Министерство общего и профессионального
образования Российской Федерации

Тверской государственный университет

Ю.С.ХОХЛОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Часть III

Тверь 2000

УДК 519.2

Хохлов Ю.С. Теория вероятностей и математическая статистика: Ч. III. Учебное пособие/ ТвГУ. — Тверь, 2000. — 61 с.

Пособие является третьей частью курса лекций по теории вероятностей и математической статистике, в которой излагаются следующие разделы математической статистики: статистическая структура, эмпирическое распределение, оценки параметров и их свойства, доверительные интервалы, проверка статистических гипотез.

Рекомендуется студентам математических специальностей, а также экономистам.

Библиогр. 10.

Рецензенты: кафедра математической статистики Московского госуниверситета им. М.В. Ломоносова;

доктор физико-математических наук
В.В. Сенатов

Содержание

Глава 1

Основная задача математической статистики. Статистическая структура

В курсе теории вероятностей мы всегда предполагали, что мы имеем некоторое фиксированное пространство $(\Omega, \mathcal{A}, P)$, т.е. исходная модель изучаемого явления задана однозначно. Задача состояла в том, чтобы изучить свойства этой модели. Основной проблемой являлось разработка методов вычисления вероятностей более сложных событий , исходя из вероятностей других, более простых событий, в рамках данной модели. В частности, при изучении случайных величин мы интересовались их распределением. Но мы оставляли в стороне вопрос о том, как была построена данная конкретная модель, как она была выбрана из множества других. Задачи такого типа относятся к другому разделу нашей науки о случайных явлениях – математической статистике. Эта наука имеет тот же предмет, что и теория вероятностей. Она изучает случайные эксперименты с неопределенным исходом, в которых выполнено свойство устойчивости частот, но интересуется несколько иными вопросами. В определенном смысле она дополняет теорию вероятностей, а та, в свою очередь, дает теоретическую основу для методов математической статистики.

Чтобы лучше понять соотношение этих двух дисциплин рассмотрим простой пример. Предположим, что мы проводим эксперимент, в котором независимо n раз подбрасываем некоторую монету. Если выпадет герб, то будем ставить 1, если цифра–то 0. Тогда исход такого эксперимента ω можно записать в виде $\omega = (\omega_1, \ldots, \omega_n)$, где ω_k–описывает результат k–ого подбрасования. Множество Ω всех таких исходов определяет нам пространство элементарных событий, а класс $\mathcal{A}$ всех подмножеств Ω это класс событий, связанных с таким экспериментом. Если мы имеем основания предполагать, что монета симметричная, то, в сочетании с предположением о независимости испытаний, мы получаем, что

$p(\omega) = 1/2^n$. Далее мы можем попытаться вычислить вероятности более сложных событий. Например, если A_m есть событие, состоящее в том, что в таком эксперименте появилось ровно m гербов, то, как мы уже знаем, вероятность этого события равна

$$(A_m) = C_n^m \cdot \frac{1}{2^n} \ .$$

Это типичная вероятностная задача: модель точно задана, необходимо что-либо вычислить в рамках этой фиксированной модели.

Посмотрим на этот экперимент с другой стороны. Обычно мы не знаем точно свойств нашей монеты и не можем предполагать, что она симметричная. Тогда мы имеем те же Ω и $\mathcal{A}$, но для вероятностей элементарных исходов получаем только

$$P(\omega) = p^m(1-p)^{n-m} \ , \tag{1.1}$$

где m–число единиц в исходе ω, а вероятность $0 \leq p \leq 1$ появления единицы нам неизвестна. Таким образом, мы имеем целый набор $\mathcal{P}$ вероятностных мер на $(\Omega, \mathcal{A})$, которые могут реализоваться в нашем эксперименте. В этой ситуации мы можем поставить следующие вопросы. Пусть, для определенности, мы производим 100 подбрасываний нашей монеты и получаем 47 гербов.

1) Как на основе этой информации оценить неизвестную вероятность p?

2) Какова точность полученной оценки?

3) Сколько нужно произвести испытаний, чтобы добиться нужной точности оценки?

4) Можно ли считать, что $p = 1/2$?

Что общего во всех перечисленных вопросах? У нас есть некоторый априорно заданный класс возможных вероятностных моделей для нашего эксперимента, т.е. мы имеем не одну, а целый набор $\mathcal{A}$ вероятностных мер на $(\Omega, \mathcal{A})$. Мы хотели бы **на основе экспериментальных данных** уточнить наши знания об истинной вероятностной мере $P \in \mathcal{P}$, которая дает описание вероятностного механизма рассматриваемого эксперимента.

Рассмотрение этого примера приводит нас к определению **статистической структуры** и основной задаче математической статистики.

Определение 1 . **Статистической структурой** *называется тройка* $(\Omega, \mathcal{A}, \mathcal{P})$, *где*

1) Ω*–произвольное множество = пространство элементарных исходов,*

2) $\mathcal{A}$*–*σ*-алгебра подмножеств* Ω *= события, доступные наблюдению,*

3) $\mathcal{P}$*–некоторый набор вероятностных мер на* $(\Omega, \mathcal{A})$.

Примеры. 1) Рассмотренный выше пример приводит нас к модели $(\Omega, \mathcal{A}, \mathcal{P})$, где ω–набор длины n, состоящий из нулей и единиц, $\mathcal{A}$-σ-алгебра всех подмножеств пространства Ω, $\mathcal{P}$–набор вероятностных мер, которые задаются на элементарных исходах по формуле (1), $0 \leq p \leq 1$.

2) Пусть мы имеем одно измерение случайной величины ξ. В этом случае естественно взять $\Omega = R^1$, $\mathcal{A} = \mathcal{B}$–$\sigma$-алгебра всех борелевских подмножеств R^1, а $\mathcal{P}$–класс всевозможных распределений вероятностей для ξ на R^1 (заданных, например, с помощью функций распределения).

3) Иногда мы имеем какую-либо дополнительную информацию о распределении X. Например, если мы знаем, что X есть сумма большого числа маленьких слагаемых, то естественно предположить, что распределение X будет нормальным. В этом случае $\mathcal{P}$ есть класс всех нормальных распределений, вид которых известен, но есть два неизвестных параметра $a \in R^1$, $\sigma^2 > 0$.

4) Если мы производим n независимых измерений некоторой количественной характеристики в одинаковых условиях, то мы приходим к случайному вектору $X = (X_1, \ldots, X_n)$, где с.в. $X_1, \ldots, X_n$– н.о.р. В этом случае распределение X задается, например, совместной функцией распределения. Тогда $\Omega = R^n$, $\mathcal{A} = \mathcal{B}_n$–класс всех борелевских подмножеств в R^n, $\mathcal{P}$–класс всевозможных n-мерных функций распределения, соответствующих случаю н.о.р.с.в.

Если существует конечное число числовых параметров $(\theta_1, \ldots, \theta_m) = \theta$, $\theta \in \Theta \subset R^m$, с помощью которых удается занумеровать все распределения P из класса $\mathcal{P}$, то статистическая структура $(\Omega, \mathcal{A}, \mathcal{P})$ (класс распределений $\mathcal{P}$) называется **параметрической**. В противном случае мы имеем **непараметрическую** структуру. В рассмотренных выше примерах в случаях 1 и 3 соответствующие структуры–параметрические, а в случаях 2 и 4–непараметрические.

Теперь мы готовы сформулировать **основную задачу математической статистики**: *на основе экспериментальных данных сузить класс $\mathcal{P}$ априорно заданных вероятностных мер до некоторого более узкого подкласса $\mathcal{P}_0 \subset \mathcal{P}$ (в идеале выбрать одно распределение).* Например, в случае 1 выше мы хотели бы, зная число m (экспериментальные данные!), оценить неизвестную вероятность p выпадения герба, т.е. выделить одно распределение.

Глава 2

Выборка и выборочные характеристики

2.1 Выборка

Напомним еще раз основную задачу: на основе экспериментальных данных уточнить наши знания о модели, т.е. о распределении вероятностей. В этом параграфе мы уточним, что понимается в нашей теории под экспериментальными данными. Всюду далее мы будем рассматривать в качестве основного примера с.в. ξ, распределение вероятностей P_ξ которой (или ее функция распределения $F_\xi(y)$) неизвестно или известно не полностью. Для того, чтобы получить информацию об этой с.в. ξ мы организуем случайный эксперимент, в котором реализуется при определенных условиях одно или несколько измерений этой случайной величины. В результате мы получаем несколько чисел. Набор $x = (x_1, \ldots, x_N)$ этих чисел называется **выборкой** . Происхождение этого термина связана с тем, что в простейших ситуациях проведение эксперимента связано с реальным выбором из конечной совокупности (выбор шара в лотерии, выбор объекта для обследования и т.п.). Число N измерений называют **объемом выборки**. Способы получения выборок из конечных совокупностей для изучения статистических свойств этих совокупностей представляют из себя отдельную и очень непростую задачу. Они изучаются в курсе так называемой *общей статистики* и не будут рассматриваться в нашем курсе.

Мы уже обсуждали в курсе теории вероятностей, что нас интересует, обычно, не результат отдельного эксперимента, а что мы получаем **в среднем**, в длинной серии экспериментов. Точно также и в математической статистике нас будут интересовать свойства предложенных нам статистических процедур не при однократном их применении, а в среднем, когда они применяются много раз. Другой вариант – мы показываем, в большистве реализаций нашего эксперимента предложенная статистическая процедура дает хороший результат. В обоих случаях утверждение но-

сит вероятностный характер. Поэтому далее мы часто будем рассматривать отдельное измерение как случайную величину X_k, а выборку – как случайный вектор $X = (X_1, \ldots, X_N)$. Чтобы отличать выборку, рассматриваемую как набор конкретных чисел, от выборки – случайного вектора, в первом случае мы будем писать $x = (x_1, \ldots, x_N)$, а во втором – $X = (X_1, \ldots, X_N)$. Множество $\mathcal{X}$ всех возможных значений выборки X называется *выборочным пространством.*

Как мы говорили выше, выборка это результат измерений случайной величины ξ, которые производились в определенных условиях. Если в разных измерениях условия сильно меняются, то это означает, что мы имеем дело, вообще говоря, с разными объектами. Если условия эксперимента фиксированы столь сильно, что ничего не меняется, то имеем, по сути дела, одно измерение. Поэтому для "хорошего" эксперимента естественно предположить, что случайные величины $X_1, \ldots, X_N$ **одинаково распределены** (**однородная** выборка) и **независимы**. Когда выполнены оба свойства выборка X называется **повторной**. Чтобы уточнить дальнейшую терминологию рассмотрим следующий

Пример. Из *большой* совокупности людей (например, жителей некоторого города) отбирают некоторое количество их представителей для статистического обследования. Нас интересует некоторая количественная характеристика, измеряемая для каждого отобранного человека (например, возраст). В пределах данной совокупности эту характеристику мы рассматриваем как случайную величину ξ с некоторым распределением $P = P_\xi$. Отобрав N человек, мы получим N измерений $x_1, \ldots, x_N$ интересующей нас характеристики. Вся совокупность людей называется **генеральной совокупностью**. Поэтому мы говорим, что имеем выборку объема N из генеральной совокупности с распределением P некоторой случайной величины ξ или, более кратко, **выборку объема N из генеральной совокупности с распределением P**.

Эта терминология будет применяться и тогда, когда никакой реальной генеральной совокупности нет, а мы имеем случайную

величину ξ с распределением P_ξ.

В реальных задачах выборка содержит очень большое число измерений. Хотелось бы как-то их упорядочить, чтобы лучше понять, что же мы имеем. Первое, что обычно делают, это располагают все наблюдения в порядке возрастания

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(N)} \tag{2.1}$$

или

$$\begin{array}{cccc} x_{(1)} < x_{(2)} < \cdots < x_{(n)} \\ m_1 \quad m_2 \quad \cdots \quad m_n \end{array}$$

где во втором случае приведены только различные измерения, а m_k есть число появлений значения $x_{(k)}$. Полученный ряд неубывающих чисел называют *вариационным рядом*. Это те же самые числа, что и в выборке, но расположенные в порядке возрастания. В дальнейшем любую функцию от выборочных значений мы будем называть *статистикой*. Таким образом, статистика в этом специальном значении термина есть любая функция от наблюдений. Число $x_{(k)}$ называется *k-ой порядковой статистикий*. $x_{(1)}$ и $x_{(N)}$ – *экстремальные значения* выборки. Они позволяют оценить интервал возможных значений с.в. ξ. *Размах* выборки $x_{(N)} - x_{(1)}$ оценивает величину разброса значений с.в. ξ. Но число различных значений в вариационном ряду может быть все еще достаточно большим. Чтобы сократить объем хранимой информации применяют *группировку* элементов выборки. Множество возможных значений с. в. ξ делят на несколько интервалов

и подсчитывают сколько измерений попало в k-ый интервал. Результаты записывают в виде таблицы

$$\begin{array}{ccccccc} < a_1 & a_1 - a_2 & \cdots & a_k - a_{k+1} & \cdots & > a_r \\ n_0 \quad & n_1 \quad & \cdots \quad & n_k \quad & \cdots & n_r \end{array}$$

2.2 Эмпирическое распределение

Основной задачей математической статистики является оценка неизвестного распределения P_ξ случайной величины ξ на основе экспериментальных данных, т. е. используя выборку $x = (x_1, \ldots, x_N)$. Рассмотрим новую случайную величину ξ^*, множеством значений которой являются числа $x_1, \ldots, x_N$, каждому из которых приписывается вероятность $1/N$ (если некоторое значение появляется несколько раз, то его вероятность увеличивается в то же число раз). Случайная величина ξ^* является дискретной и ее распределение называется *эмпирическим распределением*, построенным по выборке x. По определению с. в. ξ^* для любого борелевского множества $B \subset R^1$

$$P_N^* = \frac{\nu_N(B)}{N} \ , \tag{2.2}$$

где $\nu_N(B)$ – число элементов выборки, которые попали во множество B. Таким образом, P_N^* есть, фактически, относительная частота появления события $(\xi \in B)$ в серии из N независимых и одинаковых испытаний. Напомним, что выборку можно рассматривать и как случайный вектор $X = (X_1, \ldots, X_N)$. В этом случае эмпирическое распределение P_N^* становится случайной величиной. В силу закона больших чисел (теорема Бернулли) для каждого фиксированного B мы имеем

$$P_N^*(B) \xrightarrow{P} P_\xi(B) = P(\xi \in B) \ , N \to \infty \ . \tag{2.3}$$

Распределение случайной величины часто задают с помощью функции распределения. Для ее оценки мы будем использовать *эмпирическую функцию распределения*. По определению

$$F_\xi(y) = P(\xi < y) \ .$$

Тогда, в силу формулы (2), для эмпирической функции распределения получаем

$$F_N^*(y) = \frac{N(y)}{N} \ , \tag{2.4}$$

где $N(y)$ – число элементов x_k выборки x, для которых $x_k < y$. Вновь рассматривая выборку как случайный вектор, получаем, что

$$F_N^*(y) \xrightarrow{P} F_\xi(y) \ , \ N \to \infty \ , \tag{2.5}$$

для каждого фиксированного y. На самом деле справедлив гораздо более сильный результат, известный как *теорема Гливенко-Кантелли*:

$$P\left\{\lim_{N\to\infty} \sup_y |F_N^*(y) - F_\xi(y)| = 0\right\} = 1 \ .$$

В тех случаях, когда априори известно, что распределение является непрерывным, хотелось бы и в качестве оценки получить непрерывное распределение. Тогда применяют так называемые *ядерные оценки*. Пусть Q есть некоторое фиксированное распределение вероятностей, обладающее нужными свойствами, например, абсолютно непрерывное. Тогда в качестве оценки незвестного распределения P_ξ мы берем

$$P^*(B) = \frac{1}{N} \sum_{k=1}^{N} Q(B - x_k) \ .$$

Для оценки плотности распределения используют *гистограмму*, которая определятся следующим образом. Пусть мы сгруппировали все элементы x_i выборки x в r интервалов, длина k-го интервала равна Δ_k, а N_k есть число элементов выборки, попавших в k-ый интервал. По определению гистограмма есть функция $\rho_N^*(y)$, определяемая по правилу

$$\rho_N^*(y) := \frac{N_k}{N \cdot \Delta_k} \ , \tag{2.6}$$

если y принадлежит k-му интервалу, и равная нулю в противном случае. Легко проверить, $\rho_N^*(y)$ обладает следующими свойствами:

1) $\rho_N^*(y) \geq 0$,

2) $\int\limits_{-\infty}^{\infty} \rho_N^*(y)dy = 1$,

3) $\int\limits_a^b \rho_N^*(y)dy = P(a \leq \xi^* < b)$.

Таким образом, $\rho_N^*(y)$ обладает всеми основными свойствами плотности. Более того, если при $N \to \infty$ мы имеем $r \to \infty$, $\max\limits_k \Delta_k \to 0$, но некоторым согласованным образом, то

$$P(\lim_{N\to\infty} \max_{-\infty<y<\infty} |\rho_N^*(y) - \rho_\xi(y)| = 0) = 1 \ . \tag{2.7}$$

В некоторых задачах мы знаем, что плотность является достаточно гладкой функцией, и хотим, чтобы оценка плотности обладала такими же свойствами. В этом случае вновь используют ядерные оценки. Пусть $\rho(y)$ – некоторая заданная функция ("ядро"), которая является плотностью. По выборке $x = (x_1, \ldots, x_N)$ построим функцию

$$\rho_N(y) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{b_N} \rho\left(\frac{y - x_k}{b_N}\right) \ , \tag{2.8}$$

где $\{b_n\}$ – некоторая последовательность положительных чисел. Если ядро $\rho(y)$ и последовательность $\{b_n\}$ обладают некоторыми свойствами, то мы вновь можем гарантировать сходимость $\rho_N(y)$ к $\rho(y)$ при $N \to \infty$.

2.3 Выборочные характеристики

Ранее, при изучении случайных величин, мы отмечали, что распределение является наиболее полной вероятностной характеристикиой случайной величины. Но во многих задачах бывает затруднительно (а иногда и не нужно) дать полное описание распределения. В этом случае при описании распределения используют небольшое число числовых характеристик, отражающих те или иные стороны распределения. Обычно для этой цели используют моменты (наиболее часто математическое ожидание и дисперсию) и квантили. Эти характеристики определяются через распределение случайной величины и представляют собой некоторые функционалы от этого распределения. Подставляя в них эмпирическое распределение, можно получить их эмпирические аналоги, которые называются *выборочными* или *эмпирическими характеристиками.*

Рассмотрим несколько примеров. Пусть мы имеем случайную величину ξ. Напомним, что ее момент порядка m относительно точки a определяется по правилу

$$\nu_m(a) = M[(\xi - a)^m] = \int\limits_{-\infty}^{\infty} (y-a)^m dF_\xi(y) \ .$$

В частности, мы можем определить математическое ожидание $M\xi$, характеризующее "центр" распределения, и дисперсию

$$D(\xi) = M[(\xi - M\xi)^2] = M(\xi^2) - (M\xi)^2 \ ,$$

характеризующую "разброс" или "рассеяние" распределения относительно этого центра. Их выборочными аналогами будут: *выборочный момент* (при $a = 0$)

$$\hat{\nu}_m = \frac{1}{N} \sum_{k+1}^{N} x_k^m \ ,$$

выборочное среднее

$$\bar{x} = \frac{1}{N} \sum_{k+1}^{N} x_k$$

и *выборочная дисперсия*

$$S^2 = \frac{1}{N} \sum_{k+1}^{N} x_k^2 - \bar{x}^2 = \frac{1}{N} \sum_{k+1}^{N} (x_k - \bar{x})^2 \ .$$

Следующая теорема дает описание некоторых свойств выборочных моментов. Она есть непосредственное следствие закона больших чисел и центральной предельной теоремы.

Теорема 1 . *Если $M(|\xi|^{2m}) < \infty$ и $\hat{\nu}_m$ – выборочный момент порядка m, построенный по повторной выборке объема N, то*

1) $M(\hat{\nu}_m) = \nu_m$;

2) $\hat{\nu}_m \xrightarrow{P} \nu_m = M(\xi^m), \ N \to \infty$,

в частности,

$$\bar{X} \xrightarrow{P} a = M\xi \ , \ S^2 \xrightarrow{P} \sigma^2 = D(\xi) \ ;$$

3) $\hat{\nu}_m$ имеет асимптотически нормальное распределение со средним ν_m и дисперсией $(\nu_{2m} - \nu_m^2)/N$.

Доказательство. Напомним еще раз, что повторная выборка есть случайный вектор $(X_1, \ldots, X_N)$, состоящий из независимых и одинаково распределенных случайных величин. Учитывая это и условия теоремы, имеем

1) $(X_1^m, \ldots, X_N^m)$ – независимы,

2) $(X_1^m, \ldots, X_N^m)$ – одинаково распределены,

3) существуют конечные $M(X_1^m) = \nu_m$ и $D(X_1^m) = M(X_1^{2m}) - (M(X_1^m))^2$.

Тогда

$$M(\hat{\nu}_m) = M\left(\frac{1}{N}\sum_{k=1}^{N} X_k^m\right) = \frac{1}{N}\sum_{k=1}^{N} M(X_k^m) = \frac{1}{N}\cdot N \cdot \nu_m = \nu_m \ .$$

В силу закона больших чисел для н.о.р.с.в. получаем

$$\frac{1}{N}\sum_{k=1}^{N} X_k^m \xrightarrow{P} M(X_1^m) = \nu_m \ .$$

Далее,

$$D\left(\frac{1}{N}\sum_{k=1}^{N} X_k^m\right) = D(X_1^m)/N = (\nu_{2m} - \nu_m^2)/N \ .$$

В силу ЦПТ для н.о.р.с.в. случайная величина

$$\frac{\hat{\nu}_m - M(\hat{\nu}_m)}{\sqrt{D(\hat{\nu}_m)}} = \frac{\hat{\nu}_m - \nu_m}{\sqrt{(\nu_{2m} - \nu_m^2)/N}}$$

имеет асимптотически при $N \to \infty$ стандартное нормальное распределение $\square$.

Квантиль порядка p, $0 < p < 1$, определяется как такое число y_p, для которого

$$P(\xi \le y_p) \ge p \ , \ P(\xi \ge y_p) \ge 1 - p \ .$$

Если распределение не имеет дискретных точек, то это сводится к уравнению

$$P(\xi < y_p) = F_\xi(y_p) = p \ .$$

В частности, *медиана* $y_{1/2}$ распределения с.в. ξ определяется из соотношения

$$P(\xi < y_{1/2}) = \frac{1}{2} \ ,$$

где мы предположили, для простоты, что распределение не имеет дискретных точек. Медиана является еще одной характеристикой, оценивающей центр распределения. *Выборочная квантиль* порядка p определяется по правилу

$$\hat{y}_p = x_{([Np])} \ .$$

Чаще других квантилей используется медиана $y_{1/2}$. Здесь для улучшения свойств оценки используется следующее определение:

$$\hat{y}_{1/2} = \begin{cases} x_{(n+1)} & , \quad \text{если } N = 2n+1 \ , \\ \frac{1}{2}(x_{(n)} + x_{(n+1)}) & , \quad \text{если } N = 2n \ . \end{cases}$$

Задача. Если y_p есть точка непрерывности и точка роста ф.р. $F_\xi(y)$, то

$$\hat{y}_p \xrightarrow{P} y_p \ , \ N \to \infty \ .$$

Глава 3

Точечные оценки параметров

3.1 Параметрические статистические структуры

Как уже отмечалось выше, всюду далее мы рассматриваем случай, когда имеем некоторую случайную величину ξ, распределение которой P_ξ (или ее ф.р. F_ξ) неизвестно или известно не полностью. В этом и нескольких последующих параграфах будут рассматриваться так называемые *параметрические структуры* $(\Omega, \mathcal{A}, \mathcal{P})$, для которых семейство распределений $\mathcal{P}$ является *парметрическим*, т. е. существует конечное число вещественных параметров $\theta_1, \ldots, \theta_m$, с помощью которых можно занумеровать все распределения P из нашего семейства $\mathcal{P}$. В дальнейшем мы будем записывать параметры в виде одного многомерного параметра $\theta = (\theta_1, \ldots, \theta_m)$, который пробегает некоторое множество $\Theta \subset R^m$. Тогда мы можем записать $\mathcal{P} = \{P_\theta \ , \ \theta \in \Theta\}$. Для нашей задачи, когда мы изучаем одну случайную величину ξ, в качестве Ω мы выбираем R^1, $\mathcal{A} = \mathcal{B}$ есть σ-алгебра борелевских подмножеств. Тогда $\mathcal{P}$ есть некоторое семество возможных распределений случайной величины ξ. Ниже мы рассматриваем несколько примеров параметрических семейств распределений.

Примеры. 1. **Распределение Бернулли** – это распределение случайной величины ξ, которая принимает два значения: 1 - с вероятностью p и 0 - с вероятностью $1-p$. Число $p \in (0,1)$ является параметром распределения.

Для этого распределения

$$M(\xi) = p \ , \ D(\xi) = p(1-p) \ .$$

2. **Биномиальное распределение** – это распределение случайной величины ξ, которая принимает значения $0, 1, \ldots, n$ с вероятностями

$$P(\xi = m) = C_n^m p^m (1-p)^{n-m} \ .$$

Целое число $n \geq 1$ и $p \in (0,1)$ являются параметрами распределения.

Для этого распределения

$$M(\xi) = np \ , \ D(\xi) = np(1-p) \ .$$

3. **Распределение Пуассона** – это распределение случайной величины ξ, которая принимает значения $0, 1, \ldots$ с вероятностями

$$P(\xi = m) = \frac{\lambda^m}{m!} e^{-\lambda} \ .$$

Число $\lambda > 0$ является параметром распределения.

Для этого распределения

$$M(\xi) = \lambda \ , \ D(\xi) = \lambda \ .$$

4. **Геометрическое распределение** – это распределение случайной величины ξ, которая принимает значения $1, 2 \ldots$ с вероятностями

$$P(\xi = m) = p(1-p)^{m-1} \ .$$

Число $p \in (0,1)$ является параметром распределения.

Для этого распределения

$$M(\xi) = \frac{1}{p} \ , \ D(\xi) = \frac{1-p}{p^2} \ .$$

5. **Отрицательное биномиальное распределение** – это распределение случайной величины ξ, которая принимает значения $r, r+1 \ldots$ с вероятностями

$$P(\xi = m) = C_{m-1}^{r-1} p^r (1-p)^{m-r} \ .$$

Целое число $r \geq 1$ и $p \in (0,1)$ являются параметрами распределения. При $r = 1$ получаем геометрическое распределение.

Для этого распределения

$$M(\xi) = \frac{r}{p} \ , \ D(\xi) = r\frac{1-p}{p^2} \ .$$

6. **Равномерное распределение** на отрезке (a, b) – это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \frac{1}{b-a} & , \quad y \in (a, b) \ , \\ 0 & , \quad \text{в пр. сл.} \end{cases}$$

Числа a и b, $a < b$, являются параметрами распределения.

Для этого распределения

$$M(\xi) = \frac{a+b}{2} \ , \ D(\xi) = \frac{(b-a)^2}{12} \ .$$

7. **Показательное распределение** – это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \lambda e^{-\lambda y} & , \quad y > 0 \ , \\ 0 & , \quad \text{в пр. сл.} \end{cases}$$

Число $\lambda > 0$ является параметром распределения.

Для этого распределения

$$M(\xi) = \frac{1}{\lambda} \ , \ D(\xi) = \frac{1}{\lambda^2} \ .$$

8. **Гамма-распределение** – это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} & , \quad y > 0 \ , \\ 0 & , \quad \text{в пр. сл.} \end{cases}$$

Числа $\alpha > 0$ и $\beta > 0$ являются параметрами распределения. При $\alpha = 1$ и $\beta = \lambda$ получаем показательное распределение.

Для этого распределения

$$M(\xi) = \frac{\alpha}{\beta} \ , \ D(\xi) = \frac{\alpha}{\beta^2} \ .$$

9. **Бета-распределение** – это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \begin{cases} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} y^{r-1}(1-y)^{s-1} & , \quad 0 < y < 1 \ , \\ 0 & , \quad \text{в пр. сл.} \end{cases}$$

Числа $r > 0$ и $s > 0$ являются параметрами распределения. При $r = s = 1$ получаем равномерное на [0,1] распределение.

Для этого распределения

$$M(\xi) = \frac{r}{r+s} \ , \ D(\xi) = \frac{rs}{(r+s)^2(r+s+1)} \ .$$

10. **Нормальное распределение** – это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a)^2}{2\sigma^2}} \ , \ y \in R^1 \ .$$

Числа $a \in R^1$ и $\sigma^2 > 0$ являются параметрами распределения.

Для этого распределения

$$M(\xi) = a \ , \ D(\xi) = \sigma^2 \ .$$

11. **Распределение Коши** – это распределение случайной величины ξ, которая имеет плотность распределения следующего вида:

$$\rho_\xi(y) = \frac{1}{\pi} \frac{1}{1 + \left(\frac{y-a}{b}\right)^2} \ .$$

Числа $a \in R^1$ и $b > 0$ являются параметрами распределения.

Это распределение не имеет математического ожидания и дисперсии.

12. **Сдвиг-масштабное семейство распределений**.

Пусть мы имеем фиксированную функцию распределения $F(y)$. Семейство распределений

$$\mathcal{P} = \left\{ F(y,a,b) := F\left(\frac{y-a}{b}\right) \right\}$$

называется **сдвиг-масштабным**. Числа $a \in R^1$ и $b > 0$ называются *параметрами сдвига и масштаба* соответственно.

Если существует плотность $\rho(y) = \frac{d}{dy}F(y)$, то

$$\rho(y, a, b) = \frac{1}{b}\ \rho\left(\frac{y-a}{b}\right)\ .$$

Для этого распределения

$$M(\xi) = b \cdot a_0 + a\ ,\ D(\xi) = b^2\sigma^2\ ,$$

где a_0 и σ^2 есть математическое ожидание и дисперсия для распределения из нашего семейства с параметрами $a = 0$, $b = 1$ (если они существуют).

3.2 Точечные оценки параметров и их свойства

Основной задачей математической статистики является выбор неизвестного распределения, используя имеющиеся экспериментальные данные. Если мы имеем случайную величину ξ с неизвестным распределением из некоторого параметрического семейства $\mathcal{P} = \{P_\theta,\ \theta \in \Theta \subset \mathbf{R^m}\}$, то основная задача сводится к оценке неизвестных параметров. Экспериментальными данными в нашей задаче является повторная выборка $x = (x_1, \ldots, x_N)$. Что значит построить оценку? В конкретной ситуации мы должны для заданного набора чисел $x_1, \ldots, x_N$ указать соответствующее им значение параметра θ. Рассматривая задачу более общо, с теоретической точки зрения, мы должны указать правило, по которому каждой выборке $x = (x_1, \ldots, x_N)$ сопоставляется некоторое значение параметра θ. Это приводит нас к следующему определению.

Определение 2 . **Оценкой параметра** θ *называется произвольная функция* $\hat{\theta}$ *из выборочного пространства* X *в пространство параметров* $\Theta \subset \mathbf{R^m}$.

Выше мы любую функцию от выборки назвали статистикой. Если она используется в целях оценивания неизвестного параметра,

то она называется **оценкой**. В этом определении в качестве оценки предлагается любая функция от выборки. Ясно, что не любая такая функция будет "хорошей" оценкой. Далее мы рассмотрим некоторые свойства, которыми должны обладать "хорошие" оценки. Как уже отмечалось ранее, свойства оценок будут изучаться "в среднем", когда предложенная процедура применяется многократно в длинной серии испытаний. В этом случае мы рассматриваем выборку как случайный вектор $X = (X_1, \ldots, X_N)$. Но тогда оценка $\hat{\theta}$ является случайной величиной $\hat{\theta} = \hat{\theta}_N(X) = \hat{\theta}_N(X_1, \ldots, X_N)$.

При вычислении математического ожидания или вероятности мы должны указать какое именно распределение P из нашего семейства $\mathcal{P}$ мы используем. Для этого мы будем использовать обозначения M_θ и P_θ, которые означают, что мы вычисляем математическое ожидание и вероятность в предположении, что истинным значением параметра является θ.

Ниже мы перечислим некоторые свойства, которые обычно требуют от "хороших" оценок. Для простоты обозначений мы предполагаем, что мы имеем один неизвестный параметр θ, то есть $m = 1$.

Определение 3 . *Оценка $\hat{\theta} = \hat{\theta}_N(X_1, \ldots, X_N)$ параметра θ называется* **несмещенной**, *если*

$$M_\theta(\hat{\theta}_N(X)) = \theta \qquad \forall \theta \in \Theta.$$

В силу сформулированной в §2 теоремы мы видим, что выборочные моменты являются несмещенными оценками истинных моментов. Например, $M(\bar{x}) = a = M(\xi)$. То же самое справедливо относительно линейных функций от моментов. Но в общем случае, если параметр есть некоторая функция от начального момента, то, подставляя в нее выборочный момент, мы получим смещенную оценку. На практических занятиях мы выясним, что выборочная дисперсия

$$S^2 = \frac{1}{N} \sum_{k=1}^{N} (X_k - \bar{X})^2$$

является смещенной оценкой для $\sigma^2 = D(\xi)$, так как

$$M(S^2) = \frac{N-1}{N}\sigma^2.$$

Несмещенной оценкой для σ^2 является так называемая **исправленная выборочная дисперсия**

$$S_1^2 = \frac{1}{N-1}\sum_{k=1}^{N}(X_k - \bar{X})^2.$$

Определение 4 . *Последовательность оценок* $\{\hat{\theta}_N(X),\ N \geq N_0\}$ *параметра* θ *называется* **состоятельной**, *если*

$$\hat{\theta}_N(X) \xrightarrow{P_\theta} \theta \qquad \forall\theta \in \Theta.$$

Иногда слово последовательность опускают и говорят, что оценки $\hat{\theta}_N$ состоятельны, хотя это и не точно. Вновь обращаясь к теореме из §2, мы видим, что выборочные моменты являются состоятельными оценками истинных моментов. Например, выборочное среднее $\bar{X}$ является состоятельной оценкой математического ожидания a, а оба варианта выборочной дисперсии S^2 и S_1^2 являются состоятельными оценками для дисперсии σ^2. Приведем более сложный пример.

Задача. Пусть случайная величина ξ имеет непрерывную и строго монотонную функцию распределения. Тогда выборочная медиана $\hat{y}_{1/2}$ является несмещенной и состоятельной оценкой для истинной медианы.

Сформулированные выше два свойства оценок ничего не говорят о точности полученных оценок.

Определение 5 . *Оценка* $\hat{\theta} = \hat{\theta}(X)$ *называется* **эффективной** *или* **несмещенной оценкой с минимальной дисперсией (НОМД)**, *если*

1. $M_\theta(\hat{\theta}) \equiv \theta \qquad \forall\theta$, *то есть это несмещенная оценка;*
2. $M_\theta(\hat{\theta} - \theta)^2 \leq M_\theta(\tilde{\theta} - \theta)^2 \qquad \forall\theta,$

где $\tilde{\theta}$ — любая другая несмещенная оценка для параметра θ.

Эффективная оценка является **оптимальной в среднем квадратическом** оценкой в классе несмещенных оценок (см. тему "Гильбертово пространство случайных величин").

Позже мы покажем, что выборочное среднее $\bar{X}$ является эффективной оценкой параметра a в классе нормальных распределений.

В реальных задачах нам необходимо не просто указать оценку неизвестного параметра, но и уметь оценивать вероятности различных отклонений этой оценки от истинного параметра. В этом случае оказывается полезным следующее понятие.

Определение 6 . *Последовательность оценок $\{\hat{\theta}_N,\ N \geq N_0\}$ называется* **асимптотически нормальной**, *если $\forall N \geq N_0$ существуют константы $A_N(\theta) \in \mathbf{R^1}$ и $B_N(\theta) > 0$ такие, что случайная величина*

$$\frac{\hat{\theta}_N - A_N(\theta)}{B_N(\theta)}$$

имеет асимптотически стандартное нормальное распределение, то есть ее функция распределения сходится к функции распределения стандартного нормального закона.

Теорема из §2 говорит нам, что выборочные моменты являются асимптотически нормальными оценками истинных моментов. В частности, выборочное среднее $\bar{X}$ является асимптотически нормальной оценкой для математического ожидания.

Замечание. Иногда складывается ложное впечатление, что выборочное среднее $\bar{X}$ является некоторой универсальной оценкой центра распределения, которая хороша во всех случаях. Рассмотрим семейство распределений Коши с плотностями

$$\rho(y, \theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (y - \theta)^2}, \quad y \in \mathbf{R^1}.$$

Случайная величина ξ с таким распределением не имеет математического ожидания. Если $X = (X_1, \ldots, X_N)$ есть повторная вы-

борка из такого распределения, то выборочное среднее

$$\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k$$

одинаково распределено с ξ. Отсюда мы видим, что $\bar{X}$ не является несмещенной, а также состоятельной и асимптотически нормальной. В этом случае хорошей оценкой для центра распределения θ является выборочная медиана $\hat{y}_{1/2}$.

3.3 Неравенство Рао-Крамера

Выше мы дали определение эффективной оценки и отметили, что она является оптимальной в среднем квадратическом в классе всех несмещенных оценок. Но остался открытым вопрос о том, как в конкретной задаче найти такую оптимальную оценку. Из-за недостатка времени мы не будем рассматривать методы построения эффективных оценок. Вместо этого мы докажем некоторый результат, который позволит для конкретной оценки доказать, что она является эффективной.

Напомним определение эффективной оценки: $\hat{\theta}_N = \hat{\theta}_N(X)$ является **эффективной**, если она несмещенная и для любой другой несмещенной оценки $\tilde{\theta}_N = \tilde{\theta}_N(X)$ мы имеем

$$D_\theta(\hat{\theta}_N) \le D_\theta(\tilde{\theta}_N).$$

Предположим, что мы нашли некоторую нижнюю границу для дисперсий всех несмещенных оценок. Если эта граница достигается на некоторой конкретной несмещенной оценке, то она и будет эффективной. Ниже мы получим такую нижнюю границу.

Введем вначале некоторую вспомогательную величину. Пусть случайная величина ξ имеет плотность $\rho(y, \theta)$, где $\theta \in \Theta \subset \mathbf{R^1}$ — скалярный параметр распределения.

Определение 7 . *Величина*

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln \rho(y, \theta)}{\partial \theta} \right)^2 \rho(y, \theta)\, dy = M_\theta \left(\frac{\partial \ln \rho(\xi, \theta)}{\partial \theta} \right)^2$$

называется **информацией по Фишеру**, *содержащейся в одном измерении случайной величины xi, о параметре θ.*

Если мы имеем повторную выборку $X = (X_1, \ldots, X_N)$ из генеральной совокупности с плотностью распределения $\rho(y, \theta)$, то ее совместная плотность распределения имеет вид

$$\mathcal{L}(x, \theta) = \rho(x_1, \theta) \cdot \ldots \cdot \rho(x_N, \theta).$$

Функция $\mathcal{L}(x, \theta)$ как функция от θ при фиксированном x называется **функцией правдоподобия**. Информацией по Фишеру о параметре θ, содержащейся в выборке $X = (X_1, \ldots, X_N)$, называется величина

$$I_N(\theta) = \int\limits_{\mathbf{R^N}} \left(\frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta} \right)^2 \mathcal{L}(x, \theta)\, dx_1 \ldots dx_N = M_\theta \left(\frac{\partial \ln \mathcal{L}(X, \theta)}{\partial \theta} \right)^2.$$

Нетрудно показать, что для повторной выборки имеет место соотношение

$$I_N(\theta) = N \cdot I(\theta).$$

Если ξ имеет дискретное распределение, то в определении информации по Фишеру плотность нужно заменить на вероятность, а интеграл на сумму.

Сформулированная ниже теорема справедлива, если распределение вероятностей случайной величины ξ удовлетворяет некоторым условиям регулярности. В силу громоздкости этих условий мы не будем выписывать их в явном виде. Отметим только, что эти условия гарантируют нам возможность дифференцирования по параметру под знаком интеграла (математического ожидания).

Мы будем рассматривать несколько более общую задачу, когда оценивается некоторая вещественная функция $g(\theta)$ от скалярного параметра $\theta \in \Theta \subset \mathbf{R^1}$.

Теорема 2 . *Пусть $\hat{g}_N = \hat{g}_N(X)$ — некоторая несмещенная оценка для вещественной функции $g(\theta)$ от параметра θ, построенная*

по повторной выборке $X = (X_1, \ldots, X_N)$ *из генеральной совокупности с функцией распределения* $F(y, \theta)$*, удовлетворяющей условиям регулярности. Тогда имеет место неравенство*

$$D_\theta(\hat{g}_N) \geq \frac{[g'(\theta)]^2}{N \cdot I(\theta)}\ , \tag{1}$$

называемое **неравенством Рао-Крамера**.

Доказательство. Из свойств плотности и определения несмещенной оценки получаем:

$$M_\theta \mathcal{L}(X, \theta) = \int\limits_{\mathbf{R^N}} \mathcal{L}(x, \theta)\, dx_1 \ldots dx_N \equiv 1, \tag{2}$$

$$M_\theta(\hat{g}_N) = \int\limits_{\mathbf{R^N}} \hat{g}_N(x) \cdot \mathcal{L}(x, \theta)\, dx = g(\theta). \tag{3}$$

Дифференцируя (2) и (3) по θ (здесь нужны условия регулярности), получаем

$$\theta = \int\limits_{\mathbf{R^N}} \frac{\partial}{\partial \theta} \mathcal{L}(x, \theta)\, dx, \tag{4}$$

$$g'(\theta) = \int\limits_{\mathbf{R^N}} \hat{g}_N(x) \frac{\partial}{\partial \theta} \mathcal{L}(x, \theta)\, dx. \tag{5}$$

Умножим (4) на $g(\theta)$ и вычтем почленно из (5):

$$\begin{aligned} g'(\theta) &= \int\limits_{\mathbf{R^N}} (\hat{g}_N(x) - g(\theta)) \cdot \tfrac{\partial}{\partial \theta} \mathcal{L}(x, \theta)\, dx = \\ &= \int\limits_{\mathbf{R^N}} (\hat{g}_N(x) - g(\theta)) \cdot \tfrac{\frac{\partial}{\partial \theta} \mathcal{L}(x, \theta)}{\mathcal{L}(x, \theta)} \cdot \mathcal{L}(x, \theta)\, dx = \\ &= \int\limits_{\mathbf{R^N}} (\hat{g}_N(x) - g(\theta)) \cdot \tfrac{\partial}{\partial \theta} \ln \mathcal{L}(x, \theta) \cdot \mathcal{L}(x, \theta)\, dx = \\ &= M_\theta \left[(\hat{g}_N(X) - g(\theta)) \cdot \tfrac{\partial}{\partial \theta} \ln \mathcal{L}(X, \theta) \right]. \end{aligned} \tag{6}$$

Применим к (6) неравенство Коши-Буняковского:

$$\begin{aligned} (g'(\theta))^2 &= \left[M_\theta (\hat{g}_N(X) - g(\theta)) \left(\tfrac{\partial}{\partial \theta} \ln \mathcal{L}(X, \theta) \right) \right]^2 \leq \\ &\leq M_\theta (\hat{g}_N(X) - g(\theta))^2 \cdot M_\theta \left(\tfrac{\partial}{\partial \theta} \ln \mathcal{L}(X, \theta) \right)^2 = \\ &= D_\theta(\hat{g}_N) \cdot I_N(\theta). \end{aligned}$$

Что и требовалось доказать.

Пример. ξ имеет нормальное распределение с параметрами a и σ^2. Нас интересует параметр $\theta = a$. В качестве оценки a предлагается выборочное среднее

$$\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k.$$

Покажем, что это эффективная оценка. В нашем случае $g(\theta) = \theta$, $g'(\theta) \equiv 1$, $I(\theta) = 1/\sigma^2$. Тогда

$$D_\theta(\bar{X}) = \frac{\sigma^2}{N} = \frac{1}{N/\sigma^2} = \frac{1}{I_N(\theta)}.$$

Таким образом, нижняя граница для дисперсий несмещенных оценок достигается.

3.4 Методы построения оценок

В заключение этого параграфа рассмотрим два наиболее популярных метода построения оценок.

а) **Метод моментов.**

Пусть случайная величина ξ имеет функцию распределения $F_\xi(y, \theta_1, \ldots, \theta_m)$, которая содержит несколько неизвестных параметров, а в остальном вид этой функции известен. Тогда мы можем вычислить несколько первых моментов

$$\begin{cases} M\xi = f_1(\theta_1, \ldots, \theta_m) = \nu_1, \\ M\xi^2 = f_2(\theta_1, \ldots, \theta_m) = \nu_2, \\ \ldots, \\ M\xi^m = f_m(\theta_1, \ldots, \theta_m) = \nu_m, \end{cases}$$

которые являются функциями от неизвестных параметров θ_1, $\ldots$, θ_m. Предположим, что мы решили эти уравнения относительно параметров:

$$\begin{cases} \theta_1 = g_1(\nu_1, \ldots, \nu_m), \\ \ldots, \\ \theta_m = g_m(\nu_1, \ldots, \nu_m). \end{cases} \tag{7}$$

В начале этого параграфа мы выяснили, что хорошими оценками для моментов являются их выборочные аналоги $\hat{\nu}_1, \ldots, \hat{\nu}_m$. Подставляя их в уравнения (7) мы получим оценки $\hat{\theta}_1, \ldots, \hat{\theta}_m$, построенные по **методу моментов**.

Пример. ξ имеет нормальное распределение с параметрами $a = \theta_1$ и $\sigma^2 = \theta_2$. Мы знаем, что $a = M\xi = \nu_1$, $\sigma^2 = D(\xi) = M\xi^2 - (M\xi)^2 = \nu_2 - \nu_1^2$. Отсюда получаем

$$\hat{a} = \bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k,$$
$$\hat{\sigma}^2 = \hat{\nu}_2 - (\hat{\nu}_1)^2 = \frac{1}{N}\sum_{k=1}^{N} X_k^2 - (\bar{X})^2 = \frac{1}{N}\sum_{k=1}^{N}(x_k - \bar{X})^2 = S^2.$$

Замечание. Метод моментов как правило дает состоятельные оценки, но они могут быть смещенными и не эффективными. Этот метод является довольно простым с вычислительной точки зрения.

b) **Метод наибольшего правдоподобия.**

Этот метод основан на интуитивном представлении о том, что реализуются в случайном эксперименте те события, которые имеют большие вероятности. Обращая это утверждение, мы приходим к следующему **принципу наибольшего правдоподобия**: *при заданном наборе экспериментальных данных нужно выбирать ту модель, для которой эти данные наиболее вероятны.* Применим этот принцип к оценке параметров.

Пусть ξ — дискретная случайная величина, вероятности значений y которой вычисляются с помощью некоторой функции $p(y, \theta_1, \ldots, \theta_m) = p(y, \theta)$. Пусть мы имеем конкретную повторную выборку $x = (x_1, \ldots, x_N)$. Вычислим вероятность появления именно таких значений:

$$\mathcal{L}(x, \theta) = p(x_1, \theta) \cdot \ldots \cdot p(x_N, \theta).$$

Выше мы назвали $\mathcal{L}(x, \theta)$ функцией правдоподобия. При фиксированном x найдем такое θ, которое есть решение экстремальной задачи:

$$\mathcal{L}(x, \theta) \longrightarrow \max_{\theta}.$$

Ясно, что при разных x мы будем получать разные решения. Таким образом мы получим оценку $\hat{\theta}_N = \hat{\theta}_N(x)$ — **оценку по методу наибольшего правдоподобия** (ОМНП).

Если ξ имеет плотность распределения, то в функции правдоподобия нужно вероятности заменить на плотности.

Примеры. 1. ξ имеет нормальное распределение с параметрами $a = \theta_1$ и $\sigma^2 = \theta_2$. ОМНП в этом случае вновь будут $\bar{X}$ и S^2.

2. ξ имеет распределение Пуассона с параметром λ. Если $x = (x_1, \ldots, x_N)$ — повторная выборка, то

$$\mathcal{L}(x, \lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdot \ldots \cdot \frac{\lambda^{x_N}}{x_N!} e^{-\lambda} = \frac{\lambda^{x_1 + \ldots + x_N}}{x_1! \cdot \ldots \cdot x_N!} e^{-N\lambda}.$$

Нам нужно найти точку, где достигается экстремум, а не величину экстремума. Поэтому любая монотонная функция от функции правдоподобия дает тот же результат. В нашем случае удобно перейти к логарифмам.

$$L(x, \lambda) = \ln \mathcal{L}(x, \lambda) = (x_1 + \ldots + x_N) \cdot \ln \lambda - N \cdot \lambda - \ln(x_1! \cdot \ldots \cdot x_N!).$$

$$\begin{array}{l} \frac{\partial L(x,\lambda)}{\partial \lambda} = \frac{(x_1 + \ldots + x_N)}{\lambda} - N = 0 \\ \Longrightarrow \hat{\lambda} = \frac{1}{N}(x_1 + \ldots + x_N) = \bar{X}. \end{array}$$

Несложно показать, что это точка наибольшего значения для $\mathcal{L}(x, \lambda)$.

Перечислим без доказательства некоторые свойства ОМНП. Если выполнены некоторые условия регулярности для функции распределения $F(y, \theta)$ случайной величины ξ, то ОМНП $\hat{\theta}_N$ обладают свойствами:

1. $\hat{\theta}_N$ — состоятельны,
2. $\hat{\theta}_N$ — асимптотически нормальны,
3. $\hat{\theta}_N$ — асимптотически эффективны.

Глава 4

Интервальные оценки параметров

4.1 Определение доверительного интервала

До сих пор в качестве оценки мы указывали некоторое конкретное значение $\hat{\theta}_N$ параметра. Но в реальной задаче даже самая хорошая оценка, которая возможна в рассматриваемой ситуации, может оказаться неудовлетворительной, так как не обеспечивает нужной точности. Таким образом, мы хотели бы дополнить $\hat{\theta}_N$ оценкой точности, то есть неравенством вида

$$|\hat{\theta}_N - \theta| < \delta.$$

Покажем, что в нашей задаче это, вообще говоря, невозможно, если не сделать некоторых оговорок. Чтобы прояснить ситуацию рассмотрим

Пример. $X = (X_1, \ldots, X_N)$ — повторная выборка из генеральной совокупности с нормальным распределением, имеющим параметры a и σ^2. Интересующий нас параметр $\theta = a$. Мы уже знаем, что оптимальной оценкой является $\hat{a} = \bar{X}$. Но это случайная величина, имеющая нормальное распределение со средним a и дисперсией σ^2/N. Такая случайная величина принимает любые значения от $-\infty$ до $+\infty$. Поэтому для любого $\delta > 0$ неравенство

$$|\bar{X} - a| < \delta$$

наверняка выполняться не может. Всегда есть положительная вероятность того, что оно будет не выполнено. Но при достаточно большом δ эта вероятность будет достаточно мала. Поэтому, практически достоверно, что оно выполнено. Решая неравенство относительно a, получаем

$$\bar{X} - \delta < a < \bar{X} + \delta.$$

Таким образом, этот подход дает нам в качестве оценки некоторый интервал, который покрывает истинное значение параметра с большой вероятностью.

Определение. Пусть мы имеем повторную выборку $X = (X_1, \ldots, X_N)$ из генеральной совокупности с функцией распределения $F(y, \theta)$, где $\theta \in \Theta \subset \mathbf{R^1}$ — скалярный параметр. **Доверительным интервалом** уровня γ для параметра θ называется интервал $(\hat{\theta}^{(1)}(X), \hat{\theta}^{(2)}(X))$ со случайными концами такой, что

$$P_\theta(\hat{\theta}^{(1)}(X) < \theta < \hat{\theta}^{(2)}(X)) \geq \gamma \qquad \forall \theta \in \Theta. \tag{1}$$

Число γ называется доверительным уровнем интервала.

Чем больше число γ, тем выше вероятность того, что интервал покрывает истинное значение параметра. В этом смысле число γ характеризует **надежность** этого интервала. Увеличивая длину интервала, мы, естественно, увеличиваем и его надежность. Но при этом уменьшается **точность** нашей оценки, которую естественно характеризовать средней длиной нашего интервала:

$$l = M_\theta(\hat{\theta}^{(2)} - \hat{\theta}^{(1)}).$$

γ и l — это два "конфликтующих" показателя. Улучшая один, мы ухудшаем другой и наоборот. Это типичная ситуация в так называемых многокритериальных задачах. В этих задачах нам нужно выбрать некоторый объект, который был бы хорошим сразу по нескольким показателям. Как правило, невозможно найти объект, который был бы наилучшим по всем показателям сразу. В этом случае применяют следующий подход. Мы отбираем все объекты, которые обладают достаточно хорошими (пусть и не наилучшими) свойствами по некоторым наиболее важным для нас критериям. Затем среди отобранных объектов стараются найти такой, который оптимизирует остальные критерии. В нашей задаче мы фиксируем доверительный уровень γ. Обычно это число вида 0.9, 0.95, 0.99, то есть достаточно близкое к 1. Затем среди всех доверительных интервалов уровня γ пытаются найти такой, у которого средняя длина l будет наименьшей, то есть наиболее точный.

Единого метода построения доверительных интервалов, который срабатывал бы во всех случаях жизни, не существует. Но тем не менее есть достаточно общие методы, которые мы изложим

ниже. Начнем мы с наиболее важного примера — нормального распределения.

4.2 Некоторые распределения, связанные с нормальным

При построении доверительных интервалов и проверке гипотез о параметрах нормального распределения нам потребуются некоторые специальные распределения вероятностей.

Определение 8 . *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ есть независимые случайные величины, имеющие стандартное нормальное распределение. Распределение случайной величины*

$$\chi^2_n := \xi_1^2 + \ldots + \xi_n^2$$

называется χ^2-распределением с n степенями свободы.

Это распределение имеет плотность распределения вида

$$\rho_{\chi^2_n}(y) = \frac{1}{\Gamma(\frac{n}{2}) \cdot 2^{n/2}} y^{\frac{n}{2}-1} e^{-\frac{1}{2}y} \ , \ y > 0 \ .$$

Таким образом, это частный случай гамма-распределения с $\alpha = \frac{n}{2}$, $\beta = \frac{1}{2}$.

Определение 9 . *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}$ есть независимые случайные величины, имеющие стандартное нормальное распределение. Распределение случайной величины*

$$t_n := \frac{\xi_{n+1}}{\sqrt{\frac{1}{n}(\xi_1^2 + \ldots + \xi_n^2)}}$$

называется распределением Стьюдента с n степенями свободы.

Происхождение такого названия мы объясним позднее. Это распределение имеет плотность распределения вида

$$\rho_{t_n}(y) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}} \ , \ y \in R^1 \ .$$

Заметим, что при $n = 1$ получаем *распределение Коши.*

Определение 10 . *Пусть* $\xi_1, \xi_2, \ldots, \xi_n, \xi_{n+1}, \ldots, \xi_{n+m}$ *есть независимые случайные величины, имеющие стандартное нормальное распределение. Распределение случайной величины*

$$F_{n,m} := \frac{\frac{1}{n}(\xi_1^2 + \ldots + \xi_n)}{\frac{1}{m}(\xi_{n+1}^2 + \ldots + \xi_{n+m}^2)}$$

называется распределением Снедекора или распределением дисперсионного соотношения Фишера с n *и* m *степенями свободы.*

В дальнейшем мы будем называть его *распределением Фишера-Снедекора.* Оно имеет плотность распределения следующего вида

$$\rho_{F_{n,m}}(y) = \frac{(\frac{n}{m})^{n/2} \cdot \Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2}) \cdot \Gamma(\frac{m}{2})} \cdot y^{\frac{n}{2}} \cdot \left(1 + \frac{n}{m} \cdot y\right)^{-\frac{(n+m)}{2}} \ , \ y > 0 \ .$$

Заметим, что при $n = 1$ получаем $t_m^2 \stackrel{d}{=} F_{1,m}$, т. е. квадрат с. в. t_m, имеющей распределение Стьюдента с m степенями свободы, имеет распределение Фишера-Снедекора с 1 и m степенями свободы.

Для всех определенных выше распределений составлены таблицы (см., например, Большев Л. Н., Смирнов Н. И. "Таблицы математической статистики").

4.3 Построение доверительных интервалов для параметров нормального распределения

Всюду далее в этом разделе $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с нормальным распределением, имеющим параметры a и σ^2. Нам потребуется один результат, который мы приведем без доказательства.

Теорема 3 . *Определим*

$$\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k, \quad S_1^2 = \frac{1}{N-1} \sum_{k=1}^{N} (X_k - \bar{X})^2$$

— выборочное среднее и исправленную выборочную дисперсию. Тогда:

1. *$\bar{X}$ и S_1^2 — независимы,*

2. *$\bar{X}$ имеет нормальное распределение,*

3. *$(N-1)\cdot S_1/\sigma^2$ имеет χ^2-распределение с $N-1$ степенями свободы,*

4. *$\frac{(\bar{X}-a)}{S_1}\cdot\sqrt{N}$ имеет распределение Стьюдента с $N-1$ степенями свободы.*

Мы разберем отдельно несколько случаев.

a) $\theta = a$, σ^2 – известно.

Рассмотрим случайную величину

$$Y = \frac{\bar{X} - a}{\sigma/\sqrt{N}} \ . \tag{4.1}$$

В силу теоремы 1 с. в. Y имеет стандартное нормальное распределение. По таблицам для заданного γ найдем константу $C(\gamma)$, для которой

$$P(|Y| < C(\gamma)) = \gamma \ . \tag{4.2}$$

Тогда из соотношений (1) и (2) получаем

$$\bar{X} - \frac{C(\gamma)\cdot\sigma}{\sqrt{N}} < a < \bar{X} + \frac{C(\gamma)\cdot\sigma}{\sqrt{N}}$$

b) $\theta = a$, σ^2 – неизвестно.

Рассмотрим случайную величину

$$t_{N-1} = \frac{\bar{X} - a}{S_1}\cdot\sqrt{N} \ . \tag{4.3}$$

В силу теоремы 1 с. в. t_{N-1} имеет распределение Стьюдента с $N-1$ степенями свободы. По таблицам для заданного γ найдем константу $t_{N-1}(\gamma)$, для которой

$$P(|t_{N-1}| < t_{N-1}(\gamma)) = \gamma \ . \tag{4.4}$$

Тогда из соотношений (3) и (4) получаем

$$\bar{X} - \frac{t_{N-1}(\gamma) \cdot S_1}{\sqrt{N}} < a < \bar{X} + \frac{t_{N-1}(\gamma) \cdot S_1}{\sqrt{N}} \ .$$

c) $\theta = \sigma^2$, a – неизвестно.

Рассмотрим случайную величину

$$\chi^2_{N-1} = \frac{(N-1) \cdot S_1^2}{\sigma^2} \ . \tag{4.5}$$

В силу теоремы 1 с. в. χ^2_{N-1} имеет χ^2-распределение с $N-1$ степенями свободы. По таблицам для заданного γ найдем константы $\mathrm{C}_1(\gamma)$ и $C_2(\gamma)$ для которых

$$P(\chi^2_{N-1} < C_1(\gamma)) = \frac{1-\gamma}{2} \ , P(\chi^2_{N-1} > C_2(\gamma)) = \frac{1-\gamma}{2} \ . \tag{4.6}$$

Тогда из соотношений (5) и (6) получаем

$$\frac{(N-1)S_1^2}{C_2(\gamma)} < \sigma^2 < \frac{(N-1)S_1^2}{C_1(\gamma)} \ .$$

Замечание. Можно показать, что для случаев a) и b) построенные интервалы являются наиболее точными среди всех доверительных интервалов уровня γ, основанных на статистиках Y и t_{N-1}. Для случая c) это не так. Чтобы построить наиболее точный интервал в этой ситуации, нужно решить некоторую оптимизационную задачу. Мы используем соотношения (6), чтобы упростить задачу.

4.4 Построение доверительных интервалов с помощью центральных статистик

Все примеры, рассмотренные выше, имеют одну и ту же особенность. А именно, нам удалось найти такую функцию $T(X, \theta)$ от выборки X и оцениваемого параметра θ, распределение которой не зависит от θ, и это распределение удалось вычислить в явном виде. Ниже мы опишем общую процедуру построения доверительного

интервала, применимую в подобных ситуациях. Пусть случайная величина ξ имеет функцию распределения $F_\xi(y) = F(y, \theta, \theta')$, где θ — неизвестный скалярный параметр, подлежащий оцениванию, а $\theta' = (\theta_1, \ldots, \theta_m) \in \Theta \subset \mathbf{R}^m$ — некоторое число так называемых "мешающих параметров". Мы не интересуемся значением θ', но оно влияет на распределение случайной величины ξ. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с функцией распределения $F(y, \theta, \theta')$.

Определение. Функция $T(X, \theta)$ от выборки X и основного параметра θ называется **центральной статистикой**, если

1. случайная величина $T(X, \theta)$ имеет распределение, не зависящее от параметров θ и θ';
2. $T(X, \theta)$ есть монотонная функция от параметра θ.

Если мы имеем некоторую центральную статистику $T(X, \theta)$, то легко построить доверительный интервал для θ. Для заданного γ найдём две константы C_1 и C_2, для которых

$$P(C_1 < T(X, \theta) < C_2) = \gamma.$$

В силу монотонности по θ неравенство

$$C_1 < T(X, \theta) < C_2$$

эквивалентно неравенству

$$T_1(X, C_1, C_2) < \theta < T_2(X, C_1, C_2) \tag{8}$$

для некоторых статистик $T_1(X, C_1, C_2)$ и $T_2(X, C_1, C_2)$. Формула задаёт некоторый доверительный интервал уровня γ для параметра θ.

Заметим, что этот метод срабатывает, если распределение $T(X, \theta)$ найдено в явном виде. Например, для него составлены таблицы.

Пример. Случайная величина ξ имеет показательное распределение с параметром λ. Пусть $X = (X_1, \ldots, X_N)$ есть повторная

выборка. Необходимо построить доверительный интервал для параметра $\theta = \lambda$. Параметр λ является масштабным, поэтому случайная величина X_k/λ имеет показательное распределение с параметром $\lambda' = 1$. Показательное распределение есть частный случай гамма-распределения. В частности, X_k/λ имеет гамма-распределение с параметрами $\alpha = 1$ и $\beta = 1$. На практических занятиях было показано, что случайная величина

$$T(X, \lambda) = \frac{1}{\lambda}(X_1 + \ldots + X_N)$$

имеет гамма-распределение с параметрами $\alpha = N$ и $\beta = 1/2$. Легко видеть, что $T(X, \lambda)$ — центральная статистика. Удобнее рассмотреть случайную величину

$$T_1(X, \lambda) = \frac{2}{\lambda}(X_1 + \ldots + X_N),$$

так как она имеет гамма-распределение с параметрами $\alpha = N$ и $\beta = 1/2$, а это есть χ^2-распределение с $2N$ степенями свободы. Далее, для заданного γ по таблицам χ^2-распределения находим две константы $C_1(\gamma)$ и $C_2(\gamma)$, для которых

$$P(T_1(X, \lambda) < C_1(\gamma)) = \frac{1-\gamma}{2}, \quad P(T_1(X, \lambda) > C_2(\gamma)) = \frac{1-\gamma}{2}.$$

Тогда событие, состоящее в том, что

$$C_1(\gamma) < T_1(X, \lambda) < C_2(\gamma),$$

имеет вероятность, равную γ. Из последнего неравенства и определения $T_1(X, \lambda)$ получаем

$$\frac{2(X_1 + \ldots + X_N)}{C_2(\gamma)} < \lambda < \frac{2(X_1 + \ldots + X_N)}{C_1(\gamma)}.$$

4.5 Построение доверительных интервалов с помощью асимптотически нормальных оценок

Центральные статистики существуют только для довольно узкого класса статистических моделей. Поэтому чаще строят приближённые доверительные интервалы, используя асимптотическую

нормальность точечных оценок. Пусть $\hat{\theta}_N = \hat{\theta}_N(X)$ есть асимптотически нормальная оценка параметра θ, т.е. случайная величина

$$\frac{\hat{\theta}_N - \theta}{B_N(\theta)}$$

имеет асимптотически стандартное нормальное распределение, где $B_N(\theta) > 0$ есть некоторая нормирующая константа. Тогда для заданного γ по таблицам нормального распределения найдём константу $C(\gamma)$, для которой событие, состоящее в том, что

$$\left|\frac{\hat{\theta}_N - \theta}{B_N(\theta)}\right| < C(\gamma) \tag{9}$$

приближённо (при $N \to \infty$) имеет вероятность, равную γ. Отсюда мы получаем, что

$$\hat{\theta}_N - C(\gamma) \cdot B_N(\theta) < \theta < \hat{\theta}_N + C(\gamma) \cdot B_N(\theta). \tag{10}$$

Но интервал (10), вообще говоря, не является доверительным, так как его границы зависят от оцениваемого параметра θ. Ситуацию можно исправить следующим образом.

a) Если $B_N(\theta) \equiv B_N$ (т.е. нет зависимости от θ), то (10) задаёт настоящий доверительный интервал.

b) Если $B_N(\theta) \leq B_N$, где B_N не зависит от θ, то доверительный интервал вида

$$\hat{\theta}_N - C(\gamma) \cdot B_N(\theta) < \theta < \hat{\theta}_N + C(\gamma) \cdot B_N(\theta),$$

как более широкий, имеет доверительный уровень не менее γ.

c) Иногда удаётся разрешить непосредственно неравенство (9) и получить решение в виде интервала.

Пример. Случайная величина ξ принимает значения 1 и 0 с вероятностями p и $1 - p$ соответственно, т.е. p есть вероятность появления некоторого события A, которую необходимо оценить. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка. Тогда случайная величина $S_N = X_1 + \ldots + X_N$ есть число успехов в схеме Бернулли

и имеет биномиальное распределение с параметрами N и p. В силу интегральной теоремы Муавра-Лапласа случайная величина

$$S_N^* = \frac{S_N - N \cdot p}{\sqrt{Np(1-p)}}$$

имеет асимптотически стандартное нормальное распределение. По заданному γ по таблицам нормального распределения найдём такую константу $C(\gamma) > 0$, для которой

$$P(|S_N^*| < C(\gamma)) \approx 2\Phi_0(C(\gamma)) = \gamma.$$

Для построения доверительного интервала для p нам необходимо решить неравенство

$$\left|\frac{S_N - N \cdot p}{\sqrt{Np(1-p)}}\right| = \left|\frac{\bar{X} - p}{\sqrt{p(1-p)/N}}\right| < C(\gamma). \tag{11}$$

Это квадратичное неравенство можно решить и получить некоторый интервал. Простой (но более широкий!) интервал можно получить, если заметить, что $p(1-p) \leq \frac{1}{4}$. Тогда

$$\bar{X} - \frac{C(\gamma)}{2\sqrt{N}} < p < \bar{X} + \frac{C(\gamma)}{2\sqrt{N}}.$$

Глава 5

Критерии согласия

Формулировки задач и связанных с ними результатов, которые мы рассматривали выше, всегда были связаны с предположением, что мы имеем повторную выборку $X = (X_1, \ldots, X_N)$ из генеральной совокупности с функцией распределения $F(y)$, которая принадлежит некоторому конкретному (как правило, параметрическому) семейству $\mathcal{F}$. В этом параграфе мы рассмотрим статистические процедуры, которые позволяют проверить справедливость этого предположения. Такие процедуры называются **критериями согласия**.

5.1 Общий метод построения критериев согласия

Сформулируем более аккуратно стоящую перед нами задачу. Пусть $\mathcal{F}$ есть некоторое семейство функций распределения. ξ есть случайная величина с неизвестной функцией распределения $F_\xi(y)$. **Гипотезой о виде распределения** называется предположение о том, что $F_\xi \in \mathcal{F}$. Кратко это будет записываться в виде:

$$H: \ F_\xi \in \mathcal{F}.$$

Мы не знаем истинного распределения F_ξ, но на основе повторной выборки $X = (X_1, \ldots, X_N)$ можем построить его оценку, например, взять эмпирическую функцию распределения F_N^*. Выберем в пространстве всех функций распределения некоторую метрику (расстояние) ρ и оценим согласие эмпирических данных с нашей гипотезой по величине расстояния F_N^* от класса $\mathcal{F}$, которое определим по формуле

$$\Delta_N := \inf_{F \in \mathcal{F}} \rho(F_N^*, F).$$

Так как F_N^* построена по выборке, то величина Δ_N является случайной. Предположим, что нам удалось найти её распределение при условии, что гипотеза H верна. Для некоторого малого положительного числа α найдём положительную константу C, для которой

$$P(\Delta_N > C | H) \leq \alpha.$$

Таким образом, если гипотеза H верна, то вероятность того, что Δ_N принимает большие значения ($> C$), мала и можно считать, что такое событие "практически невозможно".

Теперь правило проверки H можно сформулировать следующим образом: если реально полученное значение Δ_N больше C, то мы **отвергаем** H, в противном случае говорим, что H **не противоречит экспериментальным данным**.

Далее мы рассматриваем несколько конкретных примеров применения этого общего подхода.

5.2 Критерий согласия Колмогорова

Мы рассматриваем гипотезу

$$H: \ F_\xi(y) \equiv F_0(y),$$

где $F_0(y)$ — точно заданная непрерывная функция распределения.

А. Н. Колмогоров предложил рассматривать следующее расстояние между функциями распределения:

$$\Delta_N := \sup_y |F_N^*(y) - F_0(y)|.$$

Теорема 4 (А. Н. Колмогоров, 1933.) *Если гипотеза H верна, то при $N \to \infty$*

$$P(D_N = \sqrt{N} \cdot \Delta_N < y) \longrightarrow K(y),$$

где $K(y)$ — некоторая явно вычисляемая функция распределения.

Для функции распределения $K(y)$ составлены таблицы (см. Большев А. Н., Смирнов Н. И., 1972).

Далее по таблицам для заданного α находим $K_\alpha > 0$:

$$P(D_N > K_\alpha) \approx \alpha.$$

Если реально полученное D_N будет больше K_α, то мы отвергаем гипотезу H, в противном случае говорим, что H не противоречит экспериментальным данным.

Главный недостаток этого критерия в том, что он не применим к ситуации, когда семейство $\mathcal{F}$ содержит более одного распределения.

5.3 χ^2-критерий согласия Пирсона

Мы вновь начинаем с простейшей ситуации

$$H: \ F_\xi(y) \equiv F_0(y),$$

где $F_0(y)$ — точно заданная функция распределения. Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка. Рассмотрим разбиение

$$-\infty = a_0 < a_1 < \ldots < a_k < a_{k+1} < \ldots < a_r < a_{r+1} = +\infty$$

вещественной прямой $\mathbf{R}^1$ (множество значений случайной величины ξ) на $r+1$ интервал. Пусть $p_k := F_0(a_{k+1}) - F_0(a_k)$ есть гипотетическая вероятность попадания случайной величины ξ в k-ый интервал, n_k — число элементов выборки, попавших в этот интервал. Для проверки гипотезы естественно сравнить гипотетические вероятности p_k и их оценки n_k/N, то есть частоты. В качестве меры сравнения Карл Пирсон предложил использовать следующую величину:

$$\Delta_N := \sum_{k=0}^{r} \frac{(n_k - N \cdot p_k)^2}{N \cdot p_k} = \sum_{k=0}^{r} \left(\frac{n_k}{N} - p_k \right)^2 / (p_k \cdot N).$$

Теорема 5 (К. Пирсон, 1900.) *Если гипотеза H верна, то при $N \to \infty$ случайная величина ΔN имеет асимптотически χ^2-распределение с $r = (r+1) - 1$ степенями свободы.*

Далее по таблицам для заданного α находим константу $\chi^2_r(\alpha) > 0$:

$$P(\Delta N > \chi^2_r(\alpha)) \approx \alpha.$$

Если реально полученное Δ_N будет больше $\chi^2_r(\alpha)$, то отвергаем H, в противном случае говорим, что H не противоречит экспериментальным данным.

Критерий Пирсона слабее критерия Колмогорова, но зато он применим в более сложной ситуации. Пусть мы рассматриваем гипотезу

$$H: \; F_\xi(y) = F(y, \theta_1, \ldots, \theta_m),$$

где вид функции распределения $F(y, \theta_1, \ldots, \theta_m)$ полностью задан, но параметры $(\theta_1, \ldots, \theta_m) = \theta \in \Theta \subset \mathbf{R}^m$ неизвестны. В этом случае гипотетические вероятности

$$P_k(\theta) = F(a_{k+1}, \theta) - F(a_k, \theta)$$

есть функции от неизвестных параметров и, потому, также неизвестны.

В этом случае предлагается рассмотреть величину

$$\Delta_N := \inf_{\theta\in\Theta} \Delta_N(\theta) = \inf_{\theta\in\Theta} \sum_{k=0}^{r} \frac{\left(n_k - N\cdot p_k(\theta)\right)^2}{N\cdot p_k(\theta)},$$

т.е. сравнивать экспериментальные данные с наиболее подходящим к ним распределением из нашего семейства $\mathcal{F} = \{F(y,\theta), \quad \theta \in \Theta\}$.

Теорема 6 (Р. Фишер, 1922.) *Если гипотеза H верна, то при $N \to \infty$ случайная величина ΔN имеет асимптотически χ^2-распределение с $r - m = (r+1) - 1 - m$ степенями свободы.*

Далее критерий согласия строится также, как и раньше.

Замечание. Можно показать, что наименьшее значение $\Delta_N(\theta)$ достигается при том же значении θ^* параметра $\theta \in \Theta$, при котором достигается наибольшее значение величины

$$L(X,\theta) = \frac{N!}{n_0!\cdot\ldots\cdot n_r!}[p_0(\theta)]^{n_0}\cdot\ldots\cdot[p_r(\theta)]^{n_r},$$

т.е. θ^* есть полиномиальная оценка наибольшего правдоподобия.

5.4 Проверка однородности двух выборок

Во многих прикладных задачах мы сталкиваемся с ситуацией, когда получены две повторные выборки $X = (X_1,\ldots,X_{N_1})$, $Y = (Y_1,\ldots,Y_{N_2})$. Важным является вопрос о том, будут ли они выборками из одной и той же генеральной совокупности. Если это так, то данные можно объединить и обрабатывать как единую выборку большего объёма. Пусть F_1 и F_2 есть функции распределения, отвечающие выборкам X и Y соответственно.

Гипотеза однородности имеет вид:

$$H: \; F_1(y) \equiv F_2(y),$$

причём F_1 и F_2 неизвестны. Существуют различные процедуры для проверки такой гипотезы. Мы рассмотрим только два критерия.

a) Критерий однородности Смирнова.

Пусть $F_{1,N_1}(y)$ и $F_{2,N_2}(y)$ — эмпирические функции распределения, построенные по выборкам X и Y соответственно. Определим величину

$$D_{N_1,N_2} := \frac{\sqrt{N_1 \cdot N_2}}{N_1 + N_2} \cdot \sup_y |F_{1,N_1}(y) - F_{2,N_2}(y)|.$$

Теорема 7 (Н. В. Смирнов, 1939.) *Если верна гипотеза H, то при $N_1, N_2 \to \infty$*

$$P(D_{N_1,N_2} < y) \to K(y), \quad y > 0,$$

где $K(y)$ — функция распределения Колмогорова.

Далее описание процедуры проверки гипотезы H такое же, как и раньше.

b) χ^2-критерий однородности.

Пусть как и ранее выбрано некоторое разбиение

$$-\infty = a_0 < a_1 < \ldots < a_k < a_{k+1} < \ldots < a_r < a_{r+1} = +\infty$$

пространства $\mathbf{R}^1$ (множества значений). Обозначим через n_{k_1} и n_{k_2} число элементов выборок X и Y соответственно, попавших в k-ый интервал. Для проверки гипотезы H естественно сравнить частоты n_{k_1}/N_1 и n_{k_2}/N_2. Сравнение производится с помощью величины

$$\Delta_{N_1,N_2} := \sum_{k=0}^{r} \sum_{j=1}^{2} \frac{(n_{kj} - N_j \cdot n_{k.}/N)^2}{N_j \cdot N_{k.}} = N \left(\sum_{k=0}^{r} \sum_{j=1}^{2} \frac{n_{kj}^2}{N_j \cdot N_{k.}} - 1 \right),$$

где

$$n_{k.} = n_{k_1} + n_{k_2}, \quad N = N_1 + N_2.$$

Теорема 8 *Если верна гипотеза H, то при $N_1, N_2 \to \infty$ случайная величина Δ_{N_1,N_2} имеет асимптотически χ^2-распределение с $r = [(r+1) - 1](2-1)$ степенями свободы.*

Далее описание процедуры проверки H, аналогично приведённым выше.

Заметим, что в отличие от критерия Смирнова, этот критерий легко обобщается на случай $s \geq 2$ выборок.

5.5 Проверка гипотезы о независимости

Пусть $\xi = (\xi_1, \xi_2)$ — двумерный случайный вектор с функцией распределения $F(z_1, z_2)$, $F_1(z_1)$ — функция распределения для ξ_1, $F_2(z_2)$ — функция распределения для ξ_2. **Гипотеза о независимости** имеет вид:

$$H: \ F(z_1, z_2) = F_1(z_1) \cdot F_2(z_2).$$

Рассмотрим два разбиения

$$-\infty = a_0 < a_1 < \ldots < a_k < a_{k+1} < \ldots < a_r < a_{r+1} = +\infty,$$
$$-\infty = b_0 < b_1 < \ldots < b_j < b_{j+1} < \ldots < b_s < b_{s+1} = +\infty.$$

Пусть мы имеем повторную выборку $(X_1, Y_1), \ldots, (X_N, Y_N)$ измерений случайного вектора (ξ_1, ξ_2). Обозначим через n_{kj} число элементов выборки, попавших в прямоугольник $[a_k, a_{k+1}) \times [b_j, b_{j+1})$, $n_{k.} = \sum\limits_j n_{kj}$, $n_{.j} = \sum\limits_k n_{kj}$, $N = \sum\limits_k n_{k.} = \sum\limits_j n_{.j}$. Если верна гипотеза H, то вероятность попадания в $[a_k, a_{k+1}) \times [b_j, b_{j+1})$ равна произведению вероятностей попадания в интервалы $[a_k, a_{k+1})$ и $[b_j, b_{j+1})$. Поэтому для проверки гипотезы H предлагается использовать величину

$$\Delta_N := N \cdot \sum_{k=0}^{r} \sum_{j=0}^{s} \frac{(n_{kj} - n_{k.} \cdot n_{.j}/N)^2}{n_{k.} \cdot n_{.j}} = N \left(\sum_{k=0}^{r} \sum_{j=0}^{s} \frac{n_{kj}^2}{n_{k.} \cdot n_{.j}} - 1 \right).$$

Теорема 9 *Если верна гипотеза H, то при $N \to \infty$ случайная величина Δ_N имеет асимптотически chi^2-распределение с $r \cdot s$ степенями свободы.*

Далее описание процедуры проверки аналогично прежнему.

Обычно экспериментальные данные представляют в виде следующей таблицы:

$Y \quad X$	$a_0 \div a_1$	$\ldots$	$a_k \div a_{k+1}$	$\ldots$	$a_r \div a_{r+1}$
$b_0 \div b_1$	n_{00}	$\ldots$	n_{k0}	$\ldots$	n_{r0}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$b_j \div b_{j+1}$	n_{0j}	$\ldots$	n_{kj}	$\ldots$	n_{rj}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$b_s \div b_{s+1}$	n_{0s}	$\ldots$	n_{ks}	$\ldots$	n_{rs}

называемой **таблицей сопряжённости**. Поэтому соответствующий критерий называется **анализом таблиц сопряжённости**.

Глава 6

Проверка статистических гипотез

В этом разделе будут изложены основные понятия теории проверки статистических гипотез.

6.1 Что такое статистическая гипотеза?

Математической моделью статистического эксперимента является статистическая структура $(\Omega, \mathcal{A}, \mathcal{P})$. Пытаясь уточнить нашу модель, мы делаем те или иные предположения о неизвестном распределении вероятностей $P \in \mathcal{P}$.

Примеры. 1. В эксперименте изучается случайная величина ξ с неизвестной функцией распределения F_ξ. Мы предполагаем, что $F_\xi \in \mathcal{F}$, где $\mathcal{F}$ — некоторое заданное семейство функций распределения, например, нормальные функции распределения.

2. Известно, что случайная величина ξ имеет нормальное распределение. Если ξ является ошибкой измерения некоторого прибора, то для хорошо отлаженного прибора естественно предположить, что $M(\xi) = a = 0$.

3. В некотором эксперименте изучаются две числовые характеристики ξ_1 и ξ_2, которые являются случайными величинами. Иногда можно предположить, что ξ_1 и ξ_2 независимы.

Все рассмотренные выше примеры имеют одну и ту же особенность, мы делаем то или иное предположение о неизвестном распределении вероятностей. Таким образом, неформально под **статистической гипотезой** понимают любое предположение о распределении вероятностей в рассматриваемом эксперименте. С формальной точки зрения, если мы что-то предположили об этом распределении, то этим выделили некоторый подкласс $\mathcal{P}_0$ в классе всех априорных возможных распределений $\mathcal{P}$.

Определение. Статистической гипотезой называется подкласс
$\mathcal{P}_0 \subset \mathcal{P}$. Условно это записывается в виде

$$H: \quad P \in \mathcal{P}_0.$$

Статистические гипотезы будут обозначаться H, H_0, $H_1, \ldots$.

6.1.1 Основные понятия теории проверки гипотез

Пусть мы имеем статистическую структуру $(\Omega, \mathcal{A}, \mathcal{P})$ и сформулировали некоторую статистическую гипотезу $H: P \in \mathcal{P}_0$. Если $\mathcal{P}_0$ содержит только один элемент, т.е. мы полностью фиксируем распределение, то гипотеза H называется **простой**. В противном случае гипотеза называется **сложной**. Если семейство $\mathcal{P}$ является параметрическим, т.е. $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, а гипотеза H имеет вид $\theta \in \Theta_0 \subset \Theta$, то она называется **параметрической**.

Обычно формулируют несколько гипотез. Одну из них выделяют в качестве основной и называют ее **нулевой**. Как правило, ее обозначают через H_0. Остальные гипотезы называют **альтернативами** и обозначают H_1, $H_2, \ldots$.

Всюду далее мы рассматриваем случай, когда нулевая гипотеза H_0 проверяется против одной альтернативы H_1. Для простоты обозначений мы предполагаем, что эти гипотезы являются параметрическими и записываем их в виде:

$$\begin{aligned} H_0: &\quad \theta \in \Theta_0, \\ H_1: &\quad \theta \in \Theta_1, \end{aligned}$$

где Θ_0, Θ_1 есть подмножества основного пространства параметров Θ и
$\Theta_0 \cap \Theta_1 = \emptyset$.

Основная задача состоит в том, чтобы на основе экспериментальных данных выбрать одну из предложенных нам гипотез H_0 и H_1. Это делается с помощью так называемых **критериев** или **решающих правил**. В качестве экспериментальных данных мы используем, как и раньше, повторную выборку $X = (X_1, \ldots, X_N)$.

Обозначим через $\mathcal{X}$ выборочное пространство, т.е. множество возможных значений x выборки X.

Определение. Статистическим критерием (тестом, решающим правилом) для проверки гипотезы H_0 против альтернативы H_1 называется произвольное отображение $\varphi : \mathcal{X} \to \{0, 1\}$. Если $\varphi(x) = 0$, то принимаем гипотезу H_0. Если $\varphi(x) = 1$, то принимаем гипотезу H_1.

Обозначим через K подмножество в $\mathcal{X}$, где мы принимаем H_1, т.е.

$$K = \{x \in \mathcal{X} : \quad \varphi(x) = 1\}.$$

Множество K называется **критической зоной** теста φ. Задание критической зоны K однозначно определяет тест φ. Действительно, $\varphi(x) = 1$ тогда и только тогда, когда $x \in K$. Очень часто критическая зона теста задается по правилу

$$K = \{x \in \mathcal{X} : \quad T(x) > c\},$$

или

$$K = \{x \in \mathcal{X} : \quad T(x) < c\},$$

где $T(x)$ — некоторая статистика, называемая **статистикой критерия** φ, а c — некоторая константа, называемая **критической константой**.

Иногда возникает такая ситуация, когда для полученной выборки x мы не можем однозначно принять решение в пользу гипотезы H_0 или H_1. В этом случае приписывают некоторую вероятность тому, что верна H_0, и дополнительную вероятность тому, что верна H_1. Такой вариант принятия решения фиксируется в следующем понятии.

Определение. Отображение $\varphi : \mathcal{X} \to [0, 1]$ называется **рандомизированным** критерием (тестом, решающим правилом). Значение $\varphi(x)$ интерпретируется как **вероятность принять гипотезу** H_1, если получена выборка x. Если φ принимает только значения 0 и 1, то критерий называется **нерандомизированным**.

Обычно необходимость в рандомизации возникает на границе критической области.

Для проверки гипотезы H_0 против альтернативы H_1 можно придумать много различных критериев. Хотелось бы выбрать среди них в определенном смысле наилучший. Интуитивно мы стремимся построить такое правило, при котором ошибки возникают редко. В нашей задаче есть два типа ошибок. **Ошибки первого рода:** мы принимаем гипотезу H_1, когда верна H_0. **Ошибки второго рода:** мы принимаем H_0, когда верна H_1. Так как выборка x получается в результате случайного эксперимента, но мы не знаем заранее, к какому решению мы придем, применяя правило φ. Поэтому при оценке качества теста φ естественно исходить из величины **вероятностей ошибок** первого и второго рода. Если θ есть истинное значение неизвестного параметра, то

$$\alpha(\theta) := P_\theta(X \in K), \quad \theta \in \Theta_0, \tag{1}$$

определяет вероятность ошибки первого рода, а

$$1 - \beta(\theta) := P_\theta(X \notin K), \quad \theta \in \Theta_1, \tag{2}$$

определяет вероятность ошибки второго рода. Нам бы хотелось минимизировать величины $\alpha(\theta)$ для всех $\theta \in \Theta_0$ и $1 - \beta(\theta)$ для всех $\theta \in \Theta_1$. Но из выражений (1) и (2) мы видим, что для уменьшения $\alpha(\theta)$ нужно сужать критическую область K, а для уменьшения $1-\beta(\theta)$ нужно, наоборот, расширять критическую область K. Т.е. мы приходим к двум "конфликтующим" критериям качества (вспомните построение доверительных интервалов!). Улучшая один, мы ухудшаем другой, и наоборот. В качестве H_0 обычно выбирают более важную для нас гипотезу, которую мы не хотим отвергать (альтернативно, не хотим принимать H_1) без достаточных на то оснований. В силу этого мы выбираем некоторое малое α и рассматриваем только такие критерии φ, для которых

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha. \tag{3}$$

Такие критерии называются критериями **уровня значимости** α. Точное значение верхней грани в (3) называется **размером**

критерия φ. Затем среди всех критериев уровня α мы ищем такой, для которого величина $1 - \beta(\theta)$, $\theta \in \Theta_1$ будет наименьшей. Эквивалентно, можно максимизировать величину $\beta(\theta)$. Функция

$$\beta(\theta := P_\theta(X \in K)), \quad \theta \in \Theta_1, \tag{4}$$

называется **функцией мощности**. Она показывает, какова вероятность принять гипотезу H_1, когда она верна, для различных распределений, входящих в H_1. Хотелось бы найти такой критерий φ_0, у которого функция мощности $\beta(\theta)$ была наибольшей среди всех критериев φ уровня α при любых $\theta \in \Theta_1$.

Определение. Критерий φ_0 уровня α называется **равномерно наиболее мощным критерием уровня** α, если для любого другого критерия φ уровня α имеет место

$$\beta_{\varphi_0}(\theta) \geq \beta_\varphi(\theta), \quad \forall \theta \in \Theta_1. \tag{5}$$

Заметим, что нам нужно максимизировать мощность при всех значениях $\theta \in \Theta_1$. Это удается сделать довольно редко.

6.1.2 Проверка простой гипотезы против простой альтернативы

Как мы отмечали выше, равномерно наиболее мощные тесты существуют не всегда. Но если мы проверяем простую гипотезу H_0 против простой альтернативы H_1, т.е.

$$\begin{aligned} H_0 : &\quad \theta = \Theta_0, \\ H_1 : &\quad \theta = \Theta_1, \end{aligned}$$

то наилучший (оптимальный) критерий существует. Чтобы понять, как он должен выглядеть, рассмотрим дискретную случайную величину ξ. Обозначим через $P_0(x)$ и $P_1(x)$ вероятности события $(X = x)$ в случае выполнения гипотез H_0 и H_1 соответственно. Для построения критерия нам необходимо описать критическую зону K. Следуя принципу наибольшего правдоподобия естественно помещать в K те точки x, для которых больше отношение $P_1(x)/P_0(x)$ (т.е. те, которые больше говорят в пользу H_1).

Лемма (Е. Нейман, Е. Пирсон, 1933). *Пусть случайная величина ξ имеет дискретное распределение и проверяется простая гипотеза H_0 против простой альтернативы H_1. Тогда среди всех тестов φ уровня α существует наиболее мощный тест φ_0, и он имеет следующий вид:*

$$\varphi_0(x) = \begin{cases} 1, & \text{если } P_1(x) > c \cdot P_0(x), \\ 0, & \text{если } P_1(x) < c \cdot P_0(x), \\ \gamma, & \text{если } P_1(x) = c \cdot P_0(x), \end{cases}$$

где константы $c > 0$ и $0 \le \gamma \le 1$ определяются из условия

$$P_0(X \in K) = \alpha.$$

Доказательство. Заметим, что для любого критерия φ его уровень и мощность вычисляются по формулам

$$\alpha_\varphi = M_0\varphi(X), \quad \beta_\varphi = M_1\varphi(X) = \sum_x \varphi(x)P_1(x).$$

Пусть критерий φ (вообще говоря, рандомизированный) имеет уровень α, т.е. $\alpha_\varphi \le \alpha$. Предположим, что $0 < \alpha < 1$, и определим множества

$$S^+ = \{x : \varphi_0(x) - \varphi(x) > 0\},$$
$$S^- = \{x : \varphi_0(x) - \varphi(x) < 0\}.$$

Если $x \in S^+$, то

$$\varphi_0(x) > 0 \quad \text{и} \quad P_1(x) \ge c \cdot P_0(x).$$

Если $x \in S^-$, то

$$P_1(x) \le c \cdot P_0(x).$$

Используя эти соотношения, получим

$$\begin{aligned} &\textstyle\sum_x [\varphi_0(x) - \varphi(x)]P_1(x) - c \cdot \sum_x [\varphi_0(x) - \varphi(x)]P_0(x) = \\ &= \textstyle\sum_x [\varphi_0(x) - \varphi(x)][P_1(x) - c \cdot P_0(x)] = \\ &= \textstyle\sum_{x \in S^+ \cup S^-} [\varphi_0(x) - \varphi(x)][P_1(x) - c \cdot P_0(x)] = \\ &= \textstyle\sum_{x \in S^+} [\varphi_0(x) - \varphi(x)][P_1(x) - c \cdot P_0(x)] + \\ &+ \textstyle\sum_{x \in S^-} [\varphi_0(x) - \varphi(x)][P_1(x) - c \cdot P_0(x)]. \end{aligned}$$

Отсюда следует, что

$$\beta_{\varphi_0} - \beta_\varphi = \sum_x \varphi_0(x) \cdot P_1(x) - \sum_x \varphi(x) \cdot P_1(x) =$$
$$= \sum_x [\varphi_0(x) - \varphi(x)] \cdot P_1(x) \geq c \cdot \sum_x [\varphi_0(x) - \varphi(x)] \cdot P_0(x) =$$
$$= c \cdot (\alpha - \alpha_\varphi) \geq 0.$$

Замечание. Доказанный выше результат справедлив и для абсолютно непрерывных распределений. В этом случае в формулировке леммы вероятности нужно заменить на плотности, а при доказательстве заменить суммы на интегралы.

Пример. Случайная величина ξ имеет нормальное распределение с параметрами (a, σ^2). Предполагая, что дисперсия $\sigma^2 = \sigma_0^2$ известна, необходимо проверить гипотезу

$$H_0: \quad a = a_0$$

против альтернативы

$$H_1: \quad a = a_1.$$

Пусть для определенности $a_1 > a_0$. Если $X = (X_1, \ldots, X_N)$ есть повторная выборка, то плотность распределения случайного вектора X в точке $x = (x_1, \ldots, x_N)$ имеет вид

$$\rho(x) = (2\pi\sigma_0^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{j=1}^N (x_j - a)^2\right\}.$$

По лемме Неймана-Пирсона нам нужно рассмотреть отношение

$$\frac{\rho_1(x)}{\rho_0(x)} = \frac{\exp\left\{-\frac{1}{2\sigma_0^2} \sum\limits_{j=1}^N (x_j - a_1)^2\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2} \sum\limits_{j=1}^N (x_j - a_0)^2\right\}} = \frac{\exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum\limits_{j=1}^N x_j^2 - 2a_1 \sum\limits_{j=1}^N x_j + N a_1^2\right]\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum\limits_{j=1}^N x_j^2 - 2a_0 \sum\limits_{j=1}^N x_j + N a_0^2\right]\right\}} =$$
$$= \exp\left\{-\tfrac{1}{2\sigma_0^2}\left[-2N(a_1 - a_0) \cdot \bar{x} + N(a_1^2 - a_0^2)\right]\right\}.$$

Критическая область оптимального критерия φ_0 имеет вид

$$\frac{\rho_1(x)}{\rho_0(x)} = \exp\left\{\frac{N}{\sigma_0^2}(a_1 - a_0) \cdot \bar{x}\right\} \cdot \exp\left\{-\frac{N}{2\sigma_0^2}(a_1^2 - a_0^2)\right\} > c.$$

Произведем последовательно несколько эквивалентных преобразований, которые дают различные варианты записи одной и той же области K.

$$\rho_1(x)/\rho_0(x) > c \quad \Longleftrightarrow \quad \tfrac{N}{\sigma_0^2}(a_1 - a_0)\cdot \bar{x} \quad \Longleftrightarrow$$
$$\bar{x} > c_2 \quad \Longleftrightarrow \quad \tfrac{\bar{x}-a_0}{\sigma_0}\cdot\sqrt{N} > c_3.$$

Случайная величина

$$y = \frac{\bar{x} - a_0}{\sigma_0}\cdot\sqrt{N}$$

в случае гипотезы H_0 имеет стандартное нормальное распределение. Таким образом, константу c_3 можно найти по таблицам нормального распределения из условия

$$P_0(y > c_3) = 1 - \Phi(c_3) = \frac{1}{2} - \Phi_0(c_3) = \alpha.$$

Окончательно оптимальный критерий φ имеет вид

$$\varphi_0(x) = \begin{cases} 1, & \text{если } \bar{x} > a_0 + \frac{c_3\cdot\sigma_0}{\sqrt{N}} = c_4, \\ 0, & \text{если } \bar{x} \le c_4. \end{cases}$$

Полезно нарисовать следующую картинку, которая наглядно иллюстрирует полученный результат.

6.1.3 Проверка сложных гипотез. Критерий отношения правдоподобия

Конечно, случай, когда мы проверяем простую гипотезу H_0 против простой альтернативы H_1, является исключительным и редко встречается в реальных задачах. Обычно нужно проверять сложные гипотезы. Но лемма Неймана-Пирсона подсказывает метод построения критериев для сложных гипотез. Этот метод был предложен Нейманом и Пирсоном в 1928 году.

Пусть случайная величина ξ имеет функцию распределения $F(y,\theta)$, где $\theta \in \Theta$ — неизвестный параметр. Пусть Θ_0 и Θ_1 — подмножества Θ, для которых $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$. Проверяется гипотеза

$$H_0: \quad \theta \in \Theta_0$$

против альтернативы

$$H_1: \quad \theta \in \Theta_1.$$

Для проверки таких гипотез предлагается использовать статистику

$$\lambda(x) = \frac{\sup\limits_{\theta\in\Theta_1} \mathcal{L}(x,\theta)}{\sup\limits_{\theta\in\Theta_0} \mathcal{L}(x,\theta)}$$

или, эквивалентно,

$$\lambda_1(x) = \frac{\sup\limits_{\theta\in\Theta_1} \mathcal{L}(x,\theta)}{\sup\limits_{\theta\in\Theta_0} \mathcal{L}(x,\theta)} = \max(\lambda(x), 1)$$

где $\mathcal{L}(x,\theta)$ есть функция правдоподобия, т.е. распределение вероятности (плотность) повторной выборки $X = (X_1, \ldots, X_N)$. Это естественное обобщение той статистики, что используется в лемме Неймана-Пирсона. Если удается найти распределение этих статистик, соответствующий критерий, называемый **критерием отношения правдоподобия**, задается с помощью критической области следующего вида

$$K = \{x: \quad \lambda_1(x) > c\},$$

где критическая константа c находится из условия

$$P_\theta(\lambda_1(x) > c) \le \alpha, \quad \theta \in \Theta.$$

Пример. Случайная величина ξ имеет нормальное распределение с параметрами (a, σ^2), причем σ^2 неизвестно. Проверяется гипотеза

$$H_0: \quad a = a_0$$

против альтернативы

$$H_1: \quad a \neq a_0.$$

Заметим, что обе гипотезы H_0 и H_1 являются сложными, т.к.

$$\Theta_0 = \{(a, \sigma^2) : a = a_0, \sigma^2 > 0\}, \quad \Theta_1 = \{(a, \sigma^2) : a \neq a_0, \sigma^2 > 0\}.$$

Пусть $X = (X_1, \ldots, X_N)$ есть повторная выборка. Тогда функция правдоподобия (плотности распределения выборки X) имеет вид

$$\mathcal{L}(x, a, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(x_j - a)^2\right\}.$$

При оценке параметров по методу наибольшего правдоподобия мы уже нашли, что

$$\sup_{a,\sigma^2} \mathcal{L}(x, a, \sigma^2) = \mathcal{L}(x, \bar{x}, S^2),$$

где $\bar{x}$ и S^2 есть выборочное среднее и выборочная дисперсия соответственно.

Аналогично можно показать, что

$$\sup_{a=a_0,\sigma^2} \mathcal{L}(x, a_0, \sigma^2) = \mathcal{L}(x, a_), S_0^2),$$

где

$$S^2 = \frac{1}{N}\sum_j (x_j - a_0)^2 = S^2 + (\bar{x} - a_0)^2.$$

В этом случае

$$\lambda_1(x) = \frac{(2\pi S^2)^{-\frac{N}{2}} \cdot e^{-\frac{N}{2}}}{(2\pi[S^2 + (\bar{x} - a_0)^2])^{-\frac{N}{2}} \exp\left\{-\frac{N}{2}\right\}} = \left(\left|\frac{\bar{x} - a_0}{S}\right|^2 + 1\right)^{\frac{N}{2}}.$$

Критическая область имеет вид

$$\lambda_1(x) > c$$

или, эквивалентно,

$$|t_{N-1}| = \left|\frac{\bar{x} - a_0}{C} \cdot \sqrt{N}\right| > C_1.$$

Как отмечалось ранее (при построении доверительных интервалов), если верна гипотеза H_0, то случайная величина t_{N-1} имеет распределение Стьюдента с $N-1$ степенями свободы. Тогда для заданного α константа c_1 находится по таблицам распределения Стьюдента из соотношения

$$P_0(|t_{N-1}| > c_1) = \alpha.$$

Задача. Дать описание критериев отношения правдоподобия для проверки следующих гипотез.

1) $X = (X_1, \ldots, X_N)$ — повторная выборка из генеральной совокупности с нормальным распределением с параметрами (a, σ^2). Проверяется гипотеза

$$H_0: \quad \sigma^2 = \sigma_0^2$$

против альтернативы

$$H_1: \quad \sigma^2 \neq \sigma_0^2.$$

2) $X = (X_1, \ldots, X_N)$ и $Y = (Y_1, \ldots, Y_N)$ — независимые повторные выборки из генеральной совокупности с нормальными распределениями с параметрами (a_1, σ^2) и (a_2, σ^2) соответственно. Проверяется гипотеза

$$H_0: \quad a_1 = a_2$$

против альтернативы

$$H_1: \quad a_1 \neq a_2.$$

6.1.4 Проверка сложных гипотез для распределений с монотонным отношением правдоподобия

В этом случае рассматривается семейство распределений с одним специальным свойством, для которого существуют равномерно наиболее мощные тесты для некоторых сложных гипотез.

Пусть случайная величина ξ имеет плотность распределения $\rho(y, \theta)$, где $\theta \in \Theta \subset \mathbf{R}^1$ есть скалярный параметр. Пусть, далее,

$X = (X_1, \ldots, X_N)$ есть повторная выборка из генеральной совокупности с таким распределением. Предположим, что функция правдоподобия

$$\mathcal{L}(x, \theta) = \rho(x_1, \theta) \cdot \ldots \cdot \rho(x_N, \theta)$$

допускает представление

$$\mathcal{L}(x, \theta) = g(T(x), \theta) \cdot h(x), \tag{6}$$

где $g(t, \theta)$ — некоторая функция от двух скалярных аргументов. Статистика $T(x)$ называется **достаточной** статистикой для параметра θ.

Определение. Семейство распределений $\mathcal{P} = \{\rho(y, \theta), \theta \in \Theta \subset \mathbf{R}^1\}$ имеет **монотонное отношение правдоподобия**, если для любых фиксированных $\theta_0 < \theta_1$ из Θ функция $g(t, \theta_1)/g(t, \theta_0)$ монотонно возрастает (убывает) по t.

Проверяется гипотеза

$$H_0 : \quad \theta = \theta_0$$

против альтернативы

$$H_1 : \quad \theta \geq \theta_1,$$

где $\theta_1 > \theta_0$. Выберем некоторое фиксированное $\theta' \geq \theta_1$ и построим критерий Неймана–Пирсона для проверки простой гипотезы

$$H_0 : \quad \theta = \theta_0$$

против простой альтернативы

$$H_0' : \quad \theta = \theta'.$$

Критическая зона этого критерия выделяется условием

$$\mathcal{L}(x, \theta')/\mathcal{L}(x, \theta_0) = g(T(x), \theta')/g(T(x), \theta_0) > c.$$

В силу монотонности отношения правдоподобия это эквивалентно условию

$$T(x) > c_1. \tag{7}$$

Константа c_1 находится из условия

$$P_{\theta_0}(T(x) > c_1) = \alpha,$$

т.е. полностью определяется значением θ_0 и не зависит от частного выбора $\theta' \geq \theta_1$.

В силу леммы Неймана-Пирсона построенный критерий является наиболее мощным для проверки H_0 против H_1' для любого $\theta' \geq \theta_1$, т.е. является равномерно наиболее мощным критерием уровня α для проверки H_0 против сложной альтернативы H_1.

Пример. ξ имеет нормальное распределение с параметрами (a, σ^2). Пусть $\sigma^2 = \sigma_0^2$ известно и проверяется гипотеза

$$H_0: \quad a = a_0$$

против альтернативы

$$H_1: \quad a \geq a_1,$$

где $a_1 > a_0$. В этой задаче функция правдоподобия равна

$$\begin{aligned}
&\mathcal{L}(x, a) = (2\pi\sigma_0^2)^{-\frac{N}{2}} \exp\left\{-\tfrac{1}{2\sigma_0^2} \sum_{j=1}^{N} (x_j - a)^2\right\} = \\
&= \exp\left\{\tfrac{1}{\sigma_0^2} \cdot a \cdot \sum_{j=1}^{N} x_j - \tfrac{N \cdot a^2}{2\sigma_0^2}\right\} \cdot (2\pi\sigma_0^2)^{-\frac{N}{2}} \exp\left\{-\tfrac{1}{2\sigma_0^2} \sum_{j=1}^{N} x_j^2\right\} = \\
&= g(T(x), a) \cdot h(x),
\end{aligned}$$

где $T(x) = \sum\limits_j x_j$, т.е. имеет вид (6). Это семейство с монотонным отношением правдоподобия, т.к. $g(t, a')/g(t, a_0)$ строго возрастает по t при любых $a' > a_0$. В силу сказанного выше существует равномерно наиболее мощный тест для проверки H_0 против альтернативы H_1, критическая область которого выделяется условием

$$T(x) = \sum_{j=1}^{N} x_j > c.$$

После эквивалентных преобразований ее можно записать в виде

$$\frac{\bar{x} - a_0}{\sigma_0} \cdot \sqrt{R} > c_1.$$

Критическая константа c_1 находится из условия, что

$$P_0(\frac{\bar{x} - a_0}{\sigma_0} \cdot \sqrt{N} > c_1) = 1 - \Phi(c_1) = \frac{1}{2} - \Phi_0(c_1) = \alpha.$$

Здесь мы использовали тот факт, что при гипотезе H_0 статистика

$$Y = \frac{\bar{x} - a_0}{\sigma_0} \cdot \sqrt{N}$$

имеет стандартное нормальное распределение.

Замечание. Аналогичные результаты справедливы и для дискретных распределений, если во всех формулировках плотности заменить на вероятности.