой непрерывности $F(x)$.

2. Состоятельность оценок максимального правдоподобия.

Рассмотрим параметрическую модель повторной неограниченно продолжающейся выборки $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$. Это означает, что мера P_θ , суженная на (R^n, \mathcal{B}_n) :

$$P_\theta^{(n)}(B_n) = P_\theta(x = (x_1, x_2, \dots) : x_n = (x_1, \dots, x_n) \in B_n), \quad B_n \in \mathcal{B}_n,$$

порождает параметрическую модель $(R^n, \mathcal{B}_n, \{P_\theta^{(n)}, \theta \in \Theta\})$ n -кратной повторной независимой выборки. Будем, для определенности,

рассматривать непрерывную модель, так что мера $P_{\theta}^{(n)}$ порождается n -мерной плотностью вида

$$f(x_1; \theta) \cdot \dots \cdot f(x_n; \theta).$$

Оценка максимума правдоподобия $\hat{\theta}^{(n)}(x_n)$ определяется условием

$$L^{(n)}(x_n; \hat{\theta}^{(n)}(x_n)) = \max_{\theta \in \Theta} L^{(n)}(x_n, \theta),$$

$$L^{(n)}(x_n; \theta) = f(x_1; \theta) \cdot \dots \cdot f(x_n; \theta), \quad \theta = (\theta_1, \dots, \theta_k).$$

Если функция правдоподобия дифференцируема по θ , а максимум достигается внутри области Θ , то он может быть найден решением уравнений правдоподобия

$$\begin{aligned} \partial L^{(n)}(x_n; \theta) / \partial \theta_i &= 0, \quad i = 1, \dots, k, \\ l^{(n)}(x_n; \theta) &= \log L(x_n, \theta) \end{aligned} \quad (8)$$

и последующим отбором среди решений системы (8) точки глобального максимума. На самом деле при условиях регулярности типа тех, которые использовались в § 18, 19, локальный максимум единствен и совпадает с глобальным. Поэтому и для теории и для приложений представляет интерес оценка, являющаяся корнем системы (8) и называемая *оценкой уравнений максимума правдоподобия*. Мы сохраним за любым корнем системы (8) обозначение $\hat{\theta}^{(n)}(x_n)$.

Теорема 1. Пусть плотность $f(x; \theta)$ отдельного наблюдения дифференцируема по $\theta \in \Theta$, где Θ – некоторый интервал прямой, $P_{\theta} \neq P_{\theta'}$ при любых $\theta_0 \neq \theta_1$. Введем при каждом $n=1, 2, \dots$ множество

$$B_n = \{x_n: \text{существует решение системы } \partial L^{(n)}(x_n; \theta) / \partial \theta_i = 0, i = 1, \dots, k\}.$$

Тогда, во-первых,

$$P_{\theta}(x: x_n \in B_n) \rightarrow 1, \quad n \rightarrow \infty,$$

и, во-вторых, для $x_n \in B_n$ можно выбрать корень $\hat{\theta}^{(n)}(x_n)$ уравнений правдоподобия (8), который является состоятельной оценкой θ .

Доказательство. Пусть $\theta_0, \theta_1 \in \Theta$ и $P_{\theta_1} \neq P_{\theta_0}$. В п. 2 § 22 было показано, что

$$-\infty < M_{\theta_0}(\log(f(X_1; \theta_1) / f(X_1; \theta_0))) < 0. \quad (9)$$

Применяя закон больших чисел к последовательности

$$l^{(n)}(X_n; \theta_1) - l^{(n)}(X_n; \theta_0) = \sum_{i=1}^n \log(f(X_i; \theta_1) / f(X_i; \theta_0)),$$

получаем

$$P_{\theta_0}(x : l^{(n)}(x_n; \theta_1) - l^{(n)}(x_n; \theta_0) < 0) \rightarrow 1, n \rightarrow \infty. \quad (10)$$

Определим

$$B'_n = \{x_n : l^{(n)}(x_n; \theta_0 + \delta) < l^{(n)}(x_n; \theta_0), l^{(n)}(x_n; \theta_0 - \delta) < l^{(n)}(x_n; \theta_0)\}, \delta > 0.$$

В соответствии с (10)

$$P_{\theta_0}(B'_n) \rightarrow 1, n \rightarrow \infty,$$

а для $x_n \in B'_n$, очевидно, функция $l(x_n; \theta)$ имеет внутри отрезка $[\theta_0 - \delta, \theta_0 + \delta]$ локальный максимум, что доказывает первое утверждение теоремы. Для решения $\hat{\theta}^{(n)}(x_n)$ уравнения

$$\partial l^{(n)}(x_n; \theta) / \partial \theta = 0,$$

принадлежащего определенному выше интервалу $(\theta_0 - \delta, \theta_0 + \delta)$, запишем

$$P_{\theta_0}(x : \partial l^{(n)}(x_n; \hat{\theta}^{(n)}(x_n)) / \partial \theta = 0, |\hat{\theta}^{(n)}(x_n) - \theta_0| < \delta) \rightarrow 1, n \rightarrow \infty,$$

что ввиду произвольности δ доказывает и второе утверждение теоремы 1.

3. Асимптотическая нормальность.

Состоятельность последовательности оценок $T^{(n)}(x_n)$, $n = 1, 2, \dots$, является предельным свойством и формально никак не связана с качеством этих оценок при конечных n . Чтобы состоятельная оценка $T^{(n)}(x_n)$ стала полезным инструментом статистики, необходимо, хотя бы приближенно, вычислять вероятности

$$P_\theta(x : |T^{(n)}(x_n) - \varphi(\theta)| < \delta). \quad (11)$$

Одно из важнейших направлений в теории вероятностей — предельные теоремы — позволяет дать оценку вероятностей (11) при больших n . Центральная предельная теорема занимает, в соответствии со своим названием, ведущее место в теории и приложениях. Речь идет о сходимости функций распределения (соответствующим образом центрированной и нормированной последовательности сл. в.) к нормальному закону

$$\mathcal{D}_\theta((T^{(n)}(X_n) - \mu_n(\theta)) / \sigma_n(\theta) \leq t) \rightarrow \Phi(t), n \rightarrow \infty. \quad (12)$$

Зачастую, но далеко не всегда в качестве $\mu_n(\theta)$ и $\sigma_n(\theta)$ в (12) можно взять соответственно среднее и корень из дисперсии сл. в. $T^{(n)}(X_n)$. Хорошо известна предельная теорема (12) для случая, когда

$$T^{(n)}(X_n) = Y_1 + \dots + Y_n,$$

где сл. в. Y_i , $i = 1, \dots, n$, — независимы, одинаково распределены

и имеют конечную дисперсию. Если исключить искусственные примеры, то, явно или неявно, сходимость в (12) связана с суммированием независимых или слабо зависимых сл. в.

Определение 2. Последовательность оценок $T^{(n)}(x_n)$, $n = 1, 2, \dots$, называется *асимптотически нормальной*, если выполнено соотношение (12); $\mu_n(\theta)$ называется при этом *асимптотическим средним*, $\sigma_n^2(\theta)$ — *асимптотической дисперсией* оценки $T^{(n)}(x_n)$. (Коротко: $T^{(n)}(x_n)$ асимптотически нормальна $N(\mu_n(\theta), \sigma_n^2(\theta))$.)

Параметры μ_n и σ_n определены в (12) неоднозначно. Нетрудно проверить, что если μ'_n и σ'_n удовлетворяют условиям $\mu'_n = \mu_n + o(1)\sigma_n$, $\sigma'_n = \sigma_n(1+o(1))$, $n \rightarrow \infty$, где $o(1)$, $n \rightarrow \infty$, обозначает последовательность, стремящуюся к нулю, то в (12) можно заменить μ_n , σ_n на μ'_n , σ'_n . Действительно, используя преобразование

$$\frac{T^{(n)}(X_n) - \mu'_n}{\sigma'_n} = \frac{\sigma_n}{\sigma'_n} \left(\frac{T^{(n)}(X_n) - \mu_n}{\sigma_n} + \frac{\mu_n - \mu'_n}{\sigma_n} \right).$$

сведем вопрос к следующему общему утверждению.

Лемма 2. Пусть последовательности сл. в. Y_n , U_n , V_n , $n = 1, 2, \dots$, таковы, что при всех y , таких, что $G(y)$ непрерывна:

$$\mathcal{P}(Y_n < y) \rightarrow G(y), \quad n \rightarrow \infty,$$

где $G(y)$ — некоторая ф. р.,

$$U_n \xrightarrow[p]{} 1, \quad V_n \xrightarrow[p]{} 0, \quad n \rightarrow \infty. \quad (13)$$

Тогда

$$\mathcal{P}(U_n(Y_n + V_n) < y) \rightarrow G(y), \quad n \rightarrow \infty, \quad (14)$$

во всех точках непрерывности функции $G(y)$.

Доказательство. Пусть $\varepsilon > 0$ произвольно. Имеем

$$\begin{aligned} \mathcal{P}(Y_n + V_n < y) &\leq \mathcal{P}(Y_n < y + \varepsilon, |V_n| < \varepsilon) + \\ &+ \mathcal{P}(|V_n| > \varepsilon) \leq \mathcal{P}(Y_n < y + \varepsilon) + \mathcal{P}(|V_n| > \varepsilon). \end{aligned}$$

Переходя здесь к верхнему пределу по n и используя (13), получаем

$$\limsup \mathcal{P}(Y_n + V_n < y) \leq \limsup \mathcal{P}(Y_n < y + \varepsilon) = G(y + \varepsilon). \quad (15)$$

Проводя преобразования

$$\begin{aligned} \mathcal{P}(Y_n + V_n < y) &\geq \mathcal{P}(Y_n < y - \varepsilon, |V_n| < \varepsilon) \geq \\ &\geq \mathcal{P}(Y_n < y - \varepsilon) - \mathcal{P}(Y_n < y - \varepsilon, |V_n| > \varepsilon) \geq \\ &\geq \mathcal{P}(Y_n < y - \varepsilon) - \mathcal{P}(|Y_n| > \varepsilon), \end{aligned}$$

аналогично (15) находим

$$\liminf \mathcal{P}(Y_n + V_n < y) \geq G(y - \varepsilon). \quad (16)$$

Из (15) и (16) ввиду произвольности ϵ и непрерывности $G(y)$ в точке y получаем

$$\lim_{n \rightarrow \infty} \mathcal{P}(Y_n + V_n \leq y) = G(y). \quad (17)$$

Далее покажем, что

$$Y_n V_n \xrightarrow{P} 0. \quad (18)$$

Действительно, при любом a

$$\begin{aligned} \mathcal{P}(|Y_n V_n| > \epsilon) &\leq \mathcal{P}(|Y_n V_n| > \epsilon, |Y_n| < a) + \mathcal{P}(|Y_n| > a) < \\ &\leq \mathcal{P}(|V_n| > \epsilon/a) + \mathcal{P}(|Y_n| > a). \end{aligned} \quad (19)$$

Переходя в (19) к верхнему пределу по n , получаем

$$\limsup \mathcal{P}(|Y_n V_n| > \epsilon) \leq G(-a) + 1 - G(a-0). \quad (20)$$

Устремляя в (20) a к ∞ , приходим к (18).

Для доказательства (14) можно сначала воспользоваться результатом (17), а затем применить к сл. в. $Y'_n = Y_n + V_n$, $V'_n = U_n - 1$ и

$$V'_n Y'_n + Y'_n = U_n (Y_n + V_n)$$

последовательно (18) и (17). ■

Когда $T^{(n)}(x_n)$ рассматривается как оценка параметрической функции $\phi(\theta)$, то хорошо, если свойство асимптотической нормальности (12) выполнено для $\mu_n(\theta) = \phi(\theta)$. В этом случае статистику $T^{(n)}(x_n)$ называют *асимптотически несмещенной*. Нормирующий множитель $\sigma_n(\theta)$ в (12) характеризует в этом случае разброс распределения статистики вокруг оцениваемой параметрической функции. В частности, можно написать

$$\begin{aligned} \mathcal{P}_{\theta}(x_n : T^{(n)}(x_n) - x_{1-\alpha}(N(0,1)) \sigma_n(\theta) \leq \phi(\theta) \leq \\ \leq T^{(n)}(x_n) - x_{\alpha}(N(0,1)) \sigma_n(\theta)) \approx \alpha, \end{aligned} \quad (21)$$

где степень приближения растет с увеличением n , $x_{\alpha}(N(0,1))$ — α -квантиль распределения $N(0, 1)$. В предыдущей главе мы видели, что при соблюдении определенных условий регулярности приближение оценки $T^{(n)}(x_n)$ к $\phi(\theta)$ имеет с ростом n порядок $1/\sqrt{n}$. Поэтому множитель $\sigma_n(\theta)$ в (12) в статистике обычно имеет вид $\sigma_n(\theta) = \sigma(\theta)/\sqrt{n}$, $\sigma(\theta) > 0$. Запишем окончательно наиболее важный для статистики вариант предельного соотношения (12):

$$\mathcal{P}_{\theta}(\sigma(\theta)^{-1} \sqrt{n} (T^{(n)}(X_n) - \phi(\theta)) \leq t) \rightarrow \Phi(t), \quad n \rightarrow \infty. \quad (22)$$

4. Преобразование статистик.

В случае, когда имеет место (22), очень полезным оказывается следующий простой результат.

Лемма 3. Пусть $T^{(n)}(X_n)$, $n=1, 2, \dots$, асимптотически нормальна $N(\theta, n^{-1}\sigma^2(\theta))$ и $g(t)$ — дифференцируемая функция, $g'(\theta) \neq 0$. Тогда $g(T^{(n)}(X_n))$, $n=1, 2, \dots$, асимптотически нормальна $N(g(\theta)), g'(\theta)^2\sigma(\theta)^2n^{-1}$.

Доказательство. По формуле Тейлора

$$g(t) - g(\theta) = (t - \theta)(g'(\theta) + \alpha(t; \theta)), \quad (23)$$

где $\alpha(t; \theta) \rightarrow 0$ при $t \rightarrow \theta$. Подставим в (23) $t = T^{(n)}(X_n)$:

$$\begin{aligned} \sqrt{n}(g(T^{(n)}(X_n)) - g(\theta)) &= g'(\theta)\sqrt{n}(T^{(n)}(X_n) - \theta) \times \\ &\times (1 + \alpha(T^{(n)}(X_n); \theta)/g'(\theta)). \end{aligned} \quad (24)$$

Так как, очевидно,

$$T^{(n)}(X_n) \xrightarrow[\theta]{} \theta, \quad n \rightarrow \infty,$$

то второй множитель в (24) сходится по вероятности к 1, и, применяя к (24) лемму 2, получаем требуемый результат. ■

Асимптотическую нормальность (22) можно использовать для построения приближенных (при больших n) доверительных интервалов. Один путь состоит в том, чтобы заменить в (22) $\sigma(\theta)$ на $\sigma(\hat{\theta}(x_n))$, где $\hat{\theta}(x_n)$ — некоторая оценка θ . Более интересен подход, основанный на применении леммы 3. Именно допустим, что функция $g(\theta)$ выбрана так, что

$$g'(\theta)\sigma(\theta) = c, \quad (25)$$

где c — постоянная. В таком случае оценка $g(T^{(n)}(x_n))$ асимптотически нормальна $N(g(\theta), c^2n^{-1})$, и, следовательно, приближенный α -доверительный интервал для $g(\theta)$ имеет вид

$$\begin{aligned} g(T^{(n)}(x_n)) - x_{1-\alpha/2}(N(0, 1))c/\sqrt{n} &\leq g(\theta) \leq \\ &\leq g(T^{(n)}(x_n)) - x_{\alpha/2}(N(0, 1))c/\sqrt{n}. \end{aligned} \quad (26)$$

Так как функция $g(\theta)$ монотонна, то, разрешая неравенства (26) относительно θ , получим α -доверительный интервал для параметра θ . Для нахождения $g(\theta)$ необходимо решить дифференциальное уравнение (25). Приведем пример.

(1) В схеме испытаний Бернулли с неизвестной вероятностью θ , $0 < \theta < 1$, из предельной теоремы Муавра — Лапласа

$$\mathcal{P}((S_n - n\theta)/\sqrt{n\theta(1-\theta)} < t) \rightarrow \Phi(t), \quad n \rightarrow \infty,$$

вытекает, что статистика

$$T^{(n)}(X_n) = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n x_i$$

является асимптотически нормальной $N(\theta, n^{-1}\theta(1-\theta))$. Преобразование $g(\theta)$ находим из уравнения (25) для $c=1/2$:

$$g(\theta) = \int_0^\theta \frac{d\theta}{2\sqrt{\theta(1-\theta)}} = \arcsin \sqrt{\theta}.$$

Таким образом, статистика $\arcsin \sqrt{T^n(x_n)}$ асимптотически нормальна $N(\arcsin \sqrt{\theta}, (4n)^{-1})$.

5. Асимптотическая нормальность выборочных квантилей.

Пользуясь представлением (5) порядковой статистики $Z_{k,n}$ повторной выборки из равномерного распределения на $(0,1)$, получаем

$$\sqrt{n}(Z_{k,n} - p) = \sqrt{n} \left(\frac{S_k}{S_{n+1}} - p \right), \quad (27)$$

где $S_m = X_1 + \dots + X_m$, $m = 1, \dots, n+1$; X_i , $i = 1, \dots, n+1$ — независимые сл. в., распределенные экспоненциально $G(0, 1)$. Проводя элементарные преобразования над (27), получаем

$$\begin{aligned} \sqrt{n} \frac{(1-p)S_k - p(S_{n+1} - S_k)}{S_{n+1}} &= \frac{n}{S_{n+1}} \left((1-p) \frac{S_k - k}{\sqrt{k}} \sqrt{\frac{k}{n}} - \right. \\ &\quad \left. - p \frac{S_{n+1} - S_k - (n+1-k)}{\sqrt{n+1-k}} \sqrt{\frac{n+1-k}{n}} \right). \end{aligned} \quad (28)$$

Допустим, что $k=k_n \rightarrow \infty$, $n \rightarrow \infty$, так что $k_n/n \rightarrow p$, $0 < p < 1$, например $k_n = [np]$. Применяя к суммам

$$(S_k - k)/\sqrt{k}, \quad (S_{n+1} - S_k - (n+1-k))/\sqrt{n+1-k} \quad (29)$$

центральную предельную теорему для независимых одинаково распределенных слагаемых, получаем, что каждая из величин (29) имеет в пределе нормальное распределение $N(0, 1)$. Принимая во внимание независимость сл. в. (29), используя лемму 2 и закон больших чисел для $S_{n+1}/(n+1)$, получаем, что (28) имеет предельное распределение $N(0, p(1-p))$. Таким образом, $Z_{k,n}$ асимптотически нормальна $N(p, p(1-p)n^{-1})$.

Пусть $F(x)$ — непрерывная ф. р., обладающая плотностью $f(x)$. Тогда $F^{-1}(Z_{k,n})$ распределена как k -я порядковая статистика повторной выборки из распределения $F(x)$ (см. конец п. 1). Вычисляя производную обратной функции $F^{-1}(z)$, находим

$$dF^{-1}(z)/dz = 1/f(F^{-1}(z)).$$

Допустим, что $f(x_p) > 0$, где x_p — p -квантиль распределения F , и применим лемму 3 с $g(t) = F^{-1}(t)$. В результате получаем следующее утверждение.

Теорема 2. Если ф. р. $F(x)$ имеет положительную плотность $f(x)$ в окрестности квантили x_p , $0 < p < 1$, то выборочная квантиль $X_{[n,p], n}$ повторной выборки из $F(x)$ асимптотически нормальна $N(x_p, p(1-p)f(x_p)^{-2}n^{-1})$.

В порядке иллюстрации рассмотрим повторную выборку из нормального распределения $N(\mu, \sigma^2)$. Среднее μ является медианой распределения. Выборочная квантиль $X_{[n/2], n}$ асимптотически нормальна $N(\mu, \sigma^2/2n)$. Выборочное среднее нормально $N(\mu, \sigma^2/n)$. Отношение дисперсии выборочного среднего к асимптотической дисперсии выборочной медианы равно $2/\pi = 0,6366$.

6. Асимптотическая нормальность оценок максимума правдоподобия.

Рассмотрим повторную независимую выборку и предположим для определенности, что модель непрерывна. Пусть $\hat{\theta}^{(n)}(x_n)$ — состоятельный корень уравнения правдоподобия

$$\frac{\partial l^{(n)}(x_n; \theta)}{\partial \theta} = 0. \quad (30)$$

Положим при любых $\theta \in \Theta$, $\delta > 0$

$$B_{\theta, \delta}^{(n)} = \{x_n : \frac{\partial l^{(n)}(x_n; \hat{\theta}^{(n)}(x_n))}{\partial \theta} = 0, |\hat{\theta}^{(n)}(x_n) - \theta| < \delta\}.$$

Точный смысл утверждения о состоятельности корня $\hat{\theta}^{(n)}(x_n)$ состоит в том, что существует, возможно не всюду определенная, статистика $\hat{\theta}^{(n)}(x_n)$, $n=1, 2, \dots$, для которой при любых $\theta \in \Theta$ и $\delta > 0$ имеет место предельное соотношение

$$P_\theta(B_{\theta, \delta}^{(n)}) \rightarrow 1, \quad n \rightarrow \infty$$

(см. п. 2). Выражаясь не точно, можно представить себе, что при каждом $\theta \in \Theta$ мера P_θ в основном сосредоточена на подмножестве $B_{\theta, \delta}^{(n)}$ и в пределах этого подмножества уравнение правдоподобия (30) имеет корень, лежащий в интервале $(\theta - \delta, \theta + \delta)$. В таком случае при $\delta > 0$ достаточно малом возникает идея линеаризовать уравнение (30) в окрестности любой точки θ и найти приближенное решение этого уравнения для $x_n \in B_\theta^{(n)}$. Фиксируем произвольное $\theta_0 \in \Theta$ и запишем по формуле Тейлора линейное приближение для левой части уравнения (30):

$$\frac{\partial l^{(n)}(x_n; \theta)}{\partial \theta} = \frac{\partial l^{(n)}(x_n; \theta_0)}{\partial \theta} + \frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2} (\theta - \theta_0) + o(x_n; \theta). \quad (31)$$

Для $x_n \in B_{\theta_0, \delta}^{(n)}$ отбросим остаточный член в (31) и найдем приближенное решение $\hat{\theta}^{(n)}(x_n; \theta_0)$ уравнения (31):

$$\hat{\theta}^{(n)}(x_n; \theta_0) - \theta_0 = - \frac{\frac{\partial l^{(n)}(x_n; \theta_0)}{\partial \theta}}{\frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2}}. \quad (32)$$

Покажем, что статистика $\widehat{\theta}^{(n)}(\mathbf{x}_n; \theta_0)$, рассматриваемая по мере P_{θ_0} , при выполнении некоторых условий регулярности асимптотически нормальна. Эти условия следующие:

(A) Плотность распределения $f(x_1; \theta)$ отдельного наблюдения дважды дифференцируема при $\theta \in \Theta$, Θ – интервал прямой.

(B) Операция дифференцирования по θ перестановочна с интегрированием по x в интегралах

$$\frac{\partial}{\partial \theta} \int f(x_1; \theta) dx_1, \quad \frac{\partial^2}{\partial \theta^2} \int f(x_1; \theta) dx_1;$$

(C) $\partial f(x_1; \theta) / \partial \theta = 0$ при $x: f(x_1; \theta) = 0$.

В этих предположениях (см. § 19, 20)

$$M_\theta \partial l(X_1; \theta) / \partial \theta = 0,$$

$$I_{X_1}(\theta) = M_\theta (\partial l(X_1; \theta) / \partial \theta)^2 = -M_\theta (\partial^2 l(X_1; \theta) / \partial \theta^2), \quad (33)$$

где $l(x_1; \theta) = \ln f(x_1; \theta)$. Используя соотношения

$$\partial l^{(n)}(\mathbf{x}_n; \theta) / \partial \theta = \sum_{i=1}^n \partial l(x_i; \theta) / \partial \theta,$$

$$\partial^2 l^{(n)}(\mathbf{x}_n; \theta) / \partial \theta^2 = \sum_{i=1}^n \partial^2 l(x_i; \theta) / \partial \theta^2,$$

представим (32) в виде

$$(I_{X_1}(\theta_0) n)^{\frac{1}{2}} (\widehat{\theta}^{(n)}(\mathbf{x}_n; \theta_0) - \theta_0) = \\ = \left(\sum_{i=1}^n \frac{\partial l(X_i; \theta_0) / \partial \theta}{\sqrt{I_{X_1}(\theta_0) n}} \right) \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 l(X_i; \theta_0) / \partial \theta^2}{-I_{X_1}(\theta_0)} \right)^{-1}, \quad (34)$$

где сл. в. X_i , $i = 1, \dots, n$, имеют распределение P_{θ_0} . Применяя центральную предельную теорему для одинаково распределенных слагаемых, получим с учетом (33), что первый сомножитель в правой части (34) имеет в пределе при $n \rightarrow \infty$ распределение $N(0, 1)$. Ко второму сомножителю в (34) применим закон больших чисел (см. теорему Хинчина, п. 1) и с учетом (33), получаем, что он сходится по вероятности к единице. Таким образом, статистика $\widehat{\theta}^{(n)}(\mathbf{x}_n; \theta_0)$ по мере P_{θ_0} асимптотически нормальна $N(\theta_0, (I_{X_1}(\theta_0) \times n)^{-1})$.

Итак, построена последовательность статистик $\widehat{\theta}^{(n)}(\mathbf{x}_n; \theta)$, $n = 1, 2, \dots$, зависящая от параметра $\theta \in \Theta$. При каждом фиксированном θ статистика $\widehat{\theta}^{(n)}(\mathbf{x}_n; \theta)$ асимптотически нормальна $N(\theta, (I_{X_1}(\theta) n)^{-1})$ по мере P_{θ} . Однако рассматриваемая по мере $P_{\theta'} \neq P_{\theta}$ статистика $\widehat{\theta}^{(n)}(\mathbf{x}_n; \theta)$ уже не обладает хорошими качествами как оценка θ' . Таким образом, никакая из совокупно-

сти статистик $\widehat{\theta}^{(n)}(x_n; \theta)$, $\theta \in \Theta$, не пригодна как оценка параметра θ . Вспомним теперь, что при каждом $\theta \in \Theta$ статистика $\widehat{\theta}^{(n)}(x_n; \theta)$ по построению «близка» в области $B_{\theta_0}^{(n)}$ к оценке уравнения максимума правдоподобия $\widehat{\theta}^{(n)}(x_n)$. Можно сказать, что статистика $\widehat{\theta}^{(n)}(x_n)$ приближает всю совокупность статистик $\widehat{\theta}^{(n)}(x_n; \theta)$, $\theta \in \Theta$, причем каждую отдельную статистику $\widehat{\theta}^{(n)}(x_n; \theta)$ — в области переменного x_n , имеющую P_θ -меру, близкую к единице. Точный смысл утверждения состоит в следующем: при каждом фиксированном $\theta_0 \in \Theta$ имеет место сходимость по вероятности

$$(\widehat{\theta}^{(n)}(X_n) - \theta_0)/(\widehat{\theta}^{(n)}(X_n; \theta_0) - \theta_0) \xrightarrow{P_{\theta_0}} 1, \quad n \rightarrow \infty. \quad (35)$$

Мы докажем (35) при одном дополнительном условии регулярности:

(D) существует третья производная по θ плотности $f(x; \theta)$,

$$|\partial^3 l(x_1; \theta)/\partial \theta^3| < h(x_1), \quad M_\theta h(X_1) < \infty, \quad \theta \in \Theta,$$

где $h(x_1)$ — некоторая независящая от θ функция.

Для доказательства соотношение (35) запишем для левой части уравнения правдоподобия (30) разложение по формуле Тейлора до квадратичного члена:

$$\frac{\partial l^{(n)}(x_n; \theta_0)}{\partial \theta} + \frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2} (\theta - \theta_0) + \frac{1}{2} \frac{\partial^3 l^{(n)}(x_n; \theta_1)}{\partial \theta^3} (\theta - \theta_0)^2, \quad (36)$$

где $\theta_1 = \theta_1(x_n; \theta)$ лежит между θ_0 и θ . Решение $\widehat{\theta}^{(n)}(x_n)$ уравнения (30) с учетом (36) и (32) представим в виде

$$\begin{aligned} \widehat{\theta}^{(n)}(x_n) - \theta &= - \frac{\partial l^{(n)}(x_n; \theta_0)}{\partial \theta} \left(\frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2} + \frac{\widehat{\theta}(x_n) - \theta_0}{2} \frac{\partial^3 l^{(n)}(x_n; \theta_1)}{\partial \theta^3} \right)^{-1} = \\ &= (\widehat{\theta}^{(n)}(x_n; \theta_0) - \theta_0) \left(1 + \left(\frac{1}{n} \frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2} \right)^{-1} \frac{\widehat{\theta}(x_n) - \theta_0}{2} \times \right. \\ &\quad \left. \times \frac{1}{n} \frac{\partial^3 l^{(n)}(x_n; \theta_1)}{\partial \theta^3} \right)^{-1}. \end{aligned} \quad (37)$$

Выше было установлено, что

$$n^{-1} \partial^2 l^{(n)}(X_n; \theta_0)/\partial \theta^2 \xrightarrow{P_{\theta_0}} -I_{X_1}(\theta_0), \quad n \rightarrow \infty. \quad (38)$$

Далее, в силу условия (D)

$$\left| \frac{1}{n} \frac{\partial^3 l^{(n)}(X_n; \theta_1)}{\partial \theta^3} \right| \leq \frac{1}{n} \sum_{i=1}^n \left| \frac{\partial^3 l(X_i; \theta_1)}{\partial \theta^3} \right| \leq$$

$$\leq \frac{1}{n} \sum_{i=1}^n h(X_i) \xrightarrow{P_{\theta_0}} \text{const}, \quad n \rightarrow \infty. \quad (39)$$

Из (37), (38), (39) и состоятельности оценки $\widehat{\theta}(x_n)$ вытекает (35). Принимая во внимание лемму 2, получаем следующий результат.

Теорема 3. Предположим, что информация по Фишеру $I_{X_1}(\theta)$, $\theta \in \Theta$, положительна и выполнены условия (A), (B), (C), (D). Тогда любая состоятельная оценка уравнений максимума правдоподобия асимптотически нормальна $N(\theta, (I_{X_1}(\theta)n)^{-1})$.

В условиях теоремы запишем по формуле Тейлора

$$\frac{\partial^2 l^{(n)}(x_n; \theta)}{\partial \theta^2} = \frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2} + (\theta - \theta_0) \frac{\partial^3 l^{(n)}(x_n; \theta_1)}{\partial \theta^3}. \quad (40)$$

Подставляя в (40) состоятельный корень $\theta = \widehat{\theta}^{(n)}(x_n)$ уравнения правдоподобия (30) и используя (D), получаем

$$\begin{aligned} & \left| \frac{1}{n} \frac{\partial^2 l^{(n)}(x_n; \widehat{\theta}^{(n)}(x_n))}{\partial \theta^2} - \frac{1}{n} \frac{\partial^2 l(x_n; \theta_0)}{\partial \theta^2} \right| \leq \\ & \leq |\widehat{\theta}(x_n) - \theta_0| \frac{1}{n} \sum_{i=1}^n h(x_i). \end{aligned} \quad (41)$$

Поскольку

$$\frac{1}{n} \sum_{i=1}^n h(X_i) \xrightarrow{P_{\theta_0}} \text{const}, \quad \widehat{\theta}^{(n)}(X_n) - \theta_0 \xrightarrow{P_{\theta_0}} 0, \quad n \rightarrow \infty,$$

то, применяя к правой части (41) лемму 2, заключаем, что

$$n^{-1} \partial^2 l^{(n)}(X_n; \widehat{\theta}^{(n)}(X_n)) / \partial \theta^2 = n^{-1} \partial^2 l^{(n)}(X_n; \theta_0) / \partial \theta^2 \xrightarrow{P_{\theta_0}} 0, \quad n \rightarrow \infty.$$

Отсюда и из (38) выводим, что

$$n^{-1} \partial^2 l^{(n)}(X_n; \widehat{\theta}^{(n)}(X_n)) / \partial \theta^2 \xrightarrow{P_{\theta}} I_{X_1}(\theta). \quad (42)$$

Соотношение (42) показывает, что при достаточно больших n на множестве P_θ -вероятности, сколь угодно близкой к единице, справедливо неравенство

$$\partial^2 l^{(n)}(x_n; \widehat{\theta}^{(n)}(x_n)) / \partial \theta^2 < 0,$$

т. е. состоятельный корень $\widehat{\theta}(x_n)$ уравнения правдоподобия на самом деле является локальным максимумом функции $l^{(n)}(x_n; \theta)$. Более того, если $\widehat{\theta}_1^{(n)}(x_n), \widehat{\theta}_2^{(n)}(x_n)$ — два состоятельных корня уравнения правдоподобия, то либо функция $l^{(n)}(x_n; \theta)$ постоянна на отрезке с концами $\widehat{\theta}_1^{(n)}(x_n), \widehat{\theta}_2^{(n)}(x_n)$, либо между ними найдет-

ся точка $\widehat{\theta}_3^{(n)}(x_n)$, являющаяся локальным минимумом функции $l^{(n)}(x_n; \theta)$. В последнем случае, очевидно,

$$\partial^2 l^{(n)}(x_n; \widehat{\theta}_3^{(n)}(x_n)) / \partial \theta^2 > 0. \quad (43)$$

Так как корень уравнения правдоподобия $\widehat{\theta}_3^{(n)}(x_n)$ заключен между состоятельными корнями $\widehat{\theta}_1^{(n)}(x_n)$ и $\widehat{\theta}_2^{(n)}(x_n)$, то он также является состоятельным, и из (42) и (43) вытекает, что

$$n^{-1} \frac{\partial^2 l^{(n)}(X_n; \widehat{\theta}_3^{(n)}(X_n))}{\partial \theta^2} \xrightarrow{P_\theta} -I_{X_1}(\theta) \geq 0.$$

Но это противоречит предположению $I_{X_1}(\theta) > 0$, $\theta \in \Theta$. Итак, получен следующий результат.

Теорема 4. Предположим, что выполнены условия теоремы 3 и производная $d\bar{l}(x; \theta)/d\theta$ не обращается в нуль тождественно ни на каком интервале значений $\theta \in \Theta$. Тогда с P_θ -вероятностью, стремящейся к единице при $n \rightarrow \infty$, и любом $\theta \in \Theta$ функция правдоподобия имеет единственный внутренний локальный максимум $\widehat{\theta}^{(n)}(x_n)$, который представляет собой асимптотически нормальную $N(\theta, (I_{X_1}(\theta)n)^{-1})$ оценку параметра θ .

Следствие 1. Если функция правдоподобия $l^{(n)}(x_n; \theta)$ достигает максимума во внутренней точке, то теорема 4 применима к оценке максимума правдоподобия.

Следствие 2. Пусть $\varphi(\theta)$ — дифференцируемая параметрическая функция и $\varphi'(\theta) \neq 0$. В условиях теоремы 3, воспользовавшись леммой 3, выводим, что $\varphi(\widehat{\theta}^{(n)}(x_n))$ асимптотически нормальна $N(\varphi(\theta), \varphi'(\theta)^2(I_{X_1}(\theta)n)^{-1})$. Если $\varphi(\theta)$ — обратимая на Θ функция, то, выбирая в качестве нового параметра $\varphi = \varphi(\theta)$, получаем для функции правдоподобия в новой параметризации

$$\tilde{l}^{(n)}(x_n; \varphi(\theta)) = l^{(n)}(x_n; \theta),$$

$$\partial l^{(n)}(x_n; \theta) / \partial \theta = (\partial \tilde{l}^{(n)}(x_n; \varphi(\theta)) / \partial \varphi) (\partial \varphi(\theta) / \partial \theta),$$

и, следовательно, оценка $\widehat{\varphi}^{(n)}(x_n)$ уравнения правдоподобия $\partial \tilde{l}^{(n)}(x_n; \varphi) / \partial \varphi = 0$ совпадает с оценкой $\varphi(\widehat{\theta}^{(n)}(x_n))$, что дает независимый вывод асимптотической нормальности последней оценки. В случае, когда $\widehat{\theta}^{(n)}(x_n)$ является оценкой максимума правдоподобия, заключаем в условиях теоремы 4, что оценка максимума правдоподобия $\varphi(\widehat{\theta}^{(n)}(x_n))$ асимптотически нормальна $N(\varphi(\theta), \varphi'(\theta)^2(I_{X_1}(\theta)n)^{-1})$.

7. Асимптотическая эффективность оценок максимума правдоподобия.

Для выборки конечного объема оценка параметрической функции $\varphi(\theta)$, дисперсия которой совпадает с нижней границей неравенства Фреше — Рао — Крамера, называется эффективной.

В § 19 было установлено, что эффективные оценки существуют только в экспоненциальных моделях и при этом только для параметрических функций вида $\phi(\theta) = a\phi_0(\theta) + \beta$, где $\phi_0(\theta)$ определяется по модели однозначно, a, β — произвольные постоянные. Роль дисперсии как меры разброса оценки вокруг среднего на самом деле не очень велика, если оценка хотя бы приближенно не является нормально распределенной с параметрами, равными ее среднему и дисперсии. Действительно, информация о разбросе распределения оценки, которую дает дисперсия, полностью описывается неравенством Чебышева, а это довольно грубое приближение по сравнению с нормальным приближением, если последнее имеет место.

Переходя к выборкам большого объема, из теорем 3, 4 и следствия 2 мы имеем, что асимптотическая дисперсия оценки максимума правдоподобия (или оценки уравнений максимума правдоподобия) параметра θ и любой дифференцируемой параметрической функции $\phi(\theta)$, $\phi'(\theta) \neq 0$ эквивалентна нижней границе неравенства Фреше — Рао — Крамера при $n \rightarrow \infty$. По этой причине такие оценки называют *асимптотически эффективными*. Хотя дисперсия оценок максимума правдоподобия $\hat{\theta}^{(n)}(x_n)$ может и не существовать или быть больше, чем $(I_{X_1}(\theta) n)^{-1}$, указанное обстоятельство практически не играет никакой роли для асимптотической задачи оценивания: оценка $\hat{\theta}^{(n)}(x_n)$ при больших n приближенно нормальна со средним 0 и дисперсией $(I_{X_1}(\theta) n)^{-1}$, так что именно асимптотическая дисперсия оценки $\hat{\theta}^{(n)}(x_n)$ определяет ее разброс (или концентрацию) вокруг оцениваемого параметра θ .

8. Асимптотическая достаточность.

В условиях регулярности (A) — (D) запишем разложение логарифмической функции правдоподобия по формуле Тейлора:

$$l^{(n)}(x_n; \theta) = l^{(n)}(x_n; \theta_0) + \frac{\partial l^{(n)}(x_n; \theta_0)}{\partial \theta} (\theta - \theta_0) + \\ + \frac{1}{2} \frac{\partial^2 l^{(n)}(x_n; \theta_0)}{\partial \theta^2} (\theta - \theta_0)^2 + \frac{1}{3!} \frac{\partial^3 l^{(n)}(x_n; \theta_0)}{\partial \theta^3} (\theta - \theta_0)^3, \quad (44)$$

где $|\theta_1 - \theta_0| < |\theta - \theta_0|$.

Подставляя в (44) $\theta_0 = \hat{\theta}^{(n)}(x_n)$ — состоятельный корень уравнения правдоподобия $\partial l^{(n)}(x_n; \hat{\theta}^{(n)}(x_n)) / \partial \theta = 0$ и проводя элементарные преобразования, получаем

$$l^{(n)}(x_n; \theta) = l^{(n)}(x_n; \hat{\theta}^{(n)}(x_n)) - \\ - \frac{1}{2} I_{X_1}(\theta) (\sqrt{n}(\theta - \hat{\theta}^{(n)}(x_n))^2 + r^{(n)}(x_n; \theta)), \quad (45)$$

где

$$r^{(n)}(\mathbf{x}_n; \theta) = \frac{1}{2} \left(\frac{1}{n} \frac{\partial^2 l^{(n)}(\mathbf{x}_n; \widehat{\theta}^{(n)}(\mathbf{x}_n))}{\partial \theta^2} + \right. \\ \left. + I_{X_1}(\theta) (\sqrt{n}(\theta - \widehat{\theta}^{(n)}(\mathbf{x}_n)))^2 + \right. \\ \left. + \frac{1}{3!} \frac{1}{n} \frac{\partial^3 l^{(n)}(\mathbf{x}_n; \theta_1)}{\partial \theta^3} (\theta - \widehat{\theta}^{(n)}(\mathbf{x}_n)) (\sqrt{n}(\theta - \widehat{\theta}^{(n)}(\mathbf{x}_n)))^2 \right).$$

Так как при $n \rightarrow \infty$

$$\frac{1}{n} \frac{\partial^2 l^{(n)}(\mathbf{X}_n; \widehat{\theta}^{(n)}(\mathbf{X}_n))}{\partial \theta^2} \xrightarrow{\text{P}_\theta} -I_{X_1}(\theta), \quad \widehat{\theta}^{(n)}(\mathbf{X}_n) \xrightarrow{\text{P}_\theta} \theta,$$

$$\left| \frac{1}{n} \frac{\partial^3 l^{(n)}(\mathbf{X}_n; \theta_1)}{\partial \theta^3} \right| \leq \frac{1}{n} \sum_{i=1}^n h(X_i) \xrightarrow{\text{P}_\theta} 0,$$

$$\mathcal{P}_\theta(\sqrt{n}|\theta - \widehat{\theta}^{(n)}(\mathbf{X}_n)| > t) \xrightarrow{n \rightarrow \infty} 2\Phi(-I_{X_1}^{U_2}(\theta)t) \xrightarrow{t \rightarrow +\infty} 0,$$

то

$$r^{(n)}(\mathbf{X}_n; \theta) \xrightarrow{\text{P}_\theta} 0. \quad (46)$$

Перепишем (45) в виде

$$f(x_1; \theta) \cdots f(x_n; \theta) = f(x_1; \widehat{\theta}^{(n)}(\mathbf{x}_n)) \cdots f(x_n; \widehat{\theta}^{(n)}(\mathbf{x}_n)) \times \\ \times \exp \left(-\frac{1}{2} I_{X_1}(\theta) (\sqrt{n}(\theta - \widehat{\theta}^{(n)}(\mathbf{x}_n)))^2 \right) \exp(r^{(n)}(\mathbf{x}_n; \theta)). \quad (47)$$

Из соотношения (46) вытекает, что при любом $\delta > 0$ P_θ -мера множества

$$\{\mathbf{x}_n : |\exp(r^{(n)}(\mathbf{x}_n; \theta)) - 1| < \delta\} \quad (48)$$

стремится к единице при $n \rightarrow \infty$. Если рассмотреть меру P_θ лишь на множестве (48), то из (47) получается, что плотность меры P_θ с точностью до множителя, близкого к единице, факторизуется и $\widehat{\theta}^{(n)}(\mathbf{x}_n)$ — достаточная статистика этой факторизации. В связи с этим статистику $\widehat{\theta}^{(n)}(\mathbf{x}_n)$, называют *асимптотически достаточной*. Напомним, что достаточность статистики означает независимость условного распределения выборки при условии статистики. Поясним, что в этом отношении можно сказать в случае асимптотической достаточности, записав дискретный вариант соотношения (47):

$$\text{P}_\theta(\mathbf{x}_n) = g(\mathbf{x}_n) \exp \left(-\frac{1}{2} I_{X_1}(\theta) (\sqrt{n}(\theta - \widehat{\theta}^{(n)}(\mathbf{x}_n)))^2 \right) \exp(r^{(n)}(\mathbf{x}_n; \theta)),$$

где $g(\mathbf{x}_n) = \text{P}_\theta(\mathbf{x}_n)$ при $\theta = \widehat{\theta}^{(n)}(\mathbf{x}_n)$. Отсюда для условной вероят-

ности выборки \mathbf{x}^0_n при условии $\hat{\theta}^{(n)}(\mathbf{x}_n) = t$ получаем выражение (ср. п. 5 § 15)

$$\frac{P_\theta(\mathbf{x}_n^0)}{P_\theta(\mathbf{x}_n : \hat{\theta}^{(n)}(\mathbf{x}_n) = t)} = \frac{g(\mathbf{x}_n^0) \exp(r^{(n)}(\mathbf{x}_n^0; \theta))}{\sum_{\hat{\theta}^{(n)}(\mathbf{x}_n)=t} g(\mathbf{x}_n) \exp(r^{(n)}(\mathbf{x}_n; \theta))}, \quad (49)$$

где \mathbf{x}_n^0 таково, что $\hat{\theta}^{(n)}(\mathbf{x}_n^0) = t$. Так как верно (46), то для выборок \mathbf{x}_n^0 , имеющих подавляющую P_θ -вероятность, условная вероятность (49) асимптотически не зависит от θ .

9. Векторный параметр.

Когда параметр $\theta = (\theta_1, \dots, \theta_k)$ — векторный, то интерес представляет совместное асимптотическое поведение оценок $(\hat{\theta}_1^{(n)}(\mathbf{x}_n), \dots, \hat{\theta}_k^{(n)}(\mathbf{x}_n)) = \hat{\theta}^{(n)}(\mathbf{x}_n)$ уравнений правдоподобия

$$\partial l^{(n)}(\mathbf{x}_n; \theta) / \partial \theta_i = 0, \quad i = 1, \dots, k. \quad (50)$$

При условиях регулярности, аналогичных (A)–(D), векторная оценка $\hat{\theta}^{(n)}(\mathbf{x}_n)$ параметра θ в повторной выборке является состоятельной (т. е. каждая ее компонента состоятельна) и асимптотически нормальной: распределение случайного вектора

$$\sqrt{n}(\hat{\theta}^{(n)}(\mathbf{X}_n) - \theta) \quad (51)$$

рассматриваемого по мере P_θ имеет пределом k -мерное нормальное распределение $N_k(\theta, I_{X_1}^{-1}(\theta))$, где $I_{X_1}(\theta)$ — информационная матрица Фишера отдельного наблюдения. Таким образом, асимптотическая матрица ковариаций $(I_{X_1}(\theta))^{-1}$ оценок $\hat{\theta}^{(n)}(\mathbf{x}_n)$ совпадает с нижней границей неравенства Фреше — Рао — Крамера, и потому векторную оценку $\hat{\theta}^{(n)}(\mathbf{x}_n)$ называют (совместно) асимптотически эффективной.

Наметим доказательство асимптотической нормальности, которое повторяет рассуждения для скалярного параметра. Запишем линеаризованную в точке $\theta^0 = (\theta^0_1, \dots, \theta^0_k)$ систему уравнений правдоподобия

$$\frac{\partial l^{(n)}(\mathbf{x}_n; \theta^0)}{\partial \theta_i} + \sum_{j=1}^k \frac{\partial^2 l^{(n)}(\mathbf{x}_n; \theta^0)}{\partial \theta_j \partial \theta_i} (\theta_j - \theta_j^0) = 0, \quad i = 1, \dots, k. \quad (52)$$

Случайный вектор

$$\text{grad}_\theta l^{(n)}(\mathbf{X}_n; \theta) = \left(\frac{\partial l^{(n)}(\mathbf{X}_n; \theta)}{\partial \theta_1}, \dots, \frac{\partial l^{(n)}(\mathbf{X}_n; \theta)}{\partial \theta_k} \right)$$

представляется в виде суммы независимых одинаково распределенных векторов:

$$\sum_{i=1}^n \text{grad}_{\theta} l^{(n)}(X_i; \theta); \quad M \text{grad}_{\theta} l(X_1; \theta) = 0.$$

$$M_{\theta} ((\text{grad}_{\theta} l^{(n)}(X_1; \theta))' (\text{grad}_{\theta} l^{(n)}(X_1; \theta))) = I_{X_1}(\theta).$$

Кроме того,

$$\begin{aligned} M_{\theta} \left[\frac{\partial^2 l^{(n)}(X_n; \theta)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, k \right] = \\ = n M_{\theta} \left[\frac{\partial^2 l^{(n)}(X_1; \theta)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, k \right] = -n I_{X_1}(\theta). \end{aligned}$$

Обозначив решение системы (52) через

$$\widehat{\theta}^{(n)}(x_n; \theta^0) = (\widehat{\theta}_1^{(n)}(x_n; \theta^0), \dots, \widehat{\theta}_k^{(n)}(x_n; \theta^0)),$$

подставляя $\widehat{\theta}^{(n)}(x_n; \theta^0)$ в (52) и переходя к случайным векторам X_n , после элементарных преобразований получаем

$$\begin{aligned} -\frac{1}{n} \sum_{i=1}^k \frac{\partial^2 l^{(n)}(X_n; \theta^0)}{\partial \theta_i \partial \theta_i} (\sqrt{n}(\widehat{\theta}_i^{(n)}(X_n; \theta^0) - \theta_i^0)) = \\ = \frac{1}{\sqrt{n}} \frac{\partial l^{(n)}(X_n; \theta^0)}{\partial \theta_i}, \quad i = 1, \dots, k. \end{aligned} \quad (53)$$

По центральной предельной теореме для сумм независимых одинаково распределенных векторов предельное распределение вектора правых частей системы (53) — нормальное $N_k(0, I_{X_1}(\theta^0))$. Каждый элемент матрицы системы (53)

$$\left[-\frac{1}{n} \frac{\partial^2 l^{(n)}(X_n; \theta^0)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, k \right] \quad (54)$$

сходится по мере P_{θ^0} в силу закона больших чисел к своему математическому ожиданию — соответствующему элементу матрицы $I_{X_1}(\theta^0)$. Матрица $I_{X_1}(\theta^0)$ предполагается положительно определенной, так что с P_{θ^0} -вероятностью, сколь угодно близкой к единице, при $n \rightarrow \infty$ матрица (54) невырождена. Разрешая систему (53), получаем

$$\sqrt{n}(\widehat{\theta}^{(n)}(X_n; \theta^0) - \theta^0) =$$

$$= \frac{1}{\sqrt{n}} \text{grad}_{\theta} l^{(n)}(X_n; \theta^0) \left[\frac{1}{n} \frac{\partial^2 l^{(n)}(X_n; \theta^0)}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, k \right]^{-1}.$$

Вектор $\sqrt{n}(\widehat{\theta}^{(n)}(X_n; \theta^0) - \theta^0)$ представляется, таким образом, в виде

произведения асимптотически нормального случайного вектора на случайную матрицу, сходящуюся по вероятности к постоянной матрице $I_{X_1}^{-1}(\theta^0)$. Обобщая лемму 2, легко вывести отсюда, что вектор $\bar{n}(\hat{\theta}^{(n)}(X_n; \theta^0) - \theta^0)$ асимптотически нормален по мере P_{θ^0} с асимптотической матрицей ковариаций, равной (ср. (1) § 10)

$$(I_{X_1}^{-1}(\theta^0))' I_{X_1}(\theta^0) (I_{X_1}^{-1}(\theta^0)) = I_{X_1}^{-1}(\theta^0),$$

где использована симметрия матрицы Фишера. Остается провести оценки остаточных членов разложения по формуле Тейлора, как это делалось для скалярного параметра, и показать, что состоятельный корень $\hat{\theta}^{(n)}(x_n)$ системы уравнений правдоподобия (50) по вероятности P_{θ^0} эквивалентен $\hat{\theta}^{(n)}(x_n; \hat{\theta}^0)$:

$$(\hat{\theta}_j^{(n)}(X_n) - \theta_j^0) / (\hat{\theta}_j^{(n)}(X_n; \theta^0) - \theta_j^0) \xrightarrow{P_{\theta^0}} 1, \quad j = 1, \dots, k.$$

10. Оценивание параметров сдвига и масштаба.

Найдем выражение для информационной матрицы Фишера повторной выборки из семейства сдвига-масштаба с плотностью $\sigma^{-1}f((x-\mu)/\sigma)$, где $f(x)$ известна. Положив $g(x) = \ln f(x)$, имеем

$$l^{(n)}(x_1; \mu, \sigma) = -\ln \sigma + g((x_1 - \mu)/\sigma),$$

$$\partial l^{(n)}/\partial \mu = -\sigma^{-1}g'((x_1 - \mu)/\sigma),$$

$$\partial l^{(n)}/\partial \sigma = -\sigma^{-1} - \sigma^{-2}(x_1 - \mu)g'((x_1 - \mu)/\sigma),$$

$$\partial^2 l^{(n)}/\partial \mu^2 = \sigma^{-2}g''((x_1 - \mu)/\sigma),$$

$$\begin{aligned} \partial^2 l^{(n)}/\partial \sigma^2 &= \sigma^{-2} + 2\sigma^{-3}(x_1 - \mu)g'((x_1 - \mu)/\sigma) + \\ &+ \sigma^{-2}(x_1 - \mu)^2g''((x_1 - \mu)/\sigma), \end{aligned}$$

$$\partial^2 l^{(n)}/\partial \mu \partial \sigma = \sigma^{-2}g'((x_1 - \mu)/\sigma) + \sigma^{-3}(x_1 - \mu)g''((x_1 - \mu)/\sigma).$$

Обозначив через $Y_1 = (X - \mu)/\sigma$ нормированную сл. в. и учитывая, что

$$0 = M_{\mu, \sigma} \partial l^{(n)}(X_1; \mu, \sigma) / \partial \mu = -\sigma^{-1}M_{\mu, \sigma}g'((X_1 - \mu)/\sigma),$$

$$0 = M_{\mu, \sigma} \partial l^{(n)}(X_1; \mu, \sigma) / \partial \sigma = \sigma^{-1}M_{\mu, \sigma}(1 + \sigma^{-1}(X_1 - \mu)g'((X_1 - \mu)/\sigma)),$$

получаем

$$I_{X_1}(\mu, \sigma) = \sigma^{-2}M_{0,1} \begin{bmatrix} g''(Y_1) & Y_1g''(Y_1) \\ Y_1g''(Y_1) & Y_1^2g''(Y_1) - 1 \end{bmatrix}. \quad (55)$$

Систему уравнений правдоподобия

$$\sum_{i=1}^n g'((x_i - \mu)/\sigma) = 0, \quad \sum_{i=1}^n (1 + \sigma^{-1}(x_i - \mu)g'((x_i - \mu)/\sigma)) = 0$$

в общем случае приходится решать численно. Оценки уравнений правдоподобия $\hat{\mu}^{(n)}(x_n)$, $\hat{\sigma}^{(n)}(x_n)$ асимптотически нормальны $N_2((\mu, \sigma), (I_{X_1}(\mu, \sigma)n)^{-1})$, где $I_{X_1}(\mu, \sigma)$ определяется формулой (55). Подчеркнем, что асимптотическая ковариационная матрица оценок зависит только от параметра σ .

Если плотность $f(x)$ симметрична относительно нуля: $f(x) = f(-x)$, то $g(x) = g(-x)$, $g'(-x) = -g'(x)$, $g''(-x) = g''(x)$, откуда

$$M_{0,1}Y_1g''(Y_1) = 0.$$

Следовательно, матрица $I_{X_1}(\mu, \sigma)$, а также матрица $I_{X_1}^{-1}(\mu, \sigma)$ диагональны. Таким образом, матрица ковариаций двумерного нормального вектора, предельного для случайного вектора

$$\sqrt{n}(\hat{\mu}^{(n)}(X_n) - \mu, \hat{\sigma}^{(n)}(X_n) - \sigma),$$

является диагональной, т. е. предельное нормальное распределение соответствует независимым компонентам. В таком случае говорят, что оценки $\hat{\mu}^{(n)}(X_n)$ и $\hat{\sigma}^{(n)}(X_n)$ асимптотически независимы.

Если плотность $f(x)$ несимметрична, то, изменяя начало отсчета, можно обратить внедиагональные элементы матрицы $I_{X_1}(0)$ в нуль. Действительно, положим

$$a = M_{0,1}Y_1g''(Y_1)/M_{0,1}g''(Y_1) \quad (56)$$

и введем новую параметризацию, положив $\tilde{f}(x) = f(x+a)$, $\tilde{g}(x) = -\ln \tilde{f}(x) = g(x+a)$, и рассмотрим повторную выборку x_1, \dots, x_n из распределения $\sigma^{-1}\tilde{f}((x-\mu)/\sigma)$. Обозначая через $\tilde{M}_{\mu, \sigma}$ операцию математического ожидания по мере с плотностью $\sigma^{-1}\tilde{f}((x-\mu)/\sigma)$, получаем

$$\begin{aligned} \tilde{M}_{0,1}Y_1\tilde{g}''(Y_1) &= M_{-a,1}Y_1\tilde{g}''(Y_1) = M_{0,1}((Y_1-a)\tilde{g}''(Y_1-a)) = \\ &= M_{0,1}(Y_1-a)g''(Y_1) = M_{0,1}Y_1g''(Y_1) - aM_{0,1}g''(Y_1) = 0, \\ \tilde{M}_{0,1}\tilde{g}''(Y_1) &= M_{-a,1}\tilde{g}''(Y_1) = M_{0,1}\tilde{g}''(Y_1-a) = M_{0,1}g''(Y_1), \\ M_{0,1}Y_1^2\tilde{g}''(Y_1) &= M_{-a,1}Y_1^2\tilde{g}''(Y_1) = M_{0,1}(Y_1-a)^2\tilde{g}''(Y_1-a) = \\ &= M_{0,1}Y_1^2g''(Y_1) - 2aM_{0,1}Y_1g''(Y_1) + a^2M_{0,1}g''(Y_1) = \\ &= M_{0,1}Y_1^2g''(Y_1) - a^2M_{0,1}g''(Y_1), \end{aligned}$$

где при последнем переходе использована формула (56). Таким образом,

$$I_{X_1}(\mu, \sigma) = -\sigma^{-2}M_{0,1} \begin{bmatrix} g''(Y_1) & 0 \\ 0 & Y_1^2g''(Y_1) - a^2g''(Y_1) - 1 \end{bmatrix}, \quad (57)$$

где a определяется в (56), так что оценки $\hat{\mu}^{(n)}(x_n)$, $\hat{\sigma}^{(n)}(x_n)$ асимптотически независимы. Точку a называют центром расположения распределения, задаваемого плотностью $f(x)$. Асимптотическая не-

зависимость оценок $\hat{\mu}^{(n)}(x_n)$, $\hat{\sigma}^{(n)}(x_n)$ облегчает их использование, в частности расчет асимптотических доверительных интервалов.

Отметим, что условие регулярности (D) (точнее, его векторный аналог) оказывается слишком ограничительным для данной модели, если рассматривать область параметров $-\infty < \mu < +\infty$, $\sigma > 0$. При $\sigma \rightarrow 0$ плотность $\sigma^{-1}f((x - \mu)/\sigma)$ стремится к бесконечности в окрестности точки $x = \mu$ и производные по μ , σ от $l(x; \mu, \sigma) = -\ln(\sigma^{-1}f((x - \mu)/\sigma))$ не будут равномерно по μ , σ ограниченны никакой функцией $h(x)$. Но если ограничить область изменения параметров неравенствами $|\mu| < c_1$, $\sigma > c_2 > 0$, где c_1, c_2 — произвольные постоянные, то отмеченных затруднений уже не возникает. Так как c_1 можно взять сколь угодно большим, а c_2 — сколь угодно малым, то сужение параметрической области не существенно с точки зрения приложений. Рассмотрим примеры.

(II) Для повторной выборки из нормального распределения с плотностью $\sigma^{-1}\varphi((x - \mu)/\sigma)$ получаем

$$g(x) = \ln \varphi(x) = -\ln \sqrt{2\pi} - x^2/2,$$

$$g'(x) = -x, \quad g''(x) = -1,$$

$$M_{0,1}g''(Y_1) = -1, \quad M_{0,1}Y_1^2g''(Y_1) = -1.$$

Оценки максимума правдоподобия равны $\hat{\mu} = \bar{x}$, $\hat{\sigma} =$
 $= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ (см. пример (II) § 20), а их асимптотическая матрица ковариаций равна по формуле (55)

$$(nI_{X_1}(\mu, \sigma))^{-1} = n^{-1}\sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

О. м. п. для $\varphi(\sigma) = \sigma^2$, равная $\varphi'(\sigma) = \hat{\sigma}^2$, имеет асимптотическую дисперсию $(\varphi'(\sigma))^2 D_{\mu, \sigma} = 4\sigma^2\sigma^2n^{-1}2^{-1} = 2\sigma^4n^{-1}$. Асимптотическая и точная матрицы ковариаций оценок $\hat{\mu}$, $\hat{\sigma}^2$ равны соответственно (см. пример (II) § 19).

$$\begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix} \text{ и } \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4(n-1)/n^2 \end{bmatrix}.$$

Отметим, что условия регулярности (A), (B), (C) в рассматриваемом случае выполняются, а для выполнения (D) надо ограничить область изменения параметров до $|\mu| < c_1$, $\sigma > c_2$, где c_1, c_2 — произвольные положительные постоянные.

(III) Для повторной выборки из гамма-распределения с известным параметром формы p и плотностью $f(x, \mu, \sigma) = \sigma^{-1}f((x - \mu)/\sigma)$, где

$$f(x) = \frac{1}{\Gamma(p)} x^{p-1} e^{-x}, \quad x > 0,$$

условия дифференцируемости плотности $f(x, \mu, \sigma)$ могут нарушаться в точке $\mu=x$. Приняв $p>3$, обеспечим существование вторых производных (условие (A)) и возможность перемены порядка дифференцирования по параметрам и интегрирования по x (условие (B)). Условие (C) также выполнено (см. § 19, соотношения (15) и далее). Если допустить, что $p>4$, то при ограничении области изменения параметров до $|\mu| < c_1 < \infty, \sigma \geq c_2 > 0$ выполняется также и условие (D).

Имеем (см. (3), (4) § 7)

$$g(x) = -\ln \Gamma(p) + (p-1) \ln x - x, \quad g'(x) = (p-1)/x - 1,$$

$$g''(x) = -(p-1)/x^2,$$

$$\mathbf{M}_{0,1}g''(Y_1) = \mathbf{M}_{0,1}(-(p-1)/Y_1^2) = -1/(p-2),$$

$$\mathbf{M}_{0,1}Y_1g''(Y_1) = -1, \quad \mathbf{M}_{0,1}Y_1^2g''(Y_1) = -(p-1).$$

Асимптотическая матрица ковариаций о. м. п. равна

$$(nI_{X_1}(\mu, \sigma))^{-1} = n^{-1}\sigma^2 \begin{bmatrix} (p-2)^{-1} & 1 \\ 1 & p \end{bmatrix}^{-1} = \frac{(p-2)\sigma^2}{2n} \begin{bmatrix} p & -1 \\ -1 & (p-2)^{-1} \end{bmatrix}.$$

Если начало отсчета перенести в центр расположения распределения

$$a = \mathbf{M}_{0,1}Y_1g''(Y_1)/\mathbf{M}_{0,1}g''(Y_1) = p-2,$$

т. е. оценивать $\mu' = \mu + a$ вместо μ , то по формуле (57) находим

$$(nI_{X_1}(\mu', \sigma))^{-1} = n^{-1}\sigma^2 \begin{bmatrix} (p-2)^{-1} & 0 \\ 0 & 2 \end{bmatrix}^{-1} = n^{-1}\sigma^2 \begin{bmatrix} p-2 & 0 \\ 0 & 2^{-1} \end{bmatrix}.$$

Как видно, асимптотическая дисперсия о. м. п. $\hat{\mu}'(x_n)$ в $p/2$ раз меньше дисперсии $\hat{\mu}(x_n)$; нетрудно проверить, что она совпадает с асимптотической дисперсией о. м. п. $\hat{\mu}'(x_n)$ в модели с известным σ .

§ 24. АСИМПТОТИЧЕСКИЕ СВОЙСТВА КРИТЕРИЯ ОТНОШЕНИЯ ПРАВДОПОДОБИЙ

1. Скалярный параметр.

Рассмотрим для повторной выборки асимптотические свойства критерия отношения правдоподобий, введенного в п. 5 § 21, полагая, что нулевая гипотеза — простая, альтернативная — сложная:

$$H_0: \theta = \theta_0, \quad H_1: \theta \in \Theta_1, \quad \Theta = \{\theta_0\} \cup \Theta_1,$$

а параметр θ — скалярный. Статистику критерия возьмем в виде

$$T^{(n)}(x_n) = L^{(n)}(x_n; \hat{\theta}^{(n)}(x_n))/L^{(n)}(x_n; \theta_0)$$

где $\widehat{\theta}^{(n)}(x_n)$ — оценка уравнения правдоподобия параметра θ в модели $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$. Будем предполагать выполнеными условия регулярности $(A) - (D)$ п. 6 § 23, так что имеется единственный (с P_θ -вероятностью, стремящейся к единице при $n \rightarrow \infty$) состоятельный корень $\widehat{\theta}^{(n)}(x_n)$ уравнения правдоподобия, являющийся асимптотически нормальным $N(\theta, (I_{X_1}(\theta)n)^{-1})$. Используя формулу Тейлора, имеем

$$\begin{aligned} \ln T^{(n)}(x_n) &= l^{(n)}(x_n; \widehat{\theta}^{(n)}(x_n)) - l^{(n)}(x_n; \theta_0) = \\ &= \partial l^{(n)}(x_n; \theta_0) / \partial \theta (\widehat{\theta}^{(n)}(x_n) - \theta_0) + \frac{1}{2} \partial^2 l^{(n)}(x_n; \theta_0) / \partial \theta^2 (\widehat{\theta}^{(n)}(x_n) - \theta_0)^2 + \\ &\quad + \frac{1}{3!} \partial^3 l^{(n)}(x_n; \theta_1) / \partial \theta^3 (\widehat{\theta}^{(n)}(x_n) - \theta_0)^3, \end{aligned} \quad (1)$$

$$\begin{aligned} \partial l^{(n)}(x_n; \widehat{\theta}^{(n)}(x_n)) / \partial \theta &- \partial l^{(n)}(x_n; \theta_0) / \partial \theta = \\ &= \partial^2 l^{(n)}(x_n; \theta_0) / \partial \theta^2 (\widehat{\theta}^{(n)}(x_n) - \theta_0) + \frac{1}{2} \partial^3 l^{(n)}(x_n; \theta_2) / \partial \theta^3 (\widehat{\theta}^{(n)}(x_n) - \theta_0)^3, \end{aligned} \quad (2)$$

где $|\theta_i - \theta_0| \ll |\widehat{\theta}^{(n)}(x_n) - \theta_0|$, $i = 1, 2$. Используя в (2) равенство

$$\partial l^{(n)}(x_n; \widehat{\theta}^{(n)}(x_n)) / \partial \theta = 0$$

и подставляя выражение для производной $\partial l^{(n)}(x_n; \theta_0) / \partial \theta$ из (2) в (1), получаем

$$\ln T^{(n)}(x_n) = -1/2 \partial^2 l^{(n)}(x_n; \theta_0) / \partial \theta^2 (\widehat{\theta}^{(n)}(x_n) - \theta_0)^2 + 2r_n(x_n), \quad (3)$$

где для остатка $2r_n(x_n)$ имеем (ср. (37), (39) § 23)

$$|r_n(X_n)| \leq \frac{1}{n} \sum_{i=1}^n h(X_i) |\widehat{\theta}^{(n)}(X_n) - \theta_0| (\sqrt{n} (\widehat{\theta}^{(n)}(X_n) - \theta_0))^2 \xrightarrow{P_{\theta_0}} 0 \quad (4)$$

при $n \rightarrow \infty$. Используя закон больших чисел и асимптотическую нормальность оценки $\widehat{\theta}^{(n)}(x_n)$, выводим, что статистика

$$-\frac{1}{n} I_{X_1}^{-1}(\theta_0) \partial^2 l^{(n)}(x_n; \theta_0) / \partial \theta^2 n I_{X_1}(\theta_0) (\widehat{\theta}^{(n)}(x_n) - \theta_0)^2 \quad (5)$$

по мере P_{θ_0} имеет предельное распределение χ^2 с одной степенью свободы. Отсюда и из (3), (4) заключаем, что статистика $2 \ln T^{(n)}(x_n)$ в предположении гипотезы H_0 имеет χ_1^2 -распределение.

Знание, хотя и приближенное, распределения статистики критерия делает возможным его практическое использование при больших n . В критическую область включают большие значения статистики критерия:

$$\{x_n : 2 \ln T^{(n)}(x_n) > \chi_{1-\alpha}^2\} \quad (6)$$

где $\chi_\alpha(\chi_1^2)$ — α -квантиль распределения χ_1^2 . Размер критерия (6)

с ростом n стремится к a . Статистику критерия (6) можно заменить при больших n статистикой (5) и получить критерий в форме

$$\{x_n : -\partial^2 l^{(n)}(x_n; \theta_0)/\partial\theta^2 (\hat{\theta}^{(n)}(x_n) - \theta_0)^2 > x_{1-\alpha}(\chi_1^2)\}. \quad (7)$$

В свою очередь, используя соотношения

$$\frac{1}{n} \partial^2 l^{(n)}(X_n; \theta_0)/\partial\theta^2 \xrightarrow{P_{\theta_0}} -I_{X_1}(\theta_0),$$

$$I_{X_1}(\hat{\theta}^{(n)}(X_n)) \xrightarrow{P_{\theta_0}} I_{X_1}(\theta_0), \quad n \rightarrow \infty,$$

можно вместо (7) предложить любую из следующих критических областей:

$$\{x_n : I_{X_1}(\theta_0) n (\hat{\theta}^{(n)}(x_n) - \theta_0)^2 > x_{1-\alpha}(\chi_1^2)\}, \quad (8)$$

$$\{x_n : I_{X_1}(\hat{\theta}^{(n)}(x_n)) n (\hat{\theta}^{(n)}(x_n) - \theta_0)^2 > x_{1-\alpha}(\chi_1^2)\}. \quad (9)$$

Используя представление (см. (37), (32) § 23)

$$\begin{aligned} \hat{\theta}^{(n)}(X_n) - \theta_0 = -\partial l^{(n)}(X_n; \theta_0)/\partial\theta (\partial^2 l^{(n)}(X_n; \theta_0)/\partial\theta^2)^{-1} \times \\ \times (1 + r_n(X_n; \theta_0)), \end{aligned}$$

где

$$r_n(X_n; \theta_0) \xrightarrow{P_{\theta_0}} 0, \quad n \rightarrow \infty,$$

имеем

$$\begin{aligned} I_{X_1}(\theta_0) n (\hat{\theta}^{(n)}(x_n) - \theta_0)^2 = I_{X_1}(\theta_0) \left(\frac{1}{\sqrt{n}} \partial l^{(n)}(x_n; \theta_0)/\partial\theta \right)^2 \times \\ \times \left(\frac{1}{n} \partial^2 l^{(n)}(x_n; \theta_0)/\partial\theta^2 \right)^{-2} (1 + r_n(x_n; \theta_0))^2. \end{aligned} \quad (10)$$

Так как предпоследний множитель в правой части (10) стремится по мере P_{θ_0} к $I_{X_1}^{-2}(\theta_0)$, то из (8) получаем еще один критерий:

$$\left\{ x_n : I_{X_1}^{-1}(\theta_0) \left(\frac{1}{\sqrt{n}} \partial l^{(n)}(x_n; \theta_0)/\partial\theta \right)^2 > x_{1-\alpha}(\chi_1^2) \right\}. \quad (11)$$

Практическое удобство критерия (11) состоит в том, что для его применения не требуется вычислять о. м. п. $\hat{\theta}^{(n)}(x_n)$.

Проведенные рассуждения показывают, что критические области (6) — (9), (11) в предположении, что справедлива гипотеза H_0 , асимптотически эквивалентны: любые две из них отличаются на множестве, P_{θ_0} -мера которого стремится к нулю при $n \rightarrow \infty$. Покажем, что функции мощности этих критериев в любой точке θ_1 , такой, что $P_{\theta_1} \neq P_{\theta_0}$, стремятся к единице при $n \rightarrow \infty$. Рассмотрим, как ведет себя статистика критерия (6) по мере P_{θ_1} , когда $n \rightarrow \infty$. Имеем

$$2 \ln T^{(n)}(\mathbf{x}_n) = 2(l^{(n)}(\mathbf{x}_{(n)}; \hat{\theta}^{(n)}(\mathbf{x}_n)) - l^{(n)}(\mathbf{x}_n; \theta_1)) + \\ + 2(l^{(n)}(\mathbf{x}_n; \theta_1) - l^{(n)}(\mathbf{x}_n; \theta_0)). \quad (12)$$

Как мы уже знаем, первое слагаемое в (12), в предположении, что верна гипотеза $\theta = \theta_1$, имеет в пределе χ^2 -распределение. Отсюда вытекает, в частности, что для любого $\varepsilon > 0$ и всех достаточно больших n

$$\mathbf{P}_{\theta_1}(\mathbf{x}_n : l^{(n)}(\mathbf{x}_n; \hat{\theta}^{(n)}(\mathbf{x}_n)) - l^{(n)}(\mathbf{x}_n; \theta_1)) > 0 > 1 - \varepsilon. \quad (13)$$

Далее,

$$\frac{1}{n}(l^{(n)}(\mathbf{x}_n; \theta_1) - l^{(n)}(\mathbf{x}_n; \theta_0)) = \frac{1}{n} \sum_{i=1}^n \ln(f(x_i; \theta_1)/f(x_i; \theta_0)),$$

и так как (см. § 22, соотношение (17а))

$$a = \mathbf{M}_{\theta_1} \ln(f(X_1; \theta_1)/f(X_1; \theta_0)) > 0,$$

то

$$n^{-1}(l^{(n)}(\mathbf{X}_n; \theta_1) - l^{(n)}(\mathbf{X}_n; \theta_0)) \xrightarrow[\mathbf{P}_{\theta_1}]{} a > 0, \quad n \rightarrow \infty.$$

Следовательно, для любого $\varepsilon > 0$ и всех достаточно больших n

$$\mathbf{P}_{\theta_1}(\mathbf{x}_n : n^{-1}(l^{(n)}(\mathbf{x}_n; \theta_1) - l^{(n)}(\mathbf{x}_n; \theta_0)) > a/2) > 1 - \varepsilon. \quad (14)$$

Из (13) и (14) получаем для достаточно больших n

$$\begin{aligned} \beta^{(n)}(\theta_1) &= \mathbf{P}_{\theta_1}(\mathbf{x}_n : 2 \ln T^{(n)}(\mathbf{x}_n) > x_{1-\alpha}(\chi^2_1)) \geqslant \\ &\geqslant \mathbf{P}_{\theta_1}(\mathbf{x}_n : 2(l^{(n)}(\mathbf{x}_n; \theta_1) - l^{(n)}(\mathbf{x}_n; \theta_0)) > na) - \varepsilon > 1 - 2\varepsilon. \end{aligned}$$

Итак, для любой простой альтернативы θ_1 , такой, что $\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_0}$, функция мощности критерия (6) стремится к единице при $n \rightarrow \infty$. Такое свойство критерия называют *состоятельностью*. Состоятельныйми являются и все остальные критерии (7)–(9), (11). Действительно, в случае критерия (8) запишем

$$\sqrt{n}(\hat{\theta}^{(n)}(\mathbf{x}_n) - \theta_0) = \sqrt{n}(\hat{\theta}^{(n)}(\mathbf{x}_n) - \theta_1) + \sqrt{n}(\theta_1 - \theta_0). \quad (15)$$

Первое слагаемое в правой части (15) асимптотически нормально $N(0, I_{X_1}(\theta_1))$ по мере \mathbf{P}_{θ_1} . Поэтому для любого $\varepsilon > 0$ найдется $t_\varepsilon > 0$, такое, что, начиная с некоторого n , множество

$$\{\mathbf{x}_n : I_{X_1}(\theta_0) n (\hat{\theta}^{(n)}(\mathbf{x}_n) - \theta_1)^2 < t_\varepsilon\} \quad (16)$$

имеет \mathbf{P}_{θ_0} -меру большую, чем $1 - \varepsilon$. Если к тому же n удовлетворяет неравенству

$$I_{X_1}^{1/2}(\theta_0) \sqrt{n} |\theta_1 - \theta_0| > (x_{1-\alpha}(\chi^2_1))^{1/2} + t_\varepsilon^{1/2},$$

то из (15) следует, что

$$\beta^{(n)}(\theta_1) = \mathbf{P}_{\theta_1}(\mathbf{x}_n : I_{X_1}(\theta_0) n (\hat{\theta}^{(n)}(\mathbf{x}_n) - \theta_0)^2 > x_{1-\alpha}(\chi^2_1)) > 1 - \varepsilon.$$

Состоятельность критерия (9) доказывается аналогично, если учесть, что

$$I_{X_1}(\widehat{\theta}^{(n)}(X_n)) \xrightarrow{P_{\theta_1}} I_{X_1}(\theta_1), \quad n \rightarrow \infty.$$

Так же нетрудно установить состоятельность критериев (7), (11).

Сравнить состоятельные критерии можно, изучая порядок близости к единице их функций мощности при различных θ . Можно сравнить их по поведению функции в точках $\theta = \theta_n$, таких, что $\theta_n \rightarrow \theta_0$, $n \rightarrow \infty$. Заметим, что асимптотически эквивалентные критерии имеют, вообще говоря, не эквивалентные при $n \rightarrow \infty$ ошибки второго рода. Это легко обнаружить, сравнивая, например, статистики критериев (8) и (9): при одном и том же члене $n(\widehat{\theta}^{(n)}(x_n) - \theta_0)^2$ стоит множитель $I_{X_1}(\theta_0)$ в (8) и множитель, близкий к $I_{X_1}(\theta)$ по мере P_θ , в (9).

Стоит отметить, что статистика критерия (11) является квадратом статистики локально наиболее мощного против односторонних альтернатив критерия из п. 4 § 21. Поскольку двусторонний локально наиболее мощный критерий существует лишь в исключительных случаях, то компромиссная область (11) может рассматриваться как «оптимальная» для двусторонних близких альтернатив.

2. Векторный параметр.

В случае векторного параметра $\theta = (\theta_1, \dots, \theta_k)$, пользуясь разложением по формуле Тейлора и асимптотической нормальностью $N_k(\theta, (I_{X_1}(\theta))^n)^{-1}$ оценок $\widehat{\theta}^{(n)}(x_n)$ и асимптотической нормальностью $N_k(0, I_{X_1}(\theta)^n)$ вектора $\text{grad}_\theta(I^n(X_n; \theta))$, получаем, что статистики векторных вариантов критериев (6)–(9), (11):

$$W_1^{(n)}(x_n) = 2 \ln T^{(n)}(x_n),$$

$$W_2^{(n)}(x_n) = n(\widehat{\theta}^{(n)}(x_n) - \theta^0) I_{X_1}(\theta^0) (\widehat{\theta}^{(n)}(x_n) - \theta^0)',$$

$$W_3^{(n)}(x_n) = n(\widehat{\theta}^{(n)}(x_n) - \theta^0) I_{X_1}(\widehat{\theta}^{(n)}(x_n)) (\widehat{\theta}^{(n)}(x_n) - \theta^0)'$$

имеют по мере P_θ предельное χ_k^2 -распределение. Все эти критерии также являются состоятельными и асимптотически эквивалентными.

3. Полиномиальное распределение.

Рассмотрим повторную выборку из распределения с исходами 1, 2, ..., m и вероятностями исходов p_1, p_2, \dots, p_m ; $p_1 + p_2 + \dots + p_m = 1$. За вектор неизвестных параметров примем $p = (p_1, p_2, \dots, p_{m-1})$. Найдем вначале о.м.п. параметра p . Ис-

пользуя индикаторное обозначение, запишем логарифмическую функцию правдоподобия отдельного наблюдения:

$$l(x_i; p) = \sum_{j=1}^{m-1} I_{\{x_i=j\}} \ln p_j + I_{\{x_i=m\}} \ln (1 - p_1 - \dots - p_{m-1}),$$

где $I_A = 1$, если A верно, $I_A = 0$, если A неверно. Далее,

$$\begin{aligned} l^{(n)}(x_n; p) &= \sum_{i=1}^n l(x_i; p) = \sum_{j=1}^{m-1} \ln p_j \sum_{i=1}^n I_{\{x_i=j\}} + \\ &+ \ln (1 - p_1 - \dots - p_{m-1}) \sum_{i=1}^n I_{\{x_i=m\}}, \\ \partial l^{(n)}(x_n; p) / \partial p_j &= p_j^{-1} \sum_{i=1}^n I_{\{x_i=j\}} - (1 - p_1 - \dots - p_{m-1})^{-1} \times \\ &\times \sum_{i=1}^n I_{\{x_i=m\}}, \quad j = 1, \dots, m-1. \end{aligned}$$

Уравнения правдоподобия запишем в виде

$$p_j / (1 - p_1 - \dots - p_{m-1}) = \sum_{i=1}^n I_{\{x_i=j\}} / \sum_{i=1}^n I_{\{x_i=m\}}, \quad j = 1, \dots, m-1, \quad (17)$$

Складывая соотношения (17), имеем

$$p_1 + \dots + p_{m-1} = 1 - \frac{1}{n} \sum_{i=1}^n I_{\{x_i=m\}}. \quad (18)$$

Подставляя (18) в (17), получаем $p_j^{(n)} = v_j^{(n)} / n$, $i = 1, \dots, \dots, m-1$, где через $v_j^{(n)} / n$ обозначена частота исхода j ; $j = 1, \dots, m-1$.

Также и о. м. п. параметрической функции p_m является частота $v_m^{(n)} / m$ исхода m . Статистика $W_1^{(n)}$ имеет вид

$$W_1^{(n)} = 2 \sum_{j=1}^m v_j^{(n)} \ln \frac{v_j^{(n)}}{n} - 2 \sum_{j=1}^m v_j^{(n)} \ln p_j^0 = 2 \sum_{j=1}^m v_j^{(n)} \ln \frac{v_j^{(n)}}{n p_j^0}.$$

Найдем матрицу Фишера отдельного наблюдения. Имеем

$$\partial^2 l(x_i; p) / \partial p_j^2 = -p_j^{-2} I_{\{x_i=j\}} - (1 - p_1 - \dots - p_{m-1})^{-2} I_{\{x_i=m\}}$$

$$\partial^2 l(x_i; p) / \partial p_j \partial p_k = -(1 - p_1 - \dots - p_{m-1})^{-2} I_{\{x_i=m\}}, \quad j \neq k.$$

$$M_p \partial^2 l(X_1; p) / \partial p_j^2 = -p_j^{-1} - p_m^{-1},$$

$$M_p \partial^2 l(X_1; p) / \partial p_j \partial p_k = -p_m^{-1}, \quad j \neq k.$$

$$I_{X_1}(\mathbf{p}) = p_m^{-1} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} + \begin{bmatrix} p_1^{-1} & 0 & \dots & 0 \\ 0 & p_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{m-1}^{-1} \end{bmatrix}.$$

Вычислим статистику $W_2^{(n)}$:

$$\begin{aligned} n(\mathbf{v}^{(n)}/n - \mathbf{p}) I_{X_1}(\mathbf{p}) (\mathbf{v}^{(n)}/n - \mathbf{p})' = \\ = n^{-1} \left(p_m^{-1} \sum_{i,j=1}^{m-1} (v_i^{(n)} - np_i)(v_j^{(n)} - np_j) + \sum_{i=1}^{m-1} p_i^{-1} (v_i^{(n)} - np_i)^2 \right). \quad (19) \end{aligned}$$

Поскольку первое слагаемое правой части (19) равно

$$\begin{aligned} p_m^{-1} \left(\sum_{i=1}^{m-1} (v_i^{(n)} - np_i) \right)^2 = p_m^{-1} \left(\sum_{i=1}^{m-1} v_i^{(n)} - \sum_{i=1}^{m-1} np_i \right)^2 = \\ = p_m^{-1} (v_m^{(n)} - np_m)^2, \quad (20) \end{aligned}$$

то получаем

$$W_2^{(n)} = \sum_{i=1}^m \frac{(v_i^{(n)} - np_i^0)^2}{np_i^0}.$$

Итак, статистика $W_2^{(n)}$ имеет предельное распределение χ^2_{m-1} . Критерий со статистикой (19) называется χ^2 -критерием Пирсона. Выпишем статистику $W_3^{(n)}$, используя (19), (20):

$$\begin{aligned} n(\mathbf{v}^{(n)}/n - \mathbf{p}) I_{X_1}(\mathbf{v}^{(n)}/n) (\mathbf{v}^{(n)}/n - \mathbf{p})' = \\ = n^{-1} \left((v_m^{(n)}/n)^{-1} (v_m^{(n)} - np_m)^2 + \sum_{i=1}^{m-1} (v_i^{(n)}/n)^{-1} (v_i^{(n)} - np_i)^2 \right) = \\ = \sum_{i=1}^m (v_i^{(n)} - np_i)^2 / v_i^{(n)}. \end{aligned}$$

Критерий со статистикой

$$W_3^{(n)}(\mathbf{x}_n) = \sum_{i=1}^m \frac{(v_i^{(n)} - np_i^0)^2}{v_i^{(n)}}$$

также называют критерием χ^2 .

Пример [5]. При считывании шкалы, когда последняя цифра оценивается на глаз, часто наблюдается предпочтение одних цифр другим. Следующие данные представляют частоты последних цифр при 200 считываниях шкалы:

j	0	1	2	3	4	5	6	7	8	9
v_j	35	16	15	17	17	19	11	16	30	24

Если считать, что наблюдатель не имеет предпочтений, то мы имеем дело с повторной выборкой объема $n=200$ из полиномиального распределения с 10 равновероятными исходами: гипотеза H_0 . Рассчитаем значение статистики (19):

$v_j - 20$	15	-4	-5	-3	-3	-1	-9	-4	-10	4
$(v_j - 20)^2 / 20$	11,25	0,80	1,25	0,45	0,45	0,05	4,05	0,80	5,00	0,80

Получаемое значение есть 24, 90. По табл. 2.1а [1] для распределения χ^2 при числе степеней свободы 9 находим с помощью линейной интерполяции, что вероятностью наблюдать такое же или большее значение равна 0,0031. Это выходит за предел минимального уровня значимости 0,005, принятого в статистических таблицах, и гипотезу H_0 следует отвергнуть.

Рассчитаем для тех же данных значение статистики $W_3^{(n)}$. Имеем

$(v_j - 20)^2$	225	16	25	9	9	1	81	16	100	16
$(v_j - 20)^2 v_j$	6,43	1,00	1,66	0,53	0,53	0,05	7,36	1,0	3,33	1,0

Значение статистики равно 22,55, вероятность наблюдать такое же или большее значение равна 0,0074.

Для статистики $W_1^{(n)}$ имеем

v_j / np_j	1,75	0,8	0,75	0,85	0,85	0,95	0,55	0,8	1,5	1,2
$\ln(v_j / np_j)$	0,56	-0,22	-0,29	-0,16	-0,16	-0,05	-0,59	-0,22	0,405	0,182
$v_j \ln(v_j / np_j)$	19,59	-3,52	-4,35	-2,72	-2,72	-0,55	-9,44	-3,52	12,15	4,32

Значение статистики равно 18,39, вероятность наблюдать такое же или большее значение равна 0,030.

ЛИТЕРАТУРА

1. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М.: Наука, 1983.
2. Феллер В. Введение в теорию вероятностей и ее приложения, т. 1, 2. М.: Мир, 1984.
3. Андерсен Т. Статистический анализ временных рядов. М.: Мир, 1976.
4. Кокс Д., Льюис П. Статистический анализ последовательностей событий. М.: Мир, 1969.
5. Хальд А. Математическая статистика с техническими приложениями. М.: ИЛ, 1956.
6. Шеффе Г. Дисперсионный анализ. М.: Наука, 1980.
7. Рао С. Р. Линейные статистические методы и их применение. М.: Наука, 1968.
8. Сархан А., Гринберг Б. Введение в теорию порядковых статистик. М.: Статистика, 1970.
9. Дейвид Г. Порядковые статистики. М.: Наука, 1979.
10. Ван дер Варден Б. Л. Математическая статистика. М.: ИЛ, 1960.
11. Леман Э. Проверка статистических гипотез. М.: Наука, 1979.
12. Боровков А. А. Математическая статистика. М.: Наука, 1984.
13. Закс Ш. Теория статистических выводов. М.: Мир, 1975.
14. Кокс Д., Хинкли Д. Теоретическая статистика. М.: Мир, 1978.
15. Колмогоров А. Н. Основные понятия теории вероятностей, изд. 2. М.: Наука, 1974.
16. Ширяев А. Н. Вероятность. М.: Наука, 1980.
17. Гнеденко Б. В. Курс теории вероятностей. М.: Наука, 1969.
18. Розанов Ю. А. Теория вероятностей, случайные процессы и математическая статистика. М.: Наука, 1985.
19. Севастьянов Б. А. Курс теории вероятностей и математической статистики. М.: Наука, 1982.
20. Крамер Г. Математические методы статистики. М.: Мир, 1975.
21. Тутубалин В. Н. Теория вероятностей. М.: Изд-во МГУ, 1972.
22. Ивченко Г. И., Медведев Ю. И. Математическая статистика. М.: Высшая школа, 1984.
23. Барра Ж.-Р. Основные понятия математической статистики. М.: Мир, 1974.
24. Кендалл М. Дж., Стьюарт А. Теория распределений. М.: Наука, 1966.
25. Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973.
26. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976.