Министерство общего и профессионального образования РФ

Санкт-Петербургский государственный
электротехнический университет "ЛЭТИ"

А. Н. Лившиц
С. В. Малов

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Санкт-Петербург
1999

Министерство общего и профессионального образования РФ

Санкт-Петербургский государственный
электротехнический университет "ЛЭТИ"

А. Н. ЛИВШИЦ С. В. МАЛОВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Санкт-Петербург
1999

УДК 519.2
ББК В172я7
Л 55

Лившиц А.Н., Малов С.В. Математическая статистика: Учеб.пособие / Под ред. В. А. Егорова; СПбГЭТУ "ЛЭТИ". СПб., 1999. 68 с.

Представляет собой вводный курс математической статистики. Помимо традиционных разделов, посвященных классическим методам математической статистики, в пособие включен специальный раздел, представляющий собой введение в теорию обработки цензурированных данных типа времени жизни.

Построено на базе курсов по математической статистике, читаемых авторами в СПбГЭТУ. Предназначено для студентов различных специальностей.

Рецензенты: кафедра высшей математики СПбГТУРП;
д-р физ.-мат. наук, проф. Я. Ю. Никитин (СПбГУ).

Утверждено
редакционно-издательским советом университета
в качестве учебного пособия

Александр Нахимович Лившиц, Сергей Васильевич Малов

Математическая статистика

Учебное пособие

Редактор И. Г. Скачек
ЛР N 020617 от 24.06.98

Подписано в печать 08.07.99. Формат 60 × 84 1/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 3,95. Уч.-изд. л. 4,25.
Тираж 125. Зак.
Издательско-полиграфический центр СПбГЭТУ "ЛЭТИ"
197376, С.-Петербург, ул. Проф. Попова, 5

ISBN 5-7629-0273-0

Введение

В жизни довольно часто приходится делать выводы и принимать решения в условиях недостаточной информации и неполной определенности. В современной науке, в частности, в физике важным является выбор вероятностной модели изучаемой системы или явления, т. е. должны быть сделаны определенные предположения о распределении некоторых случайных величин (скажем, принадлежность этих распределений к как-либо параметризованному фиксированному множеству), которые должны существенно уточниться в результате некоторых эсперимента или наблюдений. Статистическая устойчивость, обеспечивающая такую возможность, присуща физическим процессам. Разумеется, представляет интерес вопрос о точности, с которой мы можем делать наши статистические выводы.

Первыми крупными работами в области математической статистики были работы Я. Бернулли. В начале XIX в. появились работы П. Лапласа и К. Гаусса, посвященные обработке результатов астрономических наблюдений. Фундаментальные результаты о случайном рассеивании получили П. Л. Чебышев, А. А. Марков, А. М. Ляпунов. Потребности исследований по биологии были стимулом для появления работ Ф. Гальтона (конец XIX в.) и, первоначально, его последователя К. Пирсона, работавшего в начале XX в., а также В. Госсета (Стьюдента). Начало исследований Р. Фишера в области математической статистики, предложившего много важных концепций и понятий и получившего множество результатов, датируется 1912-м г. В работах Ю. Неймана и Э. Пирсона (30-е годы) было положено начало теории проверки статистических гипотез и даны формулировки соответствующих оптимизационных задач. В 40-е годы А. Вальд и его коллеги выдвинули идеи последовательного анализа и теории аналитических решений. Велик вклад в развитие математической статистики таких ученых, как А. Н. Колмогоров Б. В. Гнеденко, С. Р. Рао, Г. Крамер, М. Кендалл, А. Стюарт, Э. Леман, Ю. В. Линник, Н. В. Смирнов, Л. Н. Большев, Ю. В. Прохоров и др.

Диапазон применения статистики необычайно широк. Модели и методы обработки статистических данных интересны, но не всегда просты. Однако труд их изучения должен окупиться, ибо имеется очень много практических применений. Знание теории упростит понимание и применение практических рекомендаций.

Предметом изучения в пособии являются основные понятия и методы математической статистики. Рассматриваемый материал базируется на основе курса математической статистики, читаемого на факультете автоматики СПбГЭТУ, и может быть использован при изучении данного предмета студентами технических вузов и университетов.

Предполагается, что читатель знаком с основными понятиями теории вероятностей, а также математического анализа, теории меры и интеграла и линейной алгебры.

Следует отметить, что представленный материал может быть интересен

как для человека, желающего познать азы математической статистики, так и для "пользователя"(инженера, экспериментатора, производственника, финансиста и пр.). При этом некоторые математические выкладки и определения можно опускать, ограничиваясь лишь описательной частью и примерами использования изложенных методов математической статистики.

Материал пособия разбит на пять разделов, соответствующих различным направлениям математической статистики. В первом и втором разделах рассматриваются задачи оценивания по результатам наблюдений неизвестного распределения и параметров распределения, принадлежащего некоторому параметрическому семейству соответственно. Третий посвящен проверке статистических гипотез. В четвертом описываются некоторые методы многомерного статистического анализа. Последний представляет собой введение в теорию анализа данных типа времени жизни, бурно развивающуюся в настоящее время. Отметим, что в современных версиях статистических систем Statistica, SPSS и SAS данной тематике отведена не последняя роль.

Задача статистики состоит в сборе данных, их анализе и интерпретации. Будем считать, что данные уже получены. В некотором смысле задачи математической статистики являются обратными к задачам теории вероятностей. Если задача теории вероятностей заключается в изучении результатов наблюдений в условиях рассматриваемого эксперимента, то задача математической статистики состоит в изучении свойств эксперимента на основании полученных данных. Известный тезис, что критерий истины есть практика, имеет непосредственное отношение к математической статистике. Отметим, что при изучении свойств эксперимента помимо полученных данных могут использоваться также некоторые известные свойства эксперимента, с которыми мы приступаем к анализу данных. Эти свойства отражаются при построении статистической модели эксперимента.

Статистическим экспериментом будем называть тройку объектов $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ – семейство вероятностных мер (или распределений, если $\mathfrak{X} = \mathbb{R}^n$), Θ определяется известными свойствами эксперимента. Напомним, что при каждом фиксированном $\theta \in \Theta$ тройка $(\mathfrak{X}, \mathfrak{F}, P_\theta)$ является вероятностным пространством. Однако невозможно делать серьезные статистические выводы, базируясь на одном статистическом эксперименте, без каких-либо дополнительных предположений. Идея математической статистики состоит в рассмотрении независимого набора статистических экспериментов и, в частности, неоднократного повторения эксперимента. Иными словами, рассматриваемый статистический эксперимент $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$ допускает представление в виде прямого произведения статистических экспериментов (простых) $\mathfrak{X} = \mathfrak{X}_1 \times \cdots \times \mathfrak{X}_n$, т. е. исход эксперимента представляет собой набор $(X_1, \ldots, X_n)$, где $X_i \in \mathfrak{X}_i$, $i = 1, 2, \ldots, n$ – исходы простых экспериментов; $\mathfrak{F} = \sigma(\mathfrak{F}_1 \times \cdots \times \mathfrak{F}_n)$ – σ-алгебра (сигма-алгебра), порожденная $\mathfrak{F}_1 \times \cdots \times \mathfrak{F}_n$. Иными словами, событиями являются всевозможные наборы $(A_1, \ldots, A_n)$, $A_i \in \mathfrak{F}_i$, а также их не более чем счетные объединения и пере-

сечения. Свойство независимости простых экспериментов задается условием $\mathcal{P} = \{P_{1,\theta} \times \cdots \times P_{n,\theta},\ \theta \in \Theta\}$, т. е. $\mathbf{P}_\theta(A_1, \ldots, A_n) = P_{1,\theta}(A_1) \cdots P_{n,\theta}(A_n)$ для любого набора $A_i \in \mathfrak{F}_i$, $i = 1, 2, \ldots, n$. В частности, при повторных экспериментах $P_{i,\theta} = P_{1,\theta} = P_\theta$ при всех $i = 2, 3, \ldots, n$.

В данной модели предполагается, что все события σ-алгебры $\mathfrak{F}$ являются наблюдаемыми, т. е. на основании полученных данных можно сказать, произошло ли то или иное событие или нет. В случае потери части данных вводится наблюдаемая σ-алгебра событий $\mathfrak{A} \subset \mathfrak{F}$. Модели такого типа удобно использовать при анализе неполных или цензурированных данных, а также в случае, если информация о результатах эксперимента поступает с течением времени. Рассмотрим следующие примеры.

Пример 1. Эксперимент состоит в наблюдении за работой некоторой однородной группы технических систем в течение времени T. Задача: оценить надежность (т. е. время безотказной работы) систем данного класса. Данные представляют собой времена отказов элементов исходной группы до момента T, а также информацию о том, что остальные элементы не отказали к моменту времени T. В качестве σ-алгебры $\mathfrak{F}$ удобно выбрать σ-алгебру, порожденную временами отказов компонент. В этом случае $\mathfrak{F} \neq \mathfrak{A}$, поскольку наблюдения не содержат времен отказов, больших T.

Среди задач математической статистики можно выделить несколько типов:

1. По результатам наблюдений выбрать из исходного семейства распределений P_{θ_0}, наиболее соответствующее полученным данным в том или ином смысле (*точечное оценивание*).

2. По результатам наблюдений выбрать подсемейство распределений $\{P_\theta,\ \theta \in \Theta_0\}$, с достаточно большой степенью достоверности содержащее истинное распределение (неизвестное, но существующее), при котором проводился эксперимент (*интервальное оценивание*).

3. На основании полученных данных принять или отвергнуть некоторое утверждение об истинном распределении P_θ (*проверка статистических гипотез*).

Обычно при проведении статистических экспериментов данные поступают (или могут быть интерпретированы) в виде набора случайных величин или векторов. Поэтому без ограничения общности можно считать, что $\mathfrak{X}_i = \mathbb{R}$ или $\mathfrak{X}_i = \mathbb{R}^k$ соответственно, т. е. исход эксперимента представляет собой набор случайных величин (векторов) $X_1, \ldots, X_n$, а семейства $P_{i,\theta}$, $i = 1, \ldots, n$ – соответствующие классы распределений. Эти случайные величины (векторы) будем называть *наблюдениями*. В частности, если все наблюдения идентичны, т. е. имеют одинаковые распределения, то исходная совокупность наблюдений называется *выборкой*. Иногда вместо термина статистический эксперимент используется термин выборочное пространство. При каждом фиксированном значении параметра θ изучение свойств выборок идентично

изучению свойств независимых одинаково распределенных случайных величин с распределением P_θ.

Отображение на множестве наблюдений $T : \mathbb{R}^n \to \mathcal{S}$ ($T : \mathbb{R}^{kn} \to \mathcal{S}$) называется *статистикой*. Отметим, что если $\mathcal{S} = \mathbb{R}^m$, то статистика, как функция от случайных векторов, сама является случайным вектором. Тем самым при каждом фиксированном значении параметра для исследования свойств статистик могут быть использованы вероятностные методы.

Пример 2. Пусть $X_1, X_2, \ldots, X_n$ – набор наблюдений (случайных величин). Отображение $T : \mathbb{R}^n \to \mathbb{R}^n$, сопоставляющее исходной совокупности *вариационный ряд* $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, т. е. тот же набор величин, но упорядоченных по возрастанию, является статистикой. Элементы вариационного ряда называются *порядковыми статистиками*. Так, вариационный ряд совокупности (1, 3, 2, 3) имеет вид (1, 2, 3, 3).

Статистика $T : \mathbb{R}^n \to \Theta^*$, где $\Theta \subseteq \Theta^*$, используемая для оценивания параметра называется *оценкой* параметра. В следующих разделах будут рассмотрены задачи и методы теории точечного и интервального оценивания параметра, проверки статистических гипотез, многомерного анализа, а также элементы анализа цензурированных данных типа времени жизни.

1. Выборочный метод

1.1. Эмпирические распределения

Пусть $X_1, X_2, \ldots, X_n$ – выборка из распределения P_θ, $\theta \in \Theta$. Истинное значение P_θ будем называть *теоретическим распределением*.

По исходной выборке построим дискретное распределение, имеющее атомы $1/n$ в точках $X_1, X_2, \ldots, X_n$. Оно называется *эмпирическим распределением*, построенным по данной выборке, а соответствующая функция распределения F_n называется *эмпирической функцией распределения*. Эмпирическая функция распределения имеет вид

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i < x\},$$

где

$$\mathbb{1}_A = \begin{cases} 1, & \text{если } A \text{ имеет место;} \\ 0, & \text{в остальных случаях.} \end{cases}$$

Иными словами, значение эмпирической функции распределения в точке x есть отношение числа наблюдений меньших x к общему числу наблюдений.

Эмпирическая функция распределения является статистикой, и ее можно рассматривать в качестве оценки для истинной (или теоретической) функции распределения элементов выборки. Очевидно, что эмпирические функции распределения, построенные по исходной выборке и по соответствующе-

му вариационному ряду, совпадают. Следующая теорема оправдывает введение эмпирической функции распределения.

Теорема 1.1. /Гливенко – Кантелли/ *Пусть $X_1, X_2, \ldots, X_n$ – выборка из распределения P_θ, $\theta \in \Theta$; $F(x)$ и $F_n(x)$ – теоретическая и эмпирическая функции распределения соответственно. Тогда с вероятностью 1*

$$\lim_{n\to\infty} D_n(\vec{X}) = \lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \xrightarrow[n\to\infty]{} 0.$$

Поточечная сходимость получается непосредственно из усиленного закона больших чисел в схеме Бернулли с вероятностями успеха $F(x)$, а равномерная получается, поскольку значения на $\pm\infty$ у функций $F(x)$ и $F_n(x)$ совпадают и конечны.

Отклонение $D_n(\vec{X})$ теоретической функции распределения от эмпирической называется *статистикой Колмогорова.*[1] Ее асимптотическое поведение описывает следующая теорема.

Теорема 1.2. /Колмогорова/ *Если $F(x)$ непрерывна, то для любого положительного значения t*

$$\lim_{n\to\infty} \mathbf{P}(\sqrt{n}D_n \le t) = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2t^2}.$$

Функция $K(t)$ называется *функцией распределения Колмогорова.* На практике данная теорема дает хорошее приближение уже при $n \ge 20$. Как осуществлять это практическое применение? Забегая вперед, опишем алгоритм.

1. Задаемся малым числом $\delta > 0$ – точностью, с которой мы хотим оценить неизвестное распределение по определяемому выборкой эмпирическому и близкой к 1 вероятностью $1-\alpha$, с которой мы должны быть уверены в правильности нашей оценки.

2. Находим такое z_α, что $1 - K(z_\alpha) = \alpha$.

3. Находим такое n, что $\dfrac{z_\alpha}{\sqrt{n}} \leqslant \delta$.

Если выборка имеет объем n, то неравенство $F_n^*(x) - \delta < F(x) < F_n^*(x) + \delta$ выполняется одновременно для всех x с вероятностью $\geqslant 1 - \alpha$. То есть, если, скажем, выборка из 100 элементов получилась $0, 0.01, 0.02, \ldots, 0.99$, то распределение очень близко к равномерному на отрезке [0, 1] (эмпирическая функция распределения, соответствующая выборке, близка в каждой точке отрезка к теоретической функции равномерного распределения) с вероятностью, близкой к единице. Конечно, получение в точности такой выборки – весьма маловероятное событие.

[1] Мы подразумеваем, что теоретическая функция распределения F известна. Иначе, указанное отклонение не является статистикой, т. к. зависит от неизвестной функции F.

1.2. Выборочные характеристики

Известно, что некоторые свойства распределений задаются числовыми характеристиками. Выделим два типа числовых характеристик:

1) числовые характеристики, представимые в виде

$$H(\mathbf{E}\,G_1(X),\dots,\mathbf{E}\,G_k(X)) = H\Big(\int_{-\infty}^{\infty} G_1(x)\,dF(x),\dots,\int_{-\infty}^{\infty} G_k(x)\,dF(x)\Big),$$

где H – некоторая непрерывная функция k аргументов;

2) числовые характеристики, представимые в виде непрерывного в равномерной метрике[2] (хотя бы в окрестности теоретической функции распределения) функционала от функции распределения – $\mathbf{G}(F)$.

К характеристикам первого типа относятся моменты ($h \equiv 1$) и, в частности, $\mathbf{E}\,X$ – математическое ожидание, $\mathbf{E}\,X^2 - (\mathbf{E}\,X)^2$ – дисперсия, $\mathbf{E}\,X^k$ – момент порядка k, $\mathbf{E}\,(X - \mathbf{E}\,X)^k$ – центральные моменты, $k \in \mathbb{N}$, и т. д. К характеристикам второго типа относятся, например, квантили $\zeta_p = \min\{x \in \mathbb{R} : F(x) \geq p\}$, $p \in (0,1)$. Квантиль чаще определяется (не всегда однозначно) из соотношений $\mathbf{P}(X \leq \zeta_p) \geq p$, $\mathbf{P}(X \geq \zeta_p) \geq 1 - p$.

Конечно, некоторые числовые характеристики являются характеристиками обоих типов одновременно.

Наряду с теоретическими числовыми характеристиками, которые, вообще говоря, неизвестны, можно рассматривать их выборочные аналоги, т. е. соответствующие числовые характеристики эмпирического распределения. Для статистик первого типа получаем выборочные характеристики:

$$H\Big(\int_{-\infty}^{\infty} G_1(x)\,dF_n(x),\dots,\int_{-\infty}^{\infty} G_k(x)\,dF_n(x)\Big) = H\Big(\frac{1}{n}\sum_{i=1}^{n} G_1(X_i),\dots,\frac{1}{n}\sum_{i=1}^{n} G_k(X_i)\Big).$$

Для статистик второго типа выборочные характеристики имеют вид $\mathbf{G}(F_n)$. Для них справедлива следующая теорема.

Теорема 1.3. *Пусть $X_1, X_2, \dots, X_n$ – выборка из распределения F. Тогда, если при каждом $P_\theta \in \mathcal{P}$, числовая характеристика $\mathbf{G}(F)$ первого или второго типа существует, то с вероятностью 1 с ростом размера выборки*

$$\mathbf{G}(F_n) \xrightarrow[n\to\infty]{} \mathbf{G}(F).$$

Доказательство. Выражение $\frac{1}{n}\sum_{i=1}^{n} G_j(X_i)$ представляет собой сумму независимых одинаково распределенных случайных величин с математическим ожиданием $\int_{-\infty}^{\infty} G_j(x)\,dF(x)$, $j = 1,\dots,k$. Тогда, используя усиленный закон больших чисел, и непрерывность функции H получаем требуемое утверждение.

[2]Непрерывность в равномерной метрике означает, что если $\sup_x |F_n(x) - F(x)| \to 0$ при $n \to \infty$, то $G(F_n) \to G(F)$.

Для статистик второго типа утверждение теоремы следует непосредственно из теоремы Гливенко – Кантелли. ■[1]

Приведем выборочные аналоги основных числовых характеристик:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{(выборочное среднее)},$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \text{(выборочная дисперсия)},$$

$$\alpha_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad \text{(выборочные моменты)},$$

$$\beta_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \qquad \text{(выб. центральные моменты)}.$$

Выборочные квантили выражаются через порядковые статистики следующим образом: $\hat{\zeta}_p = X_{([np]+1)}$, если np – нецелое, и любое число из интервала $[X_{(np)}, X_{(np+1)})$, если np – целое.

Понятие эмпирического распределения естественным образом обобщается на случай векторных выборок. Эмпирическое распределение представляет собой распределение с одинаковыми атомами, совпадающими с наблюдениями. Большинство теорем переносятся (с небольшими изменениями) на многомерный случай. Среди выборочных характеристик векторов, построенных по выборке $(\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n)$, где $\vec{X}_s = (X_{s,1}, \ldots, X_{s,m})$, отметим

$$\mathbf{S} = ||S_{i,j}|| = \frac{1}{n}\sum_{k=1}^{n}(X_{k,i} - \overline{X_i})^T(X_{k,j} - \overline{Y_j}) \qquad \text{(выборочные ковариации)},$$

$$\mathbf{r} = ||r_{i,j}|| = \frac{S_{i,j}}{\sqrt{S_{i,i}}\sqrt{S_{j,j}}} \qquad \text{(выборочные корреляции)}.$$

Некоторые выборочные характеристики могут использоваться для оценивания параметров теоретического распределения. В связи с этим могут оказаться существенными моменты этих выборочных характеристик.

Имеем для выборочного среднего:

$$\mathbf{E}\,(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\,X_i = \mathbf{E}\,X_1, \quad \mathbf{D}\overline{X} = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}X_i = \frac{1}{n}\mathbf{D}X_1 = \frac{\mu_2}{n}.$$

Для выборочной дисперсии (не умаляя общности, можно считать наши случайные величины центрированными), поскольку $s^2 = \frac{1}{n}\sum_1^n X_i^2 - \overline{X}^2$,

[1] Знак ■ обозначает конец доказательства.

имеем

$$\mathbf{E}\, s^2 = \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \mathbf{E}\,\overline{X}^2 = \mu_2 - \frac{\mu_2}{n} = (n-1)\frac{\mu_2}{n}.$$

Вычисление дисперсии выборочной дисперсии основывается на том, что

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\Big(\sum_{i=1}^{n} X_i\Big)^2 = \frac{n-1}{n^2}\sum_{i=1}^{n} X_i^2 - \frac{2}{n^2}\sum_{i<j} X_i X_j.$$

Снова считая исходные величины центрированными, можно найти

$$\mathbf{D}s^2 = \mathbf{E}\,(s^2)^2 - (\mathbf{E}\, s^2)^2 = \frac{(n-1)^2}{n^3}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right).$$

1.3. Асимптотическая нормальность выборочных квантилей

Рассмотрим важный для дальнейшего вопрос о распределении порядковых статистик $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, полученных по выборке из абсолютно непрерывного распределения с функцией распределения $F(x)$ и плотностью $f(x)$. Какова плотность распределения статистики $X_{(k)}$?

Допустим, что нам известно, какой из элементов выборки X_s есть $X_{(k)}$, какие $(k-1)$ элементов меньше его и какие $(n-k)$ – больше. Зафиксировав это разбиение выборки на три части (одна – одноэлементная), зададимся вопросом о том, какова вероятность того, что $X_s = X_{(k)}$ лежит в малой окрестности длины Δt некоторой точки t. Эта вероятность есть приблизительно произведение трех частей $f(t)\Delta t$, $F(t)^{k-1}$ и $(1-F(t))^{n-k}$ по определению функции распределения и плотности распределения. Сколько существует таких разбиений выборки на три части? Одноэлементное множество можно выбрать n способами, разбить оставшееся $(n-1)$-элементное множество на две части из $(k-1)$ и $(n-k)$ элементов можно C_{n-1}^{k-1} способами. Поэтому вероятность того, что $X_{(k)}$ лежит в малой окрестности точки t длины Δt, равна приблизительно (с точностью до малых второго порядка) $n\, C_{n-1}^{k-1} F(t)^{k-1}(1-F(t))^{n-k} f(t)\Delta t$. Следовательно, плотность распределения $x_{(k)}$ есть $f_{(k)}(t) = nC_{n-1}^{k-1}F(t)^{k-1}(1-F(t))^{n-k}f(t)$. Аналогично можно вычислить совместную плотность распределения нескольких порядковых статистик. Например, плотность распределения вариационного ряда имеет вид

$$f_{(1),\ldots,(n)}(x_1,\ldots,x_n) = n!\, f(x_1)\cdots f(x_n)\, \mathbb{1}_{\{x_1 \le \ldots \le x_s\}}.$$

Далее предполагаем, что теоретическое распределение имеет плотность $f(x)$, которая непрерывна и положительна в окрестности квантили ζ_p, $f(\zeta_p) > 0$ (тем самым однозначно определенной). Рассмотрим выборочную квантиль $Z_{n,\,p}$ порядка p: $Z_{n,\,p} = X_{(np)}$ при $np \in \mathbb{N}$, и $Z_{n,\,p} = X_{([np]+1)}$ в остальных случаях.

Теорема 1.4. /об асимптотическом поведении выборочных квантилей/ *Пусть* $\eta_n = \sqrt{\frac{n}{pq}}\, f(\zeta_p)(Z_{n,p} - \zeta_p)$, $q = 1 - p$, $p \in (0,1)$. *Тогда при* $n \to \infty$ *имеет место сходимость по распределению* $\eta_n \Longrightarrow N(0,1)$, *т. е.*

$$\mathbf{P}(\eta_n < x) \xrightarrow[n\to\infty]{} \Phi(x), \quad x \in \mathbb{R},$$

где

$$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\, dt$$

есть функция стандартного нормального распределения $N(0,1)$.

Доказательство. Пусть $k = np$, если np – целое, и $k = [np] + 1$ в остальных случаях. Рассмотрим случайные величины

$$Y_i = \sqrt{\frac{n}{pq}}\; f(\zeta_p)(X_i - \zeta_p), \quad J_i(t) = \mathbb{1}_{\{Y_i < t\}},$$

$i = 1, 2, \ldots, n$, где $\mathbb{1}_A = 1$, если событие A произошло, и $\mathbb{1}_A = 0$ в противном случае. Очевидно, что $\eta_n = Y_{(k)}$. При этом $\mathbf{P}(Y_{(k)} < t) = \mathbf{P}(\sum_{i=1}^n J_i(t) \geq k)$. Отметим, что $J_1(t), \ldots, J_n(t)$ – испытания Бернулли с вероятностью успеха

$$\mathbf{P}(J_i(t) = 1) = \mathbf{P}(Y_i < t) = \mathbf{P}\left(X_i < \zeta_p + \sqrt{\frac{pq}{n}}\frac{t}{f(\zeta_p)}\right) = F\left(\zeta_p + \sqrt{\frac{pq}{n}}\frac{t}{f(\zeta_p)}\right).$$

Используя разложение Тейлора в окрестности ζ_p получаем

$$a = \mathbf{E}\, J_i = \mathbf{P}(J_i(t) = 1) = F(\zeta_p) + f(\zeta_p)\frac{t\sqrt{pq}}{\sqrt{n} f(\zeta_p)} + o(1/\sqrt{n}) =$$
$$= F(\zeta_p) + t\sqrt{pq/n} + o(1/\sqrt{n}), \qquad \mathbf{D}J_i = a(1-a).$$

Рассмотрим

$$\mathbf{P}\Big(\sum_{i=1}^n J_i \geq k\Big) = \mathbf{P}\left(\frac{\sum_{i=1}^n J_i - na}{\sqrt{na(1-a)}} \geq \frac{k - na}{\sqrt{na(1-a)}}\right).$$

Остается отметить, что

$$\frac{k - np - t\sqrt{np(1-p)} + o(\sqrt{n})}{\sqrt{np(1-p)(1+o(1))}} \xrightarrow[n\to\infty]{} -t.$$

Используя центральную предельную теорему и симметричность стандартного нормального распределения, получаем:

$$\mathbf{P}\Big(\sum_{i=1}^n J_i \geq k\Big) \xrightarrow[n\to\infty]{} 1 - \Phi(-t) = \Phi(t).$$

Что и требовалось доказать. ■

1.4. Выборка из нормального распределения

Важнейшую роль в математической статистике, как и в теории вероятностей, играет семейство нормальных распределений $N(a, \sigma^2)$. Помимо того, что нормальные модели подходят для описания широкого круга реальных процессов, нормальное распределение является центральным в асимптотической теории.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ – выборка из распределения $N(0,1)$. Введем некоторые распределения, используемые в математической статистике.

Рассмотрим случайную величину $\chi^2_n = \sum\limits_{i=1}^{n} \xi_i^2$. Говорят, что χ^2_n имеет χ^2(*хи-квадрат*)-*распределение* (или распределение Пирсона) с n степенями свободы. Плотность распределения величины χ^2_n имеет вид

$$k_n(t) = K_n'(t) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1}\exp(-x/2), & \text{если } x \geq 0; \\ 0, & \text{в остальных случаях,} \end{cases}$$

где $\Gamma(p)$, $p > 0$ – гамма-функция Эйлера, определяемая равенством

$$\Gamma(p) = \int\limits_0^\infty x^{p-1}\exp(-x)\,dx.$$

Семейство χ^2-распределений является подмножеством двухпараметрического семейства гамма-распределений $\Gamma(b,p)$, $p, b \geq 0$, с плотностями

$$\gamma_{p,\,b}(x) = \begin{cases} \dfrac{b^p}{\Gamma(p)} x^{p-1}\exp(-bx), & \text{если } x \geq 0; \\ 0, & \text{в остальных случаях,} \end{cases}$$

при $b = 1/2$; $p = n/2$, $n \in \mathbb{N}$. Известное свойство, что сумма двух независимых гамма-распределений $\Gamma(b,p)$ и $\Gamma(b,q)$ снова имеет гамма-распределение $\Gamma(b, p+q)$, здесь следует непосредственно из представления в виде суммы квадратов независимых нормальных величин.

Пусть случайная величина Y независима от χ^2_n. Рассмотрим случайную величину $T_n = \dfrac{Y}{\sqrt{n^{-1}\chi^2_n}}$. Распределение величины T_n называется распределением *Стьюдента* с n степенями свободы. Соответствующая плотность распределения имеет вид

$$s_n(x) = S_n'(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)}(1 + x^2/n)^{-(n+1)/2}.$$

Отметим, что плотность распределения Стьюдента симметрична относительно нуля.

Распределение *Фишера – Снедекора* $F(n_1, n_2)$ определяется как распределение случайной величины $\dfrac{n_2}{n_1}\zeta$, $\zeta = \eta_1/\eta_2$, где η_1, η_2 независимы и распределены как $\chi^2_{n_1}$ и $\chi^2_{n_2}$. Плотность распределения Фишера – Снедекора

представляется в виде

$$f(x) = \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{\Gamma((n_1+n_2)/2)}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{x^{n_1-1}}{(1+n_1x/n_2)^{(n_1+n_2)/2}}, \quad x>0.$$

Теорема 1.5. /Фишера/. *Пусть $X_1, X_2, \ldots, X_n$ – выборка из нормального распределения $N(a, \sigma^2)$. Тогда*

1) $\sqrt{n}\,\dfrac{\overline{X}-a}{\sigma} \in N(0,1)$;

2) $\overline{X}$ *и* s^2 *– независимые статистики;*

3) $\dfrac{ns^2}{\sigma^2}$ *имеет* χ^2*-распределение с* $(n-1)$ *степенью свободы;*

4) $\sqrt{n-1}\,\dfrac{\overline{X}-a}{s}$ *имеет распределение Стьюдента с* $n-1$ *степенями свободы.*

Доказательство. Введем случайные величины $Y_k = X_k - \mathbf{E}\,X_k$, $1 \le k \le n$. Они независимы и имеют нормальное распределение $N(0,\sigma^2)$. Рассмотрим произвольную ортонормированную $n \times n$-матрицу A, первая строка которой состоит из чисел, равных $\frac{1}{\sqrt{n}}$ (построение ортонормированного базиса). Пусть ξ_k, $1 \leqslant k \leqslant n$ – новые случайные величины, такие что $\xi = (\xi_1, \ldots, \xi_n)^{\mathbf{T}} = \frac{1}{\sigma} AY$, $Y = (Y_1, \ldots, Y_n)^{\mathbf{T}}$. В этом случае

$$\xi_1 = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} Y_k = \frac{\sqrt{n}}{\sigma}\,\overline{Y}, \quad \mathbf{E}\,\xi_k = 0, \quad \mathbf{D}\xi_k = 1,\ 1 \leqslant k \leqslant n,$$
$$\mathbf{cov}(\xi_k, \xi_l) = 0 \quad (k \neq l).$$

Кроме того, можно выразить s^2 через случайные величины ξ_k, $1 \leqslant k \leqslant n$, которые независимы, как некоррелированные нормальные $N(0,1)$. Поскольку A – ортогональна, суммы квадратов Y_k/σ^2 и ξ_k равны, так что

$$\frac{ns^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{k=1}^{n} Y_k^2 - \left(\frac{\overline{Y}}{\sigma}\right)^2 = \xi_2^2 + \ldots + \xi_n^2.$$

Отсюда уже следует независимость ns^2/σ^2 и $\overline{X} = \dfrac{\sigma}{\sqrt{n}}\,\xi_1 + a$. Утверждение о распределениях ns^2/σ^2 и $\overline{X}$ теперь очевидно. ■

Замечание 1.1. Если $X_1, \ldots, X_{n_1}$ и $Y_1, \ldots, Y_{n_2}$ – две независимые нормальные выборки из $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$ соответственно, то статистика s_1^2/s_2^2 имеет распределение Фишера – Снедекора $F(n_1, n_2)$.

2. Оценивание параметра

В данном разделе освещаются основные принципы точечного оценивания вещественнозначного (или векторно-значного) параметра и нахождения оценок оптимальных в том или ином смысле.

2.1. Постановка задачи точечного оценивания

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ – статистический эксперимент, результатом которого является набор наблюдений $X_1, X_2, \ldots, X_n$. Задача точечного оценивания заключается в том, чтобы используя результаты наблюдений, выбрать из множества параметров Θ значение, наиболее подходящее в том или ином смысле. Пусть в качестве оценки параметра θ (или функции от параметра $g(\theta)$) выбрана оценка $\delta(\vec{X})$. Для определения близости оценки к истинному значению параметра θ вводится *функция потерь* $W(\delta, \theta)$, удовлетворяющая следующим условиям:

1) неотрицательность: $W(\delta, \theta) \geq 0$;

2) если $\delta = \theta$, то потери нулевые: $W(\theta, \theta) = 0$.

Наиболее употребительными функциями потерь являются

$$W(\delta, \theta) = (\delta - \theta)^2 \qquad \text{(функция потерь Гаусса)},$$

$$W(\delta, \theta) = |\delta - \theta| \qquad \text{(функция потерь Лапласа)}.$$

Точность оценки измеряется *функцией риска*

$$R(\delta, \theta) = \mathbf{E}_\theta W(\delta(\vec{X}), \theta),$$

где $\mathbf{E}_\theta$ берется при условии, что распределение $\vec{X}$ соответствует значению параметра θ, т. е. средними потерями при оценивании с помощью δ. Хотелось бы найти оценку, минимизирующую риск при каждом значении θ. Однако в такой постановке задача неразрешима. Действительно, если выбрать в качестве оценки параметра θ некоторое постоянное значение $\delta \equiv \theta_0$, $\theta_0 \in \theta$, то при $\theta = \theta_0$ данная оценка абсолютно точна, т. е. имеет нулевой риск. Ясно, что подобная оценка с точки зрения математической статистики абсолютно бесполезна, однако приведенный пример показывает, что, за исключением тривиальных случаев (когда параметр определяется абсолютно точно), оценки, минимизирующей риск при каждом $\theta \in \Theta$, не существует. Для преодоления этой трудности можно ограничить класс рассматриваемых оценок.

Постулировав тезис о том, что с ростом числа наблюдений оценка должна становиться точнее и точность ее должна стремиться к абсолютной, введем следующее понятие.

Оценка $\delta(\vec{X})$ параметрической функции $g(\theta)$ называется *состоятельной*, если при каждом значении $\theta \in \Theta$ для любого $\epsilon > 0$

$$\mathbf{P}_\theta(|\delta(\vec{X}) - g(\theta)| > \epsilon) \xrightarrow[n\to\infty]{} 0,$$

и *сильно состоятельной*, если с вероятностью 1 $\delta(\vec{X}) \to g(\theta)$ при $n \to \infty$.

С точки зрения практических применений требования состоятельности и сильной состоятельности неразличимы (поскольку рассматриваются конечные наборы наблюдений), однако с точки зрения асимптотической теории требование сильной состоятельности более жесткое.

Ясно, что при выполнении условия состоятельности с ростом длины выборки оценка должна концентрироваться вокруг истинного значения параметра, т. е. ее среднее должно стремиться к θ.

Оценка $\delta(\vec{X})$ параметрической функции $g(\theta)$ называется *несмещенной*, если при любом значении параметра $\theta \in \Theta$

$$\mathbf{E}_\theta \delta(\vec{X}) = g(\theta),$$

и *асимптотически несмещенной*, если

$$\mathbf{E}_\theta \delta(\vec{X}) \xrightarrow[n\to\infty]{} g(\theta).$$

Смещением оценки называется величина

$$b_g(\theta) = \mathbf{E}_\theta \delta(\vec{X}) - g(\theta).$$

Понятие состоятельности оценки является аналогом сходимости в законе больших чисел в теории вероятностей. Естественно ввести понятие, близкое в том же смысле центральной предельной теореме.

Оценка δ параметрической функции $g(\theta)$ называется *асимптотически нормальной*, если при $n \to \infty$ имеет место сходимость по распределению

$$\sqrt{n}(\delta(\vec{X}) - g(\theta)) \Longrightarrow N(0, \sigma^2(\theta)).$$

Ясно, что если $X_1, X_2, \ldots, X_n$ – выборка из распределения P_θ с параметром $\theta = \mathbf{E}\, X_1$, то $\overline{X}$ является несмещенной оценкой для θ. Более того, в силу закона больших чисел она состоятельна, а при наличии $\mathbf{E}\, X_1^2 < \infty$ по центральной предельной – асимптотически нормальна.

Отметим, что выборочная дисперсия s^2 является состоятельной оценкой дисперсии $\theta = \sigma^2$, где $\sigma^2 = \mathbf{E}\,(X_1 - \mathbf{E}\, X_1)^2$ (по теореме о состоятельности выборочных характеристик), но не является несмещенной. С учетом независимости наблюдений (не умаляя общности, считаем $\mathbf{E}\, X_1 = 0$)

$$\mathbf{E}\, s^2 = \mathbf{E}\, \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \mathbf{E}\, (\overline{X})^2 = \sigma^2 - \sigma^2/n = \frac{n-1}{n}\sigma^2.$$

Тогда смещение $b(\sigma^2) = -\sigma^2/n$. Нетрудно заметить, что несмещенной оценкой дисперсии будет $s'^2 = ns^2/(n-1)$. Для асимптотической нормальности s^2 достаточно наличия $\mathbf{E}\, X_1^4 < \infty$.

Найдем наилучшую оценку дисперсии в классе всех оценок вида $s(\lambda) = \lambda s'^2$, где $\lambda > 0$, по выборке из нормального распределения $N(a, \sigma^2)$. Только при $\lambda = 1$ оценка $s(\lambda)$ является несмещенной $\mathbf{E}_\theta s'^2 = \sigma^2$. Вычислим среднеквадратичное отклонение (риск) оценок этого типа:

$$\mathbf{E}_\theta (s(\lambda) - \sigma^2)^2 = \mathbf{E}_\theta (\lambda(s'^2 - \sigma^2) + (\lambda - 1)\sigma^2)^2 = \lambda^2 \mathbf{D} s'^2 + (\lambda - 1)^2 \sigma^4.$$

По теореме Фишера $(n-1)s'^2/\sigma^2$ имеет распределение χ^2 с $(n-1)$ степенью свободы. Дисперсия распределения χ^2_{n-1} равна $2\,(n-1)$. Тогда $\mathbf{D}s'^2 = = 2\sigma^4/(n-1)$. Следовательно

$$\mathbf{E}_\theta(s(\lambda)-\sigma^2)^2 = \left(\frac{2\lambda^2}{n-1}+(\lambda-1)^2\right)\sigma^4.$$

Ясно, что минимум выражения в скобках достигается при $\lambda = \frac{n-1}{n+1}$. Таким образом, наилучшей в указанном классе оценок дисперсии нормальной выборки является смещенная оценка $\dfrac{1}{n+1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$. Отметим, что асимптотически риск данной оценки эквивалентен риску несмещенной оценки s'^2, т. е. несмещенная оценка при больших n тоже неплоха. В широком классе случаев требование несмещенности все-таки предъявляется.

Следует отметить, что несмещенная оценка не всегда может быть построена. Пусть, например, нам известно, что исходная выборка принадлежит классу распределений Пуассона, т. е. имеет дискретную плотность вида $q(k;\theta) = \mathbf{P}\{X=k\} = \dfrac{\theta^k}{k!}\,e^{-\theta}$. Ограничимся одним наблюдением и зададимся целью построить несмещенную оценку параметрической функции $g(\theta) = 1/\theta$. Несмещенность означала бы, что при каждом θ имеет место

$$E_\theta(\delta) = \sum_{x=0}^{\infty}\delta(x)\frac{\theta^x}{x!}\,e^{-\theta} = \frac{1}{\theta}, \quad \sum_{x=0}^{\infty}\delta(x)\frac{\theta^{x+1}}{x!} = e^\theta = \sum_{r=0}^{\infty}\frac{\theta^r}{r!},$$

но этого не может быть.

Риск несмещенной оценки совпадает с ее дисперсией, поэтому если существует равномерно наилучшая из несмещенных оценок (т. е. имеющая наименьший среди всех несмещенных оценок риск при каждом значении $\theta\in\Theta$), то она называется *несмещенной с равномерно минимальной дисперсией* (НРМД).

Утверждение 2.1. /о единственности НРМД-оценки/. *Может существовать не более одной НРМД-оценки параметра $\theta\in\Theta$.*

Доказательство. Допустим, что существуют две НРМД-оценки δ_1 и δ_2. Рассмотрим оценку $\delta_3 = (\delta_1+\delta_2)/2$. Имеем, по определению НРМД-оценки при каждом $\theta\in\Theta$

$$\mathbf{D}_\theta\delta_3 = \frac{\mathbf{D}_\theta\delta_1+\mathbf{D}_\theta\delta_2+2\mathbf{cov}(\delta_1,\delta_2)}{4} \geq \mathbf{D}_\theta\delta_1 = \mathbf{D}_\theta\delta_2.$$

Отсюда следует, что $\mathbf{D}_\theta\delta_1 = \mathbf{D}_\theta\delta_2 \leq \mathbf{cov}(\delta_1,\delta_2)$. Тогда $\mathbf{D}_\theta\delta_1 = \mathbf{D}_\theta\delta_2 = \mathbf{cov}(\delta_1,\delta_2)$ и, следовательно,

$$\mathbf{D}(\delta_1-\delta_2) = \mathbf{D}_\theta\delta_1+\mathbf{D}_\theta\delta_2-2\mathbf{cov}(\delta_1,\delta_2) = 0.$$

Таким образом, $\delta_1=\delta_2$ с вероятностью 1. ■

Если не удается найти в рассматриваемом классе оценок наилучшую при каждом значении θ, то можно это требование ослабить.

Оценка δ_0 называется *недопустимой*, если существует оценка δ такая, что $R(\delta_0,\theta) \geq R(\delta,\theta)$, при каждом $\theta \in \Theta$, и $R(\delta_0,\theta) > R(\delta,\theta)$, при некотором $\theta \in \Theta$. Остальные оценки называются допустимыми. Так, согласно приведенному примеру, несмещенная оценка дисперсии нормального распределения s'^2 является недопустимой как в классе оценок вида $s(\lambda)$, $\lambda > 0$, так и в классе всевозможных оценок дисперсии. Чтобы выбрать оптимальную из допустимых оценок, возможны различные подходы.

2.2. Минимаксный и байесовский подходы

Оценка δ_0 называется *минимаксной*, если она минимизирует максимальный риск, т. е.

$$R(\delta_0,\theta) = \inf_{\delta} \sup_{\theta\in\Theta} R(\delta,\theta).$$

Предположим, что на множестве Θ задано априорное распределение Q. *Байесовским риском* оценки δ, соответствующим функции потерь W, называется величина

$$R(\delta) = \int_{\Theta} R(\delta,\theta)\, Q(d\theta).$$

Оценка, минимизирующая байесовский риск, называется *байесовской*.

Таким образом, исходное параметрическое семейство может быть представлено как распределение на $\mathfrak{X} \times \Theta$. Пусть $p(\theta)f(\vec{x};\theta)$ – плотность (дискретная плотность, плотность относительно некоторой доминирующей меры $\mu^* : \mu^*(d\vec{x};\, d\theta) = \mu(d\vec{x};\, \theta)\,\mu'(d\theta)$). Тогда, с учетом известного варианта формулы Байеса, возникает апостериорное распределение с плотностью

$$p(\theta \mid \vec{X}) = \frac{p(\theta)f(\vec{X};\,\theta)}{\int\limits_{\Theta} p(\theta)f(\vec{X};\,\theta)\,\mu'(d\theta)}.$$

Рассмотрим функцию потерь Гаусса. В этом случае байесовский риск имеет вид

$$R(\delta) = \int_{\Theta} \mathbf{E}_{\,\theta}(\delta(\vec{X}) - \theta)^2\, Q(d\theta).$$

Путем дифференцирования по δ находим $\delta_*(\vec{X})$, минимизирующую байесовский риск

$$\delta_*(\vec{X}) = \frac{\int\limits_{\Theta} \theta f(\vec{X};\,\theta)\, Q(d\theta)}{\int\limits_{\Theta} f(\vec{X};\,\theta)\, Q(d\theta)},$$

т. е. условное (апостериорное) среднее параметра θ при условии $\vec{X}$.

Пример 2.1. Пусть $X_1, \ldots, X_n$ – выборка из $N(\theta, \sigma^2)$ с известным σ и надо оценить θ в предположении, что параметр θ имеет априорное нормальное распределение $N(\mu, b^2)$. Совместная плотность θ и $\vec{X} = (X_1, \ldots, X_n)$ имеет вид

$$p(\theta)f(\vec{x};\theta) = \frac{1}{2\pi\sigma b}\exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\Big)\exp\Big(-\frac{1}{2b^2}(\theta-\mu)^2\Big).$$

Тогда апостериорная плотность имеет вид

$$U(\vec{X})\exp\left(-\frac{1}{2}\theta^2\left(\frac{n}{\sigma^2}+\frac{1}{b^2}\right)+\theta\Big(\frac{n\overline{X}}{\sigma^2}+\frac{\mu}{b^2}\Big)\right) =$$

$$= U(\vec{X})\exp\left(-\frac{1}{2}\Big(\frac{n}{\sigma^2}+\frac{1}{b^2}\Big)\Big(\theta^2-2\theta\frac{n\overline{X}+\mu\sigma^2/b^2}{n+\sigma^2/b^2}\Big)\right),$$

где $U(\vec{X}) = \dfrac{\exp\Big(-\frac{1}{2}\Big(\frac{\sum_i X_i^2}{\sigma^2}+\frac{\mu^2}{b^2}\Big)\Big)}{\int_{-\infty}^{\infty}p(\theta)f(\vec{X};\theta)\,d\theta)}$. Нетрудно видеть, что это нормальная плотность (для которой, стало быть, $U(x)$ нормирующий множитель) со средним $\mathbf{E}\,(\theta|\vec{X}) = \dfrac{n\overline{X}+\mu\sigma^2/b^2}{n+\sigma^2/b^2}$ и дисперсией $\mathbf{D}(\theta\,|\,\vec{X}) = (n+\sigma^2/b^2)^{-1}$. Тогда

$$\delta(\vec{X}) = \mathbf{E}\,(\theta\,|\,\vec{X}) = \frac{n}{n+\sigma^2/b^2}\overline{X} + \frac{\mu}{nb^2/\sigma^2+1}$$

является байесовской оценкой для функции потерь Гаусса (квадратический риск). При больших n эта оценка, как легко видеть, близка к выборочному среднему. Байесовский риск в этом случае может быть вычислен по формуле $R(\delta) = \int_{-\infty}^{\infty}\mathbf{D}(\theta\,|\,\vec{X}=\vec{x})p(\theta)f(\vec{x};\theta)\,\mu(d\vec{x})$.

Следующая теорема позволяет находить минимаксные оценки.

Теорема 2.1. /Лемана/ *Пусть* $\{\delta_k\}_{k\in\mathbb{N}}$ *– последовательность байесовских оценок по отношению к априорным распределениям* $\{Q_k\}$ *соответственно;*

$$\delta : \sup_\theta R(\delta,\theta) \le \limsup_{k\to\infty}\int_\Theta R(\delta_k,\theta)\,dQ_k.$$

Тогда δ *– минимаксна.*

Доказательство. Пусть δ^* – произвольная оценка. Тогда, поскольку Q_k – вероятностная мера и поскольку δ_k – байесовская:

$$\sup_\theta R(\delta^*,\,\theta) \ge \int_\Theta R(\delta^*,\,\theta)\,Q_k(d\theta) \ge \int_\Theta R(\delta_k,\,\theta)\,Q_k(d\theta).$$

Переходим к пределу

$$\sup_\theta R(\delta^*,\theta) \ge \overline{\lim_{k\to\infty}}\int_\Theta R(\delta_k,\,\theta)\,Q_k(d\theta) \ge \sup_\theta R(\delta,\,\theta).$$

Следовательно, δ – минимаксна. ■

Пусть $X_1, \ldots, X_n$ – выборка из $N(\theta, \sigma^2)$ с известным σ и надо оценить θ. Предположим, что параметр θ имеет нормальное распределение $N(0, k)$, $k \in \mathbb{N}$. Получаем последовательность байесовских оценок:

$$\delta_k(\vec{X}) = \frac{n\overline{X}}{n + \sigma^2/k}$$

и соответствующую последовательность байесовских рисков:

$$R(\delta_k) = \int_{-\infty}^{\infty} \mathbf{E}_\theta \left(\frac{n\overline{X}}{n+\sigma^2/k} - \theta \right)^2 Q_k(d\theta) = \int_{-\infty}^{\infty} \mathbf{E}_\theta \left(\frac{n(\overline{X}-\theta) - \theta\sigma^2/k}{n+\sigma^2/k} \right)^2 Q_k(d\theta) =$$

$$= \frac{1}{(n+\sigma^2/k)^2} \left(n\sigma^2 - 2\int_{-\infty}^{\infty} \frac{\theta\sigma^2}{k} \mathbf{E}_\theta n(\overline{X}-\theta) Q_k(d\theta) + \int_{-\infty}^{\infty} \theta^2 \, Q_k(d\theta)/k^2 \right) \xrightarrow[k\to\infty]{} \sigma^2/n.$$

Далее отметим, что

$$\mathbf{E}_\theta (\overline{X} - \theta)^2 = \sigma^2/n.$$

Следовательно, по теореме Лемана $\overline{X}$ – минимаксна.

2.3. Метод максимума правдоподобия

Одним из центральных в теории точечного оценивания и в математической статистике в целом является понятие правдоподобия. Оно позволяет сравнивать вероятностные шансы тех или иных исходов эксперимента при различных значениях параметра $\theta \in \Theta$.

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta, \ \theta \in \Theta\}$ – статистический эксперимент. Если результат эксперимента представляется как некоторое утверждение (уже произошедшее событие) A, имеющее ненулевую вероятность, то правдоподобие, построенное на основании A, представляет собой просто вероятность этого события. Однако, если в результате эксперимента поступает более точная информация, то для сравнения шансов полученного исхода со всеми остальными необходимо вводить дополнительные построения. Пусть результат эксперимента представляется как значение случайного вектора. Тогда при каждом фиксированном значении $\theta \in \Theta$ естественно сравнивать достоверности исходов через значение плотности (или производной Радона – Никодима) относительно некоторой доминирующей меры. Если величины имеют абсолютно непрерывное распределение, то это – плотность относительно меры Лебега (или просто плотность). Если же распределение исходного вектора дискретно, то в качестве доминирующей можно выбрать считающую меру на множестве возможных значений рассматриваемого вектора. Тогда степень достоверности (правдоподобие) исхода $\vec{X}$ определяется значением дискретной плотности в данной точке или вероятностью того, что исход эксперимента именно такой. Для сравнения степени достоверности того или иного исхода при различных значениях параметра $\theta \in \Theta$ вводится, если это

возможно, мера, доминирующая все семейство распределений. Естественно, что отношение правдоподобий при различных θ не зависит от выбора доминирующей меры.

Пусть $X_1, X_2, \ldots, X_n$ – набор независимых наблюдений с плотностями (дискретными плотностями, плотностями относительно доминирующей меры μ) $f_{\theta,1}, f_{\theta,2}, \ldots, f_{\theta,n}$ соответственно. *Функцией правдоподобия*, построенной по исходным наблюдениям, будем называть

$$L(\vec{X};\, \theta) = \prod_{i=1}^{n} f_{\theta,\,i}(X_i), \quad \theta \in \Theta.$$

Значение $L(\vec{X};\, \theta)$ при фиксированном θ будем называть правдоподобием исходного набора наблюдений. Идея, лежащая в основе *метода максимального правдоподобия*, состоит в том, чтобы по результатам наблюдений отыскать значение $\hat{\theta}(\vec{X}) \in \Theta$, максимизирующее правдоподобие, т. е. $L(\vec{X}; \hat{\theta}(\vec{X})) \leq L(\vec{X}; \theta)$ для любого $\theta \in \Theta$. Далее отметим, что в силу монотонности логарифма задача максимизации правдоподобия сводится к задаче максимизации его логарифма (чтобы максимизировать сумму, а не произведение) по всем $\theta \in \Theta$

$$\ln L(\vec{X};\, \theta) = \sum_{i=1}^{n} \ln(f_{\theta,\,i}(X_i)).$$

Если $\theta = (\theta_1, \ldots, \theta_n)$ – n-мерный параметр и $f_{\theta,i}$ дифференцируемы по θ, то для нахождения максимума надо найти решения системы уравнений

$$U(\vec{X};\, \theta) = \frac{\partial}{\partial \theta_i} \ln L(\vec{X};\, \theta) = 0.$$

Ясно, что если $\hat{\theta}$ – оценка максимального правдоподобия параметра θ, g – параметрическая функция, то $g(\hat{\theta})$ является оценкой максимального правдоподобия для $g(\theta)$. Рассмотрим некоторые примеры.

Пример 2.2. Пусть $X_1, X_2 \ldots, X_n$ – выборка из двухпараметрического ($\theta = (a, \sigma^2)$) нормального распределения $N(a, \sigma^2)$. Логарифм функции правдоподобия имеет вид

$$\ln L(\vec{x};\, \theta) = -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(x_i - a)^2}{2\sigma^2}.$$

Тогда максимум правдоподобия находится из системы уравнений

$$\begin{cases} 2\sum\limits_{i=1}^{n} (X_i - a) = 0, \\ n/\sigma - \sum\limits_{i=1}^{n} \dfrac{(X_i - a)^2}{\sigma^3} = 0. \end{cases}$$

Получаем, что оценка максимального правдоподобия имеет вид $(\hat{a},\, \hat{\sigma^2}) = (\overline{X},\, s^2)$.

Пример 2.3. Предположим, что $X_1, X_2 \ldots, X_n$ – выборка из двухпараметрического ($\theta = (a, b)$) равномерного распределения на интервале (a, b). Функция правдоподобия имеет вид

$$L(\vec{x};\, \theta) = \frac{1}{(b-a)} \prod_{i=1}^{n} \mathbb{1}_{[a,b]}(x_i) = \frac{1}{b-a}\, \mathbb{1}_{[a,b]}(\min_i(x_i))\, \mathbb{1}_{[a,b]}(\max_i(x_i)),$$

где

$$\mathbb{1}_A(x) = \begin{cases} 1, \text{ если } x \in A; \\ 0, \text{ в остальных случаях.} \end{cases}$$

В данной задаче нет смысла логарифмировать функцию правдоподобия. Заметим, что решение исходной задачи оптимизации сводится к минимизации разности $(b-a)$ при условии $a \le \min\limits_{1\le i\le n}(X_i) \le \max\limits_{1\le i\le n}(X_i) \le b$. Ясно, что решение данной задачи $a = \min\limits_{1\le i\le n}(X_i)$, $b = \max\limits_{1\le i\le n}(X_i)$. Таким образом, оценка максимального правдоподобия имеет вид $(\hat{a}, \hat{b}) = (\min\limits_{1\le i\le n}(X_i), \max\limits_{1\le i\le n}(X_i))$.

Замечание 2.1. В случае, если $\mathcal{P}$ – семейство всевозможных распределений наборов независимых одинаково распределенных случайных величин, то правдоподобие в указанной ранее постановке ввести не удается, поскольку не существует меры, доминирующей $\mathcal{P}$. В то же время, если сравнивать вероятности полученного результата наблюдений при различных распределениях выборки, то максимум достигается при равномерном распределении на множестве полученных наблюдений. Таким образом, эмпирическое распределение, в каком-то смысле, является оценкой максимального правдоподобия теоретического распределения.

2.4. Достаточные статистики

В данном разделе вводится одно из важнейших понятий математической статистики, позволяющее выделить из полного объема информации, содержащейся в исходном эксперименте, ту ее часть, которая может быть использована при оценивании теоретического распределения. Однако мы ограничимся лишь описательным введением данного понятия и формулировкой основных результатов.

Из теории вероятностей известно понятие условного распределения одного случайного вектора относительно другого.

Пусть $X_1, \ldots, X_n$ – выборка из распределения $P_\theta \in \mathcal{P}$, $T(\vec{X})$ – статистика. Будем говорить, что статистика T *достаточна*, если условное распределение вектора $\vec{X}$ при условии T не зависит от значения параметра θ. Такое распределение можно понимать, как распределение $\vec{X}$ по поверхности $T(\vec{X}) = s$. Это означает, что знание того, где находится $\vec{X}$ на поверхности $T(\vec{X}) = s$, не несет никакой дополнительной информации о теоретическом

распределении, а следовательно, вся информация об исходном распределении содержится в статистике $T(\vec{X})$. При наличии меры μ^*, доминирующей семейство векторов $(\vec{X}, T(\vec{X}))$, достаточность означает, что условная плотность вектора $\vec{X}$ при условии $T(\vec{X})$ ($\{T(\vec{X}) = \vec{y}\}$) не зависит от θ (при каждом $\vec{y}$ с точностью до событий нулевой вероятности). Для понимания данного свойства в полной мере надо владеть понятием условного распределения относительно σ-алгебры.

Пусть $L(\vec{x};\, \theta)$ – функция правдоподобия. В этом случае справедлива теорема факторизации Неймана – Фишера, дающая простой способ нахождения достаточных статистик.

Теорема 2.2. /факторизации/ *Статистика T достаточна тогда и только тогда, когда функция правдоподобия допускает представление в виде*

$$L(\vec{x};\, \theta) = g(T(\vec{x}),\, \theta)\, h(\vec{x}),$$

где $g(t,\, \theta)$ и h – некоторые неотрицательные функции.

Тривиальным следствием данной теоремы является тот факт, что оценка максимального правдоподобия представляется как функция от достаточной статистики (т. е. зависит лишь от достаточной статистики).

Рассмотрим пример многопараметрического экспоненциального семейства распределениий с плотностями вида

$$f(x;\, \theta) = h(x) \exp\Big(\sum_{j=1}^{k} a_j(\theta)\delta_j(x) + r(\theta)\Big),$$

где θ – k-мерный параметр $\theta = (\theta_1, \ldots, \theta_k)$. Тогда функция правдоподобия имеет вид

$$L(\vec{x};\, \theta) = \prod_{i=1}^{n} h(x_i) \exp\Big(\sum_{i=1}^{n}\sum_{j=1}^{k} a_j(\theta)\delta_j(x_i) + \nu(\theta)\Big).$$

Достаточная статистика здесь тоже многомерна $\delta = (\delta_1, \ldots, \delta_k)$, т. е. при фиксации всех k значений вектора условное распределение тоже не зависит от θ. Для нормального распределения

$$L(x_1, \ldots, x_n;\, a, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\Big) =$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\Big(-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n} x_i^2 - 2a\sum_{1}^{n} x_i + na^2\Big)\Big).$$

Достаточную статистику составляет пара $(\sum_1^n X_i^2,\, \sum_1^n x_i = n\overline{x})$, т. е. в этих двух числах содержится информация, которую нам достаточно знать о выборке для оценивания и иных задач. Поскольку существует взаимно-однозначное соответствие между статистикой $(\sum_1^n X_i^2,\, \sum_1^n x_i = n\overline{x})$ и статистикой

$(\overline{X},\, s^2)$, последняя также является достаточной (с той же самой редукцией данных).

Рассмотрим еще пример распределения Пуассона:

$$L(x_1,\ldots,x_n;\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}\,e^{-\lambda}=\prod_{i=1}^{n}\frac{1}{x_i!}\;\lambda^{\sum_{i=1}^{n}x_i}\,e^{-n\lambda},$$

т. е. $\sum_1^n X_i = n\overline{X}$ – достаточная статистика. Здесь очень легко найти условные вероятности $\mathbf{P}_\lambda(X_1 < x_1,\ldots,X_n < x_n \mid \sum_1^n X_i = k)$ и убедиться в том, что они не зависят от λ.

Нетрудно видеть, что сам набор наблюдений $\vec{X}$ всегда является достаточной статистикой (поскольку условное распределение $\vec{X}$ при условии $\vec{X}$ тривиально и сконцентрировано в точке $\vec{X}$), но не дает никакой редукции данных. Достаточная статистика называется *минимальной*, если она может быть представлена в виде функции от любой достаточной статистики (т. е. редукция данных максимальна). Следующая теорема подтверждает интуитивные представления о достаточности.

Теорема 2.3. /Рао – Блэкуэлла – Колмогорова/ *Пусть T – достаточная статистика для семейства $\mathcal{P}$, δ – оценка параметра θ. Положим $\eta = \mathbf{E}\,(\delta|T)$ – условное математическое ожидание δ при условии T (по определению, это функция от T). Если функция потерь W выпукла (вниз), то $R(\delta;\,\theta) \geq R(\eta;\,\theta)$.*

Суть данного утверждения состоит в том, что наилучшие оценки всегда могут быть представлены как функции от достаточных статистик. Введем еще одно важное понятие.

Достаточная статистика T называется *полной*, если выполнение условия $\mathbf{E}_\theta g(T) = 0$, при всех $\theta \in \Theta$, влечет $g(T) = 0$ с вероятностью 1. Полная достаточная статистика всегда минимальна. В заключение приведем еще одну важную теорему.

Теорема 2.4. /Лемана – Шеффе/ *Существует не более одной несмещенной (с фиксированным смещением) оценки параметра θ, являющейся функцией от полной достаточной статистики.*

Подведем итоги:

1. Эффективные оценки в классе оценок с фиксированным смещением следует искать как функции от минимальных достаточных статистик.

2. Если минимальная достаточная статистика является полной, то при каждом фиксированном значении смещения существует единственная оценка, минимизирующая риск в классе оценок с таким смещением.

2.5. Информация по Фишеру и неравенство Рао – Крамера

Эвристическими предпосылками введения понятия информации можно считать следующие два постулата:

1. Параметры тем легче различать, чем больше различаются соответствующие распределения.

2. Информация, содержащаяся в независимых экспериментах, равна сумме информации, содержащейся в каждом из них.

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ – статистический эксперимент. Предположим, что рассматриваемое семейство доминируется мерой μ. В этом случае существует соответствующее семейство плотностей $f((\cdot);\, \theta) = \dfrac{dP_\theta}{d\mu}$, $\theta \in \Theta$. В наших моделях $\mathfrak{X} = \mathbb{R}^s$, $s \in \mathbb{N}$, а доминирующая мера – либо мера Лебега (абсолютно непрерывный случай), либо считающая мера (дискретный случай). Прежде чем ввести понятие информации, ограничим круг рассматриваемых статистических экспериментов.

Пусть $\Theta \subseteq \mathbb{R}$, т. е. параметр θ – вещественное число. Будем называть эксперимент *регулярным*, если при каждом $\theta \in \Theta$:

1) $L((\cdot);\, \theta)$ непрерывна и дифференцируема по θ;

2) допустимо дифференцирование под знаком интеграла:

$$\int\limits_{\mathfrak{X}} \frac{\partial}{\partial\theta} f(x;\, \theta)\, \mu(dx) = \frac{\partial}{\partial\theta} \int\limits_{\mathfrak{X}} f(x;\, \theta)\, \mu(dx) = 0;$$

3) существует и отличен от нуля интеграл

$$0 < I(\theta) = \mathbf{E}_\theta (U(\vec{X};\, \theta))^2 = \int\limits_{\mathfrak{X}} (U(\vec{x};\, \theta))^2 L(\vec{x};\, \theta)\, \mu(dx),$$

где

$$U(\vec{x};\, \theta) = \frac{\partial}{\partial\theta} \ln L(\vec{x};\, \theta).$$

Замечание 2.2. Условие (2) регулярности обычно нарушается, если параметр выходит на границы интеграла, поскольку в этом случае появляется производная интеграла с переменным пределом. В частности, если носитель распределения $A = \{\vec{x} : f(\vec{x};\, \theta) > 0\}$ зависит от параметра.

Величина $I(\theta)$ называется *информацией Фишера*, содержащейся в исходном наборе наблюдений. Нетрудно показать, что если функция правдоподобия дважды дифференцируема под знаком интеграла по θ, то

$$I(\theta) = -\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2} \ln L(\vec{X};\, \theta)\right).$$

Обычно при практическом вычислении информации Фишера данная формула предпочтительнее.

Отметим, что последнее равенство в условии (2) регулярности выполнено всегда, поскольку

$$\int\limits_{\mathfrak{X}} f(x;\,\theta)\,\mu(dx) = 1 \quad \text{при любом } \theta.$$

Утверждение 2.2. *Пусть* $(\mathfrak{X}_1, \mathfrak{F}_1, \mathcal{P}_1)$ *и* $(\mathfrak{X}_1, \mathfrak{F}_1, \mathcal{P}_2)$ *– регулярные эксперименты с информацией* $I_1(\theta)$ *и* $I_2(\theta)$ *соответственно. Тогда эксперимент* $(\mathfrak{X}_1 \times \mathfrak{X}_2, \sigma(\mathfrak{F}_1 \times \mathfrak{F}_2), \mathcal{P})$, *где* $P_\theta(d\vec{x}) = P_{1,\theta}(dx_1)\,P_{2,\theta}(dx_2)$, $\vec{x} = (x_1, x_2)$ (*т. е.* $f(\vec{x};\,\theta) = f_1(x_1;\,\theta) f_2(x_2;\,\theta)$, $x_1, x_2 \in \mathbb{R}$), *регулярен и* $I(\theta) = I_1(\theta) + I_2(\theta)$.

Иными словами, эксперимент, состоящий в проведении двух независимых регулярных экспериментов, регулярен, а информация Фишера равна сумме информаций Фишера, составляющих его независимых экспериментов.

Доказательство состоит в непосредственной проверке условий регулярности и не представляет интереса.

Замечание 2.3. Прямым следствием приведенного утверждения является тот факт, что информация Фишера I_θ, содержащаяся в выборке $X_1, X_2, \ldots, X_n$, в n раз больше информации, содержащейся в каждом наблюдении, т. е. $I(\theta) = nI_1(\theta)$.

Оценка δ называется разрешенной, если

$$\frac{\partial}{\partial\theta}\mathbf{E}_\theta\delta(\vec{X}) = \mathbf{E}_\theta\Big(\delta(\vec{X})\frac{\partial}{\partial\theta}\ln L(\vec{X};\,\theta)\Big),$$

иными словами, допускается дифференцирование под знаком интеграла

$$\frac{\partial}{\partial\theta}\int\limits_{\mathfrak{X}} \delta(\vec{x})L(\vec{x};\,\theta)\,\mu(d\vec{x}) = \int\limits_{\mathfrak{X}} \delta(\vec{x})\frac{\partial}{\partial\theta}L(\vec{x};\,\theta)\,\mu(d\vec{x}).$$

Теорема 2.5. /неравенство Рао – Крамера/. *Пусть эксперимент регулярен,* δ *– разрешенная оценка параметрической функции* $g(\theta)$. *Тогда,*

$$\mathbf{E}_\theta(\delta - g(\theta))^2 \geq \frac{(g'(\theta) + b'_g(\theta))^2}{I(\theta)} + b_g^2(\theta)$$

или

$$\mathbf{D}_\theta\delta \geq \frac{(g'(\theta) + b'_g(\theta))^2}{I(\theta)},$$

где $b_g(\delta) = \mathbf{E}\,\delta - g(\theta)$ *– смещение.*

Доказательство. По определению смещения $\mathbf{E}_\theta\delta = g(\theta) + b_g(\theta)$. Используя свойство разрешенности оценки, после дифференцирования получаем

$$\frac{\partial}{\partial\theta}\mathbf{E}_\theta\delta = \int\limits_{\mathfrak{X}} \delta(\vec{x})f'(\vec{x},\,\theta)\,\mu(dx) = \mathbf{E}_\theta(\delta(\vec{X})\,U(\vec{X},\theta)).$$

Тогда, с учетом условия (2) регулярности эксперимента, получаем равенство

$$\mathbf{E}_\theta\big((\delta(\vec{X}) - \mathbf{E}_\theta\delta(\vec{X}))U(\vec{X},\theta)\big) = g'(\theta) + b'_g(\theta).$$

Применяем неравенство Коши – Буняковского

$$(g'(\theta) + b'_g(\theta))^2 \le \mathbf{E}_\theta(\delta(\vec{X}) - \mathbf{E}_\theta\delta(\vec{X}))^2\mathbf{E}_\theta U^2(\vec{X},\theta) = \mathbf{D}_\theta\delta(\vec{X})\ I(\theta),$$

из которого получаем второе неравенство. Отсюда первое неравенство получается тривиальным образом, так как $\mathbf{D}_\theta\delta(\vec{X}) = \mathbf{E}_\theta(\delta-\theta)^2 - b_g^2(\theta)$. ■

Замечание 2.4. Если δ – несмещенная оценка, тогда неравенство Рао – Крамера примет вид

$$\mathbf{E}_\theta(\delta-\theta)^2 \ge \frac{g'(\theta)^2}{I(\theta)}.$$

Оценки, для которых достигается равенство в неравенстве Рао – Крамера, называются эффективными по Фишеру, или *R-эффективными* со смещением $b(\theta)$. R-эффективные оценки с нулевым смещением (несмещенные) называются просто R-эффективными (часто их называют просто эффективными). Оценка δ называется асимптотически R-эффективной оценкой параметрической функции $g(\theta)$, если

$$I(\theta)\,\mathbf{E}_\theta(\delta - g(\theta))^2 \xrightarrow[n\to\infty]{} 1.$$

Далеко не всегда существуют эффективные оценки. Зададимся вопросом: «В каких случаях в неравенстве Рао–Крамера достигается равенство?» При доказательстве неравенства Рао – Крамера использовалось неравенство Коши – Буняковского, равенство в котором достигается в том и только в том случае, если

$$a_*(\theta)(\delta(\vec{X}) - \mathbf{E}_\theta\delta(\vec{X})) = U(\vec{X};\,\theta) = \frac{\partial}{\partial\theta}\ln L(\vec{X};\,\theta).$$

В этом случае

$$L(\vec{X},\,\theta) = h(\vec{X})\exp(a(\theta)\delta(\vec{X}) + r(\theta)),$$

т. е. плотность распределения $\vec{X}$ имеет вид

$$f(\vec{x};\,\theta) = h(\vec{x})\exp(a(\theta)\delta(\vec{x}) + r(\theta)).$$

Если исходный набор наблюдений – выборка, то равенство в неравенстве Рао – Крамера достигается лишь в случае, если

$$f_1(x;\,\theta) = h_*(x)\exp(a(\theta)\delta_*(x) + r(\theta)).$$

Семейства с плотностями такого вида называются однопараметрическими *экспоненциальными* семействами.

Еще одно умозаключение, вытекающее из изложенного, состоит в том, что для каждого однопараметрического экспоненциального семейства существует единственная параметрическая функция $g(\theta) = -r'(\theta)/a'(\theta)$, допускающая R-эффективное оценивание. Установим еще одно свойство R-эффективных оценок.

Теорема 2.6. *Пусть эксперимент регулярен; δ – R-эффективная (несмещенная) оценка θ. Тогда она является оценкой максимальноо правдоподобия.*

Доказательство. Поскольку $U(\vec{X},\ \theta) = a'(\theta)(\delta(\vec{X}) - \theta)$, из несмещенности следует, что

$$a'(\theta) = \sqrt{I(\theta)/\mathbf{D}\,\delta} = 1/\mathbf{D}\,\delta > 0.$$

Следовательно, δ – точка локального максимума функции $L(\vec{X};\,\theta)$ по θ. ■

Для случая многопараметрического семейства может быть получено похожее неравенство. Условия регулярности заключаются в наличии частных производных под знаком интеграла по каждому параметру и невырожденности информационной матрицы каждого наблюдения $\mathbf{I}(\theta) = ||I_{i,j}(\theta)||$, где

$$I_{i,j}(\theta) = \mathbf{E}\left(\frac{\partial}{\partial\theta_i}\ln L(\vec{x};\,\theta)\frac{\partial}{\partial\theta_j}\ln L(\vec{x};\,\theta)\right).$$

Тогда для любой разрешенной оценки $\delta = (\delta_1, \ldots, \delta_k)$ параметра θ справедливо неравенство

$$\mathbf{D}(\delta) \geq (\mathbf{E}\ + \mathcal{B}'(\theta))\mathbf{I}^{-1}(\theta)(\mathbf{E}\ + \mathcal{B}'(\theta))^T,$$

где $\mathbf{E}$ – единичная матрица; $\mathcal{B}'(\theta)$ – матрица частных производных компонент вектора смещений по параметрам, и, в частности, если δ несмещенная, то

$$\mathbf{D}(\delta) \geq \mathbf{I}^{-1}(\theta).$$

Пример 2.4. Пусть $X_1, \ldots, X_n$ – выборка из двухпараметрического $\theta = (a, \sigma^2)$ нормального распределения $N(a, \sigma^2)$. В этом случае информационная матрица и матрица $\mathbf{I}^{-1}$ имеют вид, соответственно

$$\begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^4) \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}.$$

Рассмотрим класс оценок $\delta = (\overline{X},\ \lambda(n)s'^2)$, где $s'^2 = ns^2/(n-1)$, $\lambda(n) > 0$ (чтобы оценка была состоятельной, необходимо, чтобы $\lambda(n) \xrightarrow[n\to\infty]{} 1$). Вектор смещения имеет вид $(0,\ \lambda(n) - 1)$. Следовательно

$$\mathcal{B}'(\theta) = \begin{pmatrix} 0 & 0 \\ 0 & \lambda(n)-1 \end{pmatrix}$$

и

$$(\mathbf{E}\ + \mathcal{B}'(\theta))\mathbf{I}^{-1}(\theta)(\mathbf{E}\ + \mathcal{B}'(\theta)) = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\lambda^2(n)\sigma^4/n \end{pmatrix}.$$

В то же время ковариационная матрица оценки δ есть

$$\begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\lambda^2(n)\sigma^4/(n-1) \end{pmatrix}.$$

Таким образом, ни при каком $\lambda(n)$ равенство не достигается. В частности, несмещенная оценка $(\overline{X},\ s'^2)$ не является R-эффективной. Однако все оценки такого вида являются асимптотически эффективными оценками параметра (a, σ^2), если $\lambda(n) \to 1$ при $n \to \infty$.

В заключение отметим, что при достаточно общих предположениях уравнение $\dfrac{\partial \log L(X_1, \ldots, X_n;\, \theta)}{\partial \theta} = 0$ имеет решение, сходящееся по вероятности при $n \to \infty$ к истинному значению θ (состоятельность). Это решение является асимптотически эффективной и асимптотически нормальной оценкой для θ.

2.6. Интервальное оценивание

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ – статистический эксперимент. Помимо точечного оценивания истинного значения параметра существует другой подход, состоящий в отыскании по результатам эксперимента некоторого подмножества множества Θ, в котором с определенной (наперед заданной) долей достоверности должно находиться истинное значение параметра. Это подмножество называется доверительным, а соответствующая доля достоверности – доверительным уровнем (надежностью) данного подмножества.

Пусть Θ – некоторый интервал (конечный или бесконечный); $X_1, \ldots, X_n$ – исходный набор наблюдений. Тогда, естественной постановкой задачи является отыскание по результатам наблюдений интервала, с определенной долей достоверности содержащего истинное значение параметра. В этом случае границы интервала – статистики.

Интервал $[T_1, T_2]$, образованный парой статистик $T_1(X_1, \ldots, X_n)$ и $T_2(X_1, \ldots, X_n)$, называется *доверительным интервалом* надежности (с доверительным уровнем) $1 - \alpha$, если при всех $\theta \in \Theta$ выполняется неравенство $P_\theta(T_1 \leqslant \theta \leqslant T_2) \geqslant 1 - \alpha$.

Очевидно, что в такой постановке задача имеет множество решений. Следует отметить, что если задаться целью, то можно выбирать доверительный интервал, содержащий произвольное наперед заданное значение параметра (вовсе не обязательно истинное). При этом он с большой вероятностью должен содержать и истинное значение параметра. Поэтому его длина может быть достаточно велика, вне зависимости от длины выборки (так же, как можно было выбирать оценки, вовсе не оценивающие истинное значение параметра). Таким образом, естественной мерой качества доверительного интервала является его длина. Однако возможны и другие подходы к выбору доверительного интервала. Например, если нужно оценить истинное значение параметра снизу (или сверху), то имеет смысл рассматривать односторонние интервалы, или, если длина фиксирована, то мерой качества будет минимальный объем выборки, необходимый для того, чтобы данный интервал имел уровень $1 - \alpha$.

Рассмотрим некоторые методы построения доверительного интервала. Начнем с простейшего случая, когда найдена случайная величина $G(\vec{X};\,\theta)$, монотонно зависящая от параметра, распределение которой не зависит от θ. Поскольку функция распределения F_G не зависит от θ, можно найти значения g_1 и g_2 из условия (так чтобы интервал $[g_1, g_2]$ не содержал другой интервал, удовлетворяющий этому условию)

$$\mathbf{P}_\theta(g_1 \le G(\vec{X};\,\theta) \le g_2) = F_G(g_2+) - F_G(g_1) \ge 1-\alpha.$$

Далее, границы доверительного интервала $T_1(\vec{X})$ и $T_2(\vec{X})$ для параметра θ выбираются как решения относительно θ уравнений

$$G(\vec{X};\,\theta) = g_i, \quad i = 1, 2.$$

Тогда, $\mathbf{P}_\theta(T_1(\vec{X}) \le \theta \le T_2(\vec{X})) = \mathbf{P}_\theta(g_1 \le G(\vec{X};\,\theta) \le g_2) \ge 1-\alpha$.

Если функции распределения исходного семейства $\mathcal{P}$ непрерывны и монотонно меняются по параметру (например, параметр сдвига), то можно выбрать

$$G(\vec{X};\,\theta) = -\sum_{i=1}^{n} \ln F(X_i;\,\theta).$$

Нетрудно показать, что если случайная величина X имеет непрерывную функцию распределения F, то случайная величина $F(X)$ равномерно распределена на интервале $[0, 1]$ (данное свойство носит название *теорема Смирнова*, а соответствующее преобразование называется преобразованием Смирнова). Таким образом, $-\ln F(X_i;\,\theta)$ имеет гамма-распределение $\Gamma(1, 1)$. Следовательно, используя свойства гамма-распределения заключаем, что $G(\vec{X};\,\theta)$ имеет известное распределение $\Gamma(1, n)$. Дальнейший путь решения описан ранее. Рассмотрим примеры.

Пример 2.5. Пусть $X_1, \ldots, X_n$ – выборка из двухпараметрического нормального распределения $N(a, \sigma^2)$. Будем строить доверительные интервалы (уровня доверия $1-\alpha$) для параметров a и σ^2 с использованием теоремы 1.5 (Фишера). Согласно п. 3 теоремы Фишера величина nS^2/σ^2 имеет распределение χ^2_{n-1}. Из уравнения

$$\mathbf{P}(g_1 \le \frac{ns^2}{\sigma^2} \le g_2) = \int_{g_1}^{g_2} k_{n-1}(x)\,dx = 1-\alpha$$

находим константы g_1 и g_2. Если требуется односторонний доверительный интервал, то выбор константы (g_1 или g_2) однозначен. При построении двухстороннего интервала обычно выбирают константы по принципу:

$$\int_{-\infty}^{g_1} k_{n-1}(x)\,dx = \int_{g_2}^{\infty} k_{n-1}(x)\,dx = \alpha/2.$$

Если задаться целью построить наикратчайший интервал, то необходимо минимизировать отношение g_1/g_2 при условии $\int_{g_1}^{g_2} k_{n-1}(x)\,dx = 1-\alpha$. Для этого можно воспользоваться методом Лагранжа.

Чтобы построить доверительный интервал для среднего, воспользуемся функцией $G(\vec{X};\,\theta)=\sqrt{n-1}(\overline{X}-a)/s^2$, которая, согласно п. 4 теоремы Фишера, имеет распределение Стьюдента с $n-1$ степенью свободы. Находим константы $g_{\alpha,1}$ и $g_{\alpha,2}$ из уравнений $S_{n-1}(g_{\alpha,1})=1-S_{n-1}(g_{\alpha,2})=\alpha/2$. Отметим, что в силу симметричности распределения Стьюдента $g_{\alpha,1}=-g_{\alpha,2}=t_{\alpha/2}$. Тогда доверительный интервал для a имеет вид

$$[\overline{X}-(st_{\alpha/2})/\sqrt{n-1},\ \overline{X}+(st_{\alpha/2})/\sqrt{n-1}].$$

Пример 2.6. Пусть $X_1,\ldots,X_n$ и $Y_1,\ldots,Y_k$ – независимые выборки из двухпараметрических нормальных распределений $N(a_1,\sigma_1^2)$ и $N(a_2,\sigma_2^2)$ соответственно. Положим известно, что дисперсии удовлетворяют соотношению $\sigma_1^2=r\sigma_2^2$, $c>0$. В этом случае $\theta=(a_1,\,a_2,\,\sigma)$, где $\sigma=\sigma_1$. Построим доверительный интервал для линейной комбинации $\beta a_1+\gamma a_2$, $\beta,\gamma\in\mathbb{R}$. Из теоремы Фишера вытекает, что

$$\sqrt{n}\frac{\overline{X}-a_1}{\sigma}\in N(0,1),\ \sqrt{k}\frac{\overline{Y}-a_2}{r\sigma}\in N(0,1),\ \frac{ns_1^2}{\sigma^2}\in\chi^2_{n-1},\ \frac{ks_2^2}{r^2\sigma^2}\in\chi^2_{k-1},$$

где s_1 и s_2 – выборочные дисперсии. Тогда, в силу независимости выборок, используя свойства соответствующих распределений, заключаем, что

$$\xi=\beta\overline{X}+\gamma\overline{Y}-(\beta a_1+\gamma a_2)\in N(0,(1/n+r/k)\sigma^2)$$

и $s^2/\sigma^2=(ns_1^2+ks_2^2/r^2)/\sigma^2\in\chi^2_{n+k-2}$. Тогда отношение

$$G(\vec{X},\vec{Y};\,\theta)=\frac{\xi\,\Big/\,\left(\sqrt{1/n+r/k}\ \sigma\right)}{\sqrt{(n+k-2)^{-1}s^2/\sigma^2}}=\frac{\sqrt{nk(n+k-2)}(\beta\overline{X}+\gamma\overline{Y}-(\beta a_1+\gamma a_2))}{\sqrt{(k+nr)}\,s}$$

имеет S_{n+k-2}-распределение. Ключевым моментом здесь является сокращение неизвестного параметра σ. Затем находится значение квантили распределения Стьюдента с $n+k-2$ степенями свободы $t_{\alpha/2}$ и получаем доверительный интервал для $\beta a_1+\gamma a_2$:

$$\left[\beta\overline{X}+\gamma\overline{Y}-\frac{\sqrt{k+nr}\ s\,t_{\alpha/2}}{\sqrt{nk(n+k-2)}},\ \ \beta\overline{X}+\gamma\overline{Y}+\frac{\sqrt{k+nr}\,s\,t_{\alpha/2}}{\sqrt{nk(n+k-2)}}\right],$$

где $s=\sqrt{s^2}$. Данная процедура построения доверительного интервала не проходит, если обе дисперсии неизвестны (проблема Беренса – Фишера). Различие состоит в том, что при такой постановке пара (σ_1^2,σ_2^2) – любая точка плоскости, а в решенной задаче она должна лежать на прямой $\sigma_2^2=r\sigma_1^2$. Отчасти здесь может помочь интервальная оценка для отношения σ_1/σ_2. По теореме Фишера $ns_1^2/\sigma_1^2\in\chi^2_{n-1}$, $ks_2^2/\sigma_2^2\in\chi^2_{k-1}$. Тогда распределение $G(\vec{X},\vec{Y},\,\theta)=(\sigma_2^2S_1^2/\sigma_1^2S_2^2)\in F(n-1,\,k-1)$ не зависит от параметра и можно построить доверительный интервал по известной схеме.

Более сложная ситуация возникает, если не удается найти монотонно зависящую от параметра функцию $G(\vec{X};\,\theta)$, распределение которой не зависит от параметра. В этом случае для построения доверительных интервалов используется определенный класс статистик.

Будем говорить, что распределение статистики T монотонно зависит от параметра, если функция распределения $F_T(x;\,\theta) = \mathbf{P}_\theta(T < x)$ монотонно возрастает (или убывает) по параметру θ при каждом фиксированном $x \in \mathbb{R}$. Обычно все разумные точечные оценки параметра обладают этим свойством. В дальнейшем будем считать, что исходная статистика удовлетворяет этому свойству и является точечной оценкой параметра. Более того, будем считать, что $F_T(x;\,\theta)$ – непрерывная функция θ. В силу непрерывности функции F_T по θ уравнения $F_T(x;\,\theta) = \gamma$, $\gamma \in (0,1)$ разрешимы относительно параметра θ. Пусть $b(x,\gamma)$ – корень соответствующего уравнения. Для простоты выкладок считаем, что распределение T непрерывно.

Теорема 2.7. *Пусть распределение статистики T монотонно зависит от параметра, и $F_T(x;\,\theta)$ непрерывна по θ и по x; $\alpha_1 + \alpha_2 = \alpha$. Тогда интервал с границами $b_2 = b(T(\vec{X}),\,1-\alpha_2)$ и $b_1 = b(T(\vec{X}),\,\alpha_1)$ является доверительным уровня $1-\alpha$.*

Доказательство. По теореме Смирнова $F_T(T;\,\theta)$ имеет равномерное на интервале $[0,1]$ распределение, т. е.

$$\mathbf{P}_\theta(\alpha_1 \le F_T(T;\,\theta) \le 1-\alpha_2) = 1-\alpha.$$

В силу монотонности F_T по θ решением неравенства под знаком вероятности является указанный интервал. ■

Замечание 2.5. Данный метод остается верным, если $F_T(x;\,\theta)$ дискретны. Однако в этом случае при определении квантилей $F_T^{-1}(\gamma;\,\theta)$ следует удовлетворить неравенству

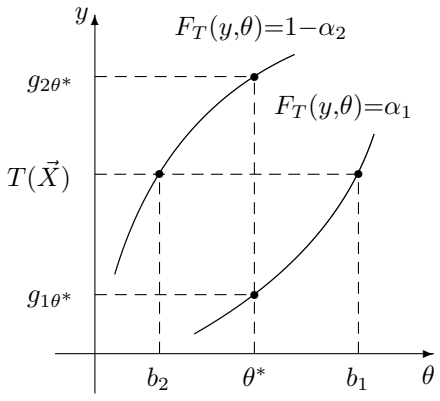
$$F_T(F_T^{-1}(1-\alpha_2)) - F_T(F_T^{-1}(\alpha_1)_-) \ge 1-\alpha.$$

Приведем графическую интерпретацию данного факта. Пусть при каждом значении параметра найдены константы $g_1(\theta)$ и $g_2(\theta)$ такие, что

$$\mathbf{P}_\theta(g_1(\theta) \le T(\vec{X}) \le g_2(\theta)) \ge 1-\alpha$$

(если это возможно, то надо добиваться равенства). Для определенности выбираем их из соотношений $F_T(g_1(\theta)) \le \alpha/2$ и $1 - F_T(g_2(\theta)) \le \alpha/2$. Рассмотрим множество $\Theta \times \Theta$ – конечный или бесконечный прямоугольник на плоскости. Рассмотрим множество точек плоскости (см. рисунок)

$$\mathrm{D} = \{(\theta',\,\theta) : g_1(\theta) \le \theta' \le g_2(\theta)\}.$$



Тогда сечение этого множества на уровне статистики T – соответствующий доверительный интервал.

Пример 2.7. Пусть $X_1, \ldots, X_n$ – выборка из распределения Бернулли Bi(1, θ) с вероятностью успеха θ. В качестве статистики, непрерывно зависящей от параметра по распределению, выберем $\overline{X}$. Очевидно, что

$$F_T(k/n;\, \theta) = \sum_{j=0}^{k-1} C_n^j \theta^j (1-\theta)^{n-j}$$

(имеет место монотонное убывание по θ). Находим значения θ_1 и θ_2 такие, что

$$\sum_{j=n\overline{X}+1}^{n} C_n^j \theta_1^j (1-\theta_1)^{n-j} \le \alpha/2 \quad \text{и} \quad \sum_{j=0}^{n\overline{X}-1} C_n^j \theta_2^j (1-\theta_2)^{n-j} \le \alpha/2$$

(надо добиваться максимальной близости к равенствам). Тогда, в силу приведенных аргументов, интервал $[\theta_1, \theta_2]$ будет доверительным уровня $1-\alpha$.

В общем случае для построения доверительного интервала (области) уровня значимости $1-\alpha$ можно использовать следующий метод. Рассмотрим набор подмножеств выборочного пространства $\{\mathfrak{D}(\theta)\}_{\theta\in\Theta}$ такой, что $P_\theta(\mathfrak{D}(\theta)) \ge 1-\alpha$, $\theta \in \Theta$. Тогда множество всех θ, при которых результат эксперимента $\vec{X}$ попадает в $\mathfrak{D}(\theta)$, будет доверительной областью (а если это интервал, то доверительным интервалом) уровня значимости $1-\alpha$. Однако данный метод слишком общий, и, вообще говоря, трудно ожидать, чтобы он давал хорошие результаты.

При больших объемах выборок можно использовать асимптотический подход к построению доверительных интервалов. Формально, разговор об асимптотических доверительных интервалах конечной выборки не имеет смысла. С другой стороны, если выборка бесконечна, то, по всей вероятности, доверительный интервал будет состоять из одной точки – теоретического значения параметра. Однако можно рассматривать последовательности интервалов, которые аппроксимируют последовательности доверительных интервалов, построенных по конечным выборкам. При больших объемах выборок их, с определенными оговорками, можно использовать вместо доверительных интервалов.

Пусть $X_1, X_2, \ldots$ – последовательность независимых одинаково распределенных случайных величин, и для заданного $\alpha > 0$ существуют статистики (точнее, последовательности статистик) $T^n_{1,\alpha}(X_1, \ldots, X_n)$ и $T^n_{2,\alpha}(X_1, \ldots, X_n)$, такие что

$$\lim_{n\to\infty} \inf_\theta \mathbf{P}_\theta\Big(T^n_{1,\alpha}(X_1, \ldots, X_n) < \theta < T^n_{2,\alpha}(X_1, \ldots, X_n)\Big) \geqslant 1-\alpha.$$

Тогда мы говорим, что задан асимптотический доверительный интервал с уровнем доверия $1-\alpha$.

Рассмотрим метод построения асимптотических доверительных интервалов на базе асимптотически нормальной оценки параметра δ, т. е. с ростом объема выборки имеет место сходимость по распределению

$$\sqrt{n}(\delta(\vec{X}) - \theta) \Longrightarrow N(a, \sigma^2(\theta)).$$

Если исходная последовательность распределений непрерывно меняется по θ, то из состоятельности оценки δ ($\delta(\vec{X}) \xrightarrow[n\to\infty]{} \theta$ по вероятности) немедленно следует, что $\sigma(\delta(\vec{X})) \xrightarrow[n\to\infty]{} \sigma(\theta)$. Тогда последовательность случайных величин

$$\frac{\sqrt{n}(\delta(\vec{X}) - \theta)}{\sigma(\delta)}$$

сходится по распределению к стандартному нормальному закону с известной функцией распределения Φ. Выбрав значение $x_{\alpha/2}$ из условия $1 - \Phi(x_{\alpha/2}) = \alpha/2$ (тогда, в силу симметричности стандартного нормального распределения, $\Phi(x_{\alpha/2}) - \Phi(-x_{\alpha/2}) = \alpha$), получаем последовательность асимптотических доверительных интервалов

$$[\delta(\vec{X}) - x_{\alpha/2}\sigma(\delta(\vec{X}))/\sqrt{n},\ \delta(\vec{X}) + x_{\alpha/2}\sigma(\delta(\vec{X}))/\sqrt{n}].$$

Пример 2.8. Построим доверительный интервал для оценки вероятности биномиального распределения по частоте. Имеет место асимптотическая нормальность. Пусть $w = m/n$ – наблюденная частота успеха. Имеем $\mathbf{P}(|w - p| \le t\sqrt{p(1-p)/n}) \approx 1 - \alpha$, где $\Phi(t) = 1 - \alpha/2$. Чтобы построить доверительный интервал, решим неравенство $|w - p| < t\sqrt{p(1-p)/n}$ или $((t^2/n) + 1)p^2 - 2(w + (t^2/n))p + w^2 < 0$ относительно p. Решением будет интервал $[p_1, p_2]$, где

$$p_1 = \frac{n}{t^2 + n}\left(w + \frac{t^2}{2n} - t\sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right),$$

$$p_2 = \frac{n}{t^2 + n}\left(w + \frac{t^2}{2n} + t\sqrt{\frac{w(1-w)}{n} + \left(\frac{t}{2n}\right)^2}\right).$$

Итак, $[p_1, p_2]$ – асимптотический доверительный интервал для p.

Мы помним, что в задачах оценивания рассматривался байесовский подход, при котором можно говорить о вероятности того, что для данной выборки параметр принимает те или иные значения. Допустим снова, что параметр имеет плотность распределения $p(\theta)$. В этом случае, если получена выборка, то информация о распределении уточняется, и, в соответствии с теоремой Байеса, мы получаем также апостериорное распределение параметра. В этом случае (но ни в коем случае в ранее обсуждавшихся) разумно говорить, что доверительный интервал значений параметра при условии знания выборки тот, который имеет большую (близкую к 1) апостериорную вероятность, т. е.

$$\int_{T_{1,\alpha}(X_1,\ldots,X_n)}^{T_{2,\alpha}(X_1,\ldots,X_n)} \frac{p(\theta)f(X_1,\ldots,X_n;\,\theta)}{f(X_1,\ldots,X_n)}\,d\theta = 1 - \alpha,$$

где $f(x_1,\ldots,x_n) = \int_{-\infty}^{\infty} q(\theta)f(x_1,\ldots,x_n;\,\theta)\,d\theta$. Таким образом, в качестве T_1 и T_2 достаточно взять квантили апостериорного распределения порядков

$(1-\alpha_1)$ и $\alpha_2 : \alpha_1 + \alpha_2 = \alpha$. Так же, как и для небайесовских интервалов, они могут выбираться по различным принципам: односторонние, минимальной длины при заданном доверительном уровне, симметричные по вероятности, симметричные по длине относительно какой-либо оценки и пр.

Итак, в чем различие? В байесовском случае утверждение: для выборки 1, 2, 3, 5 интервал [7, 15] – доверительный интервал с уровнем доверия 0,95 означает, что апостериорная вероятность этого интервала не меньше 0,95. В небайесовском случае то же утверждение фактически означает, что для некоторого семейства областей выборочного пространства $\{\mathfrak{D}(\theta)\}_{\theta\in\theta}$ (заданного априори, фиксированного) такого, что $P_\theta(\mathfrak{D}(\theta)) \geqslant 0,95$, $\theta \in \Theta$, полученная выборка 1, 2, 3, 5 накрывается $\mathfrak{D}(\theta)$, при $\theta \in [7, 15]$.

3. Проверка статистических гипотез

3.1. Постановка задачи

В обыденной жизни под гипотезой понимается некоторое предположение. В математической статистике понятие гипотезы несколько сужается. Какие предположения интересно рассматривать в условиях статистического эксперимента? Прежде всего отметим, что предположения, не имеющие отношения к эксперименту, не представляют интереса, поскольку исход эксперимента не несет никакой информации о таких предположениях. С другой стороны, предположения, касающиеся исхода эксперимента, становятся фактами, поэтому тоже не требуют исследования. Остаются предположения об условиях проведения эксперимента. Приведем примеры задач о проверке гипотез. По результатам наблюдений надо проверить утверждение о том, что параметры измерительного прибора (а это и есть параметры распределения его случайной ошибки) не выходят за пределы нормы. Проверить утверждение о том, что тренировка повысила уровень рабочего или спортсмена (это может быть утверждением о параметрах нормальных распределений каких-либо показателей).

Пусть $(\mathfrak{x}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ – статистический эксперимент. Статистической *гипотезой* будем называть любое предположение о виде или свойствах теоретического распределения.

Пусть H_0 – некоторая гипотеза. Требуется по результатам наблюдений принять гипотезу H_0 или одну из предлагаемых альтернативных (конкурирующих) гипотез. В классической постановке существует одна альтернативная гипотеза H_1, исключающая H_0. Каждая гипотеза может быть описана как множество точек параметрического множества Θ, для которых соответствующее свойство имеет место. Поэтому будем писать, например, $H_0 : \theta \in \Theta_0$, $\Theta_0 \subseteq \Theta$. Поскольку наличие основной гипотезы и альтернативы априори несет некоторую информацию о параметре, то можно считать, что $\Theta = \Theta_0 \cup \Theta_1$. Если гипотеза H имеет вид $H : \theta = \theta_*$, то она называется *простой*. В противном случае, соответствующая гипотеза *сложная*.

Конечно, по выборке конечного объема, вообще говоря, нельзя построить процедуру, дающую достоверный ответ на вопрос, какая из гипотез верна. Исследователь имеет право на ошибку. Далее в таблице приведены возможные ситуации. По вертикали указано возможное положение дел, а по горизонтали – возможные действия исследователя.

Фактическая ситуация	Решение исследователя	
	Принять H_0	Принять H_1
Верна H_0	Верное решение	Ошибка 1-го рода
Верна H_1	Ошибка 2-го рода	Верное решение

Итак, помимо верного решения возможна ошибка 1-го рода, заключающаяся в отвержении основной гипотезы при ее справедливости, а также ошибка 2-го рода, заключающаяся в принятии основной гипотезы при справедливости альтернативной. Обычно ошибка 1-го рода считается наиболее нежелательной.

Отображение $\phi : \mathfrak{X} \to [0,1]$, сопоставляющее каждому набору наблюдений соответствующую вероятность отвергнуть гипотезу H_0, называется *критерием* (или тестом).

Если множество значений критерия $\{0,1\}$ (или содержится в нем), то критерий называется *нерандомизованным* (т. е. результаты наблюдений однозначно определяют решение). Если же возможны промежуточные значения, то критерий рандомизованный (например, если результаты эксперимента не прояснили ситуации – бросить монету). Таким образом, выборочное пространство разбивается на три области – *доверительная* $\{\vec{x} \in \mathfrak{X} : \phi(\vec{x}) = 0\}$, или область принятия гипотезы, *критическая* $\{\vec{x} \in \mathfrak{X} : \phi(\vec{x}) = 1\}$, или область отвержения гипотезы, и *область сомнения* $\{\vec{x} \in \mathfrak{X} : \phi(\vec{x}) \in (0,1)\}$.

Идея подхода Пирсона состоит в том, чтобы ограничить вероятность ошибки 1-го рода некоторым наперед заданным числом α, называемым *уровнем значимости критерия*. При этом вероятность ошибки 2-го рода должна быть по возможности минимальной.

Нетрудно видеть, что если справедлива гипотеза H_0, т. е. $\theta \in \Theta_0$, то $\mathbf{P}_\theta(\text{ош. 1-го рода}) = \mathbf{E}_\theta\phi(\vec{X})$. В противном случае, если $\theta \in \Theta_1$, то $\mathbf{P}_\theta(\text{ош. 2-го рода}) = \mathbf{E}_\theta(1-\phi(\vec{X}))$. *Мощностью критерия* называется функция $\beta : \Theta_1 \to [0,1]$, задаваемая равенством $\beta(\theta) = \mathbf{E}_\theta\phi(\vec{X}) = 1-\mathbf{P}_\theta(\text{ош. 2-го рода})$. Возникает экстремальная задача

$$\sup_{\theta\in\Theta_0} \mathbf{E}_\theta\phi(\vec{X}) \le \alpha,$$
$$\beta(\theta) = \mathbf{E}_\theta\phi(\vec{X}) \to \max_\phi,\ \theta \in \Theta_1.$$

Если решение этой задачи существует, то оно называется *равномерно наиболее мощным* критерием. Само собой, равномерно наиболее мощные критерии существуют довольно редко. Однако в случае проверки простой гипотезы

против простой альтернативы наиболее мощный критерий (для простых гипотез понятие равномерности вырождается) всегда существует и строится явно с использованием статистики отношения правдоподобия.

3.2. Теорема Неймана – Пирсона

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, где $\mathcal{P} = \{P_{\theta_0}, P_{\theta_1}\}$ – статистический эксперимент. Предположим, что меры P_{θ_0} и P_{θ_1} имеют плотности $f((\cdot);\, \theta_0)$ и $f((\cdot);\, \theta_1)$, соответственно, относительно доминирующей меры μ. Обычно в качестве μ выбирается либо мера Лебега (абсолютно непрерывный случай), либо считающая мера (дискретный случай). Поставим задачу проверки простой гипотезы $H_0 : P = P_{\theta_0}$ против простой альтернативы $H_1 : P = P_{\theta_1}$ или, что то же самое, $H_0 : \theta = \theta_0$; $H_1 : \theta = \theta_1$. Введем *статистику отношения правдоподобия*

$$l(\vec{X},\, \theta_1,\, \theta_0) = \frac{L(\vec{X};\, \theta_1)}{L(\vec{X};\, \theta_0)} = \frac{f(\vec{X};\, \theta_1)}{f(\vec{X};\, \theta_0)}.$$

Теорема 3.1. /Неймана – Пирсона/ *В условиях поставленной задачи существует наиболее мощный критерий уровня значимости α. Данный критерий представляется в виде*

$$\phi(\vec{x}) = \begin{cases} 1, & \textit{при}\ \ l(\vec{x}) > c; \\ p, & \textit{при}\ \ l(\vec{x}) = c; \\ 0, & \textit{при}\ \ l(\vec{x}) < c, \end{cases}$$

где константа c и вероятность $p \in [0, 1]$ находятся из уравнения

$$\mathbf{E}_{\theta_0}\phi(\vec{X}) = \mathbf{P}_{\theta_0}(l(\vec{X}) > c) + p\,\mathbf{P}_{\theta_0}(l(\vec{X}) = c) = \alpha.$$

Замечание 3.1. Константа c однозначно находится из приведенного уравнения. Если $\mathbf{P}_{\theta_0}(l(\vec{X}) = c) > 0$, то константа p также находится однозначно. В противном случае, выбор p не имеет значения, поскольку событие $\{l(\vec{X}) = c\}$ имеет нулевую вероятность (невозможно).

Доказательство. Пусть $\overline{\phi}$ – произвольный критерий уровня значимости α. Рассмотрим множество $S = S_+ \cup S_-$, где $S_+ = \{\phi > \overline{\phi}\}$, $S_- = \{\phi < \overline{\phi}\}$. Тогда

$$\mathbf{E}_{\theta_1}\phi - \mathbf{E}_{\theta_1}\overline{\phi} - c(\mathbf{E}_{\theta_0}\phi - \mathbf{E}_{\theta_0}\overline{\phi}) = \int\limits_{S_+ \cup S_-} (\phi - \overline{\phi})(f(\vec{x};\, \theta_1) - cf(\vec{x};\, \theta_2))d\vec{x} \geq 0.$$

Далее отметим, что $\mathbf{E}_{\theta_0}\phi - \mathbf{E}_{\theta_0}\overline{\phi} \geq 0$. Следовательно, критерий ϕ наиболее мощный. ■

Замечание 3.2. Нетрудно заметить, что если взять критерий $\phi^* \equiv \alpha$, то при уровне значимости α его мощность будет равна α. Следовательно, если ϕ – наиболее мощный критерий, то его мощность $\beta(\theta_1) \geq \alpha$.

3.3. Использование правдоподобия при проверке односторонних гипотез

В определенной постановке критерий, основанный на статистике отношения правдоподобия, является наиболее мощным и при проверке сложных гипотез.

Пусть θ – скалярный параметр, $\theta_* \in \Theta$. Рассмотрим задачу проверки гипотезы $H_0 : \theta \le \theta_*$ при альтернативе $H_1 : \theta > \theta_*$. Будем говорить, что семейство $\mathcal{P}$ имеет монотонное (относительно θ_* и T) отношение правдоподобия, если при каждом $\theta \in \Theta : \theta < \theta_*$, статистика отношения правдоподобия $l(\vec{X};\, \theta,\, \theta_*)$ является монотонной функцие й некоторой одномерной статистики $T(\vec{X})$ ($l(\vec{X};\, \theta,\, \theta_0) = l^*(T(\vec{X});\, \theta,\, \theta_0)$). В этом случае решение уравнения $l(\vec{x};\, \theta,\, \theta_*) < c$ может быть записано с использованием статистики T в виде $T < c^*$, если l^* возрастает, и в виде $T > c^*$, если l^* убывает. Пусть в поставленной задаче семейство $\mathcal{P}$ имеет монотонное отношение правдоподобия относительно θ_* и некоторой статистики T (для определенности считаем, что l^* возрастает). Тогда справедлива следующая теорема.

Теорема 3.2. *При сделанных ранее предположениях существует равномерно наиболее мощный критерий проверки гипотезы H_0 при альтернативе H_1, который имеет вид (единственный T–измеримый с точностью до множеств нулевой вероятности)*

$$\phi(\vec{x}) = \begin{cases} 1, & \textit{если } T(\vec{X}) > c; \\ p, & \textit{если } T(\vec{X}) = c; \\ 0, & \textit{если } T(\vec{X}) < c, \end{cases}$$

где константы c и $p \in [0,1]$ выбираются из уравнения

$$\sup_{\theta \le \theta_*} \mathbf{E}_\theta \phi(\vec{X}) = \mathbf{E}_{\theta_*} \phi(\vec{X}) = \mathbf{P}_{\theta_*}(T(\vec{X}) > c) + p\, \mathbf{P}_{\theta_*}(T(\vec{X}) = c) = \alpha.$$

Доказательство. Исследуем поведение функции $\beta(\theta) = \mathbf{E}_\theta \phi(\vec{X})$. Вероятность ошибки 1-го рода равна $\sup_{\theta \le \theta_*} \beta(\theta)$, а при $\theta > \theta_*$ функция $\beta(\theta)$ – мощность. С учетом теоремы 3.2 (Неймана – Пирсона), критерий ϕ является наиболее мощным уровня значимости $\beta(\theta_0)$ для проверки гипотезы $H_0^* : \theta = \theta_0$ при альтернативе $H_1^* : \theta = \theta_1$ для любых $\theta_0, \theta_1 \in \Theta : \theta_0 < \theta_1$. Тогда, по замечанию 3.2, $\beta(\theta_0) \le \beta(\theta_1)$. Следовательно, в исходной постановке $\sup_{\theta \le \theta_*} \beta(\theta) = \beta(\theta_*)$. Остается отметить, что, согласно теореме Неймана – Пирсона, для любого значения $\theta_+ > \theta_*$ рассматриваемый критерий является единственным наиболее мощным уровня значимости α для проверки гипотезы $H_0^+ : \theta = \theta_*$ при альтернативе $H_1^+ : \theta = \theta_+$. Следовательно, ϕ – равномерно наиболее мощный критерий уровня значимости α для проверки H_0 при альтернативе H_1. Теорема доказана. ■

Если $\mathcal{P}$ – однопараметрическое экспоненциальное семейство с плотностями вида

$$f(\vec{x};\,\theta) = h(\vec{x})\exp(a(\theta)D(\vec{x}) + r(\theta)),$$

то оно имеет монотонное отношение правдоподобия относительно $T(\vec{X}) = D(\vec{X})$ при любом θ_*. Более того, существует равномерно наиболее мощный критерий уровня значимости α для проверки гипотезы $H_0 : \theta \in [\theta_1,\,\theta_2]$ при альтернативе $H_1 : \theta \notin [\theta_1,\,\theta_2]$, который строится по правилу

$$\phi(\vec{x}) = \begin{cases} 1, & \text{если } \; c_1 < U(\vec{X}) < c_2, \\ p_i, & \text{если } \; U(\vec{X}) = c_i, \\ 0, & \text{если } \; U(\vec{X}) \notin [c_1, c_2], \end{cases}$$

где константы c_i, p_i, $i = 1,\,2$, выбираются из уравнений

$$\mathbf{E}_{\,\theta_1}\phi(\vec{X}) = \mathbf{E}_{\,\theta_2}\phi(\vec{X}) = \alpha.$$

Пример 3.1. Для целей некоторого химического производства желательно, чтобы вода содержала не более одной бактерии на единицу объема $v = 1$. Для проверки чистоты воды отбирается n проб объема v. Каждая из этих проб добавляется в пробирку с питательной средой. Если проба была загрязнена (т. е. содержала хоть одну бактерию), то раствор в соответствующей пробирке потемнеет. Считаем, что бактерии случайным образом распределены по исходному объему жидкости. Концентрацией ν будем называть среднее число бактерий на единицу объема. Положим, что $m = \nu V$ бактерий случайным образом распределены в объеме V. Тогда вероятность того, что в отобранной пробе объема $v = 1$ будет в точности k бактерий, вычисляется по формуле Бернулли

$$\mathbf{P}(\mu_m = k) = C_m^k p^k (1-p)^{m-k},$$

где $p = v/V$. Далее отметим, что по теореме Пуассона, приближенно (почти точно, если $v << V$)

$$\mathbf{P}(\mu_m = k) \approx \frac{\lambda^k}{k!}\, e^{-\lambda},$$

где $\lambda = mp = \nu V(v/V) = \nu v = \nu$. В частности $\mathbf{P}(\mu_m = 0) = e^{-\lambda}$. Кроме того, при $v << V$ можно считать, что отбор проб происходит независимо.

Итак, исходный набор наблюдений представляет собой выборку из распределения Бернулли (успех – проба чистая) с параметром $p = e^{-\lambda}$. Построим наиболее мощный критерий проверки гипотезы $H_0 : p \geq e^{-1}$ $(\nu \leq 1)$ при альтернативе $H_1 : p < e^{-1}$ $(\nu > 1)$ для выборки из распределения Бернулли. Функция правдоподобия имеет вид

$$L(\vec{x};\,p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{(1-x_i)} = p^{\sum_{k=1}^n x_k}(1-p)^{(n-\sum_{k=1}^n x_k)}.$$

Тогда статистика отношения правдоподобия представляется в виде

$$l(\vec{X}; p, p_0) = \Big(p(1-p_0)/(p_0(1-p))\Big)^{\sum_{k=1}^n X_k}\Big((1-p)/(1-p_0)\Big)^n.$$

Очевидно, что данная статистика монотонно зависит (убывает) от статистики $T(\vec{X}) = \sum_{k=1}^n X_k$ – число зараженных проб в выборке. Тогда наиболее мощный критерий имеет вид, приведенный ранее с $D(\vec{X}) = T(\vec{X})$, только знаки неравенств будут противоположными.

3.4. Использование правдоподобия при проверке сложной гипотезы согласия

Поставим задачу проверки согласия результатов наблюдений с гипотезой $H_0 : \theta \in \Theta_0$, где $\theta \in \mathbb{R}^k$, $\dim(\Theta_0) = l$. Рассмотрим статистику отношения правдоподобия

$$\lambda_n = \frac{\sup_{\theta\in\Theta_0} L(\vec{X};\, \theta)}{\sup_{\theta\in\Theta} L(\vec{X};\, \theta)}.$$

Теорема 3.3. *При выполнении ряда ограничительных условий регулярности, в условиях H_0 имеет место асимптотическое соотношение*

$$\lim_{n\to\infty} \mathbf{P}(-2\,\ln\lambda_n > t) = \mathbf{P}(\chi^2_{k-l} > t), \quad t \geq 0.$$

Данное соотношение позволяет при больших n приближенно вычислять границу критической области с заданным уровнем значимости.

Пример 3.2. Пусть $X_1, \ldots, X_n$ – выборка из двухпараметрического нормального распределения $N(\theta, \sigma^2)$. Рассмотрим гипотезу $H_0 : \theta = \theta_0$ (при неизвестном параметре σ^2 – это сложная гипотеза). Функция правдоподобия имеет вид

$$L_n(\vec{X};\, \theta,\, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{n/2} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \theta)^2\Big).$$

Отметим, что $\sup(L_n(\vec{X};\, \theta,\, \sigma^2),\ \theta = \theta_0,\ \sigma^2 > 0)$ достигается при $\sigma^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^n (X_i - \theta_0)^2 = s_0^2$, а $\sup(L_n(\vec{X};\, \theta,\, \sigma^2),\ \theta \in \mathbb{R},\ \sigma^2 > 0)$ достигается при $\theta = \overline{X}$, $\sigma^2 = s^2$ (оценки максимального правдоподобия). Тогда

$$-2\,\ln\lambda_n = -2\,\ln\frac{(s\sqrt{2\pi})^n \exp(-n/2)}{(s_0\sqrt{2\pi})^n \exp(-n/2)} = n\,\ln(s_0^2/s^2).$$

Далее отметим, что $s^2 = s_0^2 - (\theta_0 - \overline{X})^2$. Следовательно,

$$-2\,\ln\lambda_n = n\ln\left(\frac{s^2 + (\overline{X} - \theta_0)^2}{s^2}\right) = n\ln\left(1 + \frac{(\overline{X} - \theta_0)^2}{s^2}\right).$$

Данная статистика является монотонной функцией от выражения $\frac{(\overline{X}-\theta_0)^2}{s^2}$. Таким образом, асимптотически данный критерий эквивалентен критерию Стьюдента.

3.5. Различные постановки задач проверки статистических гипотез

Прежде всего рассмотрим класс задач, в которых не предполагается наличие альтернативной гипотезы. В этом случае ошибка 2-го рода не изучается, а следовательно, результатом процедуры проверки не может быть ответ, что гипотеза справедлива с определенной долей достоверности. Результатом будет либо отвержение основной гипотезы с определенной долей достоверности, либо признание факта о том, что результаты наблюдений не противоречат (или согласуются) основной гипотезе. Решение задач в такой постановке называется проверкой значимости, а соответствующие критерии – *критериями значимости.*

Построение критерия (нерандомизованного) в случае простой гипотезы обычно сводится к выбору функции $G(\vec{X};\,\theta)$, имеющей известное распределение, не зависящее от выбора $\theta \in \Theta_0$, и выбору квантилей $g_{1,\alpha}$ и $g_{2,\alpha}$ таких, что при выполнении основной гипотезы ($\theta \in \Theta_0$)

$$\mathbf{P}_\theta(g_{1,\alpha} < G(\vec{X};\,\theta_0) < g_{2,\alpha}) = 1-\alpha.$$

При этом, в силу неоднозначности выбора $g_{1,\alpha}$ и $g_{2,\alpha}$, следует быть осторожным и следить за тем, чтобы предположение основной гипотезы никак не влияло на их выбор. Обычно имеет смысл выбирать их либо из условий

$$F_G(g_{1,\alpha}) = 1 - F_G(g_{1,\alpha}) = \alpha/2,$$

либо из условий минимизации длины интервала $[g_{1,\alpha},\, g_{2,\alpha}]$.

Значение критерия принимается равным 0 (подтверждение согласования), если $G(\vec{X};\,\theta_0)$ попадает в доверительный интервал (область), и принимается равным 1 (отвержение гипотезы) в противном случае. В случае, если распределение $G(\vec{X};\,\theta_0)$ имеет атомы, то равенство в уравнении для нахождения g_i, $i \in \{1,\,2\}$, возможно не достигается. Тогда можно ввести рандомизацию на границе. Обычно задача построения нерандомизованного критерия значимости эквивалентна (известная двойственность) задаче построения доверительного интервала (доверительной области).

По характеру основной гипотезы (определяемой природой полученных данных и целями исследования) можно выделить следующие группы задач.

Задача проверки согласия. Пусть производится n независимых наблюдений некоторой случайной величины X, имеющей неизвестное распределение P_X; $R(P_X)$ – некоторая характеристика соответствующего распределения (это может быть само распределение или функция распределения,

принадлежность тому или иному классу распределений, числовая характеристика, и т. д.). Задача состоит в проверке согласования результатов наблюдений с гипотезой $H_0 : R(P_X) = R(P_0)$. Вообще говоря, гипотезой согласия называется любое утверждение о характере P_X.

Задача проверки однородности. Рассматриваются m независимых выборок $X_{i,1}, \ldots, X_{i,n_i}$, $i=1, \ldots, m$, из неизвестных распределений $P_1, \ldots, P_m$ с соответствующими функциями распределения $F_1, \ldots, F_m$. Задача состоит в проверке согласования результатов наблюдений с гипотезой однородности $H_0 : F_1 \equiv \ldots \equiv F_m$, т. е. все выборки производились из одного и того же распределения.

Задача проверки независимости. Рассматриваются m выборок $X_{i,1}, \ldots, X_{i,n_i}$, $i = 1, \ldots, m$, из неизвестного распределения $P_{\vec{X}}$ с соответствующей функцией распределения $F_{\vec{X}}$. Пусть $F_{X_1}, \ldots, F_{X_m}$ – функции распределения компонент (также неизвестные). Задача состоит в проверке гипотезы независимости компонент исходных векторов $H_0 : F_{\vec{X}}(\vec{x}) = F_{X_1}(x_1) \cdots F_{X_m}(x_m)$.

Задача проверки случайности. Пусть $X_1, \ldots, X_n$ – исходные наблюдения (случайный вектор), имеющие неизвестные совместную функцию распределения $F_{\vec{X}}$ и функции распределения компонент F_{X_i}, $i = 1, \ldots, n$. Надо ответить, согласуются ли результаты наблюдений с гипотезой независимости и одинаковой распределенности исходных случайных величин $H_0 : F_{\vec{X}}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_1}(x_n)$.

Ясно, что гипотезы указанных типов используются и в задачах с альтернативами. Отметим, что мы ничего не можем сказать о качестве критерия, если не рассматривать альтернативной гипотезы. В частности, для исследования свойств критерия, построенного по описанной процедуре на базе $G(\vec{X}; \theta)$, необходимо исследование свойств $G(\vec{X}; \theta)$ при альтернативе.

Среди задач проверки статистических гипотез можно выделить класс параметрических задач, т. е. если рассматриваемое параметрическое семейство $\mathcal{P}$ параметризовано (или допускает параметризацию) векторно-значным параметром $\theta = (\theta_1, \ldots, \theta_m)$. Обычно $\mathcal{P}$ – некоторый класс распределений с параметром сдвига, масштаба и т. д. (например, двухпараметрический класс нормальных распределений). Задача проверки простой гипотезы при простой альтернативе также допускает параметризацию. В противном случае задача называется непараметрической. Критерий, позволяющий решать непараметрическую задачу, называется *непараметрическим*.

Приведем примеры непараметрических задач. Является ли заданная выборка выборкой из нормального распределения? Это – задача согласия. В определенных задачах можно проверять непараметрическую гипотезу о равенстве двух или более распределений, из которых получены две или более выборки. Может быть также проверена гипотеза независимости двух выборок, из которых получена двухмерная выборка. Кроме этих двух гипотез (однородности и независимости) могут быть проверены гипотезы слу-

чайности о том, что выборка $X_1, \ldots, X_n$ имеет функцию распределения $F(x_1, \ldots, x_n) = F(x_1) \cdots F(x_n)$, где $F(x)$ – функция распределения некоторой случайной величины, а также гипотеза согласия о том, что наблюдения имеют некоторое заданное распределение. Для проверки непараметрических гипотез используется самый разнообразный аппарат. Это и критерий χ^2 с группировкой наблюдений[2] и критерии типа Колмогорова, Смирнова и Крамера – Фон Мизеса – Смирнова, использующие предельные распределения различных характеристик отклонения теоретической функции распределения от эмпирической функции распределения. Эти критерии проверяют гипотезу о распределениии, а не об отдельных его характеристиках: большая выборка с вероятностью, близкой к 1, позволяет получить хорошее приближение функции распределения на всей прямой. Для проверки нормальности могут использоваться выборочные аналоги асимметрии и эксцесса – характеристик, обращающихся в 0 для нормального распределения. Чрезвычайно удобными для непараметрических задач являются так называемые *ранговые критерии.* Мы уже давали определение вариационного ряда выборки $X_1, \ldots, X_n$. Полагая для простоты, что элементы выборки попарно различны (в случае непрерывных распределениий это имеет место с вероятностью 1), упорядочим X_i в порядке возрастания $X_{(1)}, \ldots, X_{(n)}$. Для нашей выборки ранг элемента X_i есть R_i, $i = 1, \ldots, n$ – число элементов, не превосходящих X_i, т. е. номер X_i в вариационном ряду – $X_i = X_{(R_i)}$.

Критерии, использующие ранги, имеет смысл применять при проверке гипотез случайности, симметричности ($f(x) = f(-x)$, где $f(x)$ – плотность распределения), согласия, независимости, однородности. Например, при проверке гипотез однородности естественна идея рассматривать ранги элементов одной из выборок в объединенной выборке. Если гипотеза верна, то трудно ожидать, например, что все элементы одной выборки окажутся слева, а другой – справа, т. е. можно надеяться с помощью какой-либо статистики, являющейся функцией от рангов, получить хороший критерий.

У непараметрических гипотез тоже бывают альтернативы, иногда соответствующие существу задачи, иногда предлагаемые в качестве специального приема (например, удобные при вычислении мощности). Сформулируем, например, альтернативу равенству случайных величин такую, как "стохастически больше". Мы говорим, что случайная величина Y стохастически больше, чем Z, если для их функций распределения $F_Y(x)$ и $F_Z(x)$ при каждом X верно $F_Y(x) \leqslant F_Z(x)$ и хоть при одном x неравенство строгое. В частности, если график функции распределения или график плотности распределения случайной величины сдвинуть вправо, то получится стохастически большая случайная величина. Из двух нормальных распределениий с одинаковыми дисперсиями стохастически большее то, у которого больше математическое ожидание. Две случайные величины с одинаковым математическим ожида-

[2]Применение распределения χ^2 в статистике уже знакомо нам по проверкам сложных гипотез методом отношения правдоподобия и гипотезы. Оно также используется при проверке гипотез о дисперсии нормального распределения.

нием и разными дисперсиями нельзя сравнивать с этой точки зрения (две точки на графике, в которых неравенства противоположны, найти легко). В подобных случаях могут использоваться иные классы альтернатив.

Еще примеры альтернатив: $F_1(x) = F_0(x - \Delta)$ как альтернатива равенству – альтернативное распределение есть ненулевой сдвиг гипотетического (то же самое, что прибавить константу к случайной величине или сдвинуть график функции распределения). "Стохастически больше обобщение положительного сдвига. Альтернатива независимости: случайные величины X_i, Y_i представимы в виде $X_i = X_i^* + \Delta Z_i$, $Y_i = Y_i^* + \Delta Z_i$, где $\Delta > 0$, X_i^*, Y_i^*, Z_i – взаимно независимы и их распределения не зависят от i.

Разумеется, непараметрические методы могут использоваться и при решении параметрических задач. Обычно непараметрические критерии уступают параметрическим, поскольку не используют специфику класса $\mathcal{P}$. Например, если известно, что исходная выборка из нормального распределения $N(a, \sigma^2)$, то для проверки гипотезы согласия $H_0 : a = a_0$ при неизвестном σ^2 не рекомендуется использовать критерии типа Колмогорова, ω^2 или ранговые критерии, а следует использовать критерий Стьюдента (см. задачу построения доверительных интервалов для параметров нормального закона), который иногда называют t-критерием.

Рассмотрим простой непараметрический критерий – критерий знаков. Пусть $F(x)$ есть теоретическая функция распределения, и гипотеза H_0 состоит в том, что для некоторой точки v мы имеем $F(v) = p$. Альтернативная гипотеза состоит в том, что $F(v) \neq p$. В качестве статистики критерия знаков берется статистика $G(\vec{X};\, F) = G(X_1, \ldots, X_n)$ – число тех X_i, для которых $X_i - v < 0$. Каково выборочное распределение этой статистики? Если верна наша основная гипотеза H_0, то $P_{H_0}(G(\vec{X}) = k) = C_n^k p^k (1-p)^{n-k}$. Заметим, что это распределение не зависит от того, какое именно распределение с таким значением $F(v)$ имеет исходный набор наблюдений (весьма характерное свойство непараметрических критериев). Понятно, что ожидать этого следует и от ранговых критериев. Возвращаемся к нашей задаче. Критерий знаков сводит ее к обычной схеме независимых испытаний: успех – $X_i < v$, т. е. вероятность успеха p. Критическое множество – множество всех таких выборок $\vec{X}$, что $G(\vec{X}) \notin (c_1, c_2)$, где числа c_1, c_2 выбраны таким образом, что $P_{H_0}(G(\vec{X}) \in (c_1, c_2)) \geqslant 1 - \alpha$. Аналогичным образом могут быть построены и односторонние критерии для проверки нулевой гипотезы при односторонних альтернативах ($F(v) = p$ против $F(v) > p$, $F(v) < p$ и т. д.). Из этой же аналогии со схемой независимых испытаний непосредственно вытекает, что для любого альтернативного распределения мощность зависит только от $F(v)$ и при $n \to \infty$ стремится к 1. Это свойство критерия называют *состоятельностью*. Если нас интересует большая информация о распределении (значения функции распределения в нескольких точках), то задача превратится в так называемую задачу о группировке данных (по отрезкам прямой), которая изучается методом χ^2, о котором речь пойдет впоследствии.

3.6. Критерии Колмогорова, Смирнова и Крамера – Фон Мизеса – Смирнова

В данном разделе рассматривается группа непараметрических критериев, использующих свойства эмпирической функции распределения. Допустим, что надо проверить гипотезу о том, что исходная выборка имеет распределение P_0, функция распределения которого $F_0(x)$ – $H_0 : F \equiv F_0$. Рассмотрим статистику Колмогорова, которая была введена раньше $D_n(\vec{X}) = \sup_{-\infty<x<\infty} |F_n(x)–F_0(x)|$. Как было показано, если F_0 – непрерывная функция распределения, то при справедливости основной гипотезы $D_n \xrightarrow[n\to\infty]{} 0$ с вероятностью 1 (теорема Гливенко–Кантелли) и $\lim_{n\to\infty} \mathbf{P}(\sqrt{n}D_n \leqslant t) = K(t)$, где K – функция распределения Колмогорова (теорема Колмогорова).

Применение теоремы Колмогорова уже обсуждалось. Дадим теперь формулировку в терминах теорий проверки статистических гипотез и интервального оценивания.

Критерий, построенный на базе статистики D_n называется *критерием Колмогорова.* Критическая область уровня значимости α при больших n есть $\{D_n \geqslant t_\alpha/\sqrt{n}\}$, где $K(t_\alpha) = 1 - \alpha$, т. е.

$$\mathbf{P}\Big(F_n(x) - \frac{t_\alpha}{\sqrt{n}} \leqslant F(x) \leqslant F_n(x) + \frac{t_\alpha}{\sqrt{n}} \text{ при всех } x\Big) \approx 1 - \alpha.$$

Таким образом, если взять полосу шириной $2t_\alpha/\sqrt{n}$ около графика эмпирической функции распределения, то с вероятностью $1 - \alpha$ график истинной функции распределения лежит в такой полосе. Мы видим, что распределение статистики снова не зависит от рассматриваемого распределения. Применяя преобразование Смирнова, свойства которого обсуждались ранее, получаем, что при справедливости основной гипотезы $\overrightarrow{F_0}(\vec{X}) = (F_0(X_1), \ldots, F_0(X_n))$ – выборка из $U(0,1)$ распределения. Тогда $D_n(\vec{X};\, F_0) = D_n(\overrightarrow{F_0}(\vec{X});\, U)$. Действительно, $F_n(x) = U_n(F_0(x))$, ибо наше преобразование, будучи монотонным, сохраняет все неравенства. Очевидно, что

$$\sup_{-\infty<x<\infty} |F_n(x) - F_0(x)| = \sup_{-\infty<x<\infty} |U_n(F_0(x)) - F_0(x)| = \sup_{0<y<1} |U_n(y) - y|.$$

Предельная функция распределения протабулирована.

Еще один критерий – Крамера – Фон Мизеса – Смирнова. Рассмотрим вместо максимального отклонения среднеквадратичное

$$\omega^2 = \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x)$$

(если есть плотность $f(x)$, то $dF(x) = f(x)dx$). Эта статистика может быть выражена через вариационный ряд этой выборки $X_{(1)} < \ldots < X_{(n)}$:

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^{n}\Big(F(X_{(i)}) - \frac{2i-1}{2n}\Big)^2.$$

Предельное распределение ω^2 протабулировано. Построение критических областей и доверительных интервалов осуществляется аналогично предыдущему случаю (правда, представление в виде полосы на сей раз невозможно). Критические значения достаточно точны уже при $n = 3$.

Аналог статистики Колмогорова был использован Смирновым для проверки гипотезы однородности. Постановка задачи такова. Пусть у нас имеются две независимых выборки $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_m)$ из непрерывных распределений. Пусть требуется проверить гипотезу о том, что обе выборки из одного распределения. Статистика критерия –

$$D_{n,m} = \sup_{-\infty<x<\infty} |F_{1,n}(x) - F_{2,m}(x)|,$$

где $F_{1,n}$ и $F_{2,m}$ – эмпирические функции распределения, построенные по нашим выборкам. Если наша гипотеза верна, то распределение статистики $\sqrt{\frac{mn}{m+n}}\, D_{n,m}$ стремится к распределению с функцией распределения $K(x)$. Полагая $K(t_\alpha) = 1 - \alpha$, сформулируем критерий однородности Смирнова: если объемы выборок n и m достаточно велики, то, вычислив по выборкам $u = \sqrt{\frac{mn}{m+n}}\, D_{n,m}$, принимают решение отвергнуть гипотезу (приблизительно с уровнем значимости α) в том и только в том случае, если $u \geqslant t_\alpha$.

Можно предложить в качестве альтернативы любую пару не равных между собой распределений, найти для них распределение построенной нами статистики и мощность нашего критерия. Показано, что при больших n и m она близка к 1, т. е. налицо состоятельность.

Обратим внимание на одну интересную связь. Статистика критерия Смирнова характеризует попросту взаимное расположение элементов вариационных рядов наших двух выборок. То есть естественно ожидать, что наш критерий является ранговым. Это и в самом деле так. Сейчас мы в этом убедимся. Пусть $(X_{(1)} < \ldots < X_{(n)})$ – вариационный ряд выборки $X_1, \ldots, X_n$. Мы говорили о том, что рангом R_k элемента X_k называется его номер в вариационном ряду $(X_k = X_{(R_k)})$. Введем понятие антиранга – номера элемента вариационного ряда в выборке $R_{I_{(i)}} = i$. Если выборка – (11, 2, 7, 6), то вариационный ряд – (2, 6, 7, 11) и антиранги – $I_{(1)} = 2$, $I_{(2)} = 4$, $I_{(3)} = 3$, $I_{(4)} = 1$. Составим теперь из наших двух выборок объединенную выборку $Z_1, \ldots, Z_{m+n}$, первые n элементов которой – X-сы, остальные – Y-ки. Пусть $I_{(1)}, \ldots, I_{(m+n)}$ – антиранги и $Z_{(1)}, \ldots, Z_{(m+n)}$ – вариационный ряд объединенной выборки. Положим $C_i = 1$, если $i = 1, \ldots, n$; $C_i = 0$, если $i = n+1, \ldots, m+n$ и введем статистику T:

$$T = \Big(\frac{m+n}{nm}\Big)^{1/2} \max_{1\leqslant j\leqslant m+n} \Big|\frac{m+n}{nm}\, j - C_{I_{(1)}} - \ldots - C_{I_{(j)}}\Big|.$$

Ясно, что $B_j = \sum_1^j C_{I_{(i)}}$ есть число элементов первой выборки среди первых j элементов вариационного ряда. Чем больше оно, тем больше первая выборка "смещена влево"относительно второй. Далее видно, что

$$B_j/n - (j - B_j)/m = (m+n)(B_j - nj/(m+n))/(mn)$$

есть значение разности $F_{2,n}(x) - F_{1,m}(x)$ для произвольной точки x, $Z_{(k+1)} > x > Z_{(k)}$ и что в точках такого типа разность принимает все свои ненулевые значения. Поэтому мы получаем, что

$$T = \Big(\frac{mn}{n+m}\Big)^{1/2} \max_{-\infty<x<\infty} |(F_{2,n}(x) - F_{1,m}(x))|,$$

т. е. как раз статистика критерия однородности Смирнова.

3.7. Критерий хи-квадрат и его применения к проверке сложных гипотез

Пусть $\vec{X} = (X_1, \ldots, X_n)$ – выборка из распределения P с функцией распределения F. Рассмотрим задачу проверки значимости гипотезы согласия $H_0 : F = F_0$ (подразумевается, что семейство $\mathcal{P}$ непараметрическое). Предположим, что множество возможных значений X_1 (определяемое видом семейства $\mathcal{P}$) разбито на N непересекающихся подмножеств, например, само состоит из N точек (N исходов эксперимента, N состояний системы) или множество $-\infty < x < \infty$ разбито на интервалы $I_i = [t_i, t_{i+1}]$, $i = 1, \ldots, N$, с границами $-\infty = t_1 < t_2 < \ldots < t_{N-1} < t_N = \infty$. Пусть n_i – число элементов выборки $\vec{X} = (X_1, \ldots, X_n)$, попавших в интервал I_i (так называемый метод группировки наблюдений). Обозначим также вероятности $P_{H_0}(X_1 \in I_i)$ через p_i (теоретические частоты). Частота n_i/n является состоятельной оценкой p_i. Используем следующую меру отклонения выборочных значений от теоретических $S = \sum_{i=1}^N f_i(\frac{n_i}{n} - p_i)^2$, где f_i – некоторые веса. Если в качестве таких весов взять n/p_i, то получится статистика так называемого критерия *хи-квадрат*[3]

$$X^2 = \sum_{i=1}^N \frac{n}{p_i}\Big(\frac{n_i}{n} - p_i\Big)^2 = \sum_{i=1}^N \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^N \frac{n_i^2}{np_i} - n.$$

Известно, что распределение (мультиномиальное) вектора нормированных частот $(n_1^*, \ldots, n_{N-1}^*) = \mathbf{n}^*$, $n_i^* = (n_i - np_i)/\sqrt{n}$ стремится к невырожденному нормальному распределению $N(0, \Sigma = ||\sigma_{i,j}||_1^{N-1})$, где $\sigma_{i,i} = p_i(1-p_i)$, $\sigma_{i,j} = -p_ip_j$, $i \neq j$ (ковариации частот мультиномиального распределения).

Известно также, что квадратичная форма $Y^{\mathbf{T}}\Sigma^{-1}Y$, стоящая в показателе экспоненты, выражающей плотность $N(0, \Sigma)$, распределена по закону χ^2_{N-1}. Но именно такой квадратичной формой от нормированных частот

[3] Очевидно, что данный критерий не может различать распределения с одинаковыми теоретическими частотами. Это надо учитывать при выборе альтернативы.

как раз оказывается статистика критерия χ^2 (хи-квадрат):

$$X^2 = \sum_{i=1}^{N-1} \frac{1}{p_i}(n_i^*)^2 + \frac{1}{p_N}(n_1^* + \ldots + n_{N-1}^*)^2 = \mathbf{n}^{*\mathrm{T}}\Omega\mathbf{n}^*,$$

где $\Omega = ||g_{i,j}||_1^{N-1}$, $g_{i,i} = \frac{1}{p_i} + \frac{1}{p_N}$, $g_{i,j} = \frac{1}{p_N}$, $i \neq j$, и, как легко проверить, $\Omega = \Sigma^{-1}$, т. е. имеем утверждение.

Теорема 3.4. *Если* $0 < p_i < 1$, $i = 1, \ldots, N$, *то при* $n \to \infty$ *распределение* X^2 *стремится к* χ^2_{N-1}.

Теорема имеет много практических применений. Имеются рекомендации в методе группировки разбивать на интервалы, вероятности которых достаточно близки между собой. Если n и n_i достаточно велики, то критерий согласия выглядит так: если α – уровень значимости, то при $X^2(\mathbf{n}) \geqslant \chi^2_{1-\alpha,N-1}$, где $\chi^2_{1-\alpha,N-1}$ – квантиль порядка $1-\alpha$ распределения χ^2_{N-1}, гипотезу H_0 отвергают, в противном случае – нет.

Широко применяемый метод проверки сложных гипотез с помощью критерия χ^2 включает в себя оценки максимума правдоподобия для параметров.

Пусть $\vec{X} = \{X_1, \ldots, X_N\}$ – выборка из распределения с полностью неизвестной функцией распределения. Проверяется гипотеза H_0 о принадлежности теоретического распределения исходной выборки некоторому параметрическому семейству $F \in \{P_\theta,\ \theta \in \Theta,\ \Theta \in R^m\}$. Пусть осуществлена группировка данных (т. е. задача сведена к мультиномиальному распределению) или изначально речь идет о мультиномиальном распределении, зависящем от многомерного параметра. Обозначим через $\mathbf{n} = \{n_1, \ldots, n_N\}$ вектор частот $\sum_{j=1}^N n_j = n$. В указанной постановке значение X^2 при справедливости основной гипотезы зависит от θ:

$$X^2(\theta) = \sum_{i=1}^{N} \frac{(n_i - np_i(\theta))^2}{np_i(\theta)},$$

поэтому мы не можем использовать данное выражение в качестве статистики критерия. Выберем из всех значений $X^2(\theta)$ наиболее подходящее. В качестве статистики критерия будем использовать $\widetilde{X}^2 = X^2(\widetilde{\theta}) = \min_{\theta\in\Theta} X^2(\theta)$. При определенных условиях регулярности статистика $\widetilde{X}^2$ асимптотически эквивалентна статистике $\widehat{X}^2 = X^2(\widehat{\theta})$, если $\widehat{\theta}$ – мультиномиальная оценка максимального правдоподобия[4] параметра θ.

Итак, имея вектор частот $\mathbf{n} = \{n_1, \ldots, n_N\}$, мы ищем максимум $L_n(\widehat{\theta})$ функции правдоподобия соответствующего мультиномиального распределения по $\theta \in \theta$

$$L(\mathbf{n};\, \theta) = \frac{n!}{n_1! \ldots n_N!} \prod_{j=1}^{N} p_j^{n_i}(\theta),$$

[4]Мультномиальной оценкой максимального правдоподобия будем называть оценку, максимизирующую функцию правдоподобия соответствующего мультиномиального распределения по параметру $\theta \in \Theta$.

для чего решаются уравнения

$$\sum_{j=1}^{N} \frac{n_j}{p_j(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_k} = 0,\ k = 1, \ldots, m.$$

Теорема 3.5. /Неймана – Фишера/. *Пусть* $p_i(\theta)$, $i = 1, \ldots, N$, $\theta = (\theta_1, \ldots, \theta_m)$ *удовлетворяют следующим условиям:*

а) $\sum_{i=1}^{N} p_i(\theta) = 1$ *для всех* $\theta \in \Theta$;

б) $p_i(\theta) \geqslant c > 0$ *для всех* $\theta \in \Theta$, *существуют непрерывные производные* $\dfrac{\partial p_i(\theta)}{\partial \theta_k}$, $\dfrac{\partial^2 p_i(\theta)}{\partial \theta_k \partial \theta_l}$, $1 \leqslant k, l \leqslant r$;

в) *матрица* $\left\| \dfrac{\partial p_i(\theta)}{\partial \theta_k} \right\|$ *с размерами* $N \times m$ *имеет ранг* m.

Тогда, если $\widetilde{\theta}$ *– оценка, минимизирующая значение* $\chi^2(\theta)$, *и* $\widehat{\theta}_n$ *– мультиномиальная оценка максимума правдоподобия для* $\theta \in \Theta$, *то при справедливости* H_0 *величины* $\widetilde{X}^2$ *и* $\widehat{X}^2$ *сходятся по распределению к* χ^2_{N-m-1}, *при* $n \to \infty$.

Рассмотрим ряд примеров.

Пример 3.3 /генетическая модель Фишера/. При самоскрещивании кукурузы по двум характеристикам (крахмалистая, сахаристая, с белым основанием листа, с зеленым основанием листа) возникают 4 типа потомства. Если N_i – число растений типа i среди общего числа n экземпляров потомства и θ_i – вероятность такого типа, то (N_1, N_2, N_3, N_4) имеет мультиномиальное распределение с вероятностями $P(N_1, N_2, N_3, N_4) = \dfrac{n!}{N_1! N_2! N_3! N_4!}$. В модели Фишера величины θ_i выбираются так, что $\theta_1 = (2+\theta)/4$; $\theta_2 = \theta_3 = (1-\theta)/4$; $\theta_4 = \theta/4$, где θ – некий параметр. Нам надо проверить гипотезу о том, что в нашей мультиномиальной модели θ_i являются указанными функциями θ, т. е. о том, что параметры теоретического мультиномиального распределения лежат на данной прямой в трехмерной гиперплоскости $(\theta_1, \theta_2, \theta_3, \theta_4)$ с $\sum_{i=1}^{4} \theta_i = 1$.

В соответствии с описанной ранее процедурой, мы должны, располагая выборкой N_1, N_2, N_3, N_4, оценить параметр θ методом максимума правдоподобия $\max \prod_{i=1}^{4} \theta_i(\theta)^{N_i}$, т. е. $\max \sum_{i=1}^{4} N_i \ln(\theta_i(\theta))$. Находим производную $\sum_{i=1}^{4} N_i\, \theta_i'(\theta)/\theta_i(\theta)$ и приравниваем к нулю:

$$\frac{N_1}{2+\theta} - \frac{N_2 + N_3}{1-\theta} + \frac{N_4}{\theta} = 0.$$

Отсюда $n\theta^2 + (N_4 + 2N_2 + 2N_3 - N_1)\theta - 2N_4 = 0$. Но левая часть меньше нуля, при $\theta = 0$; больше нуля, при $\theta = 1$; поэтому в интервале (0, 1) уравнение

имеет единственный корень $\widehat{\theta}_n$. Таким образом, критерий согласия χ^2 при уровне значимости α отклоняет гипотезу H_0 лишь в том случае, если

$$\sum_{j=1}^{4} \frac{N_j^2}{(n\theta_j(\widehat{\theta}_n))} - n \geqslant \chi^2_{1-\alpha,\,2}$$

или, что эквивалентно,

$$\frac{N_1^2}{n(2+\widehat{\theta}_n)} + \frac{N_2^2+N_3^2}{n(1-\widehat{\theta}_n)} + \frac{N_4^2}{n\widehat{\theta}_n} \geqslant (\chi^2_{1-\alpha,\,2} + n)/4.$$

Пример 3.4 /проверка гипотезы о показательном распределении/. По выборке $\vec{X} = (X_1, \ldots, X_n)$ надо проверить гипотезу H_0 о том, что распределение имеетфункцию распределения $F_\xi = 1 - e^{-x/\theta}$ (параметр θ неизвестен). Применяя метод группировки данных с интервалами $E_j = [(j-1)a, ja]$, $j = 1, \ldots, N-1$, $E_N = [(N-1)a, \infty)$, где $a > 0$, N – заданные числа, построить критерий согласия χ^2 для гипотезы H_0. Обозначим через $p_j(\theta)$, $j = 1, \ldots, N$, вероятности $\mathbf{P}(\xi \in E_j | H_0)$. Имеем $p_j(\theta) = e^{-(j-1)a/\theta}(1 - e^{-(a/\theta)})$, $j = 1, \ldots, N-1$, $p_N(\theta) = e^{-(N-1)a/\theta}$. Как и во всех подобных задачах, метод максимума правдоподобия нужно применять к схеме независимых испытаний с N исходами и вышеописанными вероятностями. Располагая выборкой объема n, мы знаем число n_j элементов выборки, попавших в интервал E_j, $j = 1, \ldots, N$; $\sum_{j=1}^N n_j = n$. Функция правдоподобия имеет вид $\prod_{i=1}^N P_i(\theta)^{h_i}$, ее логарифм – $\sum_{i=1}^N n_i \ln P_i(\theta)$ и уравнение максимума правдоподобия есть снова $\sum\limits_{i=1}^{N} n_i \dfrac{P_i'(\theta)}{P_i(\theta)} = 0$, т. е., вынося за скобки общий множитель $(-a/\theta)'$, $\sum\limits_{j=1}^{N-1} n_j \dfrac{j - 1 - je^{-a/\theta}}{1 - e^{-a/\theta}} + (N-1)n_N = 0$. Обозначая $z = e^{-a/\theta}$, имеем $\hat{z}_N = (\sum_{j=1}^N jn_j - n)/(\sum_{j=1}^N jn_j - n_N)$, и оценки максимума правдоподобия для вероятностей $\hat{p}_j = p_j(\hat{\theta})$ имеют вид $\hat{p}_j = \hat{z}_N^{j-1}(1 - \hat{z}_N)$, $j = 1, \ldots, N-1$, $\hat{p}_N = \hat{z}_N^{N-1}$.

В соответствии с общей теорией, если n велико и $n_i \geqslant 5$; $j = 1, \ldots, N$, то соответствующий критерий согласия χ^2 отвергает гипотезу H_0 тогда и только тогда, когда $\widehat{X}^2 = \sum\limits_{j=1}^{N} \dfrac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j} \geqslant \chi^2_{1-\alpha,N-2}$, где α – уровень значимости.

Критерии типа χ^2 могут применяться также при проверке гипотез однородности и независимости.

3.8. Ранговые критерии

Рассмотрим задачу проверки гипотезы случайности H_0 вида $F(x_1,\dots,x_n)=F(x_1)\cdots F(x_n)$, где $F(x)$–некоторая (непрерывная в дальнейшем) функция распределения - обычно представимость многомерной функции распределения выборки $\vec{X}=(X_1,\dots,X_n)$ в таком виде предполагается. При проверке и построении статистики критерия будет использоваться вариационный ряд выборки $X_{(1)}<\dots<X_{(n)}$. Мы говорим, что X_i, X_j образуют инверсию, если $i<j$, но $X_j<X_i$, т. е. в вариационном ряду X_j стоит левее X_i. Пусть η_i – число инверсий, образованных X_i, т. е. в вариационном ряду левее X_i стоит η_i значений с большими номерами. Ясно, что если гипотеза верна, то η_i, $i<n-1$ не зависит от $\eta_{i+1},\dots,\eta_{n-1}$, т. е. $\eta_1,\dots,\eta_{n-1}$ взаимно независимы и вся совокупность может принимать $n!$ равновероятных наборов значений; η_i может с вероятностью $(n-i+1)^{-1}$ принимать значения $0,\dots,n-i$. В качестве статистики критерия берется суммарное число инверсий $T_n=\eta_1+\dots+\eta_{n-1}$. Зная вероятности, можно вычислить моменты. Ясно, что

$$\mathbf{E}\,(\eta_i)=\frac{0+1+\dots+(n-i)}{n-i+1}=\frac{n-i}{2},\quad \mathbf{D}(\eta_i)=\frac{\sum_{i=0}^{n-i}(i-(n-i)/2)^2}{n-i+1}.$$

Эти суммы можно вычислять с помощью производных производящих функций. Получаем

$$\mathbf{D}(\eta_i)=\frac{(n-i)(n-i+2)}{12},\ \ \mathbf{E}\,(T_n)=\sum_{i=1}^{n-1}\mathbf{E}\,(\eta_i)=\frac{n(n-1)}{4},$$

$$\mathbf{D}(T_n)=\sum_{i=1}^{n-1}\mathbf{D}(\eta_i)=\frac{2n^3+3n^2-5n}{12}.$$

Критическую область можно взять $\{|t-n(n-1)/4|>t_\alpha(n)\}$, где

$$\mathbf{P}(n(n-1)/4-t_\alpha(n)\leqslant T_n\leqslant n(n-1)/4+t_\alpha(n))\geqslant 1-\alpha.$$

Распределение этой статистики протабулировано, распределение же нормированной статистики $T_n^*=6(T_n-n(n-1)/4)/n^{3/2}$ при истинности H_0 сходится к $N(0,1)$, при $n\to\infty$. Вполне понятно, как, исходя из этого, построить критерий при больших n. Ясно, что при справедливости основной гипотезы распределения векторов $(X_1,\dots,X_n)$ и $(X_n,\dots,X_1)$ совпадают (переставляемость). Пусть S_i – *последовательный ранг* X_i или ранг X_i среди $X_1,\dots,X_i$. Тогда статистики $\sum_{i=1}^n S_i$ и $\sum_{i=1}^n \eta_i$ совпадают по распределению. Таким образом, аналогичный критерий можно построить с использованием последовательных рангов. При этом использовать последовательные ранги удобнее, поскольку они не требуют пересчета с ростом объема выборки.

Обсудим теперь ранговый критерий *Уилкоксона*. Пусть $X_1,\dots,X_m$, $Y_1,\dots,Y_n$ – две случайных выборки из распределениий с функциями рас-

пределения F_1, F_2. Задача состоит в том, чтобы проверить гипотезу однородности $H_0 : F_1 \equiv F_2$. Пусть $R'_1, \ldots, R'_m$ – ранги элементов первой выборки $X_1, \ldots, X_m$ в объединенной выборке $Z_1, \ldots, Z_{m+n}$. Критерий Уилкоксона проверки однородности основан на статистике $W_{m,n} = \sum_{i=1}^{m} R'_i$. Легко понять, что распределение $W_{m,n}$ в случае справедливости основной гипотезы симметрично относительно $\mathbf{E}\,(W_{m,n}) = m(n+m+1)/2; \mathbf{D}(W_{m,n}) = mn(m+n+1)/12$. Для конкретных m и n распределение этой статистики протабулировано. Для больших же (> 25) m или n можно пользоваться асимптотической нормальностью, т. е. тем, что распределение $\dfrac{W_{m,n} - \mathbf{E}\,(W_{m,n})}{\sqrt{\mathbf{D}W_{m,n}}}$ стремится к $N(0,1)$.

Для проверки гипотезы симметрии (случайные величины $X_1, \ldots, X_N$ – независимы и все имеют одну и ту же плотность распределения f, симметричную относительно $0 : f(x) = f(-x)$) используется одновыборочный критерий Уилкоксона. Упорядочим в порядке возрастания абсолютные величины $|X_i| : |X|_{(1)} < \ldots < |X|_{(N)}$. Тогда обозначим через R_i^+ ранг $|X_i|$ в этой возрастающей последовательности. Одновыборочный критерий Уилкоксона использует статистику $W^+ = \sum_{X_i>0} R_i^+$. Легко показать, что в предположении гипотезы симметричности $\mathbf{E}\,W^+ = N(N+1)/4; \mathbf{D}W^+ = N(N+1)(N+2)/24$.

Вычислим распределение статистики критерия Уилкоксона (например, двухвыборочного) $W_{m,n} = \sum_{i=1}^{m} R'_i$ при нулевой гипотезе, которое, очевидно, не зависит от F_1. Пусть $\pi_{m,n}(k)$ – число таких взаимных расположений $\{X_i\}$ и $\{Y_i\}$ в едином вариационном ряде $Z_{(1)}, \ldots, Z_{(m+n)}$, что $W_{m,n} = k$, $k = m(m+1)/2, \ldots, m(m+2n+1)/2$. В этом случае вероятность $\mathbf{P}(W_{m,n} = k) = \dfrac{\pi_{m,n}(k)}{(m+n)/n}$. Числа $\pi_{m,n}(k)$ удовлетворяют рекуррентному соотношению

$$\pi_{m,n}(k) = \pi_{m,n-1}(k) + \pi_{m-1,n}(k-m-n)$$

и граничным условиям

$$\pi_{m,0}(m(m+1)/2) = 1, \quad \pi_{m,0}(k) = 0, \quad k \neq m(m+1)/2;$$

$$\pi_{0,n}(k) = 0, \quad k \neq 0; \quad \pi_{m,n}(k) = 0, \quad k < m(m+1)/2.$$

Равенство для вероятности очевидно (в предположении нашей гипотезы все исходы равновероятны), а рекуррентные соотношения получатся, если рассмотреть случаи: $Z_{(m+n)}$ есть некое Y_j и $Z_{(m+n)}$ есть некое X_i – в первом случае $W_{m,n-1} = k$, во втором – $W_{m-1,n} = k - m - n$.

Заметим, что статистика критерия Уилкоксона в некотором смысле аналогична статистике критерия Стьюдента. Напишем ранговый аналог последней:

$$\sqrt{\frac{n_1 n_2}{n}} \frac{(n_2^{-1} \sum_{i=n_1+1}^{n} R_i - n_1^{-1} \sum_{i=1}^{n_1} R_i)}{((n-2)^{-1} \sum_1^n (R_i - \overline{R})^2)^{1/2}},$$

где $\overline{R} = n^{-1}\sum_1^n (R_i)$, но

$$\sum_{i=n_1+1}^{n} R_i = \frac{n(n+1)}{2} - \sum_{i=1}^{n_1} R_i, \quad \sum_{i=1}^{n}(R_i - \overline{R})^2 = \frac{n(n^2-1)}{12},$$

и наша статистика есть функция от статистики критерия Уилкоксона.

Обсудим применения ранговых критериев к проверке независимости. Пусть $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ – две выборки и надо проверить гипотезу их независимости. Пусть в первой выборке ранги – $R_1, \ldots, R_n$, во второй – $R'_1, \ldots, R'_n$. Рассмотрим ранговую статистику – аналог коэффициента корреляции (ранговый коэффициент корреляции Спирмена)

$$\rho = \frac{\sum_{i=1}^n (R_i - \overline{R})(R'_i - \overline{R'})}{(\sum_{i=1}^n (R_i - \overline{R})^2 \sum_{i=1}^n (R'_i - \overline{R'})^2)^{1/2}}.$$

Ясно, что $\overline{R} = \overline{R'} = (n+1)/2$, т. е.

$$\sum_{i=1}^{n}(R_i - \overline{R})^2 = \sum_{i=1}^{n}(R'_i - \overline{R'})^2 = \sum_{i=1}^{n} i^2 - \frac{n(n+1)^2}{2} = \frac{n(n^2-1)}{12},$$

$$\rho = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - R'_i)^2 = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(i - R_i^0)^2,$$

где $(1, R_1^0), \ldots, (n, R_n^0)$ – соответствующая перестановка пар (R_i, R'_i). Имеем $\mathbf{E}\,(\rho) = 0$, $\mathbf{D}(\rho) = (n-1)^{-1}$. Разумеется, использовалось то, что при справедливости основной гипотезы все $n!$ перестановок $(R_1^0, \ldots, R_n^0)$ равновероятны и $\mathbf{E}\, R_i^0 = \sum_{j=1}^{n} \frac{j(n-1)!}{n!}$. При полном соответствии рангов $\rho = 1$, при соответствии $(1,n), (2,n-1), \ldots, (n,1)$ – $\rho = -1$. Известно, что при справедливости H_0 предельное распределение статистики $\sqrt{n}\,\rho$ при $n \to \infty$ стандартное нормальное $N(0,1)$, т. е. следует отвергать H_0 на уровне значимости $\approx \alpha$ при $\sqrt{n}\,|\rho| \geqslant t_\alpha$, $\Phi(-t_\alpha) = \alpha/2$. Известно, что существует более точная аппроксимация с помощью распределения Стьюдента.

Имеется также коэффициент ранговой корреляции Кендалла

$$\tau = \frac{1}{n(n-1)}\sum_{i\neq j} \operatorname{sign}(R_i - R_j)\operatorname{sign}(S_i - S_j) =$$

$$= \frac{1}{n(n-1)}\sum_{i\neq j} \operatorname{sign}(i-j)\operatorname{sign}(R_i^0 - R_j^0).$$

Он обладает свойствами, аналогичными свойствам коэффициента ранговой корреляции Спирмена. Можно показать, что если верна гипотеза независимости, то $\mathbf{E}\,\tau = 0$, $\mathbf{D}\tau = 2(2n+5)/(9n(n-1))$, распределение $\tau/\sqrt{\mathbf{D}\tau}$ сходится к $N(0,1)$ при $n \to \infty$.

4. Линейные методы статистики

4.1. Метод наименьших квадратов. Теорема Гаусса – Маркова

Метод наименьших квадратов (МНК) был предложен Гауссом в начале XIX в. для оценивания в задачах астрономических измерений. Предположим в общем случае, что наблюдения Y_i, $1 \le i \le n$, представимы в виде $Y_i = g_i(\beta_1, \ldots, \beta_r) + \epsilon_i$, где g_i – известные функции, а параметры $\beta_1, \ldots, \beta_r$ надо оценить. Относительно остатков ϵ_i предполагается, что $\mathbf{E}(\epsilon_i) = 0$, $\mathbf{D}(\epsilon_i) = \sigma^2$, $1 \leqslant i \leqslant n$, $\mathbf{cov}(\epsilon_i, \epsilon_j) = 0$, $i \neq j$. Естественно, впрочем, делать и более сильные предположения, например, что ϵ_i – независимые с нулевым средним и одной и той же дисперсией или нормальные независимые с нулевым средним и одной и той же дисперсией $N(0, \sigma^2)$. Такое распределение напоминает распределение случайных ошибок измерений. Допустим, например, что Y_i зависят от некоторых известных величин X_i, которые, стало быть, участвуют в определении функций g_i. Важный частный случай, когда зависимость эта линейна и соответствующие коэффициенты являются искомыми параметрами. Модель в этом случае может иметь вид, например, $Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$, где ошибки ϵ_i независимы. Если ошибки имеют нулевое среднее и одну и ту же дисперсию, то модель называется моделью линейной регрессии (в этом примере $r = 2$).

В общей задаче нахождения параметров в соответствии МНК параметры $\widehat{\beta}_1, \ldots, \widehat{\beta}_r$ ищут таким образом, чтобы минимизировать выражение $\sum_{i=1}^n (Y_i - g_i(\beta_1, \ldots, \beta_r))^2$, когда $\beta_1, \ldots, \beta_r$ пробегает все параметрическое пространство. Будем решать задачу МНК с помощью дифференциального исчисления. Набор параметров $\beta_1, \ldots, \beta_r$ должен удовлетворять следующим уравнениям:

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^n (Y_i - g_i(\beta_1, \ldots, \beta_r))^2 = 0, \quad j = 1, \ldots, r.$$

Эти уравнения называются *нормальными уравнениями*. Они имеют вид

$$\sum_{i=1}^n (Y_i - g_i(\beta_1, \ldots, \beta_r)) \frac{\partial}{\partial \beta_j} g_i(\beta_1, \ldots, \beta_r) = 0, \quad j = 1, \ldots, r.$$

Простейшие примеры линейных задач на МНК:

1. $Y_i = \beta_1 + \epsilon_i$ (измерительный прибор). Нормальное уравнение имеет вид $\sum_{i=1}^n (Y_i - \beta_1) = 0$ и $\widehat{\beta}_1 = \overline{Y}$.

2. В регрессионной модели с двумя параметрами $Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$ имеем нормальные уравнения

$$\sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_i) = 0, \quad \sum_{i=1}^n X_i (Y_i - \beta_1 - \beta_2 X_i) = 0.$$

Если не все X_i одинаковы, то решения имеют вид

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^n (X_i - \overline{X})Y_i}{\sum_{i=1}^n (X_i - \overline{X})^2}, \quad \widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X}.$$

Если не все дисперсии ϵ_i равны (неравноточные измерения), но известны их отношения, то делением Y_i на надлежащие константы можно свести задачу к предыдущему случаю и оценивать сумму квадратов новых отклонений, которая окажется взвешенной суммой квадратов старых отклонений $\sum_{i=1}^n w_i(Y_i - g_i(\beta_1, \ldots, \beta_r))^2$.

Разумно рассматривать задачу линейной регрессии для параметрических семейств. Поскольку речь в этом случае будет идти о статистическом оценивании параметров $\beta_1, \ldots, \beta_r$ случайного вектора $\{Y_i\}$, $1 \leqslant i \leqslant n$, желательно получение (по результатам одного наблюдения компонент вектора), если возможно, оценок с равномерно-минимальной дисперсией. Поскольку речь идет о линейной регрессии, естественно интересоваться несмещенными линейными оценками и среди таких оценок искать оценки с минимальной дисперсией. Теперь мы уже не уверены заранее, что эти линейные оценки будут даваться теми же формулами, что линейные оценки обычного (не статистического) метода наименьших квадратов, хотя для линейных задач мы сможем это доказать, т. е. получить хорошее обоснование метода наименьших квадратов в статистике. Заметим, что в некоторых случаях (нормальность распределений независимых случайных величин Y_i) наилучшие линейные оценки оказываются НРМД-оценками.

Рассмотрим простейшую задачу построения линейной несмещенной оценки с минимальной дисперсией. По выборке $Y_1, \ldots, Y_n$ (результаты измерений) получим линейную несмещенную оценку среднего $\widehat{\mu} = a_1 Y_1 + \ldots + a_n Y_n$. Несмещенность имеет место если и только если $\sum_{i=1}^n a_i = 1$. Для дисперсии оценки имеем $D(\widehat{\mu}) = \sum_{i=1}^n a_i^2 \sigma^2$, где σ – дисперсия каждой из случайных величин X_i. Минимум суммы квадратов достигается при равенстве слагаемых $a_i = n^{-1}, 1 \leqslant i \leqslant n$, т. е. выборочное среднее является линейной оценкой математического ожидания с наименьшей дисперсией. Налицо совпадение с решением нормального уравнения.

Рассмотрим более общую задачу линейной регрессии (схема Гаусса–Маркова), важную, в частности, для задач планирования эксперимента. Пусть $Y{=}(Y_1, \ldots, Y_n)^{\mathbf{T}}$–наблюдаемый случайный вектор (отклики), компоненты которого независимые (можно некоррелированные) случайные величины, дисперсии (но не обязательно распределения) которых одинаковы и равны σ^2, а средние значения задаются функцией регрессии $\mathbf{E}\,(Y_l) = \beta_1 X_{1,l} + \ldots + \beta_k X_{k,l}$, $l = 1, \ldots, n$, или, в матричной записи – $\mathbf{E}\,(Y) = \mathbf{X}^{\mathbf{T}}\beta$, где $\vec{Y}$ и β – соответствующие вектор-столбцы, $\mathbf{X} = ||X_{i,j}||$ – матрица известных коэффициентов (регрессоров). Будем обозначать X_i i-ю строку матрицы $\mathbf{X}$, $i = 1, \ldots, k$.

Величина σ^2 называется *остаточной дисперсией* (обычно она неизвестна и тоже подлежит оцениванию). Оказывается, что если ввести некоторые

ограничения, то и в этом общем случае существует и может быть легко найдена несмещенная линейная оценка параметров β_i с наименьшей дисперсией и тоже налицо совпадение с оценкой наименьших квадратов.

Линейными оценками $\beta = (\beta_1, \ldots, \beta_k)^{\mathbf{T}}$ по откликам $Y = (Y_1, \ldots, Y_n)^{\mathbf{T}}$ будем называть оценки вида $\widehat{\beta}_i = \sum_l c_{i,l} Y_l$, $i = 1, \ldots, k$ или $\widehat{\beta} = \mathbf{C}Y$. Допустим, что мы имеем какую-то линейную несмещенную оценку β_i. Тогда $\mathbf{E}\,(\widehat{\beta}_i) = \sum_l \sum_i \beta_i c_{i,l} X_{j,l}$ и условие несмещенности (приравнивание коэффициентов при β_i) $\sum_l c_{i,l} X_{j,l} = \delta_{i,j} = \mathbb{1}_{\{i=j\}}$, $i = 1, \ldots, k$, $j = 1, \ldots, k$ ($\delta_{i,j}$ называются символами Кронекера). Если получены несмещенные оценки для всех β_i, то можно записать $\mathbf{C}\mathbf{X}^{\mathbf{T}} = \mathbf{I}$, где $\mathbf{I}$ – единичная матрица. Это означает, что $C_i^{\mathbf{T}}$ ортогонален всем $X_j^{\mathbf{T}}$, $i \neq j$. В этом случае решением задачи минимизации дисперсии $\sigma^2(\widehat{\beta}_i) = \sum_l c_{i,l}^2\, \sigma^2$ для каждого i по $C_i^{\mathbf{T}}$ при каждом i является нормированный надлежащим образом перпендикуляр, опущенный из точки $X_i^{\mathbf{T}}$ на линейное пространство, порожденное $X_i^{\mathbf{T}}$, $i \neq j$, т. е. $C_i^{\mathbf{T}}$ принадлежит линейному пространству, порожденному $X_j^{\mathbf{T}}$, $j = 1, \ldots, n$. Следовательно, существует представление $c_{i,l} = \sum_j \lambda_{i,j} X_{j,l}$, $1 \leq i, l \leq n$, или $\mathbf{C} = \Lambda\mathbf{X}$. Тогда $\sum_{j'} \lambda_{i,j'} a_{j',j} = \delta_{i,j}$, где $a_{j',j} = \sum_l X_{j',l} X_{j,l}$, или $\Lambda\mathbf{A} = \mathbf{I}$, где $A = \mathbf{X}\mathbf{X}^{\mathbf{T}}$. Если $A = ||a_{i,j}||$ – неособенная матрица (что имеет место тогда и только тогда, когда ранг матрицы $\mathbf{X}$ равен k, что равносильно линейной независимости столбцов матрицы $\mathbf{X}^{\mathbf{T}}$), то $||\lambda_{i,j}|| = ||a_{i,j}||^{-1} = ||a^{i,j}||$ или $\Lambda = A^{-1}$. Следовательно, оценка для β_i с минимальной дисперсией будет $\widehat{\beta}_i = \sum_j a^{i,j} b_j$, где $b_j = \sum_l X_{j,l} Y_l$ или, в матричной форме $\widehat{\beta} = (\mathbf{X}\mathbf{X}^{\mathbf{T}})^{-1}\mathbf{X}Y$. Найдем дисперсию $\widehat{\beta}_i$, для чего воспользуемся уже выведенными формулами

$$\sigma^2(\widehat{\beta}_i) = \sum_{j,\,j'=1}^{k} a^{i,j} a^{i,j'} a_{j',j} \sigma^2 = \sigma^2 \sum_j a^{i,j} \delta_{i,j} = a^{i,i} \sigma^2,$$

и, аналогично, матрица ковариаций оценок $\widehat{\beta}_i$ суть $||a^{i,j}||\sigma^2$.

С другой стороны, рассмотрим (предполагая также, что матрица $\mathbf{X}\mathbf{X}^{\mathbf{T}} = ||a_{i,j}||$ – неособенная) вопрос об оценках наименьших квадратов, т. е. о $\widehat{\beta}_i$, минимизирующих сумму квадратов остатков $Q = \sum_l (Y_l - \beta_1 X_{1,l} - \ldots - \beta_k X_{k,l})^2$. Легко вычислить частные производные и написать нормальные уравнения для этого линейного случая. Они имеют вид $\mathbf{X}\mathbf{X}^{\mathbf{T}}\beta = \mathbf{X}Y$, и, если $\mathbf{X}\mathbf{X}^{\mathbf{T}}$ неособенна, то $\beta = (\mathbf{X}\mathbf{X}^{\mathbf{T}})^{-1}\mathbf{X}Y$, т. е. снова для компонент β_i вектора β получаются те же выражения $\sum_j a^{i,j} b_j$ – линейные оценки с минимальной дисперсией совпадают с оценками наименьших квадратов.

Резюмируя, получаем для нашего случая теорему Гаусса–Маркова (подход Маркова, основанный на минимизации дисперсии, датируется 1900-м г.).

Теорема 4.1. *Пусть Y_l, $l = 1, \ldots, n$ – независимые случайные величины со средними $\sum_1^k \beta_i X_{i,l}$ и одинаковыми дисперсиями σ^2; X_j, $j = 1, \ldots, k$ – линейно независимые векторы. В этом случае как линейные оценки β_i,*

$i = 1, \ldots, k$, с минимальной дисперсией, так и оценки наименьших квадратов даются формулами

$$\widehat{\beta}_i = \sum_j a^{i,j} \sum_l X_{j,l} Y_l \quad \textit{или} \quad \widehat{\beta} = (\mathbf{X}\mathbf{X}^{\mathbf{T}})^{-1}\mathbf{X}Y,$$

где $||a_{i,j}|| = \mathbf{X}\mathbf{X}^{\mathbf{T}}$ *и* $||a^{i,j}|| = ||a_{i,j}||^{-1}$. *Матрица ковариаций* $\widehat{\beta}_i$ *есть* $||a^{i,j}\sigma^2||$.

Оценим остаточную дисперсию. Пусть

$$Z_l = Y_l - \sum_{i=1}^{k} \beta_i X_{i,l}, \quad \widetilde{Y}_l = \sum_{i=1}^{k} \widehat{\beta}_i X_{i,l}, \quad S = \sum_{l=1}^{n} Z_l^2,$$

$$S_1 = \sum_{l=1}^{n} (Y_l - \widetilde{Y}_l)^2, \quad S_2 = \sum_{i,\,j=1}^{k} a_{i,j}(\widehat{\beta}_i - \beta_i)(\widehat{\beta}_j - \beta_j).$$

Но $\mathbf{E}(S) = n\sigma^2$; $\mathbf{E}(S_2) = \sum_{i,\,j=1}^{k} a^{i,j} a_{i,j} \sigma^2 = k\sigma^2$ (по определению матрицы ковариаций) $S = S_1 + S_2$, так что $\mathbf{E}(S_1) = (n-k)\sigma^2$. В S_1 не входят β_i и поэтому $S_1/(n-k)$ может служить в качестве несмещенной оценки σ^2. Рассмотрим пример.

Пример 4.1. Пусть имеются четыре предмета и двухчашечные весы, показывающие (со случайной погрешностью, имеющей $N(a, \sigma^2)$ распределение с неизвестной дисперсией) разность весов, находящихся на первой и на второй чашах. Дальше приводится матрица плана[5] взвешиваний (8 взвешиваний): $X_{i,l} \in \{-1, 1\}$ определяют здесь, какой предмет кладется при данном взвешивании на какую чашу; Y_l – показания весов, $1 \leq l \leq 8$:

β_1	1	1	1	1	1	1	1	1
β_2	1	−1	1	−1	−1	1	1	−1
β_3	1	1	−1	−1	1	−1	1	−1
β_4	1	−1	−1	1	−1	−1	1	1
$\vec{Y}$	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8

Матрица плана обладает тем свойством, что ее строки ортогональны, матрица $A = ||a_{i,j}||$ – диагональна, $a_{j,j} = 8$. Тогда оценки по методу наименьших квадратов $\hat{\beta}_j = X_j Y/8$ имеют дисперсию $\mathbf{D}(\beta_j) = \sigma^2/a_{j,j} = \sigma^2/8$. Отметим, что для получения такой точности (дисперсии) оценки при взвешивании предметов по одному каждый из них надо взвесить по восемь раз.

Что происходит в том случае, когда матрица $||a_{i,j}|| = \mathbf{X}\mathbf{X}^{\mathbf{T}}$ – особенная? Произвольную линейную комбинацию неизвестных β_i, $\sum c_i \beta_i$ с известными коэффициентами назовем параметрической функцией, $c = (c_1, \ldots, c_k)^{\mathbf{T}}$. Параметрическую функцию ψ назовем *оцениваемой* (или допускающей линейную несмещенную оценку), если существует вектор-столбец $a = (a_1, \ldots, a_n)^{\mathbf{T}}$ такой, что $\mathbf{E}(a^{\mathbf{T}}Y) = \psi$ для любых $\beta_1, \ldots, \beta_n$ (понятие введено Бозе в 1944 г.).

[5] В данной задаче матрица регрессоров $\mathbf{X}$ называется матрицей плана.

Теорема 4.2. *Величина $c^{\mathbf{T}}\beta$ оцениваема тогда и только тогда, когда c – линейная комбинация столбцов матрицы* $\mathbf{X}$.

Доказательство. Пусть $c^{\mathbf{T}}\beta$ – оцениваема и a – соответствующий вектор. Тогда $\mathbf{E}\,(a^{\mathbf{T}}Y) = a^{\mathbf{T}}\mathbf{E}\,\mathbf{Y} = a^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\beta$. Поскольку $a^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\beta = c^{\mathbf{T}}\beta$ тождественно по β_i, получаем, что $a^{\mathbf{T}}\mathbf{X}^{\mathbf{T}} = c^{\mathbf{T}}$. Обратное утверждение очевидно, поскольку $\mathbf{E}\,(Y_i) = X_i^{\mathbf{T}}\beta$. Теорема доказана. ■

Теперь рассмотрим общий случай теоремы Гаусса – Маркова. Мы предполагаем снова, что Y_i – независимые (можно, как и ранее, ограничиться некоррелированностью) случайные величины $\mathbf{D}(Y_i) = \sigma^2$.

Теорема 4.3. *Каждая оцениваемая функция имеет линейную несмещенную оценку с наименьшей дисперсией, и эта оценка является единственной в классе несмещенных линейных оценок. Оценка $\widehat{\psi}$ может быть получена из формулы $\psi = \sum_{j=1}^{k} c_j\beta_j$ заменой $\{\beta_j\}$ на любую (вообще говоря неединственную) оценку наименьших квадратов $\widehat{\beta}_1, \ldots, \widehat{\beta}_n$.*

4.2. Дисперсионный анализ

Одноименный раздел математической статистики включает в себя широкий круг задач на построение доверительных множеств и проверку статистических гипотез о параметрах регрессионных моделей. Обычно наблюдаемые случайные величины предполагаются нормальными, хотя в ряде случаев методы дисперсионного анализа пригодны и для иных случайных величин. Мы не имеем возможности изложить теорию дисперсионного анализа в полном объеме, поэтому ограничимся рассмотрением лишь простейших задач дисперсионного анализа.

При любых экспериментах могут выявиться важные факторы, влияющие на средние значения наблюдаемых. Эти факторы могут относиться ко многим источникам изменчивости. Для выделения и оценки важности факторов осуществляют наблюдения (эффекты при действии каждого из факторов), которые разбивают на группы в соответствии с тем, какие факторы тех или иных источников изменчивости при этом действовали. Далее сравнивают группы наблюдений с целью сопоставления рассеяния внутри и между группами. В некоторых более сложных случаях проверяется также зависимость действия факторов одного источника изменчивости от действия факторов другого. Все наблюдения всех групп предполагаются независимыми. Имеются различные типы дисперсионного анализа (в зависимости от числа источников изменчивости, условий проведения экспериментов и пр.). Рас-

смотрим дисперсионный анализ с простой (один источник изменчивости) и двойной (два источника изменчивости) группировками.

1. *Простая группировка.* Пусть имеется p факторов $F_1, \ldots, F_p$ одного источника изменчивости (например, удобрения, вносимые на разных участках одинаковой площади). Пусть $n = n_1 + \ldots + n_p$, где n – общее число наблюдений (полученного урожая), n_g – число наблюдений (участков) под воздействием g-го фактора (т. е. наблюдений случайной величины X_g – урожая при данном удобрении). Нулевая гипотеза состоит в том, что все X_g распределены нормально и одинаково, т. е. факторы несущественны. Пусть $X_{g,h}$ – h-й элемент g-й группы. Среднее значение в каждой группе будет $\overline{X_g} = \frac{1}{n_g}\sum_{h=1}^{n_g} X_{g,h}$, а среднее значение для всей совокупности наблюдений будет равно $\overline{X} = \frac{1}{n}\sum_{g=1}^{p}\sum_{h=1}^{n_g} X_{g,h}$. Для осуществления дисперсионного анализа (т. е. проверки нулевой гипотезы с помощью изучения различных компонент выборочной дисперсии) надо общую сумму квадратов отклонений наблюденных значений от среднего $Q = \sum_{g=1}^{p}\sum_{h=1}^{n_g}(X_{g,h} - \overline{X})^2$ разбить на две части: сумму квадратов отклонений наблюдений от средних внутри своих групп $Q_1 = \sum_{g=1}^{p}\sum_{h=1}^{n_g}(X_{g,h} - \overline{X_g})^2$ (сумму квадратов отклонений "внутри групп") и взвешенную сумму квадратов отклонений средних значений по группам от общего среднего значения (сумму квадратов отклонений "между группами"): $Q_2 = \sum_{g=1}^{p} n_g(\overline{X_g} - \overline{X})^2$. Поскольку $\overline{X} = \frac{1}{n}\sum_{g=1}^{p}\sum_{h=1}^{n_g} X_{g,h}$, имеем $Q = \sum_{g=1}^{p}\sum_{h=1}^{n_g} X_{g,h}^2 - n\overline{X}^2$. Аналогично,

$$Q_1 = \sum_{g=1}^{p}\sum_{h=1}^{n_g} X_{g,h}^2 - \sum_{g=1}^{p} n_g \overline{X_g}^2, \quad Q_2 = \sum_{g=1}^{p} n_g \overline{X_g}^2 - n\overline{X}^2.$$

Из этих равенств следует, что, действительно, $Q = Q_1 + Q_2$. Если верна нулевая гипотеза, то, как легко проверить, статистики $S^2 = Q/(n-1)$, $S_1^2 = Q_1/(n-p)$, $S_2^2 = Q_2/(p-1)$ являются несмещенными оценками дисперсии σ^2 (знаменатели – это числа степеней свободы соответствующих χ^2-распределений).

Можно также показать, что если верна нулевая гипотеза, то случайные величины Q_1 и Q_2 независимы и частное S_2^2/S_1^2 должно иметь распределение Фишера – Снедекора со степенями свободы $(p-1)$ и $(n-p)$. С помощью этого распределения и строится критическая область для проверки нулевой гипотезы.

2. *Двойная группировка.* Если имеются факторы двух источников (например, удобрения и тип используемой сельскохозяйственной техники) A_g, $g = 1, \ldots, p$ и B_h, $h = 1, \ldots, q$, то имеем двойную группировку. Будем предполагать, что каждому сочетанию (g, h) соответствует ровно одно

наблюдение $X_{g,h}$. В этом случае

$$Q = \sum_{g=1}^{p}\sum_{h=1}^{q}(X_{g,h} - \overline{X})^2 = Q_A + Q_B + Q_R; \quad Q_A = q\sum_{g=1}^{p}(\overline{X_g} - \overline{X})^2;$$

$$Q_B = p\sum_{g=1}^{q}(\overline{X_h} - \overline{X})^2; \quad Q_{\text{К}} = \sum_{g=1}^{p}\sum_{h=1}^{q}(X_{g,h} - \overline{X_g} - \overline{X_h} + \overline{X})^2.$$

Можно проверить это равенство, а также то, что квадратичные формы Q, Q_A, Q_B, Q_R имеют ранги, соответственно, $pq-1$, $p-1$, $q-1$, $(p-1)(q-1)$, а также то, что эти случайные величины имеют распределения χ^2 с соответствующими числами степеней свободы, причем Q_A, Q_B, Q_R – независимы. С помощью этих форм могут быть получены оценки дисперсии наблюдения: $s^2 = \dfrac{Q}{pq-1}$ – общая; $s_A^2 = \dfrac{Q_A}{p-1}$ – между группами A; $s_B^2 = \dfrac{Q_B}{q-1}$ – между группами B; $s_R^2 = \dfrac{Q_R}{(p-1)(q-1)}$ – остаточная. Проверка нулевой гипотезы производится сравнением s_A^2 с s_R^2 и s_B^2 с s_R^2. Если для какого-либо из отношений s_A^2/s_R^2 и s_B^2/s_R^2 будет превышено табличное значение (распределение Фишера – Снедекора), то гипотеза однородности должна быть отвергнута. Далее для сравнения показателей групп A и B применяется критерий Стьюдента.

5. Методы обработки цензурированных данных типа времени жизни

Отправным пунктом исследования эксперимента является построение модели, с помощью которой можно делать предположения о тех или иных свойствах результата данного эксперимента. Модель может быть построена неоднозначно. При этом неверный выбор может привести к неверным результатам. При выборе излишне сложной модели задача может стать неразрешимой, а сильное ее упрощение – привести к ошибкам, которые сведут на нет всю работу исследователя. Поэтому выбор оптимальной статистической модели является первоочередной задачей исследователя. Рассмотрим некоторые задачи с цензурированными данными.

Пример 5.1. Данные получены в одной из датских клиник. В период с 1962 по 1977 гг. 225 человек перенесли операцию по удалению опухоли. Из них 20 человек отказались от наблюдения, т. е. были получены данные по 205 пациентам. Цель исследования: оценить распределение времени жизни после операции. При обработке данных должны быть учтены следующие положения:

а) пациенты поступали на операцию в различные моменты времени (время поступления для каждого больного принимается нулевым);

б) время смерти установлено лишь для погибших до 1977 г.;

в) часть пациентов погибли по причинам, не зависящим от операции (цензурированы в момент гибели);

г) возможно, что имеет смысл учесть специфические особенности больного (возраст, пол, стадия заболевания в момент обнаружения).

Пример 5.2. Предположим, что в ограниченное время необходимо оценить распределение времени работы некоторой (достаточно надежной) системы. Если тестировать систему в обычном режиме, то процент отказов за отведенное время в независимых испытаниях может быть крайне мал. При этом мы не получим никакой информации о распределении отказа за пределами времени, отведенного на испытание. Для повышения эффективности испытаний предлагается проводить их в более жестких режимах. Ясно, что данный подход ничего не дает, если нет дополнительных предположений о связи распределений отказов в различных режимах работы. Обычно жесткую связь установить сложно, поэтому предполагается наличие параметрической связи, диктуемой свойствами системы. В этом случае оценка распределения времени отказа системы строится с учетом данной параметрической зависимости.

5.1. Непараметрическая модель. Оценки Нельсона–Аалена и Каплана–Мейера

Пусть T — неотрицательная случайная величина (время безотказной работы рассматримоваемого объекта). Обозначим

$$F(t) = \mathbf{P}(T \le t) \qquad \text{и} \qquad S(t) = 1 - F(t), \quad t \in \mathbb{R},$$

функции распределения и отказа случайной величины T соответственно. При анализе выживания принято использовать функции отказа. Предположим, что T абсолютно непрерывна с плотностью f.

Интенсивностью отказа будем называть следующую функцию:

$$\lambda(t) = \lim_{\Delta t \to 0} \mathbf{P}(t \le T < t + \Delta t | T \ge t)/\Delta t = -\frac{dS(t)}{dt} \Big/ S(t) = \frac{f(t)}{S(t)}, \quad t \in \mathbb{R}.$$

Накопленной интенсивностью будем называть функцию

$$\Lambda(t) = \int_0^t \lambda(u)\, du, \quad t \in \mathbb{R}.$$

Замечание 5.1. В рассматриваемом случае абсолютно непрерывной S справедливо равенство

$$S(t) = \exp(-\Lambda(t)), \quad t \in \mathbb{R}.$$

В общем случае

$$\Lambda(t) = -\int_0^t \frac{S(du)}{S(u-)},$$

где интеграл понимается в смысле Лебега – Стильтьеса.

Пусть времена отказов $T_1, \ldots, T_n$ – выборка из распределения с функцией отказа S; $U_1, \ldots, U_n$ – соответствующие времена цензурирования. При этом в каждой паре (T_i, U_i) наблюдается лишь наименьшее значение. Таким образом, полученные данные можно записать в виде набора $(X_1, \delta_1), \ldots, (X_n, \delta_n)$, где $X_i = \min(T_i, U_i)$, $\delta_i = \mathbb{1}_{\{T_i \leq U_i\}}$. Положим L – число различных наблюдаемых отказов ($L \leq \sum_{i=1}^n \delta_i \leq n$); $T_1^* < \ldots < T_L^*$ – последовательные времена отказов, D_k – число отказов в момент времени T_k^*, $k = 1, \ldots, L$ ($\sum_{i=1}^L D_k = \sum_{i=1}^n \delta_i$). Далее, рассмотрим $0 = t_0 < t_1 < \ldots < t_m = t$ – разбиение интервала $[0, t]$; d_l — число наблюдавшихся отказов в интервале $[t_{l-1}, t_l)$; y_l — число элементов, находящихся под наблюдением ("рискующих"), т. е. не отказавших и не цензурированных к моменту времени t_{l-1}. Поскольку

$$\Lambda(t + \Delta t) - \Lambda(t) \sim \lambda(t)\Delta t$$

и

$$\Lambda(t + \Delta t) - \Lambda(t) \sim \mathbf{P}(t \leq T < t + \Delta t | T \geq t),$$

то естественной оценкой для $\Lambda(t_l) - \Lambda(t_{l-1})$ видится отношение d_l/y_l, т. е. для Λ:

$$\tilde{\Lambda} = \sum_{l: t_l \leq t} (d_l/y_l).$$

Переходя к пределу при $m \to \infty$ с условием $\max_{1 \leq l \leq m} |t_l - t_{l-1}| \to 0$ получаем, что $d_l \to D_k$, если соответствующий интервал содержит T_k^* (если соответствующий интервал не содержит T_j^*, $j \neq k$, то будет равенство $d_l = D_k$), и $d_l = 0$, если соответствующий интервал не содержит T_j^*, $j = 1, \ldots, n$. Таким образом, $\tilde{\Lambda} \to \hat{\Lambda}$, где $\hat{\Lambda}$ задается соотношением

$$\hat{\Lambda} = \sum_{l: T_l^* \leq t} (D_l/Y_l^*),$$

где Y_l^* – число элементов, не отказавших и не цензурированных до момента времени T_l^*. Выражение $\hat{\Lambda}$ называется оценкой *Нельсона – Аалена* накопленной интенсивности Λ.

Рассмотрим оценку функции отказа S:

$$\tilde{S}(t) = \exp(-\hat{\Lambda}) = \prod_{l: T_l^* \leq t} \exp(-D_l/Y_l^*).$$

При $D_l/Y_l^* \to 0$ оценка $\tilde{S}(t)$ эквивалентна оценке

$$\hat{S}(t) = \prod_{l: T_l^* \leq t} (1 - D_l/Y_l^*),$$

которая называется оценкой *Каплана – Мейера* функции отказа S.

Нетрудно проверить, что если все отказы наблюдаются, то оценка Каплана – Мейера совпадает с эмпирической функцией отказа.

5.2. Сравнение двух распределений

Предположим, что $(T_{i,j}, U_{i,j})$, $j = 1, \ldots, n_i$, $i = 1, 2$ – независимые пары моментов отказа и цензурирования, соответственно, двух выборок с интенсивностями λ_1 и λ_2. Отметим, что в каждой паре наблюдается лишь одна из компонент.

Снова положим $T_1^* < \ldots < T_L^*$ – последовательные времена отказов, $L \le n_1 + n_2$, $D_{i,k}$ — число элементов i-й выборки, отказавших в момент времени T_k^*; $Y_{i,k}^*$ — число элементов i-й выборки, не отказавших и не цензурированных до момента T_k^*, $k = 1, \ldots, L$, $i = 1, 2$. Также введем величины $D_k = D_{1,k} + D_{2,k}$ и $Y_k = Y_{1,k} + Y_{2,k}$. Отметим, что при справедливости гипотезы $H_0 : \Lambda_1 = \Lambda_2$ условное распределение $D_{1,k}$ при условии $Y_{i,k}^*$, $i = 1, 2$, и D_k имеет дискретное распределение с вероятностями

$$\mathbf{P}(D_{1,k} = l | D_k,\ Y_{i,k}^*,\ i = 1, 2) = C_{Y_{1,k}^*}^{l} C_{Y_{2,k}^*}^{D_k - l} \Big/ C_{Y_k^*}^{D_k}$$

и имеет условное среднее $E_{1,k} = D_k Y_{1,k}^* / Y_k^*$ и условную дисперсию

$$V_{1,k} = D_k \frac{Y_{1,k}^* Y_{2,k}^*}{(Y_k^*)^2} \frac{Y_k^* - D_k}{Y_k^* - 1}.$$

Принимая во внимание, что $D_{1,k} = Y_{1,k-1} - D_{1,k-1} - Y_{1,k}$, заключаем, что

$$\mathbf{P}\left(\frac{D_{1,k} - E_{1,k}}{V_{1,k}} \le s_k,\ k = 1, \ldots, L\right) = \prod_{k=1}^{L} \mathbf{P}\left(\frac{D_{1,k} - E_{1,k}}{V_{1,k}} \le s_k \,\Big|\, D_k,\ Y_{i,k}^*,\ i = 1, 2\right).$$

Тогда величина

$$Q = \sum_{k=1}^{L} (D_{1,k} - E_{1,k}) / V_{1,k}$$

имеет асимптотическое распределение $N(0, 1)$. Числитель этого выражения называется *логранг*-статистика. Она может быть использована для проверки согласия с гипотезой о равенстве распределений.

5.3. Сѐмипараметрические регрессионные модели

Рассмотрим независимые пары (T_i, U_i), $i = 1, \ldots, m$, моментов отказов и цензурирования соответственно. Предполагается, что эксперименты могут проводиться при различных условиях, задаваемых вектором ковариант $\mathbf{z} = (z_1, z_2, \ldots, z_n)$, выбираемых из множества $\mathfrak{B}$ (например, в примере 5.1 можно выбрать z_1 - возраст, z_2 - пол). Кроме того, предполагается задание параметрической связи

$$S(t|\mathbf{z}_1) = g(\beta, t, z_1(\cdot), z_2(\cdot), S(t|\mathbf{z}_2)),$$

при любых $\mathbf{z}_1, \mathbf{z}_2 \in \mathfrak{B}$, где g — некоторый функционал, удовлетворяющий естественным условиям согласования. Обычно предполагается, что зависимость передается через произведение $\beta^{\mathrm{T}}\mathbf{z}$, то есть

$$g(\beta, t, z_1(\cdot),\ z_2(\cdot),\ S(t|\mathbf{z}_2)) = g(t, \beta^{\mathrm{T}} z_1(\cdot),\ \beta^{\mathrm{T}} z_2(\cdot),\ S(t|\mathbf{z}_2)),$$

где $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$.

Регрессионную связь удобно задавать в терминах интенсивностей. Рассмотрим полупараметрическую регрессионную модель, введенную Коксом (1972), с пропорциональными интенсивностями отказа

$$\lambda(t|\mathbf{z}) = \lambda_0(t)\exp(\beta^{\mathrm{T}}\mathbf{z}),$$

где $\mathbf{z}(\cdot) \equiv \mathbf{z}$, λ_0 – некоторая неизвестная базовая интенсивность, соответствующая нулевому значению коварианты $\mathbf{z}_0 \equiv (0,\ldots,0)$. Отметим, что в терминах функций отказа данная модель задается соотношением

$$S(t) = (S_0(t))^{\exp(\beta^{\mathrm{T}}\mathbf{z})}.$$

Если S_0 известна или параметризована, то можно говорить о параметрических моделях.

Оценка параметров производится на основании принципов максимума правдоподобия. Введем эвристически понятие частичного правдоподобия. Пусть $(A_1,B_1),\ldots,(A_L,B_L)$ – набор пар событий. Обычно предполагается, что события A_i связаны с объектами одной природы, а B_i – с объектами другой природы, и есть некий хронологический порядок поступления данных $B_1,A_1,\ldots,B_L,A_L$. Правдоподобие указанного набора данных может быть представлено в виде

$$\mathbf{P}_\theta(A_L,B_L,\ldots,A_1,B_1) = \prod_{k=2}^{L}\mathbf{P}_\theta(A_kB_k|A_{k-1}B_{k-1}\ldots A_1B_1)\mathbf{P}_\theta(A_1B_1) =$$

$$= \Biggl(\prod_{k=2}^{L}\mathbf{P}_\theta(A_k|B_kA_{k-1}B_{k-1}\ldots A_1B_1)\mathbf{P}_\theta(A_1|B_1)\Biggr)\times$$

$$\times\Biggl(\prod_{k=2}^{L}\mathbf{P}_\theta(B_k|A_{k-1}B_{k-1}\ldots A_1B_1)\mathbf{P}_\theta(B_1)\Biggr).$$

Отметим, что первый сомножитель зависит от событий второй группы лишь через условие, а второй – от событий первой группы лишь через условие. Кроме того, если предположить, что указанный хронологический порядок имеет место, то произведение по k можно понимать как накопление информации.

Аналогично, если случайный вектор $Q = (V,W)$, где $V = (V_1,\ldots,V_L)$, $W = (W_1,\ldots,W_L)$ имеет плотность распределения $f_Q(\vec{x};\,\theta)$, то она может быть представлена в виде

$$f_Q(\vec{x};\,\theta) = f_{V_1,W_1,\ldots,V_L,W_L}(v_1,w_1,\ldots,v_L,w_L;\,\theta) =$$

$$= \prod_{k=1}^{L} f_{W_k|H_k}(w_k|v_1,w_1,\ldots,v_k;\,\theta)f_{V_k|P_k}(v_k|v_1,w_1,\ldots,v_{k-1},w_{k-1};\,\theta) =$$

$$= \Bigl(\prod_{k=1}^{L} f_{W_k|H_k}(w_k|h_k;\,\theta)\Bigr)\Bigl(\prod_{k=1}^{L} f_{W_k|P_k}(v_k|p_k;\,\theta)\Bigr),$$

где $P_k = (V_1, W_1, \ldots, V_{k-1}, W_{k-1})$, $H_k = (V_1, W_1, \ldots, W_{k-1}, V_k)$. Первый сомножитель функции правдоподобия будем называть *функцией частичного правдоподобия*. После подстановки в функцию частичного правдоподобия результатов эксперимента получаем частичное правдоподобие исследуемого набора данных.

Если $\theta = (\beta, \phi)$, где первый сомножитель правдоподобия в круглых скобках зависит только от β, то можно говорить об оценке β используя только этот сомножитель, называемый *частичным правдоподобием* исходного набора данных. При этом, если второй сомножитель не зависит от β, то информация о параметре β, содержащаяся в полном и в частичном правдоподобиях, совпадает.

5.4. Частичное правдоподобие по Коксу

Вернемся к рассматриваемой модели. Снова введем последовательные наблюдаемые времена отказов $T_1^* < \ldots < T_L^*$, $L \le m$; D_k – число отказов, произошедших в момент времени T_k^*, $(i)^j$ – номера компонент, отказавших в момент времени T_i^* и $\mathbf{z}_{(i)^j}$ – соответствующие коварианты, $j = 1, \ldots, D_i$; m_i – число компонент, цензурированных в интервале $[T_i^*, T_{i+1}^*)$ в моменты времени $T_{i,1}^*, \ldots, T_{i,m_i}^*$ и (i,j) — номер компоненты, цензурированной в момент времени $T_{i,j}^*$, $j = 1, \ldots, m_i$. Положим

$W_k = (k)^j$, $j = 1, \ldots, D_j$, $V_k = (T_k^*,\ D_k,\ T_{k-1,j},\ j = 1, \ldots, m_k)$, $k = 1, \ldots, L$.

Сначала рассмотрим случай $D_1 = \ldots = D_L = 1$. Тогда введенной последовательности отказов можно сопоставить вектор ковариант $(\mathbf{z}_{(1)}, \ldots, \mathbf{z}_{(L)})$, где $((1), \ldots, (L)) = ((1)^1, \ldots, (L)^1)$ – вектор антирангов. Пусть R_k – множество компонент, не отказавших и не цензурированных до момента T_k^*, $k = 1, \ldots, L$. Тогда, в силу независимости и однородности испытаний

$$\mathbf{P}((i) = j | T_k^*, D_k, (i)^s, T_{k-1,j}, (k-1,\ j),\ j = 1, \ldots, m_k, k \le i,\ s \le i-1) =$$

$$= \lim_{\Delta t \to 0} \frac{\mathbf{P}((k) = j, T \in [T_{(k)}^*, T_{(k)}^* + \Delta t) | T \ge T_k^*, R_k, \mathbf{z}_j)}{\sum\limits_{l \in R_k} \mathbf{P}((k) = l, T \in [T_{(k)}^*, T_{(k)}^* + \Delta t) | T \ge T_k^*, R_k, \mathbf{z}_l)} =$$

$$= \frac{\mathbf{P}(T \in [T_{(k)}^*, T_{(k)}^* + \Delta t) | T \ge T_k^*, \mathbf{z}_j)}{\sum\limits_{l \in R_k} \mathbf{P}(T \in [T_{(k)}^*, T_{(k)}^* + \Delta t) | T \ge T_k^*, \mathbf{z}_l)} = \frac{\lambda(T_k^* | \mathbf{z}_j)}{\sum\limits_{l \in R_k} \lambda(T_k^* | \mathbf{z}_l)}.$$

Итак, частичное правдоподобие имеет вид

$$L^c(\beta) = \prod_{k=1}^{L} \frac{\lambda(T_k^* | \mathbf{z}_{(k)})}{\sum\limits_{l \in R_k} \lambda(T_k^* | \mathbf{z}_l)}.$$

В частности, для модели Кокса частичное правдоподобие выглядит следующим образом:

$$L^c(\beta) = \prod_{k=1}^{L} \frac{\exp(\beta^{\mathrm{T}}\mathbf{z}_{(k)})}{\sum\limits_{l\in R_k}\exp(\beta^{\mathrm{T}}\mathbf{z}_l)}.$$

Прологарифмировав, получаем

$$\ln L^c(\beta) = \sum_{k=1}^{L}\Bigg(\beta^{\mathrm{T}}\mathbf{z}_{(k)} - \ln\Big(\sum_{l\in R_k}\exp(\beta^{\mathrm{T}}\mathbf{z}_l)\Big)\Bigg).$$

Для нахождения максимума данной функции продифференцируем по β:

$$U(\beta) = (\ln L^c(\beta))'_\beta = \sum_{k=1}^{L}\left(\mathbf{z}_{(k)} - \frac{\sum\limits_{l\in R_k}\mathbf{z}_l\exp(\beta^{\mathrm{T}}\mathbf{z}_l)}{\sum\limits_{l\in R_k}\exp(\beta^{\mathrm{T}}\mathbf{z}_l)}\right).$$

Вектор параметров находится из системы уравнений $U(\beta) = 0$.

В общем случае

$$L^c(\beta) = \prod_{i=1}^{L}\mathbf{P}(W_i = \{(i)^1,\dots,(i)^{D_i}\}|Q_i),$$

где

$$\mathbf{P}(W_i = \{(i)^1,\dots,(i)^{D_i}\}|Q_i) =$$

$$= \mathbf{P}\left(\begin{matrix} T_{(i)^j}\in[T_i^*, T_i^*+dt), j=1,\dots,D_i \\ T_k\notin[T_i^*, T_i^*+dt), k\neq(i)^j \end{matrix}\;\middle|\;\begin{matrix} R_i, \text{ровно } D_i \text{ компонент отказали} \\ \text{в интервале времени} [T_i^*, T_i^*+dt) \end{matrix}\right) =$$

$$= \frac{\mathbf{P}(T_{(i)^j}\in[T_i^*, T_i^*+dt), j=1,\dots,D_i, T_k\notin[T_i^*, T_i^*+dt), k\neq(i)^j|R_i)}{\sum\limits_{\substack{\sigma\subset R_i\\|\sigma|=D_i}}\mathbf{P}(T_l\in[T_i^*, T_i^*+dt), l\in\sigma;\ T_k\notin[T_i^*, T_i^*+dt), k\notin\sigma|R_i)},$$

поскольку эксперименты независимы при условии R_i:

$$\frac{\prod\limits_{j=1}^{D_i}\mathbf{P}(T_{(i)^j}\in[T_i^*, T_i^*+dt)|R_i)\prod\limits_{\substack{l\neq(i)^j\\l\in R_s}}\mathbf{P}(T_l\notin[T_i^*, T_i^*+dt)|R_i)}{\sum\limits_{\substack{\sigma\subset R_i\\|\sigma|=D_i}}\prod\limits_{l\in\sigma}\mathbf{P}(T_l\in[T_i^*, T_i^*+dt)|R_i)\prod\limits_{l\in R_i\setminus\sigma}\mathbf{P}(T_l\notin[T_i^*, T_i^*+dt)|R_i)} =$$

$$= \frac{\prod\limits_{j=1}^{D_i}d\Lambda(T_i^*|\mathbf{z}_{(i)^j})\prod\limits_{\substack{l\neq(i)^j\\l\in R_s}}(1-d\Lambda(T_i^*|\mathbf{z}_l))}{\sum\limits_{\substack{\sigma\subset R_i\\|\sigma|=D_i}}\prod\limits_{l\in\sigma}d\Lambda(T_i^*|\mathbf{z}_l)\prod\limits_{l\in R_i\setminus\sigma}(1-d\Lambda(T_i^*|\mathbf{z}_l))}.$$

Рассмотрим пример, известный под названием *модели логистической регрессии.*

Пример 5.3. Пусть T^* – случайная величина, определяемая равенством $T^* = t_i$, если $T \in [t_i, t_{i+1})$, где T – абсолютно непрерывная неотрицательная случайная величина с интенсивностью $\lambda(t|\mathbf{z})$ и соответствующей функцией отказа $S(t|\mathbf{z})$, $0 = t_0 < t_1 < \ldots$ — не более, чем счетное разбиение $[0, \infty)$. Введем величины

$$\lambda^*(t_i|\mathbf{z}) = \mathbf{P}(T^* = t_i | T^* \geq t).$$

Тогда

$$\lambda^*(t_i|\mathbf{z}) = 1 - S(t_{i+1}|\mathbf{z})/S(t_i|\mathbf{z}) = 1 - \exp\Big(-\int\limits_{t_i}^{t_{i+1}} \lambda(u|\mathbf{z})du\Big).$$

Модель с регрессионной связью, задаваемой соотношением

$$\frac{\lambda^*(t_i|\mathbf{z})}{1 - \lambda^*(t_i|\mathbf{z})} = \frac{\lambda_0^*(t_i)}{1 - \lambda_0^*(t_i)} g(\mathbf{z}),$$

называется моделью логистической регрессии. В частности, для бинарных данных ($T^* = \mathbb{1}_{\{T \geq t_1\}}$) отношения $\mathbf{P}(T^* = 0)/\mathbf{P}(T^* = 1)$ при различных значениях ковариант пропорциональны. Данная модель может быть переписана в виде

$$\frac{1 - \exp(-\int\limits_{t_i}^{t_{i+1}} \lambda(u|\mathbf{z})\, du)}{\exp(-\int\limits_{t_i}^{t_{i+1}} \lambda(u|\mathbf{z})\, du)} = \frac{1 - \exp(-\int\limits_{t_i}^{t_{i+1}} \lambda_0(u)\, du)}{\exp(-\int\limits_{t_i}^{t_{i+1}} \lambda_0(u)\, du)} g(\mathbf{z})$$

или

$$\frac{1 - \exp(-\int\limits_{t_i}^{t_{i+1}} \lambda(u|\mathbf{z})\, du)}{1 - \exp(-\int\limits_{t_i}^{t_{i+1}} \lambda_0(u)\, du)} = \frac{\exp(-\int\limits_{t_i}^{t_{i+1}} \lambda(u|\mathbf{z})\, du)}{\exp(-\int\limits_{t_i}^{t_{i+1}} \lambda_0(u)\, du)} g(\mathbf{z}).$$

Устремим $t_{i+1} - t_i \to 0$. Дробь в левой части стремится к единице, а предел правой части равен $\lambda(t_i|\mathbf{z})/\lambda_0(t_i)$. Таким образом, выбрав $g(\mathbf{z}) = \exp(\beta^{\mathrm{T}}\mathbf{z})$, получаем модель Кокса

$$\lambda(t|\mathbf{z}) = \exp(\beta^{\mathrm{T}}\mathbf{z})\lambda_0(t).$$

Вычислим частичное правдоподобие по Коксу для модели логистической регрессии:

$$\frac{d\Lambda(t|\mathbf{z})}{1 - d\Lambda(t|\mathbf{z})} = \frac{d\Lambda_0(t)}{1 - d\Lambda_0(t)} \exp\beta^{\mathrm{T}}\mathbf{z}.$$

В этом случае частичное правдоподобие по Коксу имеет вид

$$L^c(\beta) = \frac{\exp\Big(\beta^{\mathrm{T}} \sum\limits_{j=1}^{D_i} Z_{(i)^j}\Big)}{\sum\limits_{\substack{\sigma \subset R_i \\ |\sigma|=D_i}} \exp\Big(\beta^{\mathrm{T}} \sum\limits_{j\in\sigma} Z_l\Big)} = \frac{\exp(\beta^{\mathrm{T}} S_{(i)})}{\sum\limits_{\substack{\sigma \subset R_i \\ |\sigma|=D_i}} \exp(\beta^{\mathrm{T}} S_\sigma)},$$

где $S_{(i)} = \sum\limits_{j=1}^{D_i} Z_{(i)^j}$, $S_\sigma = \sum\limits_{j\in\sigma} Z_l$.

Список литературы

Боровков А. А. Математическая статистика. М.: Наука, 1984.

Ивченко Г. И., Медведев Ю. И. Математическая статистика. М.: Высш. шк., 1984.

Рао С. Р. Линейные статистические методы и их применение. М.: Наука, 1968.

Ван дер Варден. Математическая статистика. М.: Изд-во иностр. лит., 1960.

Гаек Я., Шидак З. Теория ранговых критериев. М.: Наука, 1971.

Крамер Г. Математические методы статистики. 2-е изд. М.: Мир, 1975.

Леман Э. Теория точечного оценивания. М.: Наука, 1991.

Леман Э. Проверка статистических гипотез. М.: Наука, 1964.

Уилкс С. Математическая статистика. М.: Наука, 1967.

Шеффе Г. Дисперсионный анализ. М.: Наука, 1980.

Andersen P. K., Borgan O., Gill R. D., Keiding N. Statistical models based on counting processes. N.-Y.: Springer-Verlag, 1993.

Fleming T. R., Harrington D. P. Counting processes & survival analysis. N.-Y.: Wiley, 1991.

Оглавление