МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА
Механико математический факультет

Курс лекций по математической статистике

Лектор — Александр Владимирович Прохоров

III курс, 5 семестр, поток математиков

Москва, 2019 г.

Содержание

Предисловие

Ну что Вам рассказать про Сахалин?... Это незавершённый курс, который, вообще говоря, можно доTEXать. Но у нас нет на это времени. Если у кого-то из читателей найдутся силы и желание это делать, это будет просто замечательно. Исходные тексты будут выданы всякому, кто захочет набрать остатки.

Дополнение к предисловию

Текст существенно дополнен С. Л. Кузнецовым и А. В. Харитоновым. Работа постепенно продвигается к завершению. Нам ещё много нужно сделать, но многое мы уже сделали. Спасибо Тиме Архангельскому и Владимиру Гаврилину за обнаружение и исправление лажи.

С.К., А.Х.

Дополнение к предисловию 2

Документ переработан В. В. Харламовым под курс лекций 2019 года. **Важно**, что дальнейший текст не является в полной мере конспектом, но в то же время есть хорошее приближение материала лекций. Некоторые доказательства и примеры могут отличаться от данных в аудитории.

В конце читатель может найти список экзаменационных вопросов 2019 года и примерное соответствие их с материалом. Так как DMVN не подаёт признаков жизни, то огромная просьба писать на мою почту при обнаружении ошибок. Самая актуальная версия конспекта находится на Github. В 2020 году Bitbucket прекращает поддержку репозиториев с системой контроля версий Mercurial, поэтому я скопировал репозиторий на Github.

От себя:

> *Математическая статистика – важный раздел современной математики. С одной стороны, хорошее понимание материала даёт возможность применять всевозможные статистические пакеты и понимать что происходит внутри. С другой стороны, математическая статистика красива сама по себе: в данном курсе надо будет вспомнить немного комбинаторики, действительного анализа, линейной алгебры. Будет довольно много теории вероятностей и математического анализа. Наш курс математической статистики обозревает основные статистические задачи и вводит в их суть. Уверен, что хорошее понимание курса может быть полезно во многих профессиях, куда может пойти математик.*

В.Х.

Последняя компиляция: `14 июня 2020 г. г.`
Обновления документа на сайтах `http://dmvn.mexmat.net`,
`http://dmvn.mexmat.ru`.
LaTeX исходники
`https://bitbucket.org/dmvn/mexmat.lectures`
Об опечатках и неточностях пишите на `dmvn@mccme.ru`.

1 Краткий обзор курса

Математическая статистика — это наука, посвященная разработке *наиболее верного* вывода, основанного на *неизвестных* закономерностях. В данном курсе рассматривается *параметрическая статистика*, где мы *пытаемся по имеющимся данным оценить значения параметров*. Далее в этом разделе приведены примеры типичных задач параметрической статистики, в которые мы углубимся в последующих разделах[1].

Напомним некоторые основные определения из курса теории вероятностей.

Определение. *Пространством элементарных событий* называется множество исходов некоторого эксперимента. *Элементарным событием* называется любой элемент пространства элементарных событий. *Событием* называется любое подмножество пространства элементарных событий. *Экспериментом* называется функция, принимающая значение на пространстве элементарных событий.

Определение. *Генеральной совокупностью* называется достаточно большое, быть может, бесконечное подмножество элементарных событий.

Определение. *Случайной величиной* называют функцию от элементарного события.

1.1 Модель конечного случайного выбора

Рассмотрим модель «Выбор без возвращения». Пусть N — общее число элементов генеральной совокупности, M — число отмеченных (каким-то свойством) элементов, n — размер выборки, т. е. число элементов, выбранных из генеральной совокупности, m — число отмеченных элементов в выборке.

Вероятностная задача рассматривает случай, когда n, M и N заданы, а $m \in \{0, \ldots, \min(n, M)\}$. Тогда вероятность того, что среди выборки размера n окажется ровно m отмеченных элементов, может быть вычислена по известной формуле

$$Q_{n,m}(N, M) = \frac{\mathbf{C}_M^m \mathbf{C}_{N-M}^{n-m}}{\mathbf{C}_N^n}.$$

Статистическая задача ставится несколько иначе. Например:

а) Допустим, что n, m, N известны, а M — неизвестно. Требуется оценить M. Это в некотором смысле задача, обратная вероятностной. Решить ее не так-то просто. Простейшее (но довольно грубое) приближение для M можно найти, например, из соотношений

$$\frac{M}{N} \approx \frac{m}{n}, \quad M \approx \frac{m}{n} N.$$

Для того, чтобы найти более точные оценки, нужны специальные методы, которыми и занимается математическая статистика.

б) Пусть заданы n, m и M, а N неизвестно. Требуется оценить N. Пример такой задачи — оценка числа рыб в водоеме: производится выборка размера M, помечаются все рыбы из этой выборки, а спустя некоторое время производится еще одна выборка размера n и подсчитывается число помеченных рыб m из этой выборки. По этим данным требуется оценить число рыб в водоеме. Для решения этой задачи рассматривается вероятность $\widetilde{Q}_{n,m}(N)$ как функция переменной N. Оказывается, что функция $\widetilde{Q}$ сначала возрастает, а затем убывает. В качестве оценки искомого значения N выбирается такое целое N_*, для которого $\widetilde{Q}_{n,m}(N)$ максимально. Можно показать, что

$$N_* = \left[M \frac{n}{m}\right] \leqslant M \frac{n}{m}.$$

Рассмотрим следующий эксперимент: два раза независимо друг от друга бросается монетка. Можно рассматривать две модели этого эксперимента:

1) 4 исхода: выпали последовательно орел–орел, орел–решка, решка–орел, решка–решка. Каждому исходу приписывается вероятность $\frac{1}{4}$.

2) 3 исхода: 2 орла, 2 решки, 1 орел и 1 решка; каждому исходу приписывается вероятность $\frac{1}{3}$.

Практика показывает, что первая модель более соответствует действительности, чем вторая: при большом числе испытаний каждый из четырех исходов появляется с частотой, близкой к $\frac{1}{4}$, в то время как во второй модели последний исход появляется с частотой, близкой к $\frac{1}{2}$, а первые два — с частотой $\frac{1}{4}$, что плохо соответствует приписанным вероятностям.

В некотором смысле задача математической статистики обратна задаче теории вероятностей. В теории вероятностей в каждой конкретной ситуации вероятность считается полностью определенной и основной задачей теории вероятностей является разработка методов нахождения вероятностей различных сложных событий (исходя из известных вероятностей более простых событий) для данной вероятностной модели. В математической

[1]Этот раздел явно не включён в программу, однако в нём даётся большинство основных определений, а также разбираются первые примеры главных разделов статистики — *примеч. В. Х.*

статистике рассматривается статистическая модель, которая описывает такие ситуации, когда в вероятностной модели изучаемого эксперимента имеется та или иная неопределенность в задании вероятности, и задача математической статистики состоит в том, чтобы уменьшить эту неопределенность, уточнить (выявить) структуру статистической модели по результатам проводимых наблюдений.

1.2 Статистическая модель

Фундаментальным понятием теории вероятностей является вероятностная модель (вероятностное пространство) — это тройка $(\Omega, \mathscr{A}, \mathsf{P})$, где Ω — пространство элементарных событий, $\mathscr{A}$ — σ-алгебра подмножеств этого пространства (событий), P — вероятностная мера на σ-алгебре $\mathscr{A}$.

Основным объектом исследования математической статистики является статистическая модель. Определим это понятие.

Результатом статистического эксперимента являются вещественные числа $x_1, \ldots, x_n$ — статистические данные. Это значения случайных величин $\xi_1, \ldots, \xi_n$. Их совокупность $x^{(n)} = (x_1, \ldots, x_n)$ называется *выборкой* размера (порядка) n.

Определение. *Статистической моделью* называется тройка[2] $(\mathscr{X}, \mathscr{B}(\mathscr{X}), \mathscr{P})$, где $\mathscr{X} = \{x^{(n)}\}$ — выборочное пространство, т. е. совокупность всевозможных выборок размера n, $\mathscr{B}(\mathscr{X})$ — σ-алгебра на выборочном пространстве, $\mathscr{P} = \{\mathsf{P}\}$ — некоторое *семейство* распределений вероятностей, заданное на $\mathscr{B}(\mathscr{X})$.

В простейшем случае считаем, что $\mathscr{X} \subseteq \mathbb{R}^m$, а $\mathscr{B}(\mathscr{X})$ — борелевская σ-алгебра.

Примерами семейств распределений могут служить, например, семейство бернуллиевских распределений, $p \in (0,1)$; семейство пуассоновских распределений, $\lambda \in (0, \infty)$; семейство биномиальных распределений с параметрами (n,p), где n фиксировано, а $p \in (0,1)$ и т.д.

Наша цель — выделить из семейства распределений то единственное распределение, которое наилучшим образом соответствует нашим запросам, точнее, полученной выборке (после этого мы сможем работать с вероятностной моделью).

Если $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$, где θ — параметр, $\Theta \subset \mathbb{R}^m$ — параметрическое множество, то говорят, что $\mathscr{P}$ — *параметрическое семейство распределений*, а $(\mathscr{X}, \mathscr{B}(\mathscr{X}), \mathscr{P})$ — *параметрическая модель*.

Пусть имеется случайный вектор $\xi^{(n)} = \{\xi_1, \ldots, \xi_n\}$ со значениями $(x_1, \ldots, x_n)$ в выборочном пространстве $\mathscr{X}$. В соответствии с определением статистической модели, $\mathscr{P}$ — семейство распределений случайного вектора $\xi^{(n)}$. Чтобы не путать набор случайных величин (случайный вектор) $\xi_1, \ldots, \xi_n$ с его конкретными значениями $x_1, \ldots, x_n$, говорят, что $x_1, \ldots, x_n$ — выборка, а $\xi_1, \ldots, \xi_n$ $(\xi_i = \xi_i(\omega))$ — случайная выборка.

1.3 Статистические решения

Допустим, наша цель найти функцию $\delta : \mathscr{X} \longrightarrow D$, чтобы по имеющейся выборке $x^{(n)}$ мы могли принимать решение $d = \delta(x^{(n)})$.

Определение. Такая функция $\delta : \mathscr{X} \longrightarrow D$ будет называться *решающим правилом*, а элементы D — *статистическими решениями.*

Множество статистических решений D может быть самым разным. Мы можем решать, будет ли снег на улице или нет ($D = \{\text{будет снег}, \text{снега не будет}\}$, $\mathscr{X}$ – какие-то метеорологические наблюдения); какая температура будет завтра днём ($D = \{(t_{\min}, t_{\max}) \mid t_{\min} \le t_{\max}\}$, $\mathscr{X}$ – то же самое), в какой из выходных будет самая лучшая погода для рыбалки ($D = \{\text{какой-то день} \mid \text{день является выходным}\}$, $\mathscr{X}$ – то же самое).

Чтобы формализовать "правильность"решающего правила δ, мы вводим понятие *функции потерь* $\mathscr{L}(d, \theta)$, где $\mathscr{L}$ – измерима на $D \times \Theta$. Так как при разных значениях выборки функция потерь может принимать разные значения, то вводится понятие *функции риска* как функции от решающего правила $R(\delta, \theta) := \mathsf{M}_\theta \mathscr{L}(\delta, \theta)$. Здесь $\mathscr{L}(\delta, \theta)$ уже является случайной величиной.

1.4 Типы статистических решений. Статистическая модель схемы Бернулли

Зафиксируем число $n \in \mathbb{N}$. Рассмотрим случайные величины $\xi_1(\omega)$, $\xi_2(\omega)$, $\ldots$, $\xi_n(\omega)$ на некотором общем вероятностном пространстве $(\Omega, \mathscr{A}, \mathsf{P})$, $\omega \in \Omega$. Их совместное распределение:

$$\mathsf{P}(\xi_1 = a_1, \xi_2 = a_2, \ldots, \xi_n = a_n) = p^{a_1 + \ldots + a_n} q^{n - (a_1 + \ldots + a_n)}, \quad a_k \in \{0,1\}, \quad p, q \geqslant 0, \quad p + q = 1.$$

Значение случайной величины ξ_1 — исход первого испытания, $\mathsf{P}(\xi_1 = 1) = p$, $\mathsf{P}(\xi_1 = 0) = q = 1 - p$, и

[2] В общем смысле $\mathscr{X}$ - множество значений случайных величин ξ_i, заданных на $(\Omega, \mathscr{A})$ и измеримых относительно $(\mathscr{A}, \mathscr{B}(\mathscr{X}))$. Случайные величины и параметрическое семейство вероятностных мер порождают вероятностные меры на $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, которые обозначаются как $\mathscr{P}$ — *примеч. В. Х.*

аналогично для $\xi_2, \ldots, \xi_n$. Отсюда $\mathsf{P}(\xi_1 = a_1) = p^{a_1} q^{1-a_1}$, и т.д. Значит,

$$\mathsf{P}(\xi_1 = a_1, \xi_2 = a_2, \ldots, \xi_n = a_n) = p^{a_1+\ldots+a_n} q^{n-(a_1+\ldots+a_n)} = \prod_{k=1}^{n} \left(p^{a_k} q^{1-a_k} \right) = \mathsf{P}(\xi_1 = a_1) \ldots \mathsf{P}(\xi_n = a_n).$$

Отсюда следует, что ξ_1, ξ_2, ..., ξ_n — независимые испытания.

Рассмотрим случайную величину $S_n = \xi_1 + \ldots + \xi_n$. Она имеет биномиальное распределение:

$$\mathsf{P}(\xi_1 + \ldots + \xi_n = m) = \mathbf{C}_n^m p^m q^{n-m}, \quad S_n = \xi_1 + \ldots + \xi_n, \quad \mathsf{M} S_n = np, \quad \mathsf{D} S_n = npq.$$

Задача математической статистики — оценить неизвестное значение p. Для этого используются три подхода — точечная оценка, интервальная оценка и выбор из двух гипотез. Продемонстрируем каждый из них на примере схемы Бернулли.

1.4.1 Точечная оценка

Пусть мы ищем значение параметра распределения $\theta \in \Theta$ по значениям данной нам выборки. Тогда $d := \theta$, $D := \Theta$. В этом случае δ называется *статистической оценкой.*

Типичной функцией потерь для данного примера является $\mathscr{L}(d, \theta) := (d - \theta)^2$ — квадратичная ошибка. Тогда

$$R(\delta, \theta) = \mathsf{M}_\theta \mathscr{L}(d, \theta) = \mathsf{M}_\theta (\delta - \theta)^2$$

Такая функция риска называется *среднеквадратичной ошибкой.*

Логично утверждать, что если для любого параметра $\theta \in \Theta$ верно $R(\delta_1, \theta) \leq R(\delta_2, \theta)$, то δ_1 *не хуже*, чем δ_2. Если ещё существует такой параметр $\theta_0 \in \Theta$, что $R(\delta_1, \theta) < R(\delta_2, \theta)$, то δ_1 *лучше*, чем δ_2[3].

Определение. Если в некотором классе $\widetilde{\Theta} \subset \Theta$ существует такая δ^*, что $R(\delta^*, \theta) \leq R(\delta, \theta)$ для любого параметра $\theta \in \widetilde{\Theta}$ и для любой оценки δ, то δ^* называется *оценкой с равномерно наименьшим риском.*

Вернёмся к бернуллиевскому случаю. Запишем закон больших чисел для схемы Бернулли:

$$\frac{S_n}{n} = \frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\mathsf{P}} \mathsf{M} \frac{S_n}{n} = \frac{np}{n} = p, \quad n \to \infty.$$

т. е. частота появления успешного исхода $\frac{S_n}{n}$ сходится по вероятности к параметру p: $\frac{S_n}{n} \xrightarrow{\mathsf{P}} p$, $n \to \infty$.

Возьмем в качестве оценки параметра p эту частоту $\frac{S_n}{n} =: \hat{p}_n$. Это случайная величина со значениями $\frac{m}{n}$, $m = 0, \ldots, n$. Проведём небольшое численное моделирование $\hat{p}_n$

	$n = 10$	$n = 100$	$n = 1000$	$n = 10000$
$p = 0.15$	0.2	0.21	0.146	0.1486
$p = 0.5$	0.5	0.48	0.477	0.5023
$p = 0.85$	0.8	0.87	0.844	0.8496

Как мы видим, среднее и правда приближает реальное значение параметра, причём в общем случае чем больше номер, тем больше точность (но не всегда это будет верно в силу случайности величин).

Теорема 1.1. *Эта оценка обладает следующими свойствами:*

1) Несмещенность: $\mathsf{M}\hat{p}_n = p$.

2) Состоятельность: $\hat{p}_n \xrightarrow{\mathsf{P}} p$, $n \to \infty$.

3) Оптимальность: Дисперсия частоты $\hat{p}_n$ является наименьшей среди дисперсий всех других оценок, которые обладают свойством 1)[4].

□ Выше уже было показано, что оценка $\hat{p}_n$ несмещенная (это следует из того, что $\mathsf{M}S_n = np$), а в силу закона больших чисел для схемы Бернулли она состоятельна; тем самым свойства 1) и 2) доказаны.

Докажем свойство 3) — оптимальность. Пусть $\tilde{p}_n$ — любая оценка параметра p, удовлетворяющая условиям 1) и 2) (несмещённость и состоятельность). Рассмотрим величину $\mathsf{M}(\tilde{p}_n - p)^2$.

Для несмещенных оценок средняя квадратическая ошибка совпадает с дисперсией, в частности для нашей оценки $\hat{p}_n$: $\mathsf{M}(\hat{p}_n - p)^2 = \mathsf{D}\hat{p}_n$.

Обозначим

$$\mathsf{P}(\xi_1 = a_1, \xi_2 = a_2, \ldots, \xi_n = a_n) = p^{a_1+\ldots+a_n} (1-p)^{n-(a_1+\ldots+a_n)} = g(p; a_1, \ldots, a_n).$$

[3] Пытливый читатель может заметить, что мы ввели лишь отношение частичного порядка, где не все оценки сравнимы. Но всё же хочется уметь сравнивать произвольные оценки. Ответ на то, как это делать, даётся в байесовской теории оценивания — *примеч. В. Х.*

[4] Точнее, оценка является эффективной, а доказательство этого факта идейно неотличимо от более общего неравенства Крамера-Рао — *примеч. В. Х.*

Для любого $p \in (0,1)$ имеет место равенство $\sum\limits_{(a_1,\dots,a_n)} g(p; a_1, \dots, a_n) \equiv 1$. Условие несмещенности оценки означает, что $\mathsf{M}\tilde{p}_n = \sum\limits_{(a_1,\dots,a_n)} \tilde{p}_n(a_1, \dots, a_n) g(p; a_1, \dots, a_n) = p$. Рассмотрим $g(p; a_1, \dots, a_n) = g(p)$ как функцию параметра p. Тогда наши условия могут быть записаны в следующем виде (суммирование ведется по всем наборам $(a_1, \dots, a_n)$):

$$\begin{cases} \sum g(p) \equiv 1, \\ \sum \tilde{p}_n \cdot g(p) \equiv p; \end{cases} \qquad 0 < p < 1.$$

Продифференцируем каждое из этих соотношений по p, а затем, умножив первое на p, вычтем его из второго; получим:

$$\sum (\tilde{p}_n - p) g'_p(p) \equiv 1.$$

Теперь представим $g'_p(p)$ как логарифмическую производную: $g'_p(p) = g(p)\frac{\partial \ln g(p)}{\partial p}$, а затем применим неравенство Коши–Буняковского, представив $g(p)$ в виде $g(p) = \sqrt{g}\sqrt{g}$:

$$1 \equiv \left(\sum (\tilde{p}_n - p) g(p) \frac{\partial \ln g(p)}{\partial p} \right)^2 \leqslant \left(\sum (\tilde{p}_n - p)^2 g(p) \right) \left(\sum \left[\frac{\partial \ln g(p)}{\partial p} \right]^2 g(p) \right).$$

Так как $\mathsf{M}\tilde{p}_n = p$ (условие несмещённости оценки $\tilde{p}_n$), то первый из множителей в правой части неравенства — это дисперсия $\tilde{p}_n$. Обозначим второй множитель через $I(p)$, тогда неравенство перепишется в виде $1 \leqslant \mathsf{D}\tilde{p}_n \cdot I(p)$, или $\mathsf{D}\tilde{p}_n \geqslant \frac{1}{I(p)}$. Найдем $I(p)$ в явном виде:

$$I(p) = \mathsf{M}\left[\frac{\sum \xi_k}{p} - \frac{n - \sum \xi_k}{1-p} \right]^2 = \frac{\mathsf{M}\left(\sum \xi_k - np\right)^2}{p^2(1-p)^2} = \frac{\mathsf{D}\left(\sum \xi_k\right)}{p^2(1-p)^2} = \frac{np(1-p)}{p^2(1-p)^2} = \frac{n}{p(1-p)}.$$

Подставляя найденное значение $I(p)$ в неравенство для дисперсии, получаем: $\mathsf{D}\tilde{p}_n \geqslant \dfrac{p(1-p)}{n} = \mathsf{D}\hat{p}_n$, т. е. оценка $\hat{p}_n$ действительно обладает наименьшей дисперсией из всех несмещенных состоятельных оценок $\tilde{p}_n$. Теорема доказана. ■

Если задана произвольная оценка $\hat{p}_n$ параметра p, то представим ее математическое ожидание $\mathsf{M}\hat{p}_n$ в виде $\mathsf{M}\hat{p}_n = p + \Delta_n$. Тогда Δ_n называется смещением оценки $\hat{p}_n$. Несмещенные оценки обладают нулевым смещением: $\Delta_n = 0$.

Для нашей оценки $\hat{p}_n = \frac{S_n}{n}$, очевидно, $\mathsf{D}\hat{p}_n \to 0$, $n \to \infty$, т. е. частота обладает наименьшим рассеянием, если рассеяние измеряется с помощью дисперсии.

1.4.2 Интервальная оценка

Пусть теперь мы ищем некоторый интервал значений параметра, где скорее всего и лежит параметр распределения. Тогда $d := [\theta_1, \theta_2]$, $D := \{[\theta_1, \theta_2] \mid \theta_1, \theta_2 \in \Theta\}$. В данном случае удобно рассмотреть следующую функцию потери

$$\mathscr{L}(d, \theta) := \begin{cases} 0, & \theta \in [\theta_1, \theta_2] \\ 1, & \theta \notin [\theta_1, \theta_2] \end{cases}$$

Для неё

$$R(\delta, \theta) = \mathsf{M}_\theta \mathscr{L}(\delta, \theta) = \mathsf{P}_\theta(\theta \notin \delta) = \mathsf{P}_\theta(\theta \notin [\theta_1(\delta), \theta_2(\delta)])$$

Такая оценка $\delta = [\theta_1(\delta), \theta_2(\delta)]$ называется *интервальной статистической оценкой* параметра θ (*доверительным интервалом*).

Мы хотим, чтобы риск был поменьше (прогноз погоды был верен), для этого строим такую оценку, чтобы $R(\delta, \theta) \le \alpha$, где $0 < \alpha < 1$. Такой α называется *коэффициентом доверия*, а $1 - \alpha$ – *доверительной вероятностью* (*доверительным уровнем*). При этом мы хотим, чтобы длина интервала была поменьше (прогноз погоды был конкретен), то есть чтобы

$$R_2(\delta, \theta) := \mathsf{M}_\theta(|\theta_2(\delta) - \theta_1(\delta)|) \le \beta < \alpha$$

Пример 4.1. Рассмотрим $n = 100$ бросаний правильной монеты (схема Бернулли с параметром $p = 0{,}5$), x_k — исход k-го испытания (значение бернуллиевской случайной величины ξ_k); $S_n = \xi_1 + \ldots + \xi_n$. Очевидно, $\mathsf{P}(0 \leqslant S_n \leqslant 100) = 1$. Прямой подсчет вероятностей показывает, что

$$\mathsf{P}(35 \leqslant S_n \leqslant 65) = 0{,}99822; \quad \mathsf{P}(39 \leqslant S_n \leqslant 61) \approx 0{,}98.$$

Таким образом, $\left[\frac{35}{100}, \frac{65}{100}\right]$ — доверительный интервал для параметра p с доверительной вероятностью 0,99822; а $\left[\frac{39}{100}, \frac{61}{100}\right]$ — доверительный интервал для параметра p с доверительной вероятностью $\approx 0{,}98$.

Для построения доверительного интервала для схемы Бернулли запишем для оценки $\hat{p}_n = \frac{S_n}{n}$ неравенство Чебышёва:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{M}\frac{S_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}\frac{S_n}{n}}{\varepsilon^2} \quad\Longrightarrow\quad \mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{M}\frac{S_n}{n}\right| \leqslant \varepsilon\right) \geqslant 1 - \frac{\mathsf{D}\frac{S_n}{n}}{\varepsilon^2},$$

$$\mathsf{P}\big(\left|\hat{p}_n - p\right| \leqslant \varepsilon\big) \geqslant 1 - \frac{pq}{n\varepsilon^2} \geqslant 1 - \frac{1}{4n\varepsilon^2} \geqslant 1-\alpha, \quad \alpha = \frac{1}{4n\varepsilon^2}.$$

Теперь зададим произвольное $\alpha \in (0,1)$. Тогда для $\varepsilon_\alpha = \sqrt{\frac{1}{4n\alpha}}$ получим:

$$\mathsf{P}\big(\left|\hat{p}_n - p\right| \leqslant \varepsilon_\alpha\big) \geqslant 1-\alpha, \quad \text{т.е.} \quad \mathsf{P}\big(\hat{p}_n - \varepsilon_\alpha \leqslant p \leqslant \hat{p}_n + \varepsilon_\alpha\big) \geqslant 1-\alpha.$$

Таким образом, мы построили интервал $[\hat{p}_n - \varepsilon_\alpha, \hat{p}_n + \varepsilon_\alpha]$, в котором с задаваемой нами вероятностью ошибки α находится неизвестный параметр p. Чем меньше мы выбираем α, тем больше этот интервал. Для заданного α длину интервала можно уменьшить за счёт увеличения числа испытаний n.

Оценим вероятность ошибки α в конкретном примере для разных α и p, проведя $m = 1000$ раз по $n = 1000$ испытаний. Для этого воспользуемся полученным доверительным интервалом

	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$
$p = 0.15$	0.001	0.009	0.045
$p = 0.5$	0.026	0.07	0.138
$p = 0.85$	0.002	0.01	0.056

Видно, что наши оценки были очень грубыми. Укажем еще один (более точный) способ нахождения интервальной оценки в схеме Бернулли. По теореме Муавра–Лапласа число «успехов» схемы Бернулли с ростом n стремится к нормальной случайной величине:

$$\frac{S_n - \mathsf{M}S_n}{\sqrt{\mathsf{D}S_n}} = \frac{S_n - np}{\sqrt{npq}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{1}$$

Используя это, можно оценить вероятность

$$\mathsf{P}\big(\left|\hat{p}_n - p\right| \leqslant \varepsilon\big) = \mathsf{P}\left(\left|\tfrac{S_n - np}{n}\right|\sqrt{\tfrac{n}{pq}} \leqslant \varepsilon\sqrt{\tfrac{n}{pq}}\right) \simeq \Phi\left(\varepsilon\sqrt{\tfrac{n}{pq}}\right) - \Phi\left(-\varepsilon\sqrt{\tfrac{n}{pq}}\right) = 2\Phi(u) - 1, \quad u = \varepsilon\sqrt{\tfrac{n}{pq}}.$$

Здесь $\Phi(u) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{u} e^{-\frac{t^2}{2}}\,dt$ — функция распределения стандартной нормальной случайной величины.

Фиксируем $0 < \alpha < 1$. Нам нужно, чтобы $2\Phi(u) - 1 = 1 - \alpha$, т.е. $\Phi(u) = 1 - \frac{\alpha}{2}$. Обозначим такое значение u, при котором это выполнено, через $u_{1-\frac{\alpha}{2}}$ (квантиль порядка $1 - \frac{\alpha}{2}$ нормального распределения, находится из таблицы квантилей). Тогда искомое ε найдем из условия $\varepsilon\sqrt{\frac{n}{pq}} = u_{1-\frac{\alpha}{2}}$; $\varepsilon = u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$. Итак, неравенство

$$\left|\hat{p}_n - p\right| \leqslant u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \tag{2}$$

выполняется с вероятностью $\simeq 1-\alpha$. Осталось найти границы доверительного интервала. Возведем неравенство в квадрат:

$$\big(\hat{p}_n - p\big)^2 \leqslant u^2_{1-\frac{\alpha}{2}}\frac{p(1-p)}{n}.$$

Получили квадратное уравнение на p. В качестве $\bar{p}_n$ и $\bar{\bar{p}}_n$ берут корни этого квадратного уравнения (можно показать, что всегда $D > 0$ и корней действительно два).

Как и ранее, оценим вероятность ошибки α в конкретном примере для разных α и p, проведя $m = 1000$ раз по $n = 1000$ испытаний. Для этого воспользуемся новым доверительным интервалом

	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.5$
$p = 0.15$	0.202	0.335	0.489
$p = 0.5$	0.196	0.298	0.508
$p = 0.85$	0.217	0.301	0.497

1.4.3 Выбор из двух гипотез

Пусть теперь мы имеем $\Theta := \{\theta_1, \theta_2\}, \theta_1 < \theta_2$. Мы хотим выбрать наиболее вероятный из этих параметров. Тогда $d := \theta, D := \Theta$. В этом случае введём следующую функцию потерь

$$L(d, \theta) := \begin{cases} 0, & d = \theta \\ 1, & d \neq \theta \end{cases}$$

Тогда

$$R(\delta, \theta) = \mathsf{M}_\theta \mathscr{L}(\delta, \theta) = \mathsf{M}_\theta I_{\delta = \theta}$$

Пусть задано θ_0. Рассмотрим две (взаимоисключающие) гипотезы о параметре θ: H_0 (основная, или нулевая, гипотеза) и H_1 (альтернативная, или конкурирующая, гипотеза). (Например, $H_0 \colon \theta = \theta_0$, $H_1 \colon \theta \neq \theta_0$.) Наша задача: выбрать из этих двух гипотез ту, которой соответствует наименьшая вероятность ошибки.

Определение. *Вероятность ошибки I рода* — это вероятность отклонить верную гипотезу H_0. *Вероятность ошибки II рода* — это вероятность принять неверную гипотезу H_0.

Критерий проверки гипотезы H_0 — это правило, на основании которого мы можем считать, что она верна или неверна (т. е. принимаем ее или не принимаем). Составим таблицу:

	H_0	H_1
принимаем H_0		β
отклоняем H_0	α	

Здесь вероятность ошибки I рода (отклоняем H_0 в то время как она верна) обозначена через α, а вероятность ошибки II рода (принимаем неверную гипотезу H_0) — β.

Рассмотрим две гипотезы $H_0 \colon p = p_0$, $H_1 \colon p \neq p_0$ (p_0 задано). Пусть для параметра p получена интервальная оценка для заданной вероятности ошибки α — доверительный интервал $[\bar{p}_n, \bar{\bar{p}}_n]$. Тогда можно предложить такой критерий:

1) Если $p_0 \in [\bar{p}_n, \bar{\bar{p}}_n]$, то H_0 принимаем (и соответственно, отклоняем H_1);

2) Если $p_0 \notin [\bar{p}_n, \bar{\bar{p}}_n]$, то H_0 отклоняем (и тем самым принимаем H_1).

Поскольку $[\bar{p}_n, \bar{\bar{p}}_n]$ — доверительный интервал с доверительной вероятностью $1 - \alpha$, то вероятность ошибки I рода не превосходит α.

Пример 4.2. Рассмотрим еще один пример гипотез о параметре p схемы Бернулли. Пусть $H_0 \colon p = p_0$, $H_1 \colon p = p_1$, где $p_0 < p_1$ — заданы, и пусть α — вероятность ошибки I рода, β — вероятность ошибки II рода. Как всегда, обозначаем $S_n = \xi_1 + \ldots + \xi_n$. Тогда существует такой критерий: если $S_n > m_*$, то H_0 отклоняем (тем самым принимая H_1), а если $S_n \leqslant m_*$, то H_0 принимаем (H_1 отклоняем). Число m_* называется критическим значением и находится из соображений минимизации при фиксированном n сумм (вероятностей ошибок I и II рода).

Пусть заданы $0 < \alpha_0 < 1$, $0 < \beta_0 < 1$. Попробуем в данном критерии дать нижнюю оценку n_0 на n, чтобы для $n \geqslant n_0$ гипотезы могли быть различимы с вероятностями ошибок I и II рода не больше α_0 и β_0 соответственно для некоторого m^*.

Так как $\mathsf{P}_p(S_n = m) = C_n^m p^m (1-p)^{n-m}$, то вероятности ошибок I и II рода выражаются как

$$\alpha = \sum_{m > m^*} \mathsf{P}_{p_0}(S_n = m) = \sum_{m > m^*} C_n^m p_0^m (1 - p_0)^{n-m},$$

$$\beta = \sum_{m \leqslant m^*} \mathsf{P}_{p_1}(S_n = m) = \sum_{m > m^*} C_n^m p_1^m (1 - p_1)^{n-m}.$$

Положим

$$x_m := \frac{\mathsf{P}_{p_1}(S_n = m)}{\mathsf{P}_{p_0}(S_n = m)},$$

$$\widetilde{S}_0 := \sum_{m > m^*} \mathsf{P}_{p_0}(S_n = m) \ln x_m,\ \widetilde{S}_1 := \sum_{m \leqslant m^*} \mathsf{P}_{p_0}(S_n = m) \ln x_m,\ \widetilde{S} := \widetilde{S}_0 + \widetilde{S}_1,$$

$$\widetilde{\Sigma}_0 := \sum_{m > m^*} \mathsf{P}_{p_1}(S_n = m) \ln x_m,\ \widetilde{\Sigma}_1 := \sum_{m \leqslant m^*} \mathsf{P}_{p_1}(S_n = m) \ln x_m,\ \widetilde{\Sigma} := \widetilde{\Sigma}_0 + \widetilde{\Sigma}_1.$$

Распишем выражение

$$\widetilde{S} = \sum_{m=0}^{n} \ln\left(\frac{\mathsf{P}_{p_1}(S_n=m)}{\mathsf{P}_{p_0}(S_n=m)}\right)\mathsf{P}_{p_0}(S_n=m) = \sum_{m=0}^{n}\left(m\ln\frac{p_1}{p_0} + (n-m)\ln\frac{1-p_1}{1-p_0}\right)\mathsf{P}_{p_0}(S_n=m) =$$

$$\ln\frac{p_1}{p_0}\underbrace{\sum_{m=0}^{n} m\mathsf{P}_{p_0}(S_n=m)}_{=\mathsf{E}_{p_0}S_n=np_0} + \ln\frac{1-p_1}{1-p_0}\underbrace{\sum_{m=0}^{n}(n-m)\mathsf{P}_{p_0}(S_n=m)}_{=n-\mathsf{E}_{p_0}S_n=n(1-p_0)} = np_0\ln\frac{p_1}{p_0} + n(1-p_0)\ln\frac{1-p_1}{1-p_0} =$$

$$-n\left(p_0\ln\frac{p_0}{1-(1-p_1)} + (1-p_0)\frac{1-p_0}{(1-p_1)}\right) = -n\varphi(p_0, 1-p_1),$$

где φ - функция Кульбака

$$\varphi(a,b) = a\ln\frac{a}{1-b} + (1-a)\ln\frac{1-a}{b}.$$

Эта функция положительна и убывает по каждой переменной.

Пользуясь неравенством Йенсена, получаем оценку

$$\widetilde{S}_0 = \sum_{m>m^*}\ln\frac{\mathsf{P}_{p_1}(S_n=m)}{\mathsf{P}_{p_0}(S_n=m)}\mathsf{P}_{p_0}(S_n=m) \leqslant \underbrace{\sum_{m>m^*}\mathsf{P}_{p_0}(S_n=m)}_{=\alpha}\ln\left(\sum_{m>m^*}\frac{\mathsf{P}_{p_1}(S_n=m)}{\mathsf{P}_{p_0}(S_n=m)}\frac{\mathsf{P}_{p_0}(S_n=m)}{\sum_{m>m^*}\mathsf{P}_{p_0}(S_n=m)}\right) =$$

$$\alpha\ln\left(\frac{1}{\alpha}\sum_{m>m^*}\mathsf{P}_{p_1}(S_n=m)\right) = \alpha\frac{1-\beta}{\alpha}.$$

Аналогично $\widetilde{S}_1 \leqslant (1-\alpha)\ln(\beta/(1-\alpha))$. Поэтому

$$-n\varphi(p_0,1-p_1) = \widetilde{S} = \widetilde{S}_0 + \widetilde{S}_1 \leqslant \alpha\frac{1-\beta}{\alpha} + (1-\alpha)\frac{\beta}{1-\alpha} = -\varphi(\alpha,\beta),$$

откуда

$$n \geqslant \frac{\varphi(\alpha,\beta)}{\varphi(p_0,1-p_1)} \geqslant \frac{\varphi(\alpha_0,\beta_0)}{\varphi(p_0,1-p_1)} =: \gamma_0.$$

Аналогично получаем оценку снизу

$$n \geqslant \frac{\varphi(\alpha,\beta)}{\varphi(p_1,1-p_0)} \geqslant \frac{\varphi(\alpha_0,\beta_0)}{\varphi(p_1,1-p_0)} =: \gamma_1.$$

Искомый $n_0 := \max\{\gamma_0, \gamma_1\}$.

p_0	p_1	n_0	m^*
0.4	0.5	150	70
0.2	0.4	36	11
0.01	0.02	1000	15

На этом мы завершаем обзор. Далее речь пойдёт подробнее о точечных оценках, интервальных оценках и проверке гипотез.

2 Эмпирическая функция распределения

2.1 Модель повторных испытаний

Определение. *Моделью повторных испытаний* называется статистическая модель, в которой случайные величины $\xi_1, \ldots, \xi_n$ (со значениями $x_1, \ldots, x_n$ соответственно, $(x_1, \ldots, x_n) \in \mathscr{X}$) независимы и одинаково распределены.

В дальнейшем мы будем рассматривать только модели повторных испытаний.

Пример 1.1. Рассмотренная выше статистическая модель схемы Бернулли — модель повторных испытаний. Действительно, в этом случае рассматриваются независимые испытания с одним и тем же распределением вероятности $\mathsf{P}(\xi=0)=p$, $\mathsf{P}(\xi=1)=1-p$, где $p\in(0,1)$ — параметр.

Пример 1.2. Рассмотрим эксперимент по измерению температуры. Мы считаем, что измерения независимы и результаты измерений — значения одинаково распределенных случайных величин $\xi_1, \ldots, \xi_n \sim \xi$. На практике обычно результаты измерений колеблются около некоторого постоянного значения a, поэтому удобно рассматривать ξ в виде $\xi = a + \Delta$, где $\xi = (\xi_1, \dots, \xi_n)$, $\Delta = (\Delta_1, \dots, \Delta_n)$ — случайная ошибка, или в координатах: $\xi_k = a + \Delta_k$, $k = 1, \dots, n$. Случайные величины $\Delta_1, \dots, \Delta_n$ также независимы и одинаково распределены; при этом $\mathsf{M}\Delta_k = 0$, $\mathsf{D}\Delta_k = \sigma^2 \ \forall k$. Средняя температура вычисляется как среднее арифметическое результатов измерений:

$$\bar{\xi}_n = \frac{\xi_1 + \ldots + \xi_n}{n} = a + \frac{\Delta_1 + \ldots + \Delta_n}{n}, \quad \bar{\Delta}_n = \frac{\Delta_1 + \ldots + \Delta_n}{n} \text{ — средняя ошибка}, \quad \mathsf{M}\bar{\Delta}_n = 0, \quad \mathsf{D}\bar{\Delta}_n = \frac{\sigma^2}{n}.$$

2.2 Обоснование предельного перехода при стремлении размера выборки к бесконечности

В курсе математической статистики довольно часто нас интересуют большие выборки, поэтому нам необходимо "узаконить" рассмотрение бесконечных выборок и вероятностных мер на $\mathbb{R}^\infty$.

Рассмотрим последовательность выборочных пространств

$$(\mathbb{R}^1, \mathscr{B}(\mathbb{R}^1)), \dots, (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n)), \dots$$

с вероятностными мерами $\mathsf{P}_1, \dots, \mathsf{P}_n, \dots$. Исследуем их предельные свойства при $n \to \infty$.

Определение. Вероятностные меры $\mathsf{P}_1, \dots, \mathsf{P}_n, \dots$ называются *согласованными*, если

$$\forall n \in \mathbb{N} \quad \forall B \in \mathscr{B}(\mathbb{R}^n) \quad \mathsf{P}_{n+1}(B \times \mathbb{R}) = \mathsf{P}_n(B)$$

Введём пространство $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$, где $\mathbb{R}^\infty = \{x = (x_1, \dots, x_n, \dots) \mid x_k \in \mathbb{R}\}$, $\mathscr{B}(\mathbb{R}^\infty)$ — борелевская σ-алгебра.

Определение. Пусть $B \in \mathscr{B}(\mathbb{R}^n)$. Тогда *борелевским цилиндром* называется следующее множество:

$$Z_n(B) = \{x = (x_1, \dots) \in \mathbb{R}^\infty \mid (x_1, \dots x_n) \in B\}.$$

Теорема 2.1 (Колмогоров). *Если меры на $\mathbb{R}^1, \dots, \mathbb{R}^n, \dots$ согласованы, то существует единственная вероятностная мера P на $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ такая, что $\mathsf{P}(Z_n(B)) = \mathsf{P}_n(B)$ для всех $B \in \mathscr{B}(\mathbb{R}^n)$ и для всех натуральных n.*

Эта теорема обосновывает законность перехода к пределу при $n \to \infty$ (n — размер выборки).

2.3 Определение и свойства эмпирической функции распределения

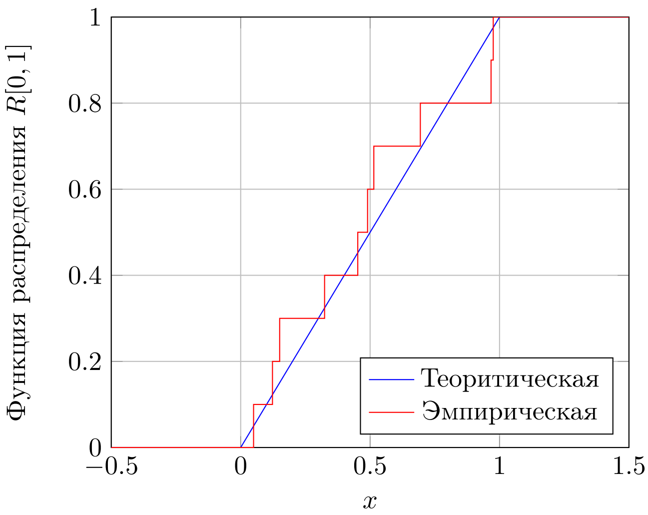
Определение. *Эмпирической функцией распределения* для данной выборки $x_1, \dots, x_n$ называется функция

$$\widehat{F}_n(x; x_1, \dots, x_n) = \frac{1}{n} \sum_{k=1}^{n} I_{(-\infty, x]}(x_k) = \frac{1}{n} \sum_{k=1}^{n} I_{(x_k \leqslant x)},$$

где $I_A(x) = \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases}$ — индикатор множества A.

Перейдем от выборки $x_1, \dots, x_n$ к вариационному ряду (совокупности порядковых статистик); иными словами, упорядочим выборку по возрастанию: $x_{(1)}, \dots, x_{(n)}$. Тогда, очевидно, эмпирическая функция распределения может быть записана в виде

$$\widehat{F}_n(x; x_1, \dots, x_n) = \widehat{F}_n(x; x_{(1)}, \dots, x_{(n)}) = \begin{cases} 0, & x < x_{(1)}; \\ \frac{k}{n}, & x_{(k)} \leqslant x < x_{(k+1)}, 1 \leqslant k \leqslant n-1; \\ 1 & x \geqslant x_{(n)}. \end{cases}$$



Если выборка $x_1, \ldots, x_n$ фиксирована, то эмпирическая функция распределения — это функция от переменной $x \in \mathbb{R}$: $\widehat{F}_n(x; x_1, \ldots, x_n) = \widehat{F}_n(x)$. Эта функция является функцией распределения случайной величины, принимающей равновероятно значения $x_1, \ldots, x_n$ [5]. Если в выборке k значений совпадают, то вероятность этого значения равна k/n.

Если же выборка не фиксирована, а случайные величины $\xi_1, \ldots, \xi_n$, породившие эту выборку, независимы и одинаково распределены с функцией распределения $F(x)$, то можно рассматривать $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$. Для каждого $x \in \mathbb{R}$ это случайная величина: $\widehat{F}_n(x; \xi_1, \ldots, \xi_n) = \widehat{F}_n(x; \omega)$.

Теорема 2.2. *1) Случайная величина $n\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ имеет биномиальное распределение с параметрами $(n, p = F(x))$ при любом фиксированном $x \in \mathbb{R}$;*

2) $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ является несмещенной состоятельной оценкой $F(x)$;

3) $\mathsf{P}\left(\lim\limits_{n \to \infty} \widehat{F}_n(x; \xi_1, \ldots, \xi_n) = F(x),\ \forall\, x \in \mathbb{R}\right) = 1$;

4) $\widehat{F}_n(x; \xi_1, \ldots, \xi_n) \xrightarrow{d} \mathcal{N}\big(F(x), \frac{F(x)(1-F(x))}{n}\big)$, $n \to \infty$.

$\square$ Найдем распределение случайной величины $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ при любом фиксированном $x \in \mathbb{R}$. Если $\xi_k \leqslant x$, то $I_{(-\infty, x]}(\xi_k) = 1$, а если $\xi_k > x$, то $I_{(-\infty, x]}(\xi_k) = 0$. Значит, при каждом $x \in \mathbb{R}$ $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ — это частота наступления события $\{\xi_k \leqslant x\}$, вероятность которого равна $\mathsf{P}(\xi_k \leqslant x) = F_{\xi_k}(x) = F(x)$. Отсюда получаем, что случайная величина $n\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ имеет биномиальное распределение с параметрами $(n, p = F(x))$ при любом фиксированном $x \in \mathbb{R}$ (утверждение 1) теоремы). Её математическое ожидание и дисперсия соответственно равны np и $np(1-p)$, поэтому получаем

$$\mathsf{M}\widehat{F}_n(x; \xi_1, \ldots, \xi_n) = F(x), \quad \mathsf{D}\widehat{F}_n(x; \xi_1, \ldots, \xi_n) = \frac{F(x)\big(1 - F(x)\big)}{n} \quad \forall\, x \in \mathbb{R}. \tag{1}$$

Таким образом, эмпирическую функцию распределения $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ можно рассматривать как оценку (теоретической) функции распределения $F(x)$. Поскольку $\mathsf{M}\widehat{F}_n(x; \xi_1, \ldots, \xi_n) = F(x)$, то эта оценка несмещенная, а в силу неравенства Чебышева $\mathsf{P}\big(|\widehat{F}_n(x; \xi_1, \ldots, \xi_n) - F(x)| \geqslant \varepsilon\big) \leqslant \frac{\mathsf{D}\widehat{F}_n}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0$, $n \to \infty$ $\forall\, x \in \mathbb{R}$, т. е. $\widehat{F}_n(x; \xi_1, \ldots, \xi_n) \xrightarrow{\mathsf{P}} F(x)$, $n \to \infty$ $\forall\, x \in \mathbb{R}$, значит эта оценка состоятельная. Таким образом, утверждение 2) также доказано.

Утверждение 3) следует из УЗБЧ для схемы Бернулли (теорема Бореля), а утверждение 4) — из формул (1) и центральной предельной теоремы для независимых одинаково распределенных случайных величин, примененной к сумме $n\widehat{F}_n(x; \xi_1, \ldots, \xi_n) = \sum\limits_{k=1}^{n} I_{(-\infty, x]}(\xi_k)$. $\blacksquare$

Задача 2.1. *Доказать, что $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ является оптимальной оценкой $F(x)$ $\forall\, x \in \mathbb{R}$.*

На самом деле имеет место еще более сильное утверждение, чем утверждение 3) доказанной теоремы, а именно равномерная сходимость $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ к $F(x)$ с вероятностью 1, что и составляет содержание следующей теоремы.

[5] Отсюда и берётся идея ЭФР. У нас есть только результаты некоторого эксперимента, но нет никаких данных о том, что какое распределение у наблюдаемой случайной величины, поэтому мы полагаем, что эти точки могли возникать равновероятно — *примеч. В. Х.*

2.4 Теорема Гливенко - Кантелли

Теорема 2.3 (Гливенко – Кантелли). *Пусть $\xi_1, \ldots, \xi_n$ — взаимно независимые, одинаково распределенные случайные величины с функцией распределения $F(x)$; $\widehat{F}_n(x; \xi_1, \ldots, \xi_n) = \frac{1}{n} \sum\limits_{k=1}^{n} I_{(-\infty, x]}(\xi_k)$ — их эмпирическая функция распределения. Тогда*

$$\mathsf{P}\left(\lim_{n \to \infty} \sup_x \left| \widehat{F}_n(x; \xi_1, \ldots, \xi_n) - F(x) \right| = 0 \right) = 1$$

(т. е. $\widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ равномерно сходится к $F(x)$ с вероятностью 1).

Замечание. Для определения эмпирической функции распределения в теореме Гливенко – Кантелли не требуется понятия выборки: она определяется для заданного (известного) набора взаимно независимых одинаково распределенных случайных величин.

□ Для краткости будем обозначать эмпирическую функцию распределения через $\widehat{F}_n(x) = \widehat{F}_n(x; \omega)$. По условию теоремы, $F(x)$ — функция распределения случайных величин $\xi_1, \ldots, \xi_n$. Рассмотрим два случая.

1) Пусть $F(x)$ — непрерывная и строго монотонная функция. Фиксируем произвольное $\varepsilon > 0$ и $k \in \mathbb{N}$, $k \neq 1$: $\frac{1}{k} \leqslant \varepsilon$. Поскольку функция $F(x)$ непрерывна и строго монотонна, то для каждого $i = 0, \ldots, k$ найдется x_i: $F(x_i) = \frac{i}{k}$ (возможно, $x_1 = -\infty$ или $x_k = +\infty$), причем такие $x_1, \ldots, x_k$ определены однозначно. Для соседних точек, по определению,

$$F(x_{i+1}) - F(x_i) = \frac{1}{k} \leqslant \varepsilon. \tag{2}$$

Зафиксируем i и произвольную точку x: $x_i < x < x_{i+1}$. В силу монотонности функций F и $\widehat{F}$ имеем:

$$\widehat{F}_n(x_i) - F(x_{i+1}) \leqslant \widehat{F}_n(x) - F(x) \leqslant \widehat{F}_n(x_{i+1}) - F(x_i), \tag{3}$$

и используя неравенство (2), отсюда получаем:

$$\widehat{F}_n(x_i) - F(x_i) - \varepsilon \leqslant \widehat{F}_n(x) - F(x) \leqslant \widehat{F}_n(x_{i+1}) - F(x_{i+1}) + \varepsilon,$$

$$\left| \widehat{F}_n(x) - F(x) \right| \leqslant \max_{0 \leqslant i \leqslant k} \left| \widehat{F}_n(x_i) - F(x_i) \right| + \varepsilon, \quad \forall\, x \in \mathbb{R}, \quad \Longrightarrow$$

$$\sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| \leqslant \max_{0 \leqslant i \leqslant k} \left| \widehat{F}_n(x_i) - F(x_i) \right| + \varepsilon, \quad \forall\, \varepsilon > 0, \text{ где } \frac{1}{k} \leqslant \varepsilon. \tag{4}$$

Рассмотрим событие $A_i = \left\{ \omega : \ \widehat{F}_n(x_i; \omega) \to F(x_i), n \to \infty \right\}$. Так как $\widehat{F}_n(x_i; \omega)$ – среднее независимых $\mathrm{Bern}(F(x_i))$, то по УЗБЧ для схемы Бернулли $\mathsf{P}(A_i) = 1$. Далее, рассмотрим событие $A^{(k)} = \bigcap\limits_{i=1}^{k} A_i$. $\mathsf{P}(\overline{A^{(k)}}) \leq \sum_{i=1}^{k} \mathsf{P}(\overline{A_i}) = 0$, поэтому событие $A^{(k)}$ происходит почти наверное. Очевидно, событие $A^{(k)}$ равносильно тому, что $\max\limits_{0 \leqslant i \leqslant k} \left| \widehat{F}_n(x_i; \omega) - F(x_i) \right| \to 0$, $n \to \infty$. Определим события

$$\tilde{A} = \bigcap_{k=2}^{\infty} A^{(k)}, \quad B = \left\{ \omega : \ \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x; \omega) - F(x) \right| \to 0, n \to \infty \right\}.$$

В силу неравенства (4) $\tilde{A} \subset B$, а поскольку $\mathsf{P}(\tilde{A}) = 1$ (аналогично конечному пересечению), то и $\mathsf{P}(B) = 1$.

2) Пусть теперь $F(x)$ — произвольная (неубывающая) непрерывная функция. Тогда определим x_i так:

$$x_i = \inf \left\{ x : \ F(x-0) \leqslant \frac{i}{k} \leqslant F(x) = F(x+0) \right\}.$$

Далее рассуждаем аналогично первому случаю. Осталось заметить, что при применении УЗБЧ для схемы Бернулли в данном случае нужно представить событие A_i в виде $A_i = A_i' \cap A_i''$, где

$$A_i' = \left\{ \omega : \ \widehat{F}_n(x_i; \omega) \to F(x_i), n \to \infty \right\}, \quad A_i'' = \left\{ \omega : \ \widehat{F}_n(x_i - 0; \omega) \to F(x_i - 0), n \to \infty \right\}.$$

Вероятность каждого из этих событий равна $\mathsf{P}(A_i') = \mathsf{P}(A_i'') = 1$, поэтому и $P(A_i) = 1$; дальнейшие рассуждения в точности такие же, как и в случае 1). ■

2.5 Свойства эмпирических моментов

Напомним некоторые определенные в п. 3.1 выборочные характеристики — выборочное среднее и выборочную дисперсию:

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}, \quad s^2 = \frac{(x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}.$$

Утверждение 2.4. *$\bar{x}$ и s^2 — соответственно математическое ожидание и дисперсия эмпирического распределения (т. е. распределения, определяемого функцией распределения $\widehat{F}_n(x; x_1, \ldots, x_n) = \widehat{F}_n(x)$).*

□ Обозначим эмпирическое распределение $\hat{\xi}$, $\widehat{F}_n(x)$ — функция распределения $\hat{\xi}$. Тогда доказательство следует из соотношений

$$\mathsf{M}\hat{\xi} = \int\limits_{\mathbb{R}} x \, d\widehat{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \int\limits_{\mathbb{R}} x \, dI_{\{x \geqslant x_k\}} = \frac{1}{n} \sum_{k=1}^{n} x_k = \bar{x};$$

$$\mathsf{D}\hat{\xi} = \int\limits_{\mathbb{R}} (x - \bar{x})^2 \, d\widehat{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \int\limits_{\mathbb{R}} (x - \bar{x})^2 \, dI_{\{x \geqslant x_k\}} = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2 = s^2.$$

Здесь мы воспользовались определением эмпирической функции распределения, линейностью интеграла Стилтьеса и формулой для интеграла Стилтьеса $\int\limits_X f(x) \, dg(x) = f(\xi)c$, где $g(x)$ — функция одного скачка (в точке ξ), $c = g(\xi + 0) - g(\xi - 0)$ — величина скачка. ■

Аналогично можно показать, что выборочные моменты порядка k являются моментами порядка k эмпирического распределения. Покажем, что выборочные моменты можно рассматривать как хорошие оценки моментов теоретического распределения.

Пусть $x_1, \ldots, x_k$ — выборка, порожденная независимыми одинаково распределенными случайными величинами $\xi_1, \ldots, \xi_n \sim \xi$, $F(x)$ — их (теоретическая) функция распределения (неизвестная, или известно в каком классе лежит, но неизвестно какая именно). Её k-тый момент равен

$$\mu_k = \int\limits_{-\infty}^{+\infty} x^k \, dF(x), \quad k \in \mathbb{N}.$$

(μ_1 — математическое ожидание, μ_2 — второй момент, $\sigma^2 = \mu_2 - \mu_1^2$ — дисперсия, и т.д.). Рассмотрим выборочные моменты — моменты эмпирического распределения:

$$\hat{\mu}_{k,n} = \frac{1}{n} \sum_{i=1}^{k} x_i^k = \frac{x_1^k + \ldots + x_n^k}{n} = \int\limits_{-\infty}^{+\infty} x^k \, d\widehat{F}_n(x).$$

Если рассматривать $\hat{\mu}_{k,n}$ как оценки μ_k, то легко получаем следующие её свойства:

1) Несмещённость: $\mathsf{M}\hat{\mu}_{k,n} = \mathsf{M}\left(\dfrac{\xi_1^k + \ldots + \xi_n^k}{n}\right) = \mathsf{M}\xi^k = \mu_k$;

2) Состоятельность: по закону больших чисел $\hat{\mu}_{k,n} \xrightarrow{\text{Pї.PS.}} \mathsf{M}\hat{\mu}_{k,n} = \mu_k \quad \Longrightarrow \quad \hat{\mu}_{k,n} \xrightarrow{\mathsf{P}} \mu_k$, $n \to \infty$.

2.6 Критерий Колмогорова

2.6.1 Статистика Колмогорова

Пусть дана случайная выборка $\xi_1, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$.

Определение. Случайная величина $D_n = D_n(\xi_1, \ldots, \xi_n) = \sup\limits_{x \in \mathbb{R}} \left| \widehat{F}_n(x; \xi_1, \ldots, \xi_n) - F(x) \right|$ называется статистикой Колмогорова.

В терминах статистики Колмогорова теорему Гливенко−Кантелли можно переформулировать так: D_n сходится к нулю при $n \to \infty$ с вероятностью 1 (т. е. P-п.н.).

Вид асимптотической функции распределения статистики $\sqrt{n}D_n$ дает следующая теорема.

Теорема 2.5 (Колмогоров). *Если функция $F(x)$ непрерывна, то при любом $y > 0$*

$$\lim_{n \longrightarrow \infty} \mathsf{P}(\sqrt{n}D_n \leqslant y) = K(y) := \sum_{m=-\infty}^{\infty} (-1)^m e^{-2m^2 y^2}.$$

Замечание. Для $y \leqslant 0$, очевидно, $\mathsf{P}(\sqrt{n}D_n \leqslant y) = 0$.

Участвующая в теореме функция $K(y) = \sum\limits_{m=-\infty}^{\infty} (-1)^m e^{-2m^2y^2}$, $y > 0$ называется функцией Колмогорова.

Мы докажем только часть теоремы Колмогорова, а именно следующую лемму:

Лемма 2.6. *Распределение статистики Колмогорова $D_n(\xi_1, \ldots, \xi_n)$, где $\xi_1, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$, не зависит от вида функции $F(x)$ (то есть для любых случайных н. о. р. $\xi_1, \ldots, \xi_n$ $D_n(\xi_1, \ldots, \xi_n)$ распределена одинаково).*

$\square$ Рассмотрим два случая.

1) Пусть $y = F(x)$ — непрерывная и строго монотонная функция. Тогда существует обратная функция: $x = F^{-1}(y)$. Рассмотрим случайные величины $Y_1, \ldots, Y_n$, $Y_k = F(\xi_k)$, $k = 1, \ldots, n$. Они независимы и имеют одинаковое равномерное распределение на отрезке $[0, 1]$:

$$R(y) = \mathsf{P}(Y_k \leqslant y) = \mathsf{P}(F(\xi_k) \leqslant y) = \mathsf{P}(\xi_k \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y, \quad 0 < y < 1.$$

Эмпирическая функция распределения $Y_1, \ldots, Y_n$:

$$\widehat{R}_n(y; Y_1, \ldots, Y_n) = \frac{1}{n}\sum_{k=1}^{n} I_{\{Y_k \leqslant y\}} = \frac{1}{n}\sum_{k=1}^{n} I_{\{\xi_k \leqslant F^{-1}(y)\}} = \widehat{F}_n(F^{-1}(y)),$$

где $\widehat{F}_n(x) = \widehat{F}_n(x; \xi_1, \ldots, \xi_n)$ — эмпирическая функция распределения случайной выборки $\xi_1, \ldots, \xi_n$.

Рассмотрим очевидное равенство

$$\sup_{x:\ 0<F(x)<1} \bigl|\widehat{F}_n(x) - F(x)\bigr| = \sup_{y:\ 0<y<1} \bigl|\widehat{R}_n(y) - R(y)\bigr|.$$

Его левая часть с вероятностью 1 совпадает с $D_n(\xi_1, \ldots, \xi_n)$, а правая часть — с $D_n(Y_1, \ldots, Y_n)$. Статистика $D_n(Y_1, \ldots, Y_n)$ от вида функции $F(x)$ не зависит, поскольку от $F(x)$ не зависит распределение случайных величин $Y_1, \ldots, Y_n$.

Для завершения доказательства осталось показать, что на множестве $C = \bigl\{x : F(x) = 0 \text{ или } F(x) = 1\bigr\}$ эмпирическая функция распределения $\widehat{F}_n(x)$ и теоретическая $F(x)$ совпадают с вероятностью 1. Для этого достаточно проверить, что $\mathsf{P}\left(\sup\limits_{x\in C}\bigl|\widehat{F}_n(x) - F(x)\bigr| = 0\right) = 1$. Проверку этого факта мы предоставляем читателю.

2) Если функция $F(x)$ — произвольная (неубывающая) непрерывная функция, то рассуждения аналогичны предыдущему случаю, только в этом случае нужно положить $F^{-1}(y) = \sup\bigl\{x : F(x) = y\bigr\}$. Читателю рекомендуется аккуратно провести рассуждения для этого случая самостоятельно. $\blacksquare$

3 Точечные оценки

3.1 Выборочные характеристики

3.1.1 Определение и свойства выборочных характеристик

Пусть в некоторой статистической модели имеется выборка порядка n: $x_1, \ldots, x_n$.

Определение. *Статистикой* называется произвольная измеримая функция $f(x_1, \ldots, x_n)$ от элементов выборки[6] $x_1, \ldots, x_n$.

Определение. Если случайная величина ξ имеет распределение $F(x)$, то *медианой* распределения называется такое число μ, что $F(\mu) = \frac{1}{2}$.

Медиана распределения обладает тем свойством, что $\mathsf{P}(\xi \geqslant \mu) = \mathsf{P}(\xi \leqslant \mu)$.

Рассмотрим примеры наиболее часто встречающихся статистик (или *выборочных характеристик*):

- Выборочное среднее:
$$\overline{x}_n := \sum_{i=1}^{n} x_i; \tag{1}$$

- Выборочная дисперсия:
$$S_n^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2; \tag{2}$$

[6] Статистика зависит **только от** элементов **выборки**, и ни в коем случае **не зависит от параметра** — *примеч. В. Х.*

- Выборочный момент порядка k:
$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \tag{3}$$
- Выборочный центральный момент порядка k:
$$\widehat{M}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^k; \tag{4}$$
- Порядковые статистики: упорядочим элементы выборки по возрастанию, получим последовательность
$$x_{(1)}, \dots, x_{(n)}. \tag{5}$$
Она называется *вариационным рядом выборки*, а её элементы — *порядковыми статистиками*. Случайные величины $\xi_{(n)}$ со значениями $x_{(n)}$ также называются порядковыми статистиками. Более формально,
$$x_{(1)} := \min(x_1, \dots, x_n), \tag{6}$$
$$x_{(2)} := \max\big[\min_k(x_1, \dots, \widehat{x_k}, \dots, x_n)\big], \tag{7}$$
$$x_{(3)} := \max\big[\min_{i\neq j}(x_1, \dots, \widehat{x}_i, \dots, \widehat{x}_j, \dots, x_n)\big], \tag{8}$$
$$\dots\dots\dots \tag{9}$$
$$x_{(n)} := \max(x_1, \dots, , x_n), \tag{10}$$
где «крышка», как обычно, означает пропуск этого элемента.
- Выборочный квантиль уровня α:
$$\widehat{\varkappa}_\alpha = \begin{cases} x_{([n\alpha]+1)}, & n\alpha \notin \mathbb{Z}; \\ \frac{1}{2}\left(x_{(n\alpha)} + x_{(n\alpha+1)}\right), & n\alpha \in \mathbb{Z}. \end{cases} \tag{11}$$
- Выборочная медиана (выборочный квантиль уровня 0.5):
$$\widehat{\mu} = \begin{cases} x_{(m)}, & n = 2m-1; \\ \frac{1}{2}\left(x_{(m)} + x_{(m+1)}\right), & n = 2m. \end{cases} \tag{12}$$

Пример 1.1. Рассмотрим равномерное распределение на $[0, \theta]$, θ — неизвестный параметр. Параметр θ можно оценить двумя способами:

1. $\widehat{\theta}^1 = 2\overline{x}_n$ — несмещённая оценка
2. $\widehat{\theta}^2 = x_{(n)}$ (оценка по крайней точке) — смещённая, но средне-квадратичная ошибка меньше, чем у $\widehat{\theta}^1$.

3.1.2 Распределение порядковых статистик

Пусть $\xi_1, \dots, \xi_n$ — случайная выборка с теоретической функцией распределения $F(x)$, $\xi_{(1)}, \dots, \xi_{(n)}$ — её порядковые статистики.

Найдем распределение $\xi_{(k)}$, $k \in \mathbb{N}$. Пусть $G_{\xi_{(k)}}(x) = \mathsf{P}(\xi_{(k)} \leqslant x)$ — функция распределения $\xi_{(k)}$. При каждом фиксированном $x \in \mathbb{R}$ имеем:
$$G_{\xi_{(k)}}(x) = \mathsf{P}\big(\xi_{(k)} \leqslant x\big) = \mathsf{P}\left(\hat{F}_n(x) \geqslant \frac{k}{n}\right) = \sum_{i=k}^{n} C_n^i \big(F(x)\big)^i \big(1 - F(x)\big)^{n-i}.$$

Пример 1.2. Пусть $F(x)$ — функция распределения равномерного распределения на отрезке $[0, 1]$: $F(x) = x$, $0 < x < 1$. Тогда
$$G_{\xi_{(k)}}(x) = \mathsf{P}\big(\xi_{(k)} \leqslant x\big) = \sum_{i=k}^{n} C_n^i x^i (1-x)^{n-i}.$$

Найдем плотность этого распределения. Для этого продифференцируем функцию распределения:
$$\big(G_{\xi_{(k)}}(x)\big)'_x = \sum_{i=k}^{n} i C_n^i x^{i-1}(1-x)^{n-i} - \sum_{i=k}^{n-1}(n-i) C_n^i x^i (1-x)^{(n-1)-i} =$$

$$= \sum_{i=k}^{n} nC_{n-1}^{i-1}x^{i-1}(1-x)^{(n-1)-(i-1)} - \sum_{i=k}^{n-1} nC_{n-1}^{i}x^{i}(1-x)^{(n-1)-i} =$$

$$= nC_{n-1}^{k-1}x^{k-1}(1-x)^{(n-1)-(k-1)} + nC_{n-1}^{k}x^{k}(1-x)^{n-k-1} + \ldots - nC_{n-1}^{k}x^{k}(1-x)^{n-k-1} - \ldots = nC_{n-1}^{k-1}x^{k-1}(1-x)^{n-k}.$$

(во втором равенстве воспользовались тождествами $iC_n^i = nC_{n-1}^{i-1}$, $(n-i)C_n^i = nC_{n-1}^i$). Таким образом, плотность распределения $\xi_{(k)}$ равна $\left[\mathsf{P}(\xi_{(k)} \leqslant x)\right]' = nC_{n-1}^{k-1}x^{k-1}(1-x)^{n-k} = \dfrac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}$, а функция распределения —

$$G_{\xi_{(k)}}(x) = \mathsf{P}(\xi_{(k)} \leqslant x) = \frac{n!}{(k-1)!(n-k)!}\int_0^x t^{k-1}(1-t)^{n-k}\,dt, \quad 0 < x < 1. \tag{13}$$

Определение. Пусть $a > 0$, $b > 0$. Распределение с плотностью

$$p(x) = \begin{cases} \dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, & 0 < x < 1; \\ 0, & x \leqslant 0 \text{ или } x \geqslant 1, \end{cases}$$

где $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ — бета-функция (эйлеров интеграл I рода), называется бета-распределением с параметрами $a > 0$, $b > 0$. Функция распределения бета-распределения $I_x(a,b) = \dfrac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}\,dt$ $(0 < x < 1)$ называется неполной бета-функцией.

Из математического анализа известно, что $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, где $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1}e^{-x}\,dx$ — гамма-функция Эйлера (эйлеров интеграл II рода), и для $n \in \mathbb{N}$ $\Gamma(n+1) = n!$, поэтому формулу (13) можно переписать в виде

$$G_{\xi_{(k)}}(x) = \mathsf{P}(\xi_{(k)} \leqslant x) = \frac{1}{B(k, n-k+1)}\int_0^x t^{k-1}(1-t)^{n-k}\,dt = I_x(k, n-k+1), \quad 0 < x < 1.$$

Таким образом, нами доказан следующий результат.

Утверждение 3.1. *Распределение порядковой статистики $\xi_{(k)}$, $1 \leqslant k \leqslant n$, для случайной выборки $\xi_1, \ldots, \xi_n$ с равномерным распределением на отрезке $[0,1]$ является бета-распределением с параметрами $a = k$, $b = n-k+1$.*

Замечание. Функция распределения порядковых статистик в случае произвольной непрерывной функции распределения $F(x)$ случайной выборки $\xi_1, \ldots, \xi_n$ имеет вид $\mathsf{P}(\xi_{(k)} \leqslant x) = I_{F(x)}(k, n-k+1)$.

3.2 Функция правдоподобия. Регулярные модели

Пусть в некоторой статистической модели $(X, \mathscr{A}_X, \mathscr{P} = \{\mathsf{P}_\theta, \theta \in \Theta\})$ имеется выборка $x = (x_1, \ldots, x_n)$, порожденная случайной выборкой $\xi_1, \ldots, \xi_n \sim \xi$, где случайные величины $\xi_1, \ldots, \xi_n$ независимы и одинаково распределены с функцией распределения $F_\theta(x)$, $\theta \in \Theta$ — параметр распределения.

Определение. Оптимальной оценкой параметра θ называется несмещенная оценка с минимальной дисперсией, т. е. такая оценка $\hat{\theta}^*$, для которой выполнены следующие свойства:

1) $\mathsf{M}_\theta\hat{\theta}^* = \theta \ \ \forall\,\theta \in \Theta$;

2) $\mathsf{M}_\theta(\hat{\theta}^* - \theta)^2 = \min\limits_{\hat{\theta}:\ \mathsf{M}_\theta\hat{\theta} = \theta} \mathsf{M}_\theta(\hat{\theta} - \theta)^2$.

Напомним, что для несмещенных оценок среднеквадратичное отклонение совпадает с дисперсией:

$$\mathsf{M}_\theta(\hat{\theta} - \theta)^2 = \mathsf{M}_\theta(\hat{\theta} - \mathsf{M}\hat{\theta})^2 = \mathsf{D}\hat{\theta}.$$

Таким образом, оптимальная оценка — это наилучшая из всех несмещенных оценок. Аналогично можно определить оптимальную оценку в классе оценок с заданным смещением Δ: $\mathsf{M}_\theta\hat{\theta}_1 = \mathsf{M}_\theta\hat{\theta}_2 = \Delta$.

Нас будут интересовать два случая: распределение ξ дискретно (с распределением вероятностей P_θ) или абсолютно непрерывно (с плотностью $p(\theta; x)$). Чтобы в дальнейшем не рассматривать эти случаи отдельно, введем следующее удобное обозначение:

$$f(\theta; x) = \begin{cases} \mathsf{P}_\theta(\xi = x), & \text{если модель дискретна;} \\ p(\theta; x), & \text{если модель абсолютно непрерывна.} \end{cases}$$

В дальнейшем придется интегрировать по выборочному пространству, поэтому отметим, что если модель дискретна, то интегрирование заменяется суммированием (для краткости, мы будем проводить все выкладки для абсолютно непрерывной модели).

Определение. Функцией правдоподобия (для данной выборки $x_1, \ldots, x_n$) называется следующая функция (параметра θ):

$$L(\theta; x_1, \ldots, x_n) := \prod_{i=1}^{n} f(\theta; x_i), \quad \theta \in \Theta.$$

В дальнейшем для краткости мы будем писать $L(\theta; x) = L(\theta; x_1, \ldots, x_n)$, $x = (x_1, \ldots, x_n)$.

Определение. Статистическая модель называется *регулярной* (по Рао–Крамеру), если выполнены следующие условия регулярности:

1) $L(\theta; x) > 0$ и дифференцируема по θ $\forall\, \theta \in \Theta$ и $\forall\, x \in X$;

2) Случайная величина $U(\theta; x) = \dfrac{\partial \ln L(\theta; x)}{\partial \theta} = \sum\limits_{i=1}^{n} \dfrac{\partial \ln f(\theta; x_i)}{\partial \theta}$ (которая называется функцией вклада выборки) имеет ограниченную дисперсию:

$$0 < \mathsf{M}_\theta^2 U(\theta; x) < \infty, \quad \forall\, \theta \in \Theta.$$

3) Для любой статистики $\hat{\theta} = \hat{\theta}(x)$ имеет место равенство

$$\frac{\partial}{\partial \theta} \int\limits_X \hat{\theta}(x) L(\theta; x)\, dx = \int\limits_X \hat{\theta}(x) \frac{\partial L(\theta; x)}{\partial \theta}\, dx$$

(Это означает, что выборочное пространство X не зависит от параметра θ).

Пример 2.1. (нерегулярной модели). Рассмотрим модель $R(0, \theta)$ (равномерное распределение на отрезке $(0, \theta)$, $\theta > 0$). Условие 3) регулярности для этой модели не выполнено:

$$\frac{\partial}{\partial \theta}\left(\int\limits_0^\theta \frac{1}{\theta}\, dx\right) = \frac{\partial}{\partial \theta}\left(\frac{1}{\theta}\theta\right) = \frac{\partial}{\partial \theta}(1) = 0, \quad \text{но} \quad \int\limits_0^\theta \frac{\partial}{\partial \theta}\left(\frac{1}{\theta}\right) dx = -\int\limits_0^\theta \frac{1}{\theta^2} dx = -\frac{1}{\theta}.$$

Таким образом, эта модель не является регулярной.

Задача 3.1. *Проверить условие регулярности 3) для экспоненциального распределения с плотностью*

$$p_\theta(x) = \begin{cases} e^{-(x-\theta)}, & x \geqslant \theta; \\ 0, & x < \theta. \end{cases}$$

3.3 Условное математическое ожидание

Определение. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — вероятностное пространство, ξ — $\mathscr{A}$-измеримая функция (случайная величина), $\mathscr{C}$ — σ-подалгебра в $\mathscr{A}$. Тогда *условным математическим ожиданием (УМО; обозначение:* $\mathsf{M}(\xi \mid \mathscr{C})$*)* ξ относительно $\mathscr{C}$ (более точно, *вариантом* УМО) называется случайная величина $\tilde{\xi}(\omega)$, удовлетворяющая следующим условиям:

1. $\tilde{\xi}$ $\mathscr{C}$-измерима;
2. для всех $C \in \mathscr{C}$

$$\int\limits_C \xi\, d\mathsf{P} = \int\limits_C \tilde{\xi}\, d\mathsf{P}, \text{ т.е. } \mathsf{M}\xi \mathrm{I}_C = \mathsf{M}\tilde{\xi}\mathrm{I}_C.$$

Замечание. ξ не обязана быть $\mathscr{C}$-измеримой (иначе определение тривиально: возьмём $\tilde{\xi} = \xi$). УМО есть осреднение случайной величины, чтобы та стала измеримой относительно более грубой σ-алгебры.

Следующие два утверждения обосновывают корректность определения:

Утверждение 3.2. *Варианты УМО существуют, если* $\mathsf{M}|\xi| < \infty$.

☐ Сначала рассмотрим случай неотрицательной случайной величины ξ. $Q(C) := \int\limits_C \xi\, d\mathsf{P}$ определяет меру на $(\Omega, \mathscr{C})$, абсолютно непрерывную относительно меры P, рассмотренной на $(\Omega, \mathscr{C})$. По теореме Радона – Никодима существует $\mathscr{C}$-измеримая функция $g(\omega) \geqslant 0$ такая, что $\int\limits_\Omega g(\omega)\, d\mathsf{P} = 1$ и $Q(C) = \int\limits_C g(\omega)\, d\mathsf{P}$ (g — производная Радона – Никодима). g по определению является вариантом УМО.

В случае знакопеременной ξ полагаем $\xi = \xi^+ - \xi^-$ (ξ^+ и ξ^- неотрицательны) и берём в качестве варианта УМО $\mathsf{M}(\xi \mid \mathscr{C}) = \mathsf{M}(\xi^+ \mid \mathscr{C}) - \mathsf{M}(\xi^- \mid \mathscr{C})$. ■

Замечание. Условие $\mathsf{M}|\xi| < \infty$ не является необходимым.

Утверждение 3.3. *Любые два варианта УМО совпадают почти наверное.*

□ От противного: пусть g_1 и g_2 — два варианта УМО $\mathsf{M}(\xi \mid \mathscr{C})$. Пусть также $C = \{\omega \mid g_1(\omega) \neq g_2(\omega)\}$, $\mathsf{P}(C) > 0$. Имеем, что $C = C_< \cup C_>$ (они измеримы как прообразы борелевских), где $C_* = \{\omega \mid g_1(\omega) * g_2(\omega)\}$, $* \in \{<, >\}$. Хотя бы одно из множеств $C_<$ и $C_>$ имеет положительную меру (иначе $\mathsf{P}(C) = 0$); без ограничения общности считаем $\mathsf{P}(C_<) > 0$. g_1 и g_2 измеримы, поэтому $D = C_< \in \mathscr{C}$. Тогда по определению варианта УМО

$$\int\limits_D g_1\, d\mathsf{P} = \int\limits_D \xi\, d\mathsf{P} = \int\limits_D g_2\, d\mathsf{P},$$

что противоречит тому, что $g_1 < g_2$ всюду на D и $\mathsf{P}(D) > 0$. ■

Поэтому УМО можно считать однозначно определённым с точностью до множеств P-меры нуль.

Определение. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — вероятностное пространство, $\mathscr{C}$ — σ-подалгебра в $\mathscr{A}$, $A \in \mathscr{A}$. Тогда *условная вероятность* A относительно $\mathscr{C}$ определяется так: $Pf(A \mid \mathscr{C}) = \mathsf{M}(I_A \mid \mathscr{C})$, где I_A — характеристическая функция A.

Определение. Пусть $(\Omega, \mathscr{A}, \mathsf{P})$ — вероятностное пространство, ξ и η — случайные величины, $\mathscr{C}_\eta$ — σ-подалгебра в $\mathscr{A}$, порождённая η (т. е. прообразы всех борелевских множеств из $\mathbb{R}$). Тогда *условным матожиданием ξ относительно η* называется $\mathsf{M}(\xi \mid \eta) = \mathsf{M}(\xi \mid \mathscr{C}_\eta)$.

3.3.1 Свойства УМО

Определение. ξ *не зависит* от σ-алгебры $\mathscr{C}$, если для любого события $C \in \mathscr{C}$ случайные величины ξ и I_C независимы.

Обозначение: $\mathscr{C}_0 = \{\Omega, \varnothing\}$ — тривиальная σ-алгебра.

Утверждение 3.4.

1. *Если* $\mathsf{P}(\xi = c) = 1$, *то* $\mathsf{M}(\xi \mid \mathscr{C}) = c$ *п.н.*
2. *Линейность:* $\mathsf{M}(\alpha_1\xi_1 + \alpha_2\xi_2 \mid \mathscr{C}) = \alpha_1\mathsf{M}(\xi_1 \mid \mathscr{C}) + \alpha_2\mathsf{M}(\xi_2 \mid \mathscr{C})$ *п.н.*
3. *Если* $\xi \leqslant \eta$ *п.н., то* $\mathsf{M}(\xi \mid \mathscr{C}) \leqslant \mathsf{M}(\eta \mid \mathscr{C})$ *п.н.*
4. $\big|\mathsf{M}(\xi \mid \mathscr{C})\big| \leqslant \mathsf{M}(|\xi| \mid \mathscr{C})$ *п.н.*
5. *Неравенство Йенсена: пусть* $g(x)$ *— непрерывная выпуклая вниз функция,* $\mathsf{M}\big|g(\xi)\big| < \infty$.
   Тогда $g(\mathsf{M}(\xi \mid \mathscr{C})) \leqslant \mathsf{M}(g(\xi) \mid \mathscr{C})$ *п.н. Равенство имеет место тогда и только тогда, когда* ξ *—* $\mathscr{C}$*-измерима (точнее, совпадает с некоторой* $\mathscr{C}$*-измеримой функцией почти всюду).*
6. $\mathsf{M}(\xi \mid \mathscr{A}) = \xi$ *п.н.*
7. *Если* ξ *не зависит от* $\mathscr{C}$, *то* $\mathsf{M}(\xi \mid \mathscr{C}) = \mathsf{M}\xi$ *п.н.*
8. *Пусть* η *является* $\mathscr{C}$*-измеримой и* $\mathsf{M}|\xi\eta| < \infty$. *Тогда* $\mathsf{M}(\xi\eta \mid \mathscr{C}) = \eta\mathsf{M}(\xi \mid \mathscr{C})$.
9. $\mathsf{M}(\xi \mid \mathscr{C}_0) = \mathsf{M}\xi$
10. *Если* $\mathscr{C}_1 \subseteq \mathscr{C}_2$, *то* $\mathsf{M}\big(\mathsf{M}(\xi \mid \mathscr{C}_2) \mid \mathscr{C}_1\big) = \mathsf{M}(\xi \mid \mathscr{C}_1)$. *Если* $\mathscr{C}_1 \supseteq \mathscr{C}_2$, *то* $\mathsf{M}\big(\mathsf{M}(\xi \mid \mathscr{C}_2) \mid \mathscr{C}_1\big) = \mathsf{M}(\xi \mid \mathscr{C}_2)$.
11. $\mathsf{M}\big(\mathsf{M}(\xi \mid \mathscr{C})\big) = \mathsf{M}\xi$
12. *Если* $\mathsf{M}\xi^2 < \infty$, *то*
$$\inf_{g(\omega)\ -\ \mathscr{C}\text{-измер.}} \mathsf{M}(\xi - g(\omega))^2 = \mathsf{M}(\xi - \mathsf{M}(\xi \mid \mathscr{C}))^2,$$
   т. е. $g_{\min}(\omega) = \mathsf{M}(\xi \mid \mathscr{C})$ *имеет наименьшее среднеквадратичное отклонение. В некотором смысмле это альтернативное определение УМО.*

□ 1 пункт по определению, $2 - -5$ пункты по свойствам математического ожидания, 6 пункт по определению.

Докажем 7-е свойство. Для любого $C \in \mathscr{C}$ ξ не зависит от I_C.

$$\int_C \mathsf{M}(\xi \mid \mathscr{C}) d\mathsf{P} = \mathsf{M}(\xi I_C) = \mathsf{M}\xi \mathsf{M} I_C = \mathsf{M}\xi \int_C d\mathsf{P} = \int_C \mathsf{M}\xi d\mathsf{P}.$$

Докажем 8-е свойство. Рассмотрим случай, когда $\eta = I_B$, $B \in \mathscr{C}$. Тогда для любого $C \in \mathscr{C}$

$$\int_C \mathsf{M}(\xi\eta \mid \mathscr{C}) d\mathsf{P} = \int_C \xi\eta d\mathsf{P} = \int_{C \cap B} \xi d\mathsf{P};$$

$$\int_C \eta \mathsf{M}(\xi \mid \mathscr{C}) d\mathsf{P} = \int_{C\cap B} \mathsf{M}(\xi \mid \mathscr{C}) d\mathsf{P} = \int_{C\cap B} \xi d\mathsf{P}.$$

Докажем 10-е свойство. Рассмотрим случай $\mathscr{C}_1 \subseteq \mathscr{C}_2$. Пусть $C \in \mathscr{C}_1$.

$$\int_C \mathsf{M}(\xi \mid \mathscr{C}_1) d\mathsf{P} = \int_C \xi d\mathsf{P}.$$

$$\int_C \mathsf{M}(\mathsf{M}(\xi \mid \mathscr{C}_2) \mid \mathscr{C}_1) d\mathsf{P} = \int_C \mathsf{M}(\xi \mid \mathscr{C}_2) d\mathsf{P} = \int_C \xi d\mathsf{P}.$$

Если в 10 пункте возьмём $\mathscr{C}_1 = \mathscr{C}_0$, $\mathscr{C}_2 = \mathscr{C}$, то по пункту 9 получим утверждение 11-го пункта.

Докажем 12-е свойство. Пусть g - $\mathscr{C}$ измеримая функция. Тогда

$$\begin{aligned}\mathsf{M}((\xi - g(\omega))^2 \mid \mathscr{C}) &= \mathsf{M}((\xi - \mathsf{M}(\xi \mid \mathscr{C}) + \mathsf{M}(\xi \mid \mathscr{C}) - g(\omega))^2 \mid \mathscr{C}) = \\ &\mathsf{M}((\xi - \mathsf{M}(\xi \mid \mathscr{C}))^2 \mid \mathscr{C}) + \mathsf{M}((\mathsf{M}(\xi \mid \mathscr{C}) - g(\omega))^2 \mid \mathscr{C}) + 2\mathsf{M}((\xi - \mathsf{M}(\xi \mid \mathscr{C}))(\mathsf{M}(\xi \mid \mathscr{C}) - g(\omega)) \mid \mathscr{C}).\end{aligned}$$

Так как $\mathsf{M}(\xi \mid \mathscr{C}) - g(\omega)$ — $\mathscr{C}$ - измерима, то произведение расписывается как

$$(\mathsf{M}(\xi \mid \mathscr{C}) - g(\omega)) \cdot \mathsf{M}(\xi - \mathsf{M}(\xi \mid \mathscr{C}) \mid \mathscr{C}) = (\mathsf{M}(\xi \mid \mathscr{C}) - g(\omega)) \cdot (\mathsf{M}(\xi \mid \mathscr{C}) - \mathsf{M}(\xi \mid \mathscr{C})) = 0.$$

Используя то, что второе слагаемое при раскрытии квадрата было неотрицательно, получаем

$$\mathsf{M}((\xi - g(\omega))^2 \mid \mathscr{C}) \geq \mathsf{M}((\xi - \mathsf{M}(\xi \mid \mathscr{C}))^2 \mid \mathscr{C}).$$

Теперь можно взять матожидание от обоих частей этого неравенство и получить искомое. ■

Определение. Вариант условного распределения $\zeta = \mathsf{P}(\xi \mid \mathscr{C})$ будем называть *регулярным*, если существует такая функция $\mathsf{P}(A, \omega)$, $A \in \mathscr{C}$, $\omega \in \Omega$, удовлетворяющая условиям:

1. $\mathsf{P}(A, \omega) = \mathsf{P}(A \mid \mathscr{C})$ (P-п.н.) для любого $A \in \mathscr{C}$;
2. $\mathsf{P}(\cdot, \omega)$ — вероятностная мера на $\mathscr{C}$ при фиксированном ω.

3.4 Достаточные статистики

$(\mathscr{X}, \mathscr{A}, \mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\})$ — параметрическая модель; $x_1, \ldots, x_n$ — выборка.

Определение. Статистика $T(x) = T(x_1, \ldots, x_n)$ называется *достаточной* (для данной модели), если для всех $\theta \in \Theta$ и для всех $A \in \mathscr{A}$ условная вероятность $\mathsf{P}_\theta(A \mid T(\xi) = t)$ не зависит от θ для всех t, для которых определена условная вероятность.

Найти достаточную статистику по определению не представляется возможным.

Пример 4.1. Схема Бернулли. $T(x) = \sum x_i$. Пусть t и x таковы, что $T(x) = t$. Тогда

$$\mathsf{P}_\theta(\xi = x \mid T(\xi) = t) = \frac{\mathsf{P}_\theta(\xi = x)}{\mathsf{P}_\theta(T(\xi) = t)} = \frac{\theta^t (1-\theta)^{n-t}}{C_n^t \theta^t (1-\theta)^{n-t}} = \frac{1}{C_n^t} \quad \text{— не зависит от } \theta.$$

Следовательно, $T(x) = \sum x_i$ — достаточная статистика.

Теорема 3.5 (критерий достаточности — теорема Неймана – Фишера — теорема о факторизации). *T — достаточная статистика тогда и только тогда, когда имеет место следующее представление функции правдоподобия:*

$$L(\theta; x) = g_\theta(T(x)) h(x) \qquad g_\theta(T(x)) = g(\theta; T(x)),$$

где g_θ и h — неотрицательные измеримые функции (каждая — в своей области).

Замечание. $L(\theta; x)$ — плотность совместного распределения $\xi_1, \ldots, \xi_n$ по некоторой мере μ; семейство $\{\mathsf{P}_\theta\}$ абсолютно непрерывно относительно меры μ, плотность — производная Радона – Никодима. Нам важны два случая: а) μ — мера Лебега, б) μ — считающая мера.

□ *[Доказательство для дискретного случая]* $x \in \mathscr{X}$ — выборка.

$$\mathsf{P}_\theta(\xi = x \mid T(\xi) = t) = \begin{cases} \frac{\mathsf{P}_\theta(\xi = x)}{\mathsf{P}_\theta(T(\xi) = t)}, & T(x) = t \\ 0, & T(x) \neq t \end{cases}$$

Необходимость. Пусть T — достаточная статистика. Тогда положим

$$h(x) = \frac{\mathsf{P}_\theta(\xi = x)}{\mathsf{P}_\theta(T(\xi) = t)} \text{ — не зависит от } \theta \text{ в силу достаточности } T,$$

откуда $L(\theta;x) = \mathsf{P}_\theta(\xi = x) = \mathsf{P}_\theta(T(\xi) = t)\cdot h(x) = g_\theta(T(x))\cdot h(x)$.

Достаточность. Пусть $T(x) = t$. Тогда

$$\mathsf{P}_\theta(\xi = x \mid T(\xi) = t) = \frac{g_\theta(t)h(x)}{\sum\limits_{T(x)=t} g_\theta(t)h(x)} = \frac{h(x)}{\sum h(x)} \text{ — не зависит от } \theta,$$

поэтому T — достаточная статистика. ■

3.5 Количество информации Фишера

3.5.1 Информация Фишера

Пусть модель регулярна. Рассмотрим тождество $\int\limits_X L(\theta;x)\,dx \equiv 1$ (оно выполнено, так как $L(\theta;x)$ — плотность распределения случайного вектора $\xi_1,\dots,\xi_n$). Продифференцируем его по θ:

$$\frac{\partial}{\partial\theta}\int\limits_X L(\theta;x)\,dx = \int\limits_X \frac{\partial \ln L(\theta;x)}{\partial\theta} L(\theta;x)\,dx = 0.$$

Отсюда следует, что математическое ожидание функции вклада равно 0: $\mathsf{M}_\theta U(\theta;X) = \mathsf{M}_\theta \frac{\partial \ln L(\theta;x)}{\partial\theta} = 0$.

Определение. Количеством информации Фишера (или просто информацией Фишера) называется дисперсия функции вклада:

$$I_n(\theta) := \mathsf{D}_\theta U(\theta;\xi_1,\dots,\xi_n) = \mathsf{M}_\theta U^2(\theta;\xi_1,\dots,\xi_n) = \int\limits_X \left(\frac{\partial \ln L(\theta;x)}{\partial\theta}\right)^2 L(\theta;x)\,dx.$$

Количество информации для одного наблюдения равно $I_1(\theta) = \int\limits_X \left(\frac{\partial \ln f(\theta;x)}{\partial\theta}\right)^2 f(\theta;x)\,dx$, а поскольку наблюдения (случайные величины) $\xi_1,\dots,\xi_n$ независимы, то информация Фишера о выборке размера n равна $I_n(\theta) = nI_1(\theta)$.

3.5.2 О свойствах информации Фишера

$\mathscr{P} = \{\mathsf{P}_\theta \mid \theta\in\Theta\}$, $x_1,\dots,x_s$ — выборка. $L(\theta;x_1,\dots,x_n)$ — функция правдоподобия. Информация Фишера:

$$I_n(\theta) = \mathsf{M}_\theta\left(\frac{\partial \ln L(\theta;\xi_1,\dots,\xi_n)}{\partial\theta}\right)^2 =: I^\xi(\theta).$$

$I^{\xi_1}(\theta) = I_1(\theta)$ ($I^\xi(\theta)$ — информация в выборке ξ — это просто такое обозначение). Для информации Фишера выполняются все обычные свойства информации:

Утверждение 3.6 (свойства информации Фишера).

1. *$I^\xi(\theta) = I^{\xi_1}(\theta) + \dots + I^{\xi_n}(\theta)$, если $\xi_1,\dots,\xi_n$ независимы.*
2. *Для повторной выборки $I^\xi(\theta) = nI^{\xi_1}(\theta)$.*
3. *Пусть $(\mathfrak{T},\mathscr{C})$ — измеримое пространство значений статистики T, Q_θ — распределение статистики T, $\widetilde{L}(\theta;T(x))$ — функция правдоподобия для $T(x)$ (статистика T не обязательно одномерна), $I^{T(\xi)}(\theta) = \mathsf{M}_\theta\left(\frac{\partial \ln \widetilde{L}(\theta;T(\xi))}{\partial\theta}\right)^2$ — информация Фишера в статистике T. Тогда $I^\xi(\theta) \geqslant I^{T(\xi)}(\theta)$.*

□ Докажем только пункт 3 (остальные тривиальны), да и то лишь в дискретном случае.
$L(\theta;x) = \mathsf{P}_\theta(\xi = x)$.

$$\widetilde{L}(\theta;\,T(x)) = \widetilde{L}(\theta;\,t) = \sum_{x:T(x)=t} \mathsf{P}_\theta(\xi = x)$$

$$0 \leqslant \mathsf{M}_\theta\left(\frac{\partial \ln L(\theta;x)}{\partial\theta} - \frac{\partial \ln \widetilde{L}(\theta;T(x))}{\partial\theta}\right)^2 = I^\xi(\theta) - 2\mathsf{M}_\theta\left(\frac{\partial \ln L(\theta;x)}{\partial\theta}\cdot\frac{\partial \ln \widetilde{L}(\theta;T(x))}{\partial\theta}\right) + I^{T(\xi)}(\theta).$$

Далее (здесь штрих означает производную по θ),

$$\widetilde{L}'(\theta;t) = \sum_{x:T(x)=t} \mathsf{P}'_\theta(\xi = x) = \sum_{x:T(x)=t} L'(\theta;x),$$

откуда

$$\mathsf{M}_\theta\left(\frac{\partial \ln L}{\partial\theta}\cdot\frac{\partial \ln \widetilde{L}}{\partial\theta}\right) = \sum_t \sum_{x:T(x)=t} \frac{L'(\theta;\, x)}{L(\theta;\, x)}\cdot\frac{\widetilde{L}'(\theta;\, T(x))}{\widetilde{L}(\theta;\, T(x))} L(\theta;\, x) = \sum_t \frac{\widetilde{L}'(\theta;\, t)}{\widetilde{L}(\theta;\, t)}\cdot\sum_{x:T(x)=t} L'(\theta;\, x) =$$

$$= \sum_t \frac{\widetilde{L}'(\theta;\, t)}{\widetilde{L}(\theta;\, t)}\widetilde{L}'(\theta;\, t) = \sum_t \left(\frac{\widetilde{L}'(\theta;\, t)}{\widetilde{L}(\theta;\, t)}\right)^2 \widetilde{L}(\theta;\, t) = \mathsf{M}\left(\frac{\widetilde{L}'(\theta;\, T(\xi))}{\widetilde{L}(\theta;\, T(\xi))}\right)^2 = I^{T(\xi)}(\theta),$$

поэтому $0 \leqslant I^\xi(\theta) - 2I^{T(\xi)}(\theta) + I^{T(\xi)}(\theta) = I^\xi(\theta) - I^{T(\xi)}(\theta)$, то есть $I^\xi(\theta) \geqslant I^{T(\xi)}(\theta)$. ■

3.6 Эффективные оценки

3.6.1 Неравенство Рао - Крамера

Пусть имеется регулярная модель с параметрическим семейством распределений $\mathscr{P} = \{\mathsf{P}_\theta,\ \theta\in\Theta\}$. Следующая теорема дает нижнюю границу дисперсий оценок произвольной дифференцируемой функции от параметра θ в классе оценок с заданным смещением.

Теорема 3.7 (Неравенство Рао – Краме́ра[7]). *Пусть модель с параметрическим семейством распределений $\mathscr{P}$ регулярна, $x = (x_1, \ldots, x_n)$ — выборка, и пусть некоторая статистика $\hat\theta(x)$ оценивает дифференцируемую функцию $\tau(\theta)$ параметра θ. Обозначим $b(\theta) = \mathsf{M}_\theta\hat\theta(x) - \tau(\theta)$ — смещение оценки $\hat\theta(x)$. Если $b(\theta)$ — дифференцируемая функция, то справедливо неравенство*

$$\mathsf{D}_\theta\big(\hat\theta(x) - \tau(\theta) - b(\theta)\big) \geqslant \frac{[\tau'(\theta) + b'(\theta)]^2}{I_n(\theta)},$$

где $I_n(\theta)$ — количество информации Фишера о выборке x. При этом неравенство обращается в равенство тогда и только тогда, когда оценка $\hat\theta(x)$ является линейной функцией вклада выборки, т. е. $\hat\theta(x) - \tau(\theta) - b(\theta) = a(\theta)U(\theta;x)$.

В частности, если оценка $\hat\theta(x)$ несмещенная, $\mathsf{M}_\theta\hat\theta(x) = \tau(\theta)$, то $b(\theta) = 0$, и с учетом $I_n(\theta) = nI_1(\theta)$ неравенство принимает вид $\mathsf{D}_\theta\hat\theta(x) \geqslant \dfrac{\big(\tau'(\theta)\big)^2}{nI_1(\theta)}$.

□ По определению смещения оценки $\hat\theta(x)$ имеем $\tau(\theta) + b(\theta) = \mathsf{M}_\theta\hat\theta(x)$. Продифференцируем это равенство, записав $\mathsf{M}_\theta\hat\theta(x)$ в виде интеграла и пользуясь условием регулярности 3):

$$\tau'(\theta) + b'(\theta) = \frac{\partial}{\partial\theta}\int\limits_X \hat\theta(x)L(\theta;x)\,dx = \int\limits_X \hat\theta(x)\frac{\partial\ln L(\theta;x)}{\partial\theta}L(\theta;x)\,dx = \int\limits_X \hat\theta(x)U(\theta;x)L(\theta;x)\,dx = \mathsf{M}_\theta\big(\hat\theta(x)U(\theta;x)\big).$$

Учитывая, что $\mathsf{M}_\theta U(\theta;x) = 0$, получаем:

$$\tau'(\theta) + b'(\theta) = \mathsf{M}_\theta\big(\hat\theta(x)U(\theta;x)\big) = \mathsf{M}_\theta\big[\big(\hat\theta(x) - \tau(\theta) - b(\theta) + \tau(\theta) + b(\theta)\big)U(\theta;x)\big] = \mathsf{M}_\theta\big[\big(\hat\theta(x) - \tau(\theta) - b(\theta)\big)U(\theta;x)\big] +$$

$$+ \big(\tau(\theta) + b(\theta)\big)\mathsf{M}_\theta U(\theta;x) = \mathsf{M}_\theta\Big[\big(\hat\theta(x) - \mathsf{M}_\theta\hat\theta(x)\big)\big(U(\theta;x) - \mathsf{M}_\theta U(\theta;x)\big)\Big] = \operatorname{cov}_\theta(\hat\theta(x), U(\theta;x)).$$

Применим к $\operatorname{cov}_\theta(\hat\theta(x), U(\theta;x))$ неравенство Коши – Буняковского $(\operatorname{cov}(\xi,\eta))^2 \leqslant \mathsf{D}\xi\cdot\mathsf{D}\eta$:

$$\big(\tau'(\theta) + b'(\theta)\big)^2 = \big(\operatorname{cov}_\theta(\hat\theta(x), U(\theta;x))\big)^2 \leqslant \mathsf{D}_\theta\hat\theta(x)\mathsf{D}_\theta U(\theta;x) = \mathsf{D}_\theta\big(\hat\theta(x) - \tau(\theta) - b(\theta)\big)\cdot I_n(x).$$

Отсюда следует требуемое неравенство. А так как неравенство Коши – Буняковского превращается в равенство тогда и только тогда, когда функции (в нашем случае случайные величины) линейно связаны, то и наше неравенство превращается в равенство в том и только том случае, когда (при каждом θ) $\hat\theta(x)$ и $U(\theta;x)$ линейно связаны: $\hat\theta(x) - \tau(\theta) - b(\theta) = a(\theta)U(\theta;x)$. Теорема доказана. ■

Определение. Оценка, для которой в неравенстве Крамера-Рао имеет место равенство, называется *эффективной.*

[7] а не Кра́мера!

3.6.2 Многомерный случай

Пусть $\theta = (\theta_1, \ldots, \theta_s)^T$ — вектор-параметр, $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$. Рассматривается оценка $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_s)$, $\hat{\theta}_k = \hat{\theta}_k(x_1, \ldots, x_n)$. Положим

$$\lambda_j = \frac{\partial \ln L(\theta; x_1, \ldots, x_n)}{\partial \theta_j}, \qquad \Lambda = (\lambda_1, \ldots, \lambda_s)^T.$$

В многомерном случае роль дисперсии играет *ковариационная матрица* $\Sigma = \Sigma_{\hat{\theta}(\xi)} = \mathsf{M}_\theta\big((\hat{\theta}(\xi) - \mathsf{M}\hat{\theta}(\xi))(\hat{\theta}(\xi) - \mathsf{M}\hat{\theta}(\xi))^T\big)$; $\Sigma = (\sigma_{ij})$, $\sigma_{ij} = \mathsf{M}\big((\hat{\theta}_i - \mathsf{M}\hat{\theta}_i)(\hat{\theta}_j - \mathsf{M}\hat{\theta}_j)\big) = \operatorname{cov}(\hat{\theta}_i, \hat{\theta}_j)$; $\sigma_{ii} = \sigma_i^2$ — дисперсия, остальные — попарные ковариации. Несмещённость оценки задаётся равенством $\mathsf{M}_\theta\hat{\theta}(\xi) = \theta$.

Аналогом количества информации Фишера является *информационная матрица Фишера* $J(\theta) = \mathsf{M}_\theta(\Lambda\Lambda^T)$. Предположим, что эта матрица обратима (существует $J^{-1}(\theta)$). Тогда имеет место аналог теоремы Крамера–Рао: матрица $\Sigma_{\hat{\theta}(\xi)} - J^{-1}(\theta)$ является неотрицательно определённой.

3.6.3 Теорема Колмогорова - Блекуэлла - Рао

$(\mathscr{X}, \mathscr{A}, \mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\})$ — параметрическая модель. $T(x)$ — достаточная статистика, если существует вариант регулярной условной вероятности $P_\theta(A \mid T(\xi)) \, \forall A \in \mathscr{A}$, не зависящий от параметра θ.

Теорема 3.8 (Колмогоров – Блекуэлл – Рао). *Пусть $T(x)$ — достаточная статистика; $\hat{\theta}(x)$ — оценка параметра θ с $\mathsf{M}_\theta\hat{\theta}^2 < \infty$. Тогда оценка $\theta^*(\xi) = \mathsf{M}_\theta(\hat{\theta}(\xi) \mid T(\xi))$ обладает следующими свойствами:*

1. *$\hat{\theta}^*(\xi)$ является статистикой и её распределение не зависит от θ,*
2. *$\mathsf{M}_\theta\theta^*(\xi) = \mathsf{M}_\theta\hat{\theta}(\xi)$,*
3. *$\mathsf{M}_\theta(\theta^*(\xi) - \theta)^2 \leqslant \mathsf{M}(\hat{\theta}(\xi) - \theta)^2$. Равенство имеет место тогда и только тогда, когда $\hat{\theta}$ измерима относительно $\mathscr{C}_T$ (является измеримой функцией достаточной статистики).*

$\square$ В силу определения достаточной статистики $\theta^* = \mathsf{M}_\theta(\hat{\theta} \mid T) = \int_{\mathscr{X}} \hat{\theta}(X) d(\mathsf{P}(X \mid T))$ (регулярность варианта нужна для того, чтобы тут взять интеграл по мере) не зависит от θ, поэтому θ^* — статистика. θ^* — $\mathscr{C}_T$-измерима (в силу определения УМО), поэтому θ^* есть функция от T. $\mathsf{M}\theta^* = \mathsf{M}\big(\mathsf{M}(\hat{\theta} \mid T)\big) = \mathsf{M}\hat{\theta}$. В силу неравенства Йенсена $\mathsf{M}\big((\hat{\theta} - \theta)^2 \mid T\big) \geqslant \big(\mathsf{M}(\hat{\theta} \mid T) - \theta\big)^2 = (\theta^* - \theta)^2$. Возьмём математическое ожидание: $\mathsf{M}(\hat{\theta} - \theta)^2 = \mathsf{M}\big(\mathsf{M}\big((\hat{\theta} - \theta)^2 \mid T\big)\big) \geqslant \mathsf{M}(\theta^* - \theta)^2$, причём равенство в неравенстве Йенсена достигается тогда и только тогда, когда $\hat{\theta}$ $\mathscr{C}_T$-измерима, т.е. есть функция от T. $\blacksquare$

Пример 6.1. Рассмотрим случаные величины ξ_i из схемы Бернулли с параметром $\theta \in (0, 1)$, $T(\xi_1, \ldots, \xi_n) := \xi_1 + \ldots + \xi_n$. Найдём $\zeta := \mathsf{M}(\xi_1 \mid T)$.

Так как $\mathsf{M}(\xi_i \mid T) = \mathsf{M}(\xi_j \mid T)$ для любых i, j, то по свойству линейности

$$n\zeta = \mathsf{M}(\xi_1 \mid T) + \ldots + \mathsf{M}(\xi_n \mid T) = \mathsf{M}(T \mid T) = T$$

Из чего следует, что $\zeta = T/n$.

3.6.4 Полная достаточная статистика

Определение. Достаточная статистика $T(x)$ называется *полной* (для данного параметрического семейства распределений), если для любой измеримой функции $f(t)$ из условия $\mathsf{M}_\theta f(T(\xi)) = 0$ (P-п.н.) для любого $\theta \in \Theta$ следует, что $f(T(\xi)) = 0$ почти наверное.

Полная достаточная статистика T обладает следующим свойством: если

$$\mathsf{M}_\theta f_1(T(\xi)) = \mathsf{M}_\theta f_2(T(\xi)),$$

то $f_1(T(\xi)) = f_2(T(\xi))$ почти наверное. Доказательство очевидно.

Теорема 3.9 (Колмогоров). *Пусть $T(x)$ — полная достаточная статистика и f и g таковы, что $\mathsf{M}_\theta f(T(\xi)) = g(\theta)$. Тогда $f(T)$ является оптимальной оценкой для $g(\theta)$.*

$\square$ Докажем теорему в частном случае $g(\theta) = \theta$.

Дано, что $\mathsf{M}_\theta f(T) = \theta$, т.е. $\hat{\theta}(\xi) := f(T(\xi))$ — несмещённая оценка параметра θ. Докажем, что эта оценка имеет наименьшую дисперсию среди всех возможных несмещённых оценок. От противного. Пусть существует такая оценка $\widetilde{\theta}$, что $\mathsf{M}\widetilde{\theta}(\xi) = \theta$ ($\widetilde{\theta}$ — несмещённая оценка) и $\mathsf{D}_{\theta_0}\widetilde{\theta}(\xi) < \mathsf{D}_{\theta_0}\hat{\theta}(\xi)$ хотя бы для одного $\theta_0 \in \Theta$. Улучшим оценку $\widetilde{\theta}$ при помощи теоремы Колмогорова – Блекуэлла – Рао: $\theta^* := \mathsf{M}_\theta(\widetilde{\theta} \mid T(\xi))$. Такая оценка обладает следующими свойствами:

1. $\theta^* = \theta^*(T(x))$;
2. $\mathsf{M}_\theta \theta^* = \theta$;
3. $\mathsf{D}_\theta \theta^* \leqslant \mathsf{D}_\theta \widetilde{\theta}\ \forall \theta \in \Theta$.

Имеем: $\mathsf{D}_{\theta_0}\theta^* \leqslant \mathsf{D}_{\theta_0}\widetilde{\theta} < \mathsf{D}_{\theta_0}\hat{\theta}$. Но θ^* и $\hat{\theta}$ суть функции полной достаточной статистики T, $\mathsf{M}_\theta \theta^* = \theta = \mathsf{M}_\theta \hat{\theta}$, откуда $\theta^*(\xi) = \hat{\theta}(\xi)$ почти наверное, а поэтому $\mathsf{D}_{\theta_0}\theta^* = \mathsf{D}_{\theta_0}\hat{\theta}$. Противоречие. ■

Замечание. Полная достаточная статистика существует не всегда.

3.6.5 Замечания о минимальной достаточной статистике

1. Сама выборка является достаточной статистикой (например, по критерию факторизации).
2. Любая измеримая взаимно-однозначная функция от достаточной статистики сама является достаточной статистикой.
3. Если для любой достаточной статистики T существует такая измеримая функция ψ, что $T_{\min}(x) = \psi(T(x))$ ($T_{\min}$ *подчинена* T), то $T_{\min}$ называется *минимальной* достаточной статистикой. Для любой биективной измеримой функции ϕ статистика $\phi \circ T_{\min}$ тоже является минимальной достаточной. Минимальная достаточная статистика существует всегда (для любой параметрической модели). Можно проводить редукцию (без потери информации) от выборки к минимальной достаточной статистике. Дальнейшая редукция невозможна.
4. Как правило, в разложении, которое даёт теорема Неймана – Фишера, появляется минимальная достаточная статистика.
5. Полная достаточная статистика (если она существует) является минимальной.

3.7 Асимптотические свойства оценок

1. *Состоятельность оценки.* $\{\mathsf{P}_\theta \mid \theta \in \Theta\}$ — параметрическая модель, $\hat{\theta}_n$ — оценка по выборке длины n, θ_0 — истинное значение параметра. Оценка состоятельна, если $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathsf{P}} \theta_0$.
2. *Асимптотическая несмещённость.* $\mathsf{M}_\theta \hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$ (т. е. смещение $b_n(\theta) = \mathsf{M}_\theta \hat{\theta}_n - \theta \xrightarrow[n\to\infty]{} 0$).
3. *Асимптотическая нормальность.* $\hat{\theta}_n$ асимптотически нормальна, если существует монотонно сходящаяся к нулю последовательность положительных чисел $\{c_n\}_{n=1}^{\infty}$ такая, что
$$\frac{\hat{\theta}_n - \theta}{c_n} \xrightarrow{\mathrm{d}} Z \sim \mathcal{N}(0,1)$$
(сходимость по распределению к некоторой случайной величине со стандартным нормальным распределением). Говорят, что $\frac{\hat{\theta}_n - \theta}{c_n}$ асимптотически нормальна с $\mathcal{N}(0,1)$, а $\hat{\theta}_n$ асимптотически нормальна с $\mathcal{N}(\theta, c_n^2)$; θ называется асимптотическим средним, а c_n^2 — асимптотической дисперсией оценки $\hat{\theta}_n$. Асимптотически нормальная оценка автоматически является асимптотически несмещённой. В качестве c_n обычно берут $c_n^2 = \frac{\sigma^2(\theta)}{n}$ (σ^2 не зависит от выборки).

   Пример 7.1. Схема Бернулли с параметром p (вероятность успеха). $\xi_1, \ldots, \xi_n$ — выборка.
$$\widehat{p} = \frac{\xi_1 + \cdots + \xi_n}{n}, \qquad \mathsf{M}\widehat{p} = p, \qquad \mathsf{D}\widehat{p} = \frac{p(1-p)}{n} = \frac{\sigma^2(p)}{n}.$$
$\widehat{p}$ асимптотически нормальна в силу теоремы Муавра – Лапласа (ЦПТ). (Обычно через эту теорему асимптотическую нормальность и доказывают.)

   Пример 7.2. $\xi_1, \ldots, \xi_n$, $\xi_i \sim \mathcal{N}(a, \sigma^2)$. $\mathsf{D}\overline{\xi} = \frac{\sigma^2}{n}$.
4. *Асимптотическая эффективность.* Рассмотрим семейство распределений, подчинённых условиям регулярности, для оценок параметров которого имеет место неравенство Крамера – Рао. $1/I_n(\theta)$ — нижняя граница дисперсий всех оценок. Коэффициент эффективности оценки: $e_n(\hat{\theta}_n) = \frac{1/I_n(\theta)}{\mathsf{D}\hat{\theta}_n}$. $0 < e_n(\hat{\theta}_n) \leqslant 1$. Если $e_n(\hat{\theta}_n) = 1$, то оценка называется эффективной. $e_\infty(\hat{\theta}_n) = \lim\limits_{n\to\infty} e_n(\hat{\theta}_n)$. Если $e_\infty(\hat{\theta}_n) = 1$, то оценка называется асимптотически эффективной.

   Асимптотическая эффективность в рамках асимптотической нормальности. Пусть $\hat{\theta}_n$ асимптотически нормальна с $\mathcal{N}(\theta, \frac{\sigma^2(\theta)}{n})$. Она называется асимптотически эффективной (в рамках асимптотической нормальности), если
$$\frac{1/I_n(\theta)}{\sigma^2(\theta)/n} \xrightarrow[n\to\infty]{} 1.$$

Лемма 3.10. *Пусть* $\xi_n := (\xi_n^{(1)}, \ldots, \xi_n^{(s)})$ *- случайный вектор,* $c := (c_1, \ldots, c_s)$*. Если* $\xi_n^{(i)} \xrightarrow{P} c_i$ *для любого* $i \in \{1, \ldots, s\}$ *и* f *- непрерывна в точке* c*, то* $f(\xi_n) \xrightarrow{P} f(c)$*.*

□ Надо доказать, что для любого $\varepsilon > 0$

$$\mathsf{P}(\omega \mid |f(\xi_n(\omega)) - f(c)| > \varepsilon) \xrightarrow{n\to\infty} 0$$

Зафиксируем $\varepsilon > 0$. Положим $A_{n,\varepsilon} := \{\omega \mid |f(\xi_n(\omega)) - f(c)| > \varepsilon\}$, $B_{n,\delta}^{(i)} := \{\omega \mid |\xi_n^{(i)}(\omega) - c| > \delta\}$, $B_{n,\delta} := \bigcap_{i=1}^s B_{n,\delta}^{(i)}$. В силу непрерывности f в точке c существует $\delta > 0$ такое, что из того, что $|\xi_n(\omega) - c| \le \delta$ следует $|f(\xi_n(\omega)) - f(c)| \le \varepsilon$, то есть $\overline{B}_{n,\delta} \subset \overline{A}_{n,\varepsilon}$ или $A_{n,\varepsilon} \subset B_{n,\delta}$. Используя тот факт, что

$$\mathsf{P}(B_{n,\delta}) \le \sum_{i=1}^s \mathsf{P}(B_{n,\delta}^{(i)}) \xrightarrow{n\to\infty} 0$$

получаем, что

$$\mathsf{P}(A_{n,\varepsilon}) \le \mathsf{P}(B_{n,\delta}) \xrightarrow{n\to\infty} 0$$

■

Лемма 3.11 (Слуцкого).

1. $X_n \xrightarrow[n\to\infty]{\mathrm{d}} X$, $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} 0 \Rightarrow X_n + Y_n \xrightarrow[n\to\infty]{\mathrm{d}} X$.
2. $X_n \xrightarrow[n\to\infty]{\mathrm{d}} X$, $Z_n \xrightarrow[n\to\infty]{\mathrm{d}} 1 \Rightarrow X_n Z_n \xrightarrow[n\to\infty]{\mathrm{d}} X$, $X_n/Z_n \xrightarrow[n\to\infty]{\mathrm{d}} X$.

□ Если $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} 0$, $Z_n \xrightarrow[n\to\infty]{\mathrm{d}} 1$, то $Y_n \xrightarrow[n\to\infty]{\mathsf{P}} 0$, $Z_n \xrightarrow[n\to\infty]{\mathsf{P}} 1$. ■

Лемма 3.12. $\hat{\theta}_n$ *- оценка* θ*, имеет асимптотически нормальное распределение* $\mathcal{N}(\theta, \sigma(\theta)^2/n)$*. Пусть* $f(\theta)$ *— дифференцируемая ,* $f'(\theta) \neq 0$*.*

□ Запишем определение дифференцируемости f в точке θ

$$f(t) = f(\theta) + (t-\theta)(f'(\theta) + \gamma(t,\theta)).$$

Преобразуем выражение к

$$\frac{f(t) - f(\theta)}{f'(\theta)} = (t-\theta)\left(1 + \frac{\gamma(t,\theta)}{f'(\theta)}\right)$$

Заменим t на $\hat{\theta}_n$ и умножим выражения на $\sqrt{n}/\sigma(\theta)$

$$\sqrt{n}\frac{f(\hat{\theta}_n) - f(\theta)}{f'(\theta)\sigma(\theta)} = \sqrt{n}\frac{\hat{\theta}_n - \theta}{f'(\theta)\sigma(\theta)}\left(1 + \frac{\gamma(\hat{\theta}_n,\theta)}{f'(\theta)}\right)$$

Так как $\sqrt{n}(\hat{\theta}_n - \theta)/\sigma(\theta) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$, а $\gamma(\hat{\theta},\theta) \xrightarrow[n\to\infty]{\mathsf{P}} 0$, то по свойствам сходимости $f(\hat{\theta}_n)$ является асимптотически нормальной оценкой $f(\theta)$ с асимптотической СКО $|f'(\theta)|\sigma(\theta)$.

■

3.8 Методы получения оценок

Пример 8.1. Схема Бернулли. $p = \mathsf{M}\xi_i$ — параметр (вероятность удачи); ξ_i — 1 или 0. $\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ — первый выборочный момент. $\widehat{p} = \overline{x}$ — оценка p по методу моментов. Эта оценка является несмещенной, так как

$$\mathsf{M}\overline{\xi} = \frac{1}{n}(\mathsf{M}\xi_1 + \ldots + \mathsf{M}\xi_n) = \frac{1}{n}(\underbrace{p + \ldots + p}_{n}) = p.$$

По закону больших чисел $\overline{\xi}$ сходится к p по вероятности. Информация Фишера одного испытания равна

$$I_1(p) = \mathsf{M}_p\left(\frac{\partial \ln\left(p^\xi(1-p)^{1-\xi}\right)}{\partial p}\right)^2 = \mathsf{M}_p\left(\frac{\xi}{p} - \frac{1-\xi}{1-p}\right)^2 = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)},$$

Поэтому $\mathsf{D}_p\xi = 1/I_1(p)$ и оценка эффективная.

3.8.1 Метод моментов

Рассматривается статистическая модель с s-мерным параметром $\theta = (\theta_1, \dots, \theta_s)$. $m_k(\theta) = \mathsf{M}\xi_i^k$ ($i = 1, \dots, n$, выборка повторная) — k-й момент (истинный); $\widehat{m}_k(\theta) = \frac{\xi_1^k + \dots + \xi_n^k}{n}$ — эмпирический момент. Предположим, что $\mathsf{M}\xi_i^s = m_s(\theta) < \infty$. Эмпирические моменты являются оценками для истинных. Запишем систему *моментных уравнений*:

$$\begin{cases} m_k(\theta_1, \dots, \theta_s) = \widehat{m}_k \\ 1 \leqslant k \leqslant s \end{cases}$$

Рассмотрим полученную систему относительно переменных $\theta_1, \dots, \theta_s$. Пусть существует единственное решение $\hat{\theta}_j = f_j(\widehat{m}_1, \dots, \widehat{m}_s)$ ($1 \leqslant j \leqslant s$). Мы получили некоторую оценку для θ. Следующая теорема сообщает, что оценка не совсем плохая.

Теорема 3.13 (О состоятельности статистических оценок, полученных методом моментов). *Пусть $\hat{\theta}_1, \dots, \hat{\theta}_s$ есть решение системы моментных уравнений и пусть функции f_j непрерывны. Тогда оценки $\hat{\theta}_j = f_j(\widehat{m}_1, \dots, \widehat{m}_s)$ для всех j являются состоятельными оценками параметров θ_j (т.е. $\hat{\theta}_j \xrightarrow[n\to\infty]{\mathsf{P}} \theta_j$ для всех j).*

□ В силу асимптотического свойства моментов имеет $\widehat{m}_k(\theta) \xrightarrow[n\to\infty]{\mathsf{P}} m_k(\theta)$. Так как f_j непрерывны, то они сохраняют сходимость по вероятности и $f_j(\widehat{m}_1(\theta), \dots, \widehat{m}_s(\theta)) \xrightarrow[n\to\infty]{\mathsf{P}} \theta_j$. ■

Пример 8.2. $\xi_1, \dots, \xi_n \sim \mathcal{N}(a, \sigma^2)$, $\mathsf{M}\xi_i = a$, $\mathsf{D}\xi_i = \sigma^2$. $\overline{x} = \widehat{a}$ — хорошая оценка. $\widehat{\sigma^2} = s^2 = \frac{1}{n}((x_1 - \overline{x})^2 + \dots + (x_n - \overline{x})^2)$ — выборочная дисперсия — смещённая оценка σ^2. Это оценки, полученные по методу моментов. А вот такая оценка: $\widehat{\widehat{\sigma^2}} = s^2 \frac{n}{n-1}$ является несмещённой.

3.8.2 Метод максимального (наибольшего) правдоподобия

Принцип МП в простейшем случае Принцип МП был рассмотрен ещё Гауссом в следующей форме: найдём такие значения параметров, чтобы вероятность получить данную выборку была максимальной (берём $\max\limits_{\theta\in\Theta} \mathsf{P}_\theta(\xi_1 = x_1, \dots, \xi_n = x_n) = \mathsf{P}_{\theta^*}(\xi_1 = x_1, \dots, \xi_n = x_n)$). θ^* — оценка МП (наиболее правдоподобное значение θ).

Общая ситуация $x_1, \dots, x_n$ — выборка в $(\mathscr{X}, \mathscr{A}, \mathscr{P})$, $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$. $L(\theta; x_1, \dots, x_n)$ — функция правдоподобия. Возьмём θ^* такое, что

$$\sup_{\theta\in\Theta} L(\theta; x_1, \dots, x_n) = L(\theta^*; x_1, \dots, x_n)$$

(если точная верхняя грань достигается). θ^* называется *оценкой максимального правдоподобия (ОМП)*.

Пример 8.3. $\xi_1, \dots, \xi_n$ взаимно независимы и равномерно распределены на $[0, \theta]$; $\theta \in \Theta = (0, +\infty)$. $L(\theta; x_1, \dots, x_n) = f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n)$, где $f_\theta(x) = \frac{1}{\theta}\mathrm{I}_{[0,\theta]}(x)$ — плотность равномерного распределения. $L \neq 0$ тогда и только тогда, когда для всех i $x_i \in [0, \theta]$.

$$L = \begin{cases} \frac{1}{\theta^n} & x_{(n)} \leqslant \theta \\ 0 & \text{в ином случае} \end{cases}$$

Ясно, что максимум L достигается при $\theta = \theta^* = x_{(n)} = \max\limits_i x_i$. θ^* — ОМП.

Предположим, что функция $L(\theta; x_1, \dots, x_n)$ дифференцируема по параметру θ ($\theta \in \Theta \subset \mathbb{R}$). Тогда ОМП можно найти, решая *уравнение правдоподобия*: $\frac{\partial L(\theta)}{\partial \theta} = 0$ (или переходим к логарифмам: $\frac{\partial \ln L}{\partial \theta} = 0$, если это возможно). В случае $\Theta \subset \mathbb{R}^k$ получаем систему уравнений.

Условия регулярности $(*)$:

1. Θ — невырожденный замкнутый интервал на $\mathbb{R}$; существуют $\frac{\partial \ln f_\theta}{\partial \theta}$, $\frac{\partial^2 \ln f_\theta}{\partial \theta^2}$, $\frac{\partial^3 \ln f_\theta}{\partial \theta^3}$ для любого $\theta \in \Theta$.
2. Для всех $\theta \in \Theta$:
$$\left|\frac{\partial \ln f_\theta}{\partial \theta}\right| \leqslant g_1(x), \qquad \left|\frac{\partial^2 \ln f_\theta}{\partial \theta^2}\right| \leqslant g_2(x), \qquad \left|\frac{\partial^3 \ln f_\theta}{\partial \theta^3}\right| \leqslant H(x),$$
где g_1 и g_2 интегрируемы на $\mathbb{R}$, а H обладает следующим свойством:
$$\mathsf{M}_\theta H = \int_{-\infty}^{+\infty} H(x) f_\theta(x)\, dx < M$$
(M не зависит от θ).

3.
$$0 < I_1(\theta) = \int\limits_{-\infty}^{+\infty} \left(\frac{\partial \ln f_\theta}{\partial \theta}\right)^2 f_\theta(x)\,dx < \infty.$$

Эти условия содержат условия Крамера – Рао.

Теорема 3.14 (о свойствах ОМП — теорема Дюге). *Пусть выполнены условия регулярности* $(*)$. *Зафиксируем* $\theta_0 \in \Theta$. *Пусть для любых* $\varepsilon > 0, \delta > 0$ *существует такой* N, *что для любых* $n > N$ *существует и единственно решение* θ^* *уравнение правдоподобия, принадлежащее интервалу* $(\theta_0 - \delta, \theta_0 + \delta)$ *с вероятностью не меньшей чем* $1 - \varepsilon$. *Тогда ОМП обладает следующими асимптотическими свойствами:*

1. *состоятельность;*
2. *асимптотическая нормальность;*
3. *асимптотическая эффективность.*

□
$$\frac{\partial \ln f_\theta(x)}{\partial \theta} = \{\text{разложение Тейлора}\} = \left.\frac{\partial \ln f_\theta}{\partial \theta}\right|_{\theta=\theta_0} + (\theta - \theta_0)\left.\frac{\partial^2 \ln f_\theta}{\partial \theta^2}\right|_{\theta=\theta_0} + \frac{1}{2}\tau(\theta-\theta_0)^2 H(x)$$

($x \in \mathbb{R}$, $|\tau| < 1$). Вспоминаем, что $L(\theta; x_1, \ldots, x_n) = f_\theta(x_1) \cdot \cdots \cdot f_\theta(x_n)$, и переписываем уравнение правдоподобия

$$\frac{1}{n}\frac{\partial \ln L(\theta; x_1, \ldots, x_n)}{\partial \theta} = 0$$

в виде:

$$\mathrm{F}(\theta) = B_0 + (\theta - \theta_0)B_1 + \frac{1}{2}\tau(\theta-\theta_0)^2 B_2 = 0,$$

где

$$B_0 = \frac{1}{n}\sum_{i=1}^n \left.\left(\frac{\partial \ln f_\theta(\xi_i)}{\partial \theta}\right)\right|_{\theta=\theta_0} \qquad B_1 = \frac{1}{n}\sum_{i=1}^n \left.\left(\frac{\partial^2 \ln f_\theta(\xi_i)}{\partial \theta^2}\right)\right|_{\theta=\theta_0} \qquad B_2 = \frac{1}{n}\sum_{i=1}^n H(\xi_i)$$

По ЗБЧ Хинчина (нужны н.о.р. сл. вел. с конечным матожиданием) $B_i \xrightarrow[n\to\infty]{\mathsf{P}} \mathsf{M}_\theta B_i$, поэтому найдём $\mathsf{M}_\theta B_i$.
В условиях регулярности меняем дифференцирование по θ и интегрирование:

$$\mathsf{M}_\theta \frac{\partial \ln f_\theta(\xi)}{\partial \theta} = \int\limits_{-\infty}^{+\infty} \frac{\partial \ln f_\theta(x)}{\partial \theta} f_\theta(x)\,dx = \int\limits_{-\infty}^{+\infty} \frac{1}{f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta} \cdot f_\theta(x)\,dx = \frac{\partial}{\partial \theta} \underbrace{\int\limits_{-\infty}^{+\infty} f_\theta(x)\,dx}_{=1} = 0$$

Отсюда $\mathsf{M}_\theta B_0 = \frac{1}{n}\sum\limits_{i=1}^n \mathsf{M}_\theta \frac{\partial \ln f_\theta(\xi_i)}{\partial \theta} = 0$.

Утверждение $\mathsf{M}_\theta B_1 = -I_1(\theta)$ вытекает из следующей выкладки[8]:

$$\mathsf{M}_\theta \frac{\partial^2 \ln f_\theta(\xi)}{\partial \theta^2} = \int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial \theta}\frac{\partial \ln f_\theta(x)}{\partial \theta} f_\theta(x)\,dx = \int\limits_{-\infty}^{+\infty} \frac{\partial}{\partial \theta}\left(\frac{1}{f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta}\right) f_\theta(x)\,dx =$$
$$= \int\limits_{-\infty}^{+\infty} \frac{-1}{f_\theta^2(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta} \cdot \frac{\partial f_\theta(x)}{\partial \theta} f_\theta(x)\,dx + \int\limits_{-\infty}^{+\infty} \frac{\partial^2 f_\theta(x)}{\partial \theta^2} \cdot \frac{1}{f_\theta(x)} f_\theta(x)\,dx = -\int\limits_{-\infty}^{+\infty} \left(\frac{\partial f_\theta(x)}{\partial \theta}\right)^2 \frac{1}{f_\theta(x)}\,dx +$$
$$+ \underbrace{\frac{\partial^2}{\partial \theta^2}\int\limits_{-\infty}^{+\infty} f_\theta(x)\,dx}_{=0} = -\int\limits_{-\infty}^{+\infty} \left(\frac{\partial \ln f_\theta(x)}{\partial \theta} f_\theta(x)\right)^2 \frac{1}{f_\theta(x)}\,dx = -\int\limits_{-\infty}^{+\infty} \left(\frac{\partial \ln f_\theta(x)}{\partial \theta}\right)^2 f_\theta(x)\,dx =$$
$$= -\mathsf{M}\left(\frac{\partial \ln f_\theta(\xi)}{\partial \theta}\right)^2 = -I_1(\theta).$$

[8] Лектор на неё забил, но я всё же приведу её. — *примеч. С. К.*

Про B_2 нам известно из условий регулярности, что $\mathsf{M}_\theta B_2 = \mathsf{M}_\theta H(\xi_i) < M$.

Докажем асимптотическую нормальность и асимптотическую эффективность. Имеем $\mathrm{F}(\theta^*) = 0$. Отсюда

$$\theta^* - \theta_0 = \frac{B_0}{-B_1 - \frac{1}{2}\tau(\theta^* - \theta_0)B_2}.$$

Далее ($k^2 = I_1(\theta)$),

$$\frac{\theta^* - \theta_0}{\frac{1}{\sqrt{nk^2}}} = \frac{\frac{1}{\sqrt{nk^2}} \sum\limits_{i=1}^{n} \frac{\partial \ln f(\xi_i)}{\partial \theta}}{-\frac{B_1}{k^2} - \frac{1}{2}\tau \frac{\theta^* - \theta_0}{k^2} B_2}.$$

$Y_i = \frac{\partial \ln f(\xi_i)}{\partial \theta}$ — независимые и одинаково распределённые с $\mathsf{M}_\theta Y_i = 0$ и $\mathsf{D}_\theta Y_i = \mathsf{M}_\theta Y_i^2 = k^2$. В силу ЦПТ числитель

$$\frac{1}{\sqrt{nk^2}} \sum_{i=1}^{n} \frac{\partial \ln f(\xi_i)}{\partial \theta} \xrightarrow[n\to\infty]{\mathrm{d}} Z \sim \mathcal{N}(0,1).$$

$B_1 \xrightarrow{\mathsf{P}} -k^2$, $B_2 \xrightarrow{\mathsf{P}} \mathsf{M}H(\xi_1) < M$, откуда знаменатель

$$-\frac{B_1}{k^2} - \frac{1}{2}\tau \frac{\theta^* - \theta_0}{k^2} B_2 \xrightarrow{\mathsf{P}} 1$$

(второе слагаемое оценивается через $\frac{1}{2}\frac{M}{k^2}(\theta^* - \theta_0)$, а $\theta^* - \theta_0 < \delta \to 0$). Тогда (в силу леммы Слуцкого 3.11) $\sqrt{nk^2}(\theta^* - \theta_0)$ асимптотически нормальна с $\mathcal{N}(0,1)$. Отсюда θ^* асимптотически нормальна с $\mathcal{N}(\theta_0, \frac{1}{nk^2})$, а т. к. асимптотическая дисперсия $nk^2 = I_n(\theta)$, то θ^* асимптотически эффективна (в рамках асимптотической нормальности). ■

Задача 3.2. *Пусть $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ — регулярное параметрическое семейство оценка $\hat\theta$, то $\hat\theta$ есть ОМП.*

4 Байесовские статистические оценки

4.1 Байесовские точечные оценки

4.1.1 Функция риска

Рассмотрим функцию u, называемую функцией штрафа (потерь), обладающую следующими свойствами:

1. u чётна;
2. u возрастает на $(0; +\infty)$;
3. $u(0) = 0$.

Функцией риска оценки δ параметра θ называется функция $R(\theta; \delta) = \mathsf{M}\big(u(\delta - \theta)\big)$. В частном случае $u(x) = x^2$ (обычно рассматриваемом на практике) функция риска $R(\theta; \delta) = \mathsf{M}(\delta - \theta)^2$ называется *квадратичной функцией риска*. В теории эффективных оценок статистики сравниваются в смысле равномерной минимизации риска.

Пример 1.1. В схеме Бернулли оцениваем параметр p. Эффективная оценка — частота $\hat p = \frac{1}{n}(x_1 + \dots + x_n)$. Для неё функция риска (квадратичная) равна $R_1 = R(p; \hat p) = \frac{p(1-p)}{n}$. Пусть теперь $n = 2$ (выборка состоит из двух элементов x_1, x_2). Рассмотрим странную оценку $\hat{\hat p} = \frac{1}{2}$ для всех значений x_i. Её функция риска есть $R_2 = R(p, \hat{\hat p}) = (\frac{1}{2} - p)^2$. Если нам откуда-то известно (имеется априорная информация), что $\frac{1}{3} \leqslant p \leqslant \frac{2}{3}$, то странная оценка $\hat{\hat p}$ оказывается лучше.

4.1.2 Байесовский подход

Параметр θ рассматривается как случайная величина $\boldsymbol{\theta}$ со значениями $\theta \in \Theta$ и распределением вероятностей Π на измеримом пространстве $(\Theta, \mathscr{B}_\Theta)$. Распределение Π называется *априорным распределением* параметра.

Усредним функцию риска $R(\theta; \delta)$ по распределению параметра:

$$R(\Pi; \delta) = \int\limits_{\Theta} R(\theta; \delta)\, \Pi(d\theta).$$

$R(\Pi; \delta)$ называется *априорным риском*. Оценка δ^*, минимизирующая априорный риск, называется *байесовской оценкой* параметра θ: $R(\Pi; \delta^*) = \inf\limits_{\delta \in \Delta} R(\Pi; \delta)$ (Δ — множество оценок). Можно схитрить: взять априорное распределение так, чтобы получить лучшую оценку.

4.2 Теорема о байесовской оценке для квадратичной функции риска

Рассмотрим функцию $\mathsf{P}_\theta(A)$, зависящую от $A \in \mathscr{A}$ и $\theta \in \Theta$. При каждом фиксированном A $\mathsf{P}_\theta(A)$ есть $\mathscr{B}_\Theta$-измеримая функция от θ, а при каждом фиксированном θ — задаёт распределение вероятностей на $(\mathscr{X}, \mathscr{A})$. Рассмотрим на декартовом произведении $\Theta \times \mathscr{X}$ σ-алгебру, порождённую прямоугольниками $B \times A$, где $B \in \mathscr{B}_\Theta$, $A \in \mathscr{A}$. На этих прямоугольниках введём меру Q:

$$Q(B \times A) = \int\limits_B \mathsf{P}_\theta(A)\,\Pi(d\theta).$$

Эта мера σ-аддитивна на полукольце прямоугольников. Значит, она продолжается по Лебегу на всю σ-алгебру, которую мы назовём $\mathscr{B}_\Theta \times \mathscr{A}$. Будем рассматривать квадратичную функцию потерь и предполагать, что $R(\Pi; \delta) = \mathsf{M}_Q(\boldsymbol{\theta} - \delta(\xi))^2 < \infty$.

Теорема 4.1 (о байесовской оценке для квадратичной функции риска). *Пусть априорное распределение* Π *и оценка* $\delta(x)$ *таковы, что* $R(\Pi; \delta) = \mathsf{M}_Q(\boldsymbol{\theta} - \delta(\xi))^2 < \infty$. *Тогда байесовской оценкой параметра* θ *является* $\delta^*(\xi) = \mathsf{M}_Q(\boldsymbol{\theta} \mid \xi)$.

$\square$

1. δ^* является статистикой, т.к. она зависит от $\boldsymbol{\theta}$ посредством распределения, а оно фиксировано.
2. $\mathsf{M}_Q\big((\boldsymbol{\theta} - \delta(\xi))^2 \mid \xi\big)$, $\mathsf{M}_Q(\boldsymbol{\theta}^2 \mid \xi)$ и $\mathsf{M}_Q(\boldsymbol{\theta} \mid \xi)$ существуют.
3. $\mathsf{M}_Q(\boldsymbol{\theta} - \delta^*(\xi))^2 = \mathsf{M}_Q\big(\mathsf{M}_Q((\boldsymbol{\theta} - \delta^*(\xi))^2 \mid \theta)\big) \le \mathsf{M}_Q\big(\mathsf{M}_Q((\boldsymbol{\theta} - \delta(\xi))^2 \mid \theta)\big) = \mathsf{M}_Q(\boldsymbol{\theta} - \delta(\xi))^2$ в силу свойств УМО ($\delta^* = \mathsf{M}_Q(\boldsymbol{\theta} \mid \xi)$). Равенство — при $\delta^*(\xi) = \mathsf{M}_Q(\boldsymbol{\theta} \mid \xi)$.

$\blacksquare$

Следствие 4.1 (единственность байесовской оценки). *Пусть* δ_1^*, δ_2^* *— две байесовские оценки (относительно квадратичной функции потерь). Тогда* $Q(\delta_1^* \neq \delta_2^*) = 0$.

$\square$ Положим $C := \{(\theta, x) \mid \delta_1^*(x) \neq \delta_2^*(x)\}$. Так как значение δ_i^* не зависят от параметра, то $C = \Theta \times A$, где A — некоторое подмножество $\mathscr{A}$. Так как УМО определено с точностью до меры нуль, то

$$Q(C) = \int_\Theta \mathsf{P}_\theta(A)\,\Pi(d\theta) = \int_\Theta 0\,\Pi(d\theta) = 0.$$

$\blacksquare$

4.3 Апостериорный риск

Пусть семейство $\mathscr{P}$ абсолютно непрерывно относительно лебеговой меры. Тогда для любого события $A \in \mathscr{A}$

$$\mathsf{P}_\theta(A) = \int\limits_A p_\theta(x)\,dx.$$

Утверждение 4.2. *Пусть* δ^* *— байесовская оценка относительно квадратичной функции потерь. Тогда* P_θ*-п. н. имеем*

$$\delta^* = \frac{\int\limits_\Theta \theta p_\theta(x)\,\Pi(d\theta)}{\int\limits_\Theta p_\theta(x)\,\Pi(d\theta)}.$$

$\square$ Пусть $A \in \mathscr{A}$ — произвольное событие. Положим $C = \Theta \times A \in \mathscr{B}_\Theta \times \mathscr{A}$. По определению УМО и теореме о байесовской оценке

$$\int\limits_C \delta^*\,dQ = \mathsf{M}_Q(I_{\xi\in A}\mathsf{M}_Q(\theta \mid \xi)) = \mathsf{M}_Q(\mathsf{M}_Q(I_{\xi\in A}\theta \mid \xi)) = \mathsf{M}_Q(I_{\xi\in A}\theta) = \int\limits_C \boldsymbol{\theta}\,dQ.$$

Вспоминая определение меры Q и теорему Фубини, получаем:

$$\int\limits_A \delta^*(x)\left(\int\limits_\Theta p_\theta(x)\,\Pi(d\theta)\right)dx = \int\limits_C \delta^*\,dQ = \int\limits_C \boldsymbol{\theta}\,dQ = \int\limits_A \left(\int\limits_\Theta \theta p_\theta(x)\,\Pi(d\theta)\right)dx,$$

откуда в силу произвольности выбора A получаем, что

$$\delta^*(x)\int\limits_\Theta p_\theta(x)\,\Pi(d\theta) = \int\limits_\Theta \theta p_\theta(x)\,\Pi(d\theta) \quad \mathsf{P}_\theta\text{-п.н.}$$

$\blacksquare$

Если априорное распределение Π имеет плотность $\pi(\theta)$, то

$$\delta^*(x) = \frac{\int\limits_{\Theta} \theta p_\theta(x)\pi(\theta)\,d\theta}{\int\limits_{\Theta} p_\theta(x)\pi(\theta)\,d\theta} = \int\limits_{\Theta} \theta q(\theta \mid x)\,d\theta,$$

где

$$q(\theta \mid x) = \frac{p_\theta(x)\pi(\theta)}{\int\limits_{\Theta} p_\theta(x)\pi(\theta)\,d\theta}$$

называется *апостериорной плотностью*[9] (плотностью апостериорного распределения параметра). Это условное распределение $\boldsymbol{\theta}$ при условии ξ. Последняя формула называется формулой Байеса для плотностей.

Замечание. Байесовская оценка минимизирует апостериорный риск $\mathsf{M}_Q((\boldsymbol{\theta}-\delta(\xi))^2 \mid \xi)$ (без доказательства).

Пример 3.1. Рассмотрим биномиальное распределение $\xi_i \sim \mathrm{Binom}(m_i, p)$. Оценим параметр p для априорного распределения $\pi \sim \mathrm{Beta}(\alpha, \beta)$ с квадратичной функцией потери. Из выше доказанной теоремы следует, что

$$\delta^*(x) = \int_0^1 pq(p \mid x)dp,\ q(p \mid x) = \frac{f_p(x)\pi(p)}{\int_0^1 f_p(x)\pi(p)dp} = \frac{f_{1,p}(x_1)\cdots f_{n,p}(x_n)\pi(p)}{\int_0^1 f_{1,p}(x_1)\cdots f_{n,p}(x_n)\pi(p)dp}.$$

Пусть $k := \sum_{i=1}^n x_i$, тогда

$$\int_0^1 f_{1,p}(x_1)\cdots f_{n,p}(x_n)\pi(p)dp = \int_0^1 C_{m_1}^{x_1} p^{x_1}(1-p)^{1-x_1}\cdots C_{m_n}^{x_n} p^{x_n}(1-p)^{1-x_n}\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}dp =$$
$$C_{m_1}^{x_1}\cdots C_{m_n}^{x_n}\int_0^1 p^k(1-p)^{n-k}\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}dp = C_{m_1}^{x_1}\cdots C_{m_n}^{x_n}\frac{B(k+\alpha, n-k+\beta)}{B(\alpha,\beta)},$$

откуда апостериорное распределение равно

$$q(p \mid x) = \frac{C_{m_1}^{x_1}\cdots C_{m_n}^{x_n} p^{k+\alpha-1}(1-p)^{n-k+\beta-1}\mathrm{Beta}(\alpha,\beta)}{\mathrm{Beta}(\alpha,\beta)C_{m_1}^{x_1}\cdots C_{m_n}^{x_n} B(k+\alpha, n-k+\beta)} = \frac{p^{k+\alpha-1}(1-p)^{n-k+\beta-1}}{B(k+\alpha, n-k+\beta)}.$$

Заключаем

$$\delta^*(x) = \int_0^1 p\frac{p^{k+\alpha-1}(1-p)^{n-k+\beta-1}}{B(k+\alpha,n-k+\beta)}dp = \frac{B(k+\alpha+1, n-k+\beta)}{B(k+\alpha, n-k+\beta)} = \frac{\Gamma(n+\alpha+\beta)\Gamma(k+\alpha+1)\Gamma(n-k+\beta)}{\Gamma(k+\alpha)\Gamma(n-k+\beta)\Gamma(n+\alpha+\beta+1)} =$$
$$\frac{k+\alpha}{n+\alpha+\beta} = \frac{\sum_{i=1}^n x_i + \alpha}{n+\alpha+\beta}.$$

Пример 3.2. Рассмотрим нормальное распределение $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$, где σ известно. Оценим параметр μ для априорного распределения $\pi \sim \mathcal{N}(\mu_0, \sigma_0^2)$ с квадратичной функцией потери. Из выше доказанной теоремы следует, что

$$\delta^*(x) = \int_{-\infty}^{\infty} \mu q(\mu \mid x)d\mu,\ q(\mu \mid x) = \frac{f_\mu(x)\pi(\mu)}{\int_{-\infty}^{+\infty} f_\mu(x)\pi(\mu)d\mu} = \frac{f_{1,\mu}(x_1)\cdots f_{n,\mu}(x_n)\pi(\mu)}{\int_{-\infty}^{+\infty} f_{1,\mu}(x_1)\cdots f_{n,\mu}(x_n)\pi(\mu)d\mu}.$$

Распишем знаменатель:

$$\int_{-\infty}^{+\infty} f_{1,\mu}(x_1)\cdots f_{n,\mu}(x_n)\pi(\mu)d\mu = \int_{-\infty}^{+\infty}\frac{1}{(\sigma\sqrt{2\pi})^n}e^{-\sum_{i=1}^n (x_i-\mu)^2/(2\sigma^2)}\frac{1}{\sigma_0\sqrt{2\pi}}e^{-(\mu-\mu_0)^2/(2\sigma_0^2)}d\mu.$$

Под экспонентой стоит многочлен от μ. Выделим полный квадрат:

$$-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\sigma_0^2} = -\frac{1}{2\sigma^2\sigma_0^2}\left((n\sigma_0^2+\sigma^2)\mu^2 - 2\left(\sigma_0^2\sum_{i=1}^n x_i + \sigma^2\mu_0\right)\mu + \sigma_0^2\sum_{i=1}^n x_i^2 + \sigma^2\mu_0^2\right) =$$
$$-\frac{n\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}\left(\left(\mu - \frac{\sigma_0^2\sum_{i=1}^n x_i + \sigma^2\mu_0}{n\sigma_0^2+\sigma^2}\right)^2 - b\right) = C\left(-(\mu-a)^2/2 + b\right).$$

[9] Для многих распределений существуют априорные и соответствующие им апостериорные распределения, входящие в одни и те же семейства. Такие априорные распределения называют *сопряженными*. Таблицу сопряженных распределений можно посмотреть на википедии. — *примеч. В. Х.*

Подставляя в интеграл и делая замену переменных, получаем

$$D \int_{-\infty}^{+\infty} \exp\left\{C\left(-(\mu-a)^2/2+b\right)\right\} d\mu = De^{Cb} \int_{-\infty}^{+\infty} e^{C-\widetilde{\mu}^2/2} d\widetilde{\mu} = De^{Cb}\sqrt{2\pi/C}.$$

Апостериорная плотность равна

$$q(\mu \mid x) = \frac{D\exp\left\{C(-(\mu-a)^2/2+b)\right\}}{De^{Cb}\sqrt{2\pi C}} = \frac{1}{\sqrt{2\pi/C}} e^{-C(\mu-a)^2/2},$$

где

$$a = \frac{\sigma_0^2 \sum_{i=1}^n x_i + \sigma^2 \mu_0}{n\sigma_0^2 + \sigma^2},\ C = \frac{n\sigma_0^2+\sigma^2}{2\sigma^2\sigma_0^2}.$$

По формуле плотности видно, что апостериорная плотность суть плотность нормального распределения с параметрами a, C^{-1} (C^{-1} - дисперсия), поэтому байесовская оценка как матожидание случайной величины с апостериорной плотностью равна

$$\delta^*(x) = a = \frac{\sigma_0^2 \sum_{i=1}^n x_i + \sigma^2 \mu_0}{n\sigma_0^2 + \sigma^2}.$$

4.4 Свойства байесовских оценок

1. Байесовская оценка смещена для любого априорного распределения Π, то есть $\mathsf{M}_\theta \delta^*(\xi) \neq \theta$ для любого C такого, что $\Pi(C) > 0$.
2. Апостериорное распределение параметра относительно любого априорного распределения является функцией достаточной статистики. Действительно, пусть $T(x)$ — достаточная статистика. Тогда по теореме Неймана – Пирсона $p_\theta(x) = g(\theta; T(x))h(x)$. При условии существования плотности π априорного распределения для плотности $q(\theta \mid x)$ условного распределения имеет место следующее:

$$q(\theta \mid x) = \frac{p_\theta(x)\pi(\theta)}{\int\limits_\Theta p_\theta(x)\pi(\theta)\,d\theta} = \frac{g(\theta; T(x))\pi(\theta)}{\int\limits_\Theta g(\theta; T(x))\pi(\theta)\,d\theta} = f(T(x)),$$

   т.е. плотность апостериорного распределения есть функция достаточной статистики.
3. При некоторых дополнительных условиях на Π (например, если $\pi(\theta) \neq 0$ при всех $\theta \in \Theta$) верно и обратное: если апостериорное распределение $\boldsymbol{\theta}$ зависит от x посредством статистики $T(x)$, то $T(x)$ является достаточной статистикой. Таким образом мы получили байесовский критерий достаточности статистики.

4.5 Минимаксные оценки

Пусть теперь нет ни оптимальной оценки, ни априорной информации. $R_M(\delta) = \sup\limits_{\theta\in\Theta} R(\theta; \delta)$ — наихудшее значение функции риска.

Определение. δ^* называется *минимаксной* оценкой θ, если $R_M(\delta^*) = \inf\limits_{\delta\in\Delta} R_M(\delta)$ (т.е. она минимизирует максимальное значение функции риска).

Теорема 4.3. *Пусть задано априорное распределение* Π. *Если* δ^* *является байесовской и* $R(\Pi, \delta^*) = \mathrm{const}$, *то* δ^* *— минимаксна.*

□ Если δ^* такая, как в теореме, но не является минимаксной, то найдется такая оценка δ, что

$$\sup_{\theta\in\Theta} R(\theta, \delta) < \sup_{\theta\in\Theta} R(\theta, \delta^*) = R(\theta, \delta^*).$$

Тогда при всех $\theta \in \Theta$ справедливо неравенство

$$R(\theta, \delta) < R(\theta, \delta^*).$$

Но поскольку δ^* – байесовская, то

$$\int_\Theta R(\theta, \delta^*)\pi(\theta) d\theta \leq \int_\Theta R(\theta, \delta)\pi(\theta) d\theta.$$

Эти два неравенства противоречат друг другу, поэтому предположение было неверно и δ^* – минимаксная. ■

Пример 5.1. Рассмотрим $\xi_i \sim \mathrm{Bern}(p)$, $p \in (0,1)$. Найдём ОМП в данной модели.

$$L(p; x_1, \dots, x_n) = p^m (1-p)^{n-m} I_{x \in \{0,1\}^n},$$

где $m = \sum_{i=1}^n x_i$. Уравнение ОМП

$$\frac{\partial L(p;x)}{\partial p} = (m(1-p) + (n-m)p)p^{m-1}(1-p)^{n-m-1} I_{x \in \{0,1\}^n} = 0,$$

откуда $\hat{p} = m/n$. По уже доказанному частота является эффективной оценкой параметра p. Но эта оценка не является минимаксной.

Минимаксной оценкой для схемы Бернулли будет оценка Ходжеса-Лемана

$$\widetilde{p} = \overline{x} + \frac{1/2 - \overline{x}}{1 + \sqrt{n}}.$$

□ Посчитаем риск оценки

$$R(p, \hat{p}) = \mathsf{E}_p (p - \hat{p})^2 = p^2 - 2p \mathsf{E}_p \left(\frac{1/2 + \sqrt{n}\overline{X}}{1 + \sqrt{n}} \right) + \mathsf{E}_p \left(\frac{1/2 + \sqrt{n}\overline{X}}{1 + \sqrt{n}} \right)^2 =$$
$$p^2 - \frac{p + 2\sqrt{n}p^2}{1 + \sqrt{n}} + \frac{1/4 + \sqrt{n}p + p(1-p) + np^2}{(1 + \sqrt{n})^2} = \frac{1}{4(1 + \sqrt{n})^2}.$$

Как мы видим, риск не зависит от параметра p. Для квадратичной функции потери мы знаем вид байесовской оценки. Подберём априорное распределение π так, чтобы оценка $\widetilde{p}$ стала байесовской.

$$\frac{m + \sqrt{n}/2}{n + \sqrt{n}} = \widetilde{p} = \frac{\int_0^1 p^{1+m}(1-p)^{n-m}\pi(p)dp}{\int_0^1 p^m (1-p)^{n-m}\pi(p)dp},$$

где $m = \sum_{i=1}^n x_i$. Как можно заметить, стоит подбирать априорную плотность π как плотность случайной величины, имеющей распределение $\mathrm{Beta}(\alpha, \beta)$. Тогда последнее равенство преобразуется к

$$\frac{m + \sqrt{n}/2}{n + \sqrt{n}} = \frac{B(1 + m + \alpha, n - m + \beta)}{B(m + \alpha, n - m + \beta)} = \frac{\Gamma(1 + m + \alpha)\Gamma(n - m + \beta)\Gamma(n + \alpha + \beta)}{\Gamma(1 + n + \alpha + \beta)\Gamma(m + \alpha)\Gamma(n - m + \beta)} = \frac{m + \alpha}{n + \alpha + \beta}.$$

Тогда для $\alpha = \beta = \sqrt{n}/2$ оценка Ходжеса-Лемана является байесовской. Так как риск постоянной, то оценка является минимаксной. ■

4.6 Байесовские интервальные оценки

Предположим, что распределения P_θ и Π абсолютно непрерывны относительно лебеговой меры, т.е. существуют плотности $p_\theta(x)$ и $\pi(\theta)$. $q(\theta \mid x)$ — апостериорная плотность. Возьмём произвольное $\alpha \in (0,1)$ и произвольным образом разобьём его на две части: $\alpha = \alpha_1 + \alpha_2$, $\alpha_{1,2} \in (0,1)$. Найдём α_1- и $(1-\alpha_2)$-квантили апостериорного распределения и обозначим их $\underline{\theta} = \underline{\theta}(\alpha_1, x)$ и $\overline{\theta} = \overline{\theta}(\alpha_2, x)$ соответственно:

$$\int\limits_{-\infty}^{\underline{\theta}} q(\theta \mid x)\, d\theta = \alpha_1, \qquad \int\limits_{-\infty}^{\overline{\theta}} q(\theta \mid x)\, d\theta = 1 - \alpha_2.$$

$$\mathsf{P}(\underline{\theta} \leqslant \boldsymbol{\theta} \leqslant \overline{\theta}) = \int\limits_{\underline{\theta}}^{\overline{\theta}} q(\theta \mid x)\, d\theta = 1 - \alpha_2 - \alpha_1 = 1 - \alpha,$$

где вероятность рассматривается в смысле меры Q (и по априорному распределению $\boldsymbol{\theta}$, и по распределениям из семейства $\mathscr{P}$). Получили интервал $[\underline{\theta}, \overline{\theta}]$ для параметра — это и есть *байесовская интервальная оценка* (этот интервал определяется по α неоднозначно). α — вероятность того, что $\boldsymbol{\theta}$ не попадёт в интервал (вероятность ошибки при использовании этого интервала).

5 Многомерное нормальное распределение

5.1 Эквивалентные определения

Пусть $\xi = (\xi_1, \ldots, \xi_n)^T$ — случайный вектор[10], $a = (a_1, \ldots, a_n)^T$ — некоторый постоянный вектор, $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, A — положительно определённая матрица $n \times n$.

Первое определение. Если плотность (совместного) распределения вектора ξ имеет вид

$$p_\xi(x, a, A) = \frac{1}{\left(\sqrt{2\pi}\right)^n} (\det A)^{\frac{1}{2}} \exp\left(-\frac{1}{2}(x-a)^T A (x-a)\right),$$

то говорят, что ξ имеет *многомерное* (n-мерное) *нормальное распределение.* Вектор $a = \mathsf{M}\xi$ ($a_i = \mathsf{M}\xi_i$) называется средним значением, а матрица $\Sigma = \Sigma_\xi = A^{-1} = \mathsf{M}(\xi - a)(\xi - a)^T$ — *ковариационной матрицей.* $\Sigma = (\sigma_{ij})$, $\sigma_{ij} = \mathsf{M}(\xi_i - a_i)(\xi_j - a_j)$. Ковариационная матрица является квадратной, симметричной и положительно определённой. Распределение, заданной такой плотностью, обозначается $\mathcal{N}(a, \Sigma)$.

Второе определение. Определим многомерное нормальное распределение через характеристические функции. $t = (t_1, \ldots, t_n)^T \in \mathbb{R}^n$. $\varphi_\xi(t) = \mathsf{M}e^{it^T\xi}$ — *характеристическая функция* случайного вектора. Будем считать, что $\xi \sim \mathcal{N}(a, \Sigma)$, если

$$\varphi_\xi(t) = \exp\left(it^T a - \frac{1}{2} t^T \Sigma t\right).$$

В отличие от предыдущего определения, сюда включается случай вырожденной ковариационной матрицы.

Замечание. Характеристическая функция вычисляется по плотности, а по многомерной формуле обращения плотность однозначно восстанавливается по характеристической функции. Поэтому первое и второе определения равносильны (по теореме единственности).

Третье определение. Если любая линейная комбинация компонент вектора ξ имеет (одномерное) нормальное распределение, то ξ имеет многомерное нормальное распределение.

Третье определение будут использоваться в теории случайных процессов.

5.2 Свойства многомерного нормального распределения

1. $\xi \sim \mathcal{N}(a, \Sigma)$, $\eta = A\xi$. Тогда $\eta \sim \mathcal{N}(Aa, A\Sigma A^T)$.

   □ Найдём характеристическую функцию η:

   $$\phi_\eta(t) = \mathsf{M}e^{it^T\eta} = \mathsf{M}e^{it^T A\xi} = \phi_\xi(A^T t) = \exp\{i(A^T t)^T a - ((A^T t)^T \Sigma (A^T t)\} = \exp\{it^T(Aa) - t^T(A\Sigma A^T)t\}.$$

   В силу теоремы единственности получаем искомое. ■

2. $\xi_1, \ldots, \xi_n$ — нормальные, $\xi := (\xi_1, \ldots, \xi_n)$. Тогда $\xi_1, \ldots, \xi_n$ — независимые тогда и только тогда, когда Σ — диагональная.

   □ Слева направо очевидно так как из независимости следует некоррелированность. Справа налево пишем явно характеристическую функцию, видим, что переменные разделяются в сумму квадратов, по теореме единственности получаем, что это многомерное нормальное с диагональной матрицей. ■

3. $\xi \sim \mathcal{N}(a, \Sigma)$. Положим $Q := (\xi - a)^T \Sigma^{-1} (\xi - a)$. Тогда $Q \sim \chi^2_n$, где $\chi^2_n \sim Z_1^2 + \ldots + Z_n^2$, $Z_i^2 \sim \mathcal{N}(0, 1)$ — независимы.

   □ Так как Σ — положительная симметричная матрица, то существует такая матрица C, что $\Sigma = C^T C$. Тогда

   $$Q := (\xi - a)^T \Sigma^{-1} (\xi - a) = (C(\xi - a))^T (C(\xi - a)) = \eta^T \eta$$

   Заметим, что

   $$\eta \sim \mathcal{N}(C(a - a), C^T \Sigma C) = \mathcal{N}(0, C^T (C^{-1})^T C^{-1} C) = \mathcal{N}(0, E).$$

   Поэтому η имеет стандартное многомерное нормальное распределение, а Q — распределение хи-квадрат. ■

[10] Здесь и далее буква T обозначает транспонирование. Векторы-столбцы записаны как транспонированные строки для экономии места. — *примеч. С. К.*

5.3 Лемма Фишера

Лемма 5.1 (Фишер). *Пусть $\eta_1, \dots, \eta_n$ — независимые (одномерные) случайные величины с $\mathcal{N}(a, \sigma^2)$. Тогда:*

1. *существуют ортогональная матрица C и случайный вектор ξ такие, что $\xi = C\eta$, $\mathsf{M}\xi_1 = \sqrt{n}a$, $\mathsf{M}\xi_i = 0$ при $i > 1$ и $\mathsf{D}\xi_i = \sigma^2$ при всех i;*
2. *$\overline{\eta} = \dfrac{\eta_1 + \dots + \eta_n}{n}$ и $\sum\limits_{i=1}^{n} (\eta_i - \overline{\eta})^2$ независимы;*

□

1. Положим $c_{11} = c_{12} = \dots = c_{1n} := \dfrac{1}{\sqrt{n}}$. $\sum\limits_{j=1}^{n} c_{1j}^2 = 1$. $\xi_1 = \sum\limits_{j=1}^{n} c_{1j}\eta_j = \sqrt{n}\overline{\eta}$, $\mathsf{M}\xi_1 = \sqrt{n}a$, $\mathsf{D}\xi_1 = n\frac{\sigma^2}{n} = \sigma^2$.

   Остальные c_{ij} $(2 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant n)$ подбираются из условий: $\sum\limits_{j=1}^{n} c_{ij}^2 = 1$ (при всех $i > 1$) и $\frac{1}{\sqrt{n}} \sum\limits_{j=1}^{n} c_{ij} = 0$.

   Из этих условий получаем, что матрица $C = (c_{ij})$ ортогональна, и при $i > 1$ $\mathsf{M}\xi_i = 0$ и $\mathsf{D}\xi_i = \sigma^2$.
2. В силу ортогональности C имеем

$$\sum_{i=1}^{n} \xi_i^2 = \xi^T \xi = \eta^T C^T C \eta = \eta^T \eta = \sum_{i=1}^{n} \eta_i^2.$$

Более того, так как C – ортогональная матрица, то

$$\operatorname{cov} \eta = C \operatorname{cov} \xi C^T = C \sigma^2 E C^T = \sigma^2 E,$$

поэтому $\xi_1, \dots, \xi_n$ – независимы. Отсюда

$$\sum_{i=2}^{n} \xi_i^2 = \sum_{i=1}^{n} \eta_i^2 - \xi_1^2 = \sum_{i=1}^{n} \eta_i^2 - n\overline{\eta}^2 = \sum_{i=1}^{n} (\eta_i - \overline{\eta})^2,$$

то есть $\sum\limits_{i=1}^{n} (\eta_i - \overline{\eta})^2 = \sum\limits_{i=2}^{n} \xi_i^2$ — не зависит от $\overline{\eta} = \xi_1/\sqrt{n}$.

■

Замечание. Если $\eta_1, \dots, \eta_n$ независимы и имеют одинаковое нормальное распределение, то $\overline{\eta}$ и $\frac{1}{n} \sum (\eta_i - \overline{\eta})^2$ независимы. На самом деле это характеристическое свойство нормальных распределений.

5.4 Связанные с нормальным распределения

5.4.1 Хи-квадрат распределение

Пусть случайные величины $Z_1, \dots, Z_n$ взаимно независимы и имеют стандартное нормальное распределение. Тогда распределение случайной величины $Z_1^2 + \dots + Z_n^2$ называется *хи-квадрат* распределением с n степенями свободы и обозначается χ_n^2. При χ_2^2 имеет показательное распределение с параметром $\lambda = \frac{1}{2}$. Плотность хи-квадрат распределения даётся формулой:

$$p_{\chi_n^2}(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2} - 1} e^{-\frac{x}{2}}, \qquad \text{где } \Gamma(\lambda) = \int\limits_0^{\infty} x^{\lambda - 1} e^{-x} \, dx \text{ — гамма-функция Эйлера.}$$

$\mathsf{M}\chi_n^2 = n$, $\mathsf{D}\chi_n^2 = 2n$. Характеристическая функция: $\varphi_{\chi_n^2} = (1 - 2it)^{-\frac{n}{2}}$.

□ Выведем формулу плотности. Очевидно, что

$$p_{\chi_1^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}.$$

Для характеристической функции верно $\varphi_{\chi_1^2}(t) = (1 - 2it)^{-1/2}$. Тогда пользуясь независимостью слагаемых для хи-квадрат случайной величины, получаем

$$\varphi_{\chi_n^2}(t) = \frac{1}{(1 - 2it)^{n/2}}.$$

Заметим, что это есть характеристическая функция гамма распределения с параметрами $n/2, 2$. В силу теоремы единственности хи-квадрат распределена как $\Gamma(n/2, 2)$. ■

5.4.2 Распределение Стьюдента

Если $Z_0, Z_1, \ldots, Z_n$ взаимно независимы и имеют стандартное нормальное распределение, то распределение случайной величины

$$\frac{Z_0}{\sqrt{\dfrac{Z_1^2 + \cdots + Z_n^2}{n}}} = \frac{Z_0}{\sqrt{\chi_n^2/n}}$$

называется *распределением Стьюдента* (t-распределением) и обозначается t_n. Плотность распределения Стьюдента даётся формулой

$$p_{t_n}(x) = \frac{C_n}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \qquad \text{где } C_n = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}}.$$

При $n = 1$ распределение Стьюдента совпадает с распределением Коши, имеющим плотность $\dfrac{1}{\pi(1+x^2)}$.

$\square$ Рассмотрим случайную величину $\xi = (Z_0, \chi_n^2)^T$, где Z_0 имеет стандартное нормальное распределение, χ_n^2 имеет хи-квадрат распределение с n степенями свободы, Z_0 и χ_n^2 – независимы. Пусть

$$\eta = \begin{pmatrix} \frac{Z_0}{\sqrt{\chi_n^2/n}} \\ \chi_n^2 \end{pmatrix} = g\begin{pmatrix} Z_0 \\ \chi_n^2 \end{pmatrix} = g(\xi).$$

Плотность случайного вектора ξ нам известна. Найдём плотность вектора η по формуле замены плотности. Заметим, что

$$J_g(z, x) = \begin{pmatrix} \sqrt{n/x} & * \\ 0 & 1 \end{pmatrix},$$

откуда при $z = t\sqrt{x/n}$

$$f_\eta(t, x) = \frac{f_\xi(z, x)}{|J_g(z, x)|} = \sqrt{\frac{x}{n}} f_{Z_0}(t\sqrt{x/n}) f_{\chi_n^2}(x) = \sqrt{\frac{x}{n}} \frac{1}{\sqrt{2\pi}} e^{-xt^2/(2n)} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}.$$

Положим $D_n := (2^{n/2}\Gamma(n/2)\sqrt{2\pi n})^{-1}$. Первая координата вектора η обладает искомым распределением Стьюдента с n степенями свободы, и её плотность находится как

$$f_{t_n}(t) = \int_0^\infty f_\eta(t, x)dx = D_n \int_0^\infty x^{(n-1)/2} e^{-(t^2/n+1)x/2} dx = D_n \frac{2^{(n+1)/2}}{(1 + t^2/n)^{(n+1)/2}} \int_0^\infty y^{(n-1)/2} e^{-y} dy =$$

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

$\blacksquare$

5.4.3 Распределение Фишера

Распределение определяется как

$$F_{n,m} := \frac{\chi_n^2/n}{\chi_m^2/m},$$

где χ_n^2, χ_m^2 — хи-квадрат распределённые случайные величины. Говорят, что распределение Фишера имеет n и m степени свободы. Плотность:

$$p_{n,m}(x) = \begin{cases} \frac{1}{B(n/2, m/2)} (\frac{n}{m})^{n/2} x^{n/2-1} (1 + \frac{n}{m}x)^{-\frac{m+n}{2}}, & x \geqslant 0 \\ 0, & x \leq 0 \end{cases}$$

Функция распределения:

$$P(F_{m,n} \leq x) = \frac{I_{\frac{m}{n}x}(m/2, n/2)}{1 + \frac{m}{n}x}, \quad \text{где } I_x(a, b) \text{ - неполная } \beta\text{-функция с параметрами } a \text{ и } b.$$

$\square$ Рассмотрим случайную величину $\xi = (\chi_n^2, \chi_m^2)^T$, где χ_n^2 имеет хи-квадрат распределение с n степенями свободы, χ_m^2 имеет хи-квадрат распределение с m степенями свободы, χ_n^2 и χ_m^2 – независимы. Пусть

$$\eta = \begin{pmatrix} \frac{\chi_n^2/n}{\chi_m^2/m} \\ \chi_m^2 \end{pmatrix} = g\begin{pmatrix} \chi_n^2 \\ \chi_m^2 \end{pmatrix} = g(\xi).$$

Плотность случайного вектора ξ нам известна. Найдём плотность вектора η по формуле замены плотности. Заметим, что

$$J_g(x_1, x_2) = \begin{pmatrix} m/(nx_2) & * \\ 0 & 1 \end{pmatrix},$$

откуда при $x_1 = nx_2 y/m > 0$, $y > 0$

$$\begin{aligned} f_\eta(y, x_2) = \frac{f_\xi(x_1, x_2)}{|J_g(x_1, x_2)|} &= \frac{nx_2}{m} f_{\chi^2_n}(nx_2 y/m) f_{\chi^2_m}(x_2) = \\ &= \frac{nx_2}{m} \frac{1}{2^{n/2}\Gamma(n/2)} (nx_2 y/m)^{n/2-1} e^{-nx_2 y/(2m)} \frac{1}{2^{m/2}\Gamma(m/2)} x_2^{m/2-1} e^{-x_2/2} = \\ &= \frac{n^{n/2}}{m^{n/2} 2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} y^{n/2-1} x_2^{(n+m)/2-1} e^{-(ny/m+1)x_2/2}. \end{aligned}$$

Положим $D_n := n^{n/2}/(m^{n/2} 2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2))$. Первая координата вектора η обладает искомым распределением Фишера с n и m степенями свободы, и её плотность находится как

$$\begin{aligned} f_{F_{n,m}}(y) &= \int_0^\infty f_\eta(y, x_2) dx_2 = D_n y^{n/2-1} \int_0^\infty x_2^{(n+m)/2-1} e^{-(ny/m+1)x_2/2} dx_2 = \\ &= D_n y^{n/2-1} \frac{2^{(n+m)/2}}{(1+ny/m)^{(n+m)/2}} \int_0^\infty z^{(n+m)/2-1} e^{-z} dz = \frac{1}{B(n/2, m/2)} \left(\frac{n}{m}\right)^{n/2} y^{n/2-1} \frac{1}{(1+ny/m)^{(n+m)/2}}. \end{aligned}$$

■

5.5 Статистики для нормальных выборок

5.5.1 Распределение статистик для нормальных выборок

Утверждение 5.2. *Пусть $\eta_1, \ldots, \eta_n$ — независимые (одномерные) случайные величины с $\mathcal{N}(a, \sigma^2)$. Тогда:*

$$\frac{1}{\sigma^2}\sum_{i=1}^n (\eta_i - \overline{\eta})^2 \sim \chi^2_{n-1}, \qquad \frac{\sqrt{n}(\overline{\eta} - a)}{s} \sim t_{n-1}, \quad \text{где } s^2 = \frac{1}{n-1}\sum_{i=1}^n (\eta_i - \overline{\eta})^2.$$

□ $Z_i = \dfrac{\xi_i}{\sigma} \sim \mathcal{N}(0,1)$ при $i > 1$, откуда

$$\frac{1}{\sigma^2}\sum_{i=1}^n (\eta_i - \overline{\eta})^2 = \sum_{i=2}^n \frac{\xi_i^2}{\sigma^2} = Z_2^2 + \cdots + Z_n^2 \sim \chi^2_{n-1}, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^n (\eta_i - \overline{\eta})^2,$$

откуда $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$.

$\dfrac{\overline{\eta} - a}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$, $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ и эти случайные величины независимы по лемме Фишера, поэтому

$$\frac{\sqrt{n}(\overline{\eta} - a)}{s} = \frac{\frac{\overline{\eta}-a}{\sqrt{\sigma^2/n}}}{\sqrt{s^2/\sigma^2}} \sim t_{n-1}.$$

■

$x_1, \ldots, x_n$ — повторная выборка из $\mathcal{N}(a, \sigma^2)$.

5.5.2 Оценка дисперсии при неизвестном среднем

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})}{\sigma^2}$$

Пусть η_i как в доказательстве леммы Фишера, тогда

$$\frac{\sum_{i=1}^n (\xi_i - \overline{\xi})}{\sigma^2} = \frac{\sum_{i=2}^n \eta_i}{\sigma^2} \sim \chi^2_{n-1}.$$

5.5.3 Оценка среднего при неизвестной дисперсии

$\mathsf{M}\overline{\xi} = a$, $\mathsf{D}\overline{\xi} = \sigma^2/n$.

$$\frac{\overline{\xi} - a}{\sqrt{s^2/n}} \sim \frac{\sqrt{n}(\overline{\xi} - a)}{\sigma\sqrt{\chi^2_{n-1}/(n-1)}} \sim t_{n-1}.$$

6 Доверительные интервалы

$(\mathscr{X}, \mathscr{A}, \mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\})$ — параметрическая модель, θ — скалярный параметр (иначе интервалы заменяются более сложными областями). $(\xi_1, \ldots, \xi_n)$ — случайная выборка, $x = (x_1, \ldots, x_n)$ — её реализация. По заданному $\alpha \in (0, 1)$ и по выборке находим такие статистики $\underline{\theta}(\alpha; x) < \overline{\theta}(\alpha; x)$, что $\mathsf{P}_\theta([\underline{\theta}, \overline{\theta}] \ni \theta) \geqslant 1 - \alpha$. В этом случае интервал $[\underline{\theta}, \overline{\theta}]$ называется (точным) *доверительным интервалом* для θ с *доверительной вероятностью* (*доверительным уровнем*) $1 - \alpha$. α — вероятность ошибки при использовании данного интервала. $\underline{\theta}$ и $\overline{\theta}$ называются *доверительными границами*.

Доверительный интервал — частный случай интервальной оценки. В отличие от байесовского подхода здесь θ — не случайная величина; зато у Байеса границы постоянные, а у нас — случайные (зависят от выборки).

Сравнение двух доверительных интервалов производится по двум характеристикам:

1. по α (или по доверительной вероятности $1 - \alpha$);
2. по средней длине доверительного интервала.

6.1 Общие способы получения доверительных интервалов

6.1.1 Метод центральной статистики

Пусть $\xi_1, \ldots, \xi_n$ взаимно независимы и распределены с абсолютно непрерывным распределением вероятностей P_θ. Найдём функцию $S(\xi; \theta)$ — так называемую *центральную статистику*, хотя на самом деле она статистикой не является, — со следующими свойствами:

1. распределение $S(\xi; \theta)$ не зависит от θ и имеет плотность $f_S(y)$,
2. при каждом фиксированном $x \in \mathscr{X}$ функция $S(x; \theta)$ (как функция аргумента θ) является непрерывной и строго монотонной.

Зададим $\alpha \in (0, 1)$ и найдём $u_1 < u_2$ так, чтобы

$$\mathsf{P}(u_1 \leqslant S(\xi; \theta) \leqslant u_2) = \int\limits_{u_1}^{u_2} f_S(y)\, dy = 1 - \alpha.$$

(u_1 и u_2 находятся неоднозначно.) Заметим, что распределение вероятностей в правой части не зависит от параметра θ. Обратим неравенство относительно θ и получим, что $\mathsf{P}(T_1(\xi) \leqslant \theta \leqslant T_2(\xi)) = 1 - \alpha$, т. е. $[T_1(x), T_2(x)]$ есть доверительный интервал для θ с доверительной вероятностью $1 - \alpha$.

6.1.2 Метод монотонного преобразования

Пусть $T(x)$ — некоторая статистика, $G_\theta(t) = \mathsf{P}_\theta(T(\xi) \leqslant t)$ — её функция распределения. Пусть $G_\theta(t)$ при любом фиксированном θ непрерывна и строго монотонна относительно t и при любом фиксированном t является непрерывной строго убывающей функцией от θ. Тогда $\mathsf{P}_\theta(G_\theta(t_1) \leqslant G_\theta(T(\xi)) \leqslant G_\theta(t_2)) = \mathsf{P}_\theta(t_1 \leqslant T(\xi) \leqslant t_2) = G_\theta(t_2) - G_\theta(t_1)$ при любых $t_1 < t_2$. Если $G_\theta(t_1) = \alpha_1$, $G_\theta(t_2) = 1 - \alpha_2$, $\alpha_1 + \alpha_2 = \alpha$, то $\mathsf{P}_\theta(\alpha_1 \leqslant G_\theta(T) \leqslant 1 - \alpha_2) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha$. Положим $\overline{\theta}$: $G_{\overline{\theta}}(T(x)) = \alpha_1$; $\underline{\theta}$: $G_{\underline{\theta}}(T(x)) = \alpha_2$. Тогда $\mathsf{P}(\underline{\theta} \leqslant \theta \leqslant \overline{\theta}) = 1 - \alpha$, т. е. $[\underline{\theta}, \overline{\theta}]$ — доверительный интервал с вероятностью ошибки α.

6.2 Точный доверительный интервал для параметра биномиального распределения

Пусть дана случайная выборка $(\xi_1, \ldots, \xi_n)$ со значениями $x_1, \ldots, x_n$, ξ_k взаимно независимы и распределены следующим образом:

$$\xi_k = \begin{cases} 0 & \text{с вероятностью } 1 - p, \\ 1 & \text{с вероятностью } p, \end{cases}$$

где p — неизвестный параметр. $S_n = \sum\limits_{i=1}^{n} \xi_i$, $\mathsf{P}(S_n = m) = C_n^m p^m (1 - p)^{n-m}$.

Положим

$$f_m(p) := \sum_{k=0}^{m} C_n^k p^k (1-p)^{n-k} = 1 - \sum_{k=m+1}^{n} C_n^k p^k (1-p)^{n-k}.$$

Вычитаемое нам знакомо: раньше мы нашли, что для $\eta_1, \dots, \eta_n$ – взаимно независимых и равномерно распределенных на $[0,1]$ верно

$$\mathsf{P}(\eta_{(m)} \leqslant x) = \sum_{k \geqslant m} C_n^k x^k (1-x)^{n-k}.$$

При фиксированном $n > m = S_n$ это функция $f_m(p)$ от аргумента $p \in [0,1]$ является непрерывной и строго убывающей функцией (так как вычитаемое из 1 это $\mathsf{P}(\eta_{(m+1)} \leqslant p)$, которое очевидно строго возрастает по p), причём $f_m(0) = 1$, $f_m(1) = 0$. Поэтому для каждого $\beta \in (0,1)$ уравнение $f_m(p) = \beta$ имеет единственный корень $p = \overline{p} = \overline{p}_m(\beta)$. $\sum\limits_{k=m}^{n} C_n^k p^k (1-p)^{n-k} = 1 - f_{m-1}(p)$. Для каждого $\beta \in (0,1)$ уравнение $1 - f_{m-1}(p) = \beta$ имеет единственный корень $p = \underline{p} = \underline{p}_{m-1}(\beta)$.

Полагаем $\beta = \alpha/2$ и получаем, что

$$\mathsf{P}_p(p \geqslant \overline{p}) = \mathsf{P}_p(f_m(p) \leqslant \beta) = \mathsf{P}_p\left(\sum_{k=0}^{S_n} C_n^k p^k (1-p)^{n-k} \leqslant \beta\right) = \mathsf{P}_p\left(\sum_{k=0}^{S_n} \mathsf{P}_p(\widetilde{S}_n = k) \leqslant \beta\right) =$$

$$\sum_{j=0}^{n} \mathsf{P}_p\left(\sum_{k=0}^{S_n} \mathsf{P}_p(\widetilde{S}_n = k) \leqslant \beta, \mathsf{P}_p(S_n = j)\right) = \sum_{j\,:\, \sum\limits_{k \leqslant j} \mathsf{P}_p(S_n = k) \leqslant \beta} \mathsf{P}(S_n = j) \leqslant \beta,$$

то есть $\mathsf{P}(p > \overline{p}) \leqslant \beta$. Аналогично $\mathsf{P}(p < \underline{p}) \leqslant \beta$. Окончательно,

$$\mathsf{P}_p(\underline{p} \leqslant p \leqslant \overline{p}) \geqslant 1 - 2\beta = 1 - \alpha,$$

то есть интервал $[\underline{p}, \overline{p}]$ есть доверительный интервал для p с доверительной вероятностью $1 - \alpha$.

6.3 Доверительные интервалы для параметров нормального распределения

6.3.1 Доверительный интервал для среднего при известной дисперсии

$\xi_1, \dots, \xi_n$ — повторная выборка, $\xi \sim \mathcal{N}(a, \sigma^2)$, $a = \mathsf{M}\xi_i$, $\sigma^2 = \mathsf{D}\xi_i$; σ^2 известно. $\overline{\xi}$ — лучшая точечная оценка для a.

$\overline{\xi} \sim \mathcal{N}(a, \frac{\sigma^2}{n})$. Отсюда $\dfrac{\overline{\xi} - a}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0,1)$ — распределение не зависит от неизвестного параметра (заметим, что это не статистика, т.к. она зависит от параметра a). Поэтому

$$\mathsf{P}_a\left(\left|\frac{\overline{\xi} - a}{\sqrt{\sigma^2/n}}\right| \leqslant u\right) = \Phi(u) - \Phi(-u) = 2\Phi(u) - 1, \qquad \text{где } \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \text{ — ф. р. } \mathcal{N}(0,1).$$

Пусть задано $\alpha \in (0,1)$. По таблицам стандартного нормального распределения находим $(1 - \frac{\alpha}{2})$-квантиль $\mathcal{N}(0,1)$. Обозначим эту квантиль $u_{1-\frac{\alpha}{2}}$. Тогда $\Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$, откуда $2\Phi(u_{1-\frac{\alpha}{2}}) - 1 = 1 - \alpha$. В силу предыдущего

$$\mathsf{P}_a\left(\left|\frac{\overline{\xi} - a}{\sigma/\sqrt{n}}\right| \leqslant u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

поэтому $\left[\overline{x} - u_{1-\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}, \overline{x} + u_{1-\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}\right]$ — доверительный интервал для a с доверительной вероятностью $1 - \alpha$.

6.3.2 Доверительный интервал для дисперсии при известном среднем

$s_0^2 = \frac{1}{n} \sum\limits_{k=1}^{n} (x_k - a)^2$ — несмещённая ($\mathsf{M}s_0^2 = \sigma^2$) оценка для σ^2.

$$Z_k = \frac{\xi_k - a}{\sigma} \sim \mathcal{N}(0,1), \qquad \frac{ns_0^2}{\sigma^2} = \sum_{k=1}^{n} \left(\frac{\xi_k - a}{\sigma}\right)^2 = Z_1^2 + \dots + Z_n^2.$$

Z_k взаимно независимы и одинаково распределены с $\mathcal{N}(0,1)$. Распределение случайной величины $Z_1^2 + \dots + Z_n^2$ называется *хи-квадрат* распределением с n степенями свободы и обозначается χ_n^2. Для квантилей этого распределения составлены таблицы. Используя эти таблицы, находим квантили $g_1 = g_1(\frac{\alpha}{2})$ — $\frac{\alpha}{2}$-квантиль и $g_2 = g_2(\frac{\alpha}{2})$ — $(1-\frac{\alpha}{2})$-квантиль χ^2_n. Получаем, что

$$\mathsf{P}_{\sigma^2}\left(g_1 \leqslant \frac{ns_0^2}{\sigma^2} \leqslant g_2\right) = \left(1-\frac{\alpha}{2}\right) - \frac{\alpha}{2} = 1-\alpha,$$

то есть $\left[\frac{ns_0^2}{g_2}, \frac{ns_0^2}{g_1}\right]$ является доверительным интервалом для σ^2 с доверительной вероятностью $1-\alpha$.

6.4 Асимптотические доверительные интервалы

Как правило, они основаны на асимптотической нормальности некоторых статистик.

Пример 4.1. В схеме Бернулли с параметром $p \in [0,1]$ рассматривается оценка $\widehat{p}_n = \frac{1}{n}(x_1 + \dots + x_n)$. $\mathsf{M}\widehat{p}_n = p, \mathsf{D}\widehat{p}_n = \frac{p(1-p)}{n}$. В силу ЦПТ $\widehat{p}_n$ асимптотически нормальна с $\mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$, т.е.

$$\frac{\widehat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n\to\infty]{\mathrm{d}} Z \sim \mathcal{N}(0,1).$$

Положим $u = u_{1-\frac{\alpha}{2}}$ — $(1-\frac{\alpha}{2})$-квантиль стандартного нормального распределения. Получаем, что

$$\mathsf{P}\left(\left|\frac{\widehat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leqslant u\right) \approx \Phi(u) - \Phi(-u) = 1-\alpha,$$

где Φ — функция стандартного нормального распределения. Точность вычисляется из оценок скорости сходимости в ЦПТ. Решая неравенство

$$n(\widehat{p}_n - p)^2 \leqslant p(1-p)u^2,$$

получаем интервал для p: $\mathsf{P}(\underline{p}_n \leqslant p \leqslant \overline{p}_n) \approx 1-\alpha$ ($\underline{p}_n$ и $\overline{p}_n$ — оценки).

Лемма 6.1. *Пусть оценка $\hat{\theta}_n(x)$ параметра θ асимптотически нормальна с $\mathcal{N}(\theta;\sigma^2(\theta)/n)$. Пусть $f(\theta)$ дифференцируема и $f'(\theta) \neq 0$ в рассматриваемой области. Тогда $f(\hat{\theta}_n)$ асимптотически нормальна с распределением*

$$\mathcal{N}\left(f(\theta); \frac{(f'(\theta))^2\sigma^2(\theta)}{n}\right).$$

Замечание. Функцию $f(\theta)$ можно выбрать так, чтобы $f'(\theta)\sigma(\theta) = \mathrm{const}$, и таким образом избавится от зависимости дисперсии от параметра.

$\square$ Разложим f по формуле Тейлора: $f(t) = f(\theta) + (t-\theta)(f'(\theta) + \gamma(t;\theta))$, где $\gamma(t;\theta) \to 0$ при $t \to 0$. Подставим $t = \hat{\theta}_n$:

$$f(\hat{\theta}_n) - f(\theta) = (\hat{\theta}_n - \theta) f'(\theta)\left(1 + \frac{\gamma(\hat{\theta}_n;\theta)}{f'(\theta)}\right).$$

Положим

$$Y_n := \frac{\hat{\theta}_n - \theta}{\sigma(\theta)/\sqrt{n}}, \qquad Z_n := \frac{\gamma(\hat{\theta}_n;\theta)}{f'(\theta)}, \qquad W_n := \frac{f(\hat{\theta}_n) - f(\theta)}{f'(\theta)\sigma(\theta)/\sqrt{n}}$$

и получим, что $W_n = Y_n(1+Z_n)$. Но из асимптотической нормальности $\hat{\theta}_n$ $Y_n \xrightarrow{\mathrm{d}} Y \sim \mathcal{N}(0,1)$, а $\gamma(\hat{\theta}_n;\theta) \xrightarrow{\mathsf{P}} 0$ в силу состоятельности, поэтому $W_n \xrightarrow{\mathrm{d}} Y \sim \mathcal{N}(0,1)$, что и даёт асимптотическую нормальность $f(\hat{\theta}_n)$. $\blacksquare$

Пример 4.2. В схеме Бернулли оценка $\widehat{p}_n = \frac{1}{n}(x_1 + \dots + x_n)$ асимптотически нормальна с $\mathcal{N}(p; p(1-p)/n)$. Положим $f(p) = \arcsin\sqrt{p}$, $p \in (0,1)$. $f'(p) = \frac{1}{2\sqrt{p(1-p)}}$, $f'(p)\sigma(p) = \frac{\sqrt{p(1-p)}}{2\sqrt{p(1-p)}} = \frac{1}{2} = \mathrm{const}$. В силу предыдущей леммы оценка $\arcsin\sqrt{\widehat{p}_n}$ асимптотически нормальна с $\mathcal{N}(\arcsin\sqrt{p}, \frac{1}{4n})$. Пользуясь монотонностью, получаем асимптотический доверительный интервал для p.

Замечание. Для любого фиксированного $x \in \mathbb{R}$ можно построить доверительный интервал для функции распределения $F(x)$, используя эмпирическую функцию распределения $F_n(x; x_1, \dots, x_n)$).

7 Проверка статистических гипотез

7.1 Задача различения двух статистических гипотез

$(\mathscr{X}, \mathscr{A}, \mathscr{P})$ — статистическая модель, $x = (x_1, \ldots, x_n) \in \mathscr{X}$ — выборка. Пусть семейство распределений разбито на два подсемейства: $\mathscr{P} = \mathscr{P}_1 \sqcup \mathscr{P}_2$. Хотим построить критерий для выбора одной из двух гипотез: H_0: $\mathsf{P} \in \mathscr{P}_0$, H_1: $\mathsf{P} \in \mathscr{P}_1$. Рассмотрим произвольное множество $S \in \mathscr{A}$ и назовём его *критическим множеством* (*критической областью*) критерия. Сам критерий в терминах S формулируется так: если $x \in S$, то H_0 *отклоняется* (H_1 принимается); если $x \notin S$, то H_0 принимается (H_1 отклоняется). Критерий с критическим множеством S называется S-критерием.

Ошибка первого рода — отклонить H_0, когда она верна. *Ошибка второго рода* — принять H_0, когда верна H_1. Нужно принять решение так, чтобы вероятности ошибок были минимальными. (Пока это не вполне понятно: как считать вероятность ошибки, если в $\mathscr{P}_0$ и $\mathscr{P}_1$ много разных распределений?)

Пример 1.1. $\mathscr{P} = \{\mathsf{P}_0, \mathsf{P}_1\}$ — два разных распределения, $\mathscr{P}_i = \{\mathsf{P}_i\}, i \in \{0, 1\}$. Пусть заданы малые числа $\alpha, \beta \in (0, 1)$. Потребуем $\mathsf{P}_0(S) \leqslant \alpha$ и $\mathsf{P}_1(\bar{S}) \leqslant \beta$ (или, что то же самое, $\mathsf{P}_0(\bar{S}) \geqslant 1 - \alpha$ и $\mathsf{P}_1(S) \geqslant 1 - \beta$), то есть числа α и β ограничивают сверху ошибки первого и второго рода соответственно. Нужно найти такое множество S, чтобы различить распределения P_0 и P_1.

Перейдём к параметрической модели: $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$, причём распределения попарно различны ($\theta' \neq \theta \Rightarrow \mathsf{P}_{\theta'} \neq \mathsf{P}_\theta$). $\Theta = \Theta_0 \sqcup \Theta_1$; H_0: $\theta \in \Theta_0$, H_1: $\theta \in \Theta_1$.

Пусть дан критерий, заданный критическим множеством S (S-критерий). Тогда функция $g_S(\theta) := \mathsf{P}_\theta(S)$ от переменной θ называется *функцией мощности* S-критерия. При $\theta \in \Theta_0$ $g_S(\theta) = \mathsf{P}_\theta(S)$ есть вероятность ошибки первого рода, а для $\theta \in \Theta_1$ вероятность ошибки второго рода выражается так: $\mathsf{P}_\theta(\bar{S}) = 1 - \mathsf{P}_\theta(S) = 1 - g_S(\theta)$. $\sup\limits_{\theta \in \Theta_0} g_S(\theta)$ — *размер критерия*. Мы требуем, чтобы $\sup\limits_{\theta \in \Theta_0} g_S(\theta) \leqslant \alpha$; в этом случае α есть *уровень значимости* критерия. $\inf\limits_{\theta \in \Theta_1} g_S(\theta)$ — *мощность критерия* (мощность есть наименьшая вероятность принять альтернативную гипотезу, когда она верна). Идеальной (недостижимой) функцией мощности является индикатор множества Θ_1.

Критерий можно определить также при помощи *критической функции* φ_S — индикатора множества S. Тогда $g_S(\theta) = \mathsf{P}_\theta(S) = \mathsf{M}_\theta \varphi_S(\xi)$.

Статистическая гипотеза называется *простой*, если она имеет вид $\mathsf{P} = \mathsf{P}_0$, где P_0 — заданное распределение (или, что то же самое, $F(x) = F_0(x)$ или $\theta = \theta_0$; F_0 и θ_0 известны), т.е. речь идёт о конкретном распределении вероятностей. Остальные гипотезы — сложные.

7.2 Разделение гипотез

Пусть теперь $\Theta = \Theta_0 \cup \Theta_1, \Theta_0 \cap \Theta_1 = \varnothing$. $\sup\limits_{\theta \in \Theta_0} \mathsf{P}_\theta(S)$ – *размер критерия* S. Если $\sup\limits_{\theta \in \Theta_0} \leq \alpha$, то α – уровень значимости S. *Мощность критерия* S: $\inf\limits_{\theta \in \Theta_1} g_s(\theta)$, где $g_s(\theta) = \mathsf{P}_\theta(S)$ – *функция мощности критерия* S. Положим

$$\phi(x) = \begin{cases} 1, & x \in S, \\ 0, & x \notin S. \end{cases}$$

Такая функция называется *критической функцией* критерия S.

7.2.1 Сравнение двух простых гипотез. Теорема Неймана - Пирсона

Пусть $\Theta = \{\theta_0, \theta_1\} \subset \mathbb{R}$, $\theta_0 < \theta_1$. H_0: $\theta = \theta_0$, H_1: $\theta = \theta_1$. $\mathsf{P}_{\theta_0}(S)$ — вероятность ошибки первого рода (*размер критерия*). $\mathsf{P}_{\theta_1}(\bar{S})$ — вероятность ошибки второго рода; $1 - \mathsf{P}_{\theta_1}(\bar{S}) = \mathsf{P}_{\theta_1}(S)$ — мощность критерия. Критерий S с критическим значением α называется *несмещённым*, если $g_S(\theta_1) \geq \alpha$.

Сравнение критериев. Зададим $\alpha \in (0, 1)$ — уровень значимости (ограничитель размера критерия). Рассмотрим все критерии (критические функции) $\varphi_S(x)$ с $\mathsf{M}_{\theta_0} \varphi_S(\xi) = \alpha$ (если P_{θ_0} — дискретное распределение, то такого S может и не найтись — тогда требуем $\leqslant \alpha$ или применяем рандомизацию — см. ниже). Среди этих критериев выделяем $\varphi_{S^*} = \varphi_S^*$ — такой, что $\mathsf{M}_{\theta_1} \varphi_S^*(\xi) = \sup\limits_{\varphi_S \,:\, \alpha = \mathsf{M}_{\theta_0} \varphi_S(\xi)} \mathsf{M}_{\theta_1} \varphi_S(\xi)$ (в общем случае берём sup по $\theta \in \Theta_1$) — если такой критерий существует, то это *наиболее мощный критерий* уровня значимости α.

Рандомизированные критерии. Расширим множество рассматриваемых критериев. Пусть $\varphi(x)$ — произвольная функция, принимающая значения из $[0, 1]$ (не обязательно индикатор). Если $\varphi(x) = 1$, то H_0 отклоняется, если $\varphi(x) = 0$, то H_0 принимается, а вот если $0 < \varphi(x) < 1$, то бросается жульническая монетка и H_0 отклоняется в вероятностью $\varphi(x)$. Такая процедура называется *рандомизацией*, а сам критерий — *рандомизированным*. (Бросание монетки — вспомогательный эксперимент.) В этом случае функция мощности критерия задаётся формулой: $g(\theta) = \mathsf{M}_\theta \varphi(\xi)$.

Теорема 7.1 (теорема Неймана–Пирсона; фундаментальная лемма). *Пусть* $\mathscr{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$,

$\Theta = \{\theta_0, \theta_1\} \subset \mathbb{R}$, $\theta_0 < \theta_1$; семейство $\mathscr{P}$ абсолютно непрерывно относительно некоторой меры (например, меры Лебега в $\mathbb{R}^n$), т.е. существует плотность $p_\theta(x) = L(\theta; x)$ — функция правдоподобия; $L_i(x) = L(\theta_i; x), i \in \{0, 1\}$. Пусть $p_\theta(x) > 0$ (для всех $x \in \mathbb{R}, \theta \in \Theta$). H_0: $\theta = \theta_0$, H_1: $\theta = \theta_1$. Задано $\alpha \in (0, 1)$ — уровень значимости (вероятность ошибки первого рода).

Тогда критерий с критической функцией

$$\varphi^*(x) = \begin{cases} 1, & L_1(x) > c_\alpha L_0(x), \\ \varepsilon_\alpha, & L_1(x) = c_\alpha L_0(x), \\ 0, & L_1(x) < c_\alpha L_0(x), \end{cases}$$

где c_α и ε_α определяются из условия $\mathsf{M}_{\theta_0} \varphi^*(\xi) = \alpha$, таков, что для любой критической функции φ с $\mathsf{M}_{\theta_0} \varphi(\xi) = \alpha$ имеет место неравенство $\mathsf{M}_{\theta_1} \varphi^*(\xi) \geqslant \mathsf{M}_{\theta_1} \varphi(\xi)$ (т.е. этот критерий обладает максимальной мощностью среди критериев уровня значимости α).

□ Сначала найдём c_α и ε_α. Рассмотрим функцию $f(c) = \mathsf{P}_{\theta_0}\{L_1(\xi) > cL_0(\xi)\}$, $c \in [0, +\infty)$. f — невозрастающая функция $f(0) = 1$, $f(+\infty) = 0$ (последняя запись понимается как предел). $1 - f(c) = \mathsf{P}_{\theta_0}\left\{\dfrac{L_1(\xi)}{L_0(\xi)} \leqslant c\right\}$ — функция распределения отношения значений функций правдоподобия (*отношения правдоподобия*) — неубывающая непрерывная слева функция.

Положим $c_\alpha = \min\{c \mid f(c) \leqslant \alpha < f(c-0)\}$ (если c_α — точка непрерывности f, то $f(c_\alpha) = \alpha$). Определим ε_α:

$$\varepsilon_\alpha = \begin{cases} 0, & \text{если } c_\alpha \text{ — точка непрерывности } f \ (f(c_\alpha) = f(c_\alpha - 0)), \\ \dfrac{\alpha - f(c_\alpha)}{f(c_\alpha - 0) - f(c_\alpha)} & \text{в ином случае.} \end{cases}$$

Критерий определён. Будем доказывать его свойства.

$$\mathsf{M}_{\theta_0} \varphi^*(\xi) = \int\limits_{\mathscr{X}} \varphi^*(x) p_{\theta_0}(x)\, dx = \int\limits_{\{x \mid L_1(x) > c_\alpha L_0(x)\}} p_{\theta_0}(x)\, dx + \varepsilon_\alpha \int\limits_{\{x \mid L_1(x) = c_\alpha L_0(x)\}} p_{\theta_0}(x)\, dx =$$

$$= \begin{cases} f(c_\alpha) = \alpha, & \text{если } c_\alpha \text{ — точка непрерывности } f, \\ f(c_\alpha) + \dfrac{\alpha - f(c_\alpha)}{f(c_\alpha - 0) - f(c_\alpha)}(f(c_\alpha - 0) - f(c_\alpha)) = \alpha & \text{в ином случае.} \end{cases}$$

Итак, ошибка первого рода равна α.

Теперь возьмём любой другой критерий с критической функцией φ с условием $\mathsf{M}_{\theta_0} \varphi(\xi) = \alpha$. Покажем, что $\mathsf{M}_{\theta_1} \varphi^*(\xi) \geqslant \mathsf{M}_{\theta_1} \varphi(\xi)$. $\mathscr{X}^+ := \{x \in \mathscr{X} \mid \varphi^*(x) - \varphi(x) \geqslant 0\}$, $\mathscr{X}^- := \{x \in \mathscr{X} \mid \varphi^*(x) - \varphi(x) < 0\}$. $\mathscr{X} = \mathscr{X}^+ \sqcup \mathscr{X}^-$. Для всех $x \in \mathscr{X}^+$ имеем, что $\varphi^*(x) - \varphi(x) \geqslant 0 \Rightarrow \varphi^*(x) \geqslant \varphi(x) \geqslant 0 \Rightarrow \varphi^*(x) \geqslant 0 \Rightarrow L_1(x) \geqslant c_\alpha L_0(x)$, откуда $(\varphi^*(x) - \varphi(x))(p_{\theta_1}(x) - c_\alpha p_{\theta_0}(x)) \geqslant 0$. Если же $x \in \mathscr{X}^-$, то $\varphi^*(x) - \varphi(x) < 0$, откуда $\varphi^*(x) < 1$, то есть $L_1(x) \leqslant c_\alpha L_0(x)$, и опять $(\varphi^*(x) - \varphi(x))(p_{\theta_1}(x) - c_\alpha p_{\theta_0}(x)) \geqslant 0$. Поэтому

$$\mathsf{M}_{\theta_1}(\varphi^*(\xi) - \varphi(\xi)) = \int\limits_{\mathscr{X}} (\varphi^*(x) - \varphi(x)) p_{\theta_1}(x)\, dx = \int\limits_{\mathscr{X}} (\varphi^* - \varphi)(p_{\theta_1}(x) - \underbrace{c_\alpha p_{\theta_0}(x)}_{\text{вычли нулевой интеграл}})\, dx \geqslant 0.$$

■

Если $\mathsf{P}_{\theta_0}(S) \leqslant \mathsf{P}_{\theta_1}(S)$, то S-критерий называется *несмещённым*.

Теорема 7.2. *Критерий Неймана-Пирсона является несмещённым.*

□ Положим

$$D_c^* := \{L_1 * cL_0\},$$

где $* \in \{<, \leq, =, \geq, >\}$. По определению ошибки второго рода

$$\beta = \mathsf{E}_{\theta_1}(1 - \phi^*) = \mathsf{P}_{\theta_1}(X \in D_{c_\alpha}^{\leq}) - \varepsilon_\alpha \mathsf{P}_{\theta_1}(X \in D_{c_\alpha}^{=}).$$

Пользуясь тем, что $\mathsf{P}_{\theta_1}(X \in D_{c_\alpha}^{=}) = c_\alpha \mathsf{P}_{\theta_0}(X \in D_{c_\alpha}^{=})$, получаем

$$\beta = \mathsf{P}_{\theta_1}(X \in D_{c_\alpha}^{\leq}) - \varepsilon_\alpha c_\alpha \mathsf{P}_{\theta_0}(X \in D_{c_\alpha}^{=}).$$

По определению ошибки первого рода

$$\alpha = \mathsf{E}_{\theta_0} \phi^* = \mathsf{P}_{\theta_0}(X \in D_{c_\alpha}^{>}) + \varepsilon_\alpha \mathsf{P}_{\theta_0}(X \in D_{c_\alpha}^{=}).$$

Выражая последнее слагаемое в предыдущее выражение получаем

$$\beta = 1 - \mathsf{P}_{\theta_1}(X \in D^{>}_{c_\alpha}) - c_\alpha(\alpha - \mathsf{P}_{\theta_0}(X \in D^{>}_{c_\alpha})). \tag{1}$$

Для непрерывных функций правдоподобий имеем

$$c_\alpha \mathsf{P}_{\theta_0}(X \in D^{>}_{c_\alpha}) = \int_{L_1(x) > c_\alpha L_0(x)} c_\alpha L_0(x)dx \le \int_{L_1(x) > c_\alpha L_0(x)} L_1(x)dx = \mathsf{P}_{\theta_1}(X \in D^{>}_{c_\alpha}).$$

$$\mathsf{P}_{\theta_1}(X \in D^{\le}_{c_\alpha}) = \int_{L_1(x) \le c_\alpha L_0(x)} L_1(x)dx \le \int_{L_1(x) \le c_\alpha L_0(x)} c_\alpha L_0(x)dx = c_\alpha \mathsf{P}_{\theta_0}(X \in D^{\le}_{c_\alpha}).$$

Для дискретных интегралы заменяются на суммы.

Используя 2 оценки, получаем

$$\frac{1 - \mathsf{P}_{\theta_1}(X \in D^{>}_{c_\alpha})}{1 - \mathsf{P}_{\theta_0}(X \in D^{>}_{c_\alpha})} \le c_\alpha \le \frac{\mathsf{P}_{\theta_1}(X \in D^{>}_{c_\alpha})}{\mathsf{P}_{\theta_0}(X \in D^{>}_{c_\alpha})}.$$

Преобразуя эти неравенства, приходим к

$$\mathsf{P}_{\theta_0}(X \in D^{>}_{c_\alpha}) \le \mathsf{P}_{\theta_1}(X \in D^{>}_{c_\alpha}).$$

Подставим эту оценку в 1

$$\beta = 1 - c_\alpha \alpha + \mathsf{P}_{\theta_0}(X \in D^{>}_{c_\alpha})(c_\alpha - 1) \le 1 - c_\alpha\alpha + \alpha(c_\alpha - 1) = 1 - \alpha.$$

Несмещённость доказана. ■

7.2.2 Равномерно наибольший критерий

Определение. Если существует критерий с критическим множеством D, такой что его уровень значимости не превосходит α и для любого другого критерия D^* с уровнем значимости не более α при всех $\theta \in \Theta_1$ справедливо соотношение

$$\mathsf{P}_\theta(D) > \mathsf{P}_\theta(D^*),\ \theta \in \Theta_1,$$

то критерий D называют *равномерно наиболее мощным* критерием для проверки H_0 против H_1. Если H_1 простая, то критерий называют *наиболее мощным.*

Определение. Говорят, что для дискретных или абсолютно-непрерывных функций распределения F_θ выполнено свойство *монотонности отношения правдоподобий*, если все плотности имеют одинаковый носитель и существует статистика $T(X_1, \dots, X_n)$ такая, что при любых $\theta_0 < \theta_1$

$$\frac{f_{\theta_1}(x_1) \cdots f_{\theta_1}(x_n)}{f_{\theta_0}(x_0) \cdots f_{\theta_0}(x_n)} = g_{\theta_0,\theta_1}(T(x_1, \dots, x_n)),$$

где g_{θ_0,θ_1} – некоторая монотонная функция.

Теорема 7.3. *Пусть F_θ – семейство распределений с монотонным отношением правдоподобий, причём упомянутая в определении функция монотонно возрастает по T. Тогда РНМ критерий уровня α для проверки гипотезы $H_0 : \theta \le \theta_0$ с альтернативой $H_1 : \theta > \theta_0$ задаётся критической функцией вида*

$$\psi(x_1, \dots, x_n) = \begin{cases} 1, & T(x_1, \dots, x_n) > c_\alpha, \\ \varepsilon_\alpha, & T(x_1, \dots, x_n) = c_\alpha, \\ 0, & T(x_1, \dots, x_n) < c_\alpha, \end{cases}$$

где $C_\alpha, \varepsilon_\alpha$ определяются соотношением $\mathsf{E}_{\theta_0}\psi(X_1, \dots, X_n) = \alpha$.

□ Положим $\beta_\psi(\theta) := \mathsf{M}_\theta\psi(X)$. Докажем, что $\beta_\psi(\theta)$ монотонна по θ. Рассмотрим критерий $H_0' : \theta = \theta_1$ с альтернативой $H_1' : \theta = \theta_2 > \theta_1$. По теореме Неймана-Пирсона существует наиболее мощный критерий

$$\phi_{\theta_1,\theta_2}(x) = \begin{cases} 1, & L_2(x) > d_\alpha L_1(x), \\ \delta_\alpha, & L_2(x) = d_\alpha L_1(x), \\ 0, & L_2(x) < d_\alpha L_1(x), \end{cases} = \begin{cases} 1, & g_{\theta_1,\theta_2}(T(x)) > d_\alpha, \\ \delta_\alpha, & g_{\theta_1,\theta_2}(T(x)) = d_\alpha, \\ 0, & g_{\theta_1,\theta_2}(T(x)) < d_\alpha, \end{cases} = \begin{cases} 1, & T(x) > \widetilde{c}_\alpha, \\ \widetilde{\varepsilon}_\alpha, & T(x) = \widetilde{c}_\alpha, \\ 0, & T(x) < \widetilde{c}_\alpha, \end{cases}.$$

Вообще говоря, $\widetilde{\varepsilon}_\alpha$ и $\widetilde{c}_\alpha$ могли зависеть от θ_1 и θ_2, но так как эти параметры определяются исходя из того, что $\mathsf{M}\phi_{\theta_1,\theta_2}(X_1, \dots, X_n) = \alpha$ – функции, не зависящей от θ_1 и θ_2, то $\varepsilon_\alpha := \widetilde{\varepsilon}_\alpha$ и $c_\alpha := \widetilde{c}_\alpha$, $\psi := \varphi_{\theta_1,\theta_2}$.

Рассмотрим критерий $\widetilde{\varphi} \equiv \beta_\psi(\theta_1)$ – критерий, вне зависимости от выборки отклоняющий гипотезу с вероятностью $\beta_\psi(\theta_1)$. Так как $\beta_\psi(\theta_1) = \beta_{\widetilde{\varphi}}(\theta_1)$, то в силу НМ критерий ψ имеем

$$\beta_\psi(\theta_2) \geq \beta_{\widetilde{\varphi}}(\theta_2) = \beta_\psi(\theta_1),$$

откуда $\beta_\psi(\theta)$ – монотонна по θ.

В силу монотонности

$$\sup_{\theta \leq \theta_0} \beta_\psi(\theta) = \beta_\psi(\theta_0) = \alpha.$$

Для любого критерия ψ_1 такого, что уровень значимости ψ_1 не более α верно, что $\beta_{\psi_1}(\theta_0) \leq \alpha$, поэтому в силу НМ ψ для $H_0' : \theta = \theta_0$, $H_1' : \theta = \theta_1 > \theta_0$

$$\beta_{\psi_1}(\theta_1) \geq \beta_\psi(\theta_1),$$

то есть критерий является РНМ. ■

Пример 2.1. Проверим гипотезу о среднем нормального распределения $H_0 : \theta = \theta_0$ с альтернативой $H_1 : \theta = \theta_1$ при известной дисперсии σ^2. Выпишем функцию правдоподобия

$$L(\theta, \sigma^2; x) = L(\theta, \sigma^2; x_1, \ldots, x_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}\right\}.$$

Тогда

$$D = \left\{\frac{L(\theta_0, \sigma^2; x)}{L(\theta_1, \sigma^2; x)} > c\right\} = \left\{\exp\left\{-\frac{\sum_{i=1}^n ((x_i - \theta_0)^2 - (x_i - \theta_1)^2)}{2\sigma^2}\right\} > c\right\} =$$
$$\left\{(\theta_0 - \theta_1)\frac{\sum_{i=1}^n (2x_i - \theta_0 - \theta_1)}{\sigma^2} > 2\ln c\right\} = \left\{2(\bar{x} - \theta_0) > \frac{2\sigma^2 \ln c}{(\theta_0 - \theta_1)n} + \theta_1 - \theta_0\right\}.$$

Обозначим выражение справа в неравенстве за $\widetilde{c}$. Пусть верна гипотеза и $X_i \sim \mathcal{N}(\theta_0, \sigma^2)$. тогда $\bar{X} \sim \mathcal{N}(\theta_0, \sigma^2/n)$, $Z := \sqrt{n}(\bar{X} - \theta_0)/\sigma \sim \mathcal{N}(0, 1)$ и

$$D = \left\{\sqrt{n}\frac{\bar{x} - \theta_0}{\sigma} > \frac{\sqrt{n}\widetilde{c}}{2\sigma}\right\} = \{Z > \hat{c}\}$$

По теореме Неймана-Пирсона $P_{\theta_0}(D) = \alpha$, поэтому $\hat{c}$ – квантиль уровня значимости $1-\alpha$ стандартного нормального распределения. Критерий будет выглядеть следующим образом: если $\bar{x} \leq \theta_0 + \sigma\hat{c}/\sqrt{n}$, тогда принимаем гипотезу H_0, иначе отклоняем. Такой критерий по теореме будет наиболее мощным и несмещённым.

7.3 Проверка гипотез о параметрах нормального распределения

Пусть дана выборка $(x_1, \ldots, x_n)$ из $\mathcal{N}(a, \sigma^2)$, H_0: $a = a_0$ (a_0 фиксировано); H_1: $a \neq a_0$, σ неизвестно. Тогда по следствию из леммы Фишера $\sqrt{n}(\overline{\eta} - a)/s \sim t_{n-1}$. Рассмотрим критическое множество

$$S = \left\{\left|\frac{\overline{x} - a_0}{s/\sqrt{n}}\right| > t_{1-\alpha/2}\right\},$$

где $t_{1-\alpha/2}$ – $1 - \alpha/2$ квантиль t_{n-1}.

Критерий: если x не принадлежит нашему критическому множеству, то H_0 принимается, иначе отклоняется. Вероятность ошибки первого рода (отклонить H_0, когда она верна) равна α, так как при гипотезе $a = a_0$.

Пусть теперь $H_0 : \sigma = \sigma_0$ (σ_0 фиксировано), $H_1 : \sigma \neq \sigma_0$, a неизвестно. При гипотезе имеем $(n-1)s/\sigma_0^2 \sim \chi^2_{n-1}$, поэтому рассмотрим критическое множество

$$S = \left\{\frac{(n-1)s}{\sigma_0^2} > \varkappa_{1-\alpha}\right\} = \left\{\frac{\sum_{i=1}^n (x_i - \overline{x})^2}{\sigma_0^2} > \varkappa_{1-\alpha}\right\},$$

где $\varkappa_{1-\alpha}$ – $1 - \alpha$ квантиль χ^2_{n-1}.

Критерий: если x не принадлежит нашему критическому множеству, то H_0 принимается, иначе отклоняется. Вероятность ошибки первого рода (отклонить H_0, когда она верна) равна α.

7.4 Проверка гипотез о параметрах нормального распределения с помощью доверительного интервала

Пусть дана выборка $(x_1, \ldots, x_n)$ из $\mathcal{N}(a, \sigma^2)$, H_0: $a = a_0$ (a_0 фиксировано); H_1: $a \neq a_0$, σ неизвестно. Тогда по следствию из леммы Фишера $\sqrt{n}(\overline{\eta} - a)/s \sim t_{n-1}$. Можем построить доверительный интервал для параметра как

$$a_0 \in \left(\overline{x} - \frac{t_{1-\alpha/2}s}{\sqrt{n}}, \overline{x} + \frac{t_{1-\alpha/2}s}{\sqrt{n}}\right),$$

где $t_{1-\alpha/2}$ – $1-\alpha/2$ квантиль t_{n-1}.

Критерий: если x принадлежит нашему доверительному интервалу, то H_0 принимается, иначе отклоняется. Вероятность ошибки первого рода (отклонить H_0, когда она верна) равна α, так как при гипотезе $a = a_0$.

Пусть теперь $H_0 : \sigma = \sigma_0$ (σ_0 фиксировано), $H_1 : \sigma \neq \sigma_0$, a неизвестно. При гипотезе имеем $(n-1)s/\sigma_0^2 \sim \chi^2_{n-1}$, поэтому построим доверительный интервал для параметра как

$$\sigma_0^2 \in \left(\frac{(n-1)s}{\varkappa_{1-\alpha/2}}, \frac{(n-1)s}{\varkappa_{\alpha/2}}\right),$$

где $\varkappa_\beta$ – β квантиль χ^2_{n-1}.

Критерий: если x принадлежит нашему доверительному интервалу, то H_0 принимается, иначе отклоняется. Вероятность ошибки первого рода (отклонить H_0, когда она верна) равна α.

7.5 Проверка гипотезы однородности нормальных выборок

Пусть имеется две выборки $x_1, \ldots, x_m; y_1, \ldots, y_n$, порожденные соответственно независимыми в совокупности случайными величинами $\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_n$, причём $\xi_1, \ldots, \xi_m \sim \mathcal{N}(a_1, \sigma_1^2), \eta_1, \ldots, \eta_n \sim \mathcal{N}(a_2, \sigma_2^2)$ и все параметры $a_1, \sigma_1^2, a_2, \sigma_2^2$ неизвестны.

$H_0 : a_1 = a_2, \sigma_1 = \sigma_2$ — *гипотеза однородности*. $H_1 = \neg H_0$. Разобьём задачу проверки гипотезы однородности на две задачи:
I. $H_0' : \sigma_1^2 = \sigma_2^2$ (при любых a), $H_1' : \sigma_1^2 \neq \sigma_2^2$ (если H_0' отклоняется, то отклоняем H_0; в противном случае движемся дальше).
II. Если H_0' принимается, то $H_0'' : a_1 = a_2, H_1'' : a_1 \neq a_2$. Теперь H_0 принимается тогда и только тогда, когда принимается H_0''.

Критерий Фишера. Если $f_{\frac{\alpha}{2}} \leqslant \hat{F}_{m,n} \leqslant f_{1-\frac{\alpha}{2}}$, то H_0' принимается, иначе отвергается. Вероятность ошибки первого рода в точности равна α. (Можно использовать и односторонний критерий)

7.5.1 Критерий Стьюдента равенства средних значений при условии равенства дисперсий

Если H_0' отклоняется, то отклоняется и гипотеза однородности H_0. Пусть теперь H_0' принята (т.е. $\sigma_1^2 = \sigma_2^2 = \sigma^2$). Будем проверять гипотезу $H_0'' : a_1 = a_2$ ($H_1'' = \neg H_0''$).

Мы знаем, что $\mathsf{M}(\bar{x} - \bar{y}) = a_1 - a_2$ и $\mathsf{D}(\bar{x} - \bar{y}) = \sigma^2(1/m + 1/n)$ ($\bar{x}$ и $\bar{y}$ независимы и $\bar{x} - \bar{y} \sim \mathcal{N}(a_1 - a_2, \sigma^2(1/m + 1/n))$.)

$$\frac{(\bar{x} - \bar{y}) - (a_1 - a_2)}{\sqrt{\sigma^2(1/m + 1/n)}} \sim \mathcal{N}(0, 1),$$

но σ^2 неизвестно. Подставим для неё несмещённую оценку

$$\hat{\sigma}^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m + n - 2}$$

При условии H_0'': $a_1 - a_2 = 0$ статистика $\hat{t}_{m+n-2} = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma^2(1/m+1/n)}}$ имеет распределение Стьюдента с $m + n - 2$ степенями свободы (следует из леммы Фишера).

Зададим $\alpha \in (0, 1)$. $t_{\alpha/2} = -t_{1-\alpha/2}$ — соответствующие квантили распределения Стьюдента (равенство имеет место в силу симметричности распределения). Критическим множеством будет

$$S = \left\{\left|\frac{(\bar{x} - \bar{y}) - (a_1 - a_2)}{\sqrt{\sigma^2(1/m + 1/n)}}\right| > t_{1-\alpha/2}\right\}$$

Критерий Стьюдента. Если $|\hat{t}_{m+n-2}| > t_{1-\alpha/2}$, то H_0'' отклоняется, иначе — принимается. Вероятность ошибки первого рода равна α. Если при этом H_0'' принимается, то принимается и исходная гипотеза однородности H_0.

7.5.2 Критерий Фишера равенства дисперсий

Рассмотрим две статистики:

$$s_1^2 = \frac{1}{m-1}\sum_{k=1}^{m}(x_k-\bar{x})^2, \qquad s_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j-\bar{y})^2.$$

Отношение $\hat{F}_{m,n} = \dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ при условии H_0' $(\sigma_1^2=\sigma_2^2)$ равно s_1^2/s_2^2 и имеет F-распределение с $(m-1)$ и $(n-1)$ степенями свободы:

$$\frac{s_1^2}{s_2^2} = \frac{\frac{(m-1)s_1^2}{\sigma^2}\cdot\frac{1}{m-1}}{\frac{(n-1)s_2^2}{\sigma^2}\cdot\frac{1}{n-1}} = \frac{\chi^2_{m-1}\cdot\frac{1}{m-1}}{\chi^2_{n-1}\cdot\frac{1}{n-1}} = F_{m-1,n-1},$$

где предпоследнее равенство справедливо в силу леммы Фишера.

Для заданного α находим квантили $f_{\frac{\alpha}{2}}(m-1,n-1)$ и $f_{1-\frac{\alpha}{2}}(m-1,n-1)$ F-распределения с $(m-1)$ и $(n-1)$ степенями свободы и берем их в качестве критических значений. Критическим множеством будет

$$S = \left\{\frac{1}{f_{1-\alpha/2}} < \frac{S_1^2}{S_2^2} < f_{\alpha/2}\right\}$$

7.6 Дисперсионный анализ однофакторной модели

Пусть $x_1,\dots,x_n$ — выборка, где все элементы получены независимо. A priori разобьём её на $k, k\geqslant 3$ подвыборок: $x_{11},\dots,x_{n_1 1};\dots;x_{1k},\dots,x_{n_k k}$, $n=n_1+\dots+n_k$. $\bar{x}_i=(x_{1i}+\dots+x_{n_i i})/n_i$ — средние значения по подвыборкам, $\bar{\bar{x}}=(x_{11}+\dots+x_{n_k k})/n$ — среднее значение.

Эти данные удобно представлять в виде *таблицы дисперсионного анализа*:

1	2	$\dots$	k	
x_{11} x_{21} $\vdots$	x_{12} x_{22} $\vdots$	$\dots$ $\dots$	x_{1k} x_{2k} $\vdots$	
n_1	n_2	$\dots$	n_k	n
a_1	a_2	$\dots$	a_k	$\bar{a}$
$\bar{x}_1$	$\bar{x}_2$	$\dots$	$\bar{x}_k$	$\bar{\bar{x}}$

Предполагается, что существуют постоянные (но неизвестные нам) $a_1,\dots,a_k$ такие, что $x_{ij}=a_j+e_{ij}$ при всех i и j, где e_{ij} — случайные ошибки, взаимно независимые при разных i и j, $\mathsf{M}e_{ij}=0$, $\mathsf{D}e_{i,j}=\sigma^2$ (считаем измерения равноточными), σ^2 неизвестно.

Эта модель называется *однофакторной.* Считается, что на эксперименты, порождающие данные в столбцах, отличаются влиянием некоторого фактора; $1,2,\dots,k$ — номера уровней фактора, a_j — характеристики уровня фактора. Различия между элементами одних столбцов „чисто случайны". Существуют и многофакторные модели.

H_0: гипотеза об отсутствии фактора $(a_1=\ldots=a_k)$ (гипотеза однородности подвыборок), $H_1=\neg H_0$. Если $e_{ij}\sim\mathcal{N}(0,\sigma^2)\Leftrightarrow x_{ij}\sim\mathcal{N}(a_j,\sigma^2)$, то H_0 есть гипотеза о равенстве средних значений k нормальных выборок. [Случай, когда x_{ij} имеют непрерывные распределения, существенно отличающиеся от нормального, относится к непараметрическому анализу (ранговые критерии, etc). Этим мы заниматься не будем.]

Изменчивость данных — отклонение от среднего (измеряется как выборочная дисперсия)

$$\sum_{i,j}(x_{ij}-\bar{\bar{x}})^2 = \sum_{i,j}(x_{ij}-\bar{x}_j)^2 + \sum_j n_j(\bar{x}_j-\bar{\bar{x}})^2$$

$$\sum_{i,j}x_{ij} = n\bar{\bar{x}} = \sum_{j=1}^{k}n_j\bar{x}_j$$

Найдём распределение первой суммы. По лемме Фишера

$$\frac{\sum\limits_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2}{\sigma^2} = \chi^2_{n_j-1},$$

откуда
$$\frac{1}{\sigma^2}\sum_{i,j}(x_{ij}-\bar{x}_j)^2=\sum_{j=1}^{k}\chi^2_{n_j-1}=\chi^2_{\sum_j(n_j-1)}=\chi^2_{n-k},$$
т.к. все $\chi^2_{n_j-1}$ у нас независимы.

$\bar{x}_j$ не зависит от $\sum(x_{ij}-\bar{x}_j)^2$ по лемме Фишера, следовательно, $\sum_j(x_{ij}-\bar{x}_j)^2$ и $\sum_j(\bar{x}_j-\bar{\bar{x}})^2$ независимы.

Если H_0 верна, то $\frac{1}{\sigma^2}\sum(x_{ij}-\bar{\bar{x}})^2=\chi^2_{n-1}$:
$$\frac{1}{\sigma^2}\sum_j n_j(\bar{x}_j-\bar{\bar{x}})^2=\chi^2_{k-1}(\text{по лемме Фишера})$$
Так как $\bar{\xi}_j$ не зависит от $\sum_{i=1}^{n_j}(\xi_{i,j}-\bar{\xi}_j)^2$, $\bar{\bar{\xi}}$ не зависит от $\sum_{i=1}^{k}(\bar{\xi}_i-\bar{\bar{\xi}})^2$, то
$$\chi^2_{n-1}=\chi^2_{n-k}+\chi^2_{k-1}$$
$\hat{\sigma}_1^2=\frac{1}{n-k}\sum_{i,j}(x_{ij}-\bar{x}_j)^2$ — оценка, не зависящая от гипотезы H_0, $\hat{\sigma}_2^2=\frac{1}{k-1}\sum_j n_j(\bar{x}_j-\bar{\bar{x}})^2$ — реагирует на H_0

Дисперсионное отношение $\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}$ имеет распределение Фишера–Снедекора с $k-1$, $n-k$ степенями свободы.

Правосторонний критерий Фишера. Пусть $\alpha:\quad f_{1-\alpha}(k-1,n-k)$ — $(1-\alpha)$-квантиль F-распределения с $k-1,n-k$ степенями свободы. Если $\hat{F}_{k-1,n-k}>f_{1-\alpha}$, то H_0 отклоняется (фактор существует); иначе — принимается.

7.7 Парные и множественные сравнения

Рассмотрим случай, когда в дисперсионном анализе однофакторной модели гипотеза отсутствия фактора (H_0) отвергнута. Проведём более тонкое исследование.

7.7.1 Парное сравнение

Проверяем гипотезу $H_{0(j,l)}$: $a_j=a_l\Leftrightarrow a_j-a_l=0$; $H_1=\neg H_0$. (однородности двух подвыборок).

Строим доверительный интервал для разности a_j-a_l. $\bar{x}_j,\bar{x}_l$ — оценки для a_j,a_l.
$$\frac{(\bar{x}_j-\bar{x}_l)-(a_j-a_l)}{\sqrt{\hat{\sigma}_1^2\left(\frac{1}{n_j}+\frac{1}{n_l}\right)}}=\hat{t}_{n-k}$$
$\hat{t}_{n-k}$ имеет распределение Стьюдента с $(n-k)$ степенями свободы, фиксируем α, строим доверительный интервал для a_j-a_l стандартным способом:
$$\left[(\bar{x}_j-\bar{x}_l)-t_{1-\frac{\alpha}{2}}(n-k)\sqrt{\hat{\sigma}_1^2\left(\frac{1}{n_j}+\frac{1}{n_l}\right)},\,(\bar{x}_j-\bar{x}_l)+t_{1-\frac{\alpha}{2}}(n-k)\sqrt{\hat{\sigma}_1^2\left(\frac{1}{n_j}+\frac{1}{n_l}\right)}\right].$$
Критерий: если 0 лежит внутри этого интервала, то принимаем $H_{0(j,l)}$, иначе — отклоняем. Если левая граница больше нуля, то делаем вывод, что $a_j>a_l$ с вероятностью как минимум $1-\alpha$. Аналогично, если правая граница меньше нуля, то делаем вывод, что $a_j<a_l$ с вероятностью как минимум $1-\alpha$.

7.7.2 Множественные сравнения

$$\Psi=\sum_{j=1}^{k}c_ja_j,\quad\sum_{j-1}^{k}c_j=0$$
Ψ — „*сравнение*" $a_1,\ldots,a_k$ („контраст"), парное сравнение — частный случай такого. Оценим: $\hat{\Psi}=\sum_{j=1}^{k}c_j$
$$\mathsf{M}\hat{\Psi}=\Psi,\qquad\mathsf{D}\hat{\Psi}=\sigma^2\sum_{j=1}^{k}c_j^2/n_j$$
$$\frac{\hat{\Psi}-\Psi}{\sqrt{\hat{\sigma}_1^2\sum_j\frac{c_j^2}{n_j}}}=\hat{t}_{n-k}$$
Далее стандартным образом строим доверительный интервал для Ψ ($\hat{t}_{n-k}$ имеет распределение Стьюдента). Проводя множественные сравнения между разными параметрами, мы можем определить, на какие однородные подгруппы можно разбить имеющуюся выбору.

7.8 Критерий Пирсона χ^2

7.8.1 Биномиальный критерий

Рассмотрим схему Бернулли с параметром $p \in (0,1)$. Проверяем гипотезу $H_0 : p = p_0, \quad H_1 : p \neq p_0$. $T(x) = x_1 + \ldots + x_n$ — достаточная статистика, α — уровень значимости (вероятность ошибки первого рода), m^*_α — некоторая критическая граница.

Критерий: если $T(x) > m^*_\alpha$, то H_0 отклоняем, иначе принимаем. Запишем вероятность ошибки первого рода: $\sum\limits_{m > m^*_\alpha} \mathbf{C}_n^m p_0^m (1-p_0)^{n-m} \leqslant \alpha$

Заменим точный критерий на приближенный с помощью ЦПТ.

$$\frac{T(\xi) - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1).$$

Вероятность ошибки первого рода:

$$\mathsf{P}_{p_0}\left(\frac{T(\xi) - np_0}{\sqrt{np_0(1-p_0)}} > \frac{m^*_\alpha - np_0}{\sqrt{np_0(1-p_0)}}\right) = \alpha$$

$$1 - \Phi\left(\frac{m^*_\alpha - np_0}{\sqrt{np_0(1-p_0)}}\right) = \alpha$$

$$u_{1-\alpha} = \frac{m^*_\alpha - np_0}{\sqrt{np_0(1-p_0)}} \text{ — } (1-\alpha)\text{-квантиль } \mathcal{N}(0,1)$$

Из последнего условия находится граница $m^*_\alpha = np_0 + u_{1-\alpha}\sqrt{np_0(1-p_0)}$.

7.8.2 Критерий χ^2 для схемы Бернулли (предисловие к критерию Пирсона)

Есть и другой критерий. Пусть $m = x_1 + \ldots + x_n$. $\mathsf{P}(S_n = m) = \mathbf{C}_n^m p^m (1-p)^{n-m}$. Рассмотрим статистику

$$\widehat{X}_1^2 = \frac{(m - np_0)^2}{np_0} + \frac{(n - m - n(1-p_0))^2}{n(1-p_0)}$$

$\widehat{X}_1^2$ есть сумма относительных квадратичных отклонений эмпирических результатов от ожидаемых. В условиях H_0 $\mathsf{M}S_n = np_0$.

$$\widehat{X}_1^2 = \frac{(m - np_0)^2}{np_0} + \frac{(m - np_0)^2}{n(1-p_0)} = \frac{(m - np_0)^2}{np_0(1-p_0)} = \left(\frac{m - np_0}{\sqrt{np_0(1-p_0)}}\right)^2$$

Следовательно, так как в силу ЦПТ $\frac{S_n - np_0}{\sqrt{np_0(1-p_0)}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$, то $\widehat{X}_1^2$, как квадрат этой величины, сходится по распределению к χ_1^2 при условии H_0. $\widehat{X}_1^2$ называется *статистикой Пирсона (хи-квадрат)*. Критерий строится стандартным образом при помощи квантилей хи-квадрат распределения.

7.8.3 Критерий знаков

$(x_1, y_1), \ldots, (x_n, y_n)$ — набор парных наблюдений, порождённый парой многомерных случайных величин (ξ, ξ') (ξ и ξ' могут быть зависимыми, а вот $\xi_1, \ldots, \xi_n$ независимы в совокупности и одинаково распределены, равно как и $\xi'_1, \ldots, \xi'_n$). Пусть $\xi_1, \ldots, \xi_n$ распределены с функцией распределения $F(x)$, а $\xi'_1, \ldots, \xi'_n$ — с функцией распределения $G(x) = F(x - \theta)$, $\theta \in \mathbb{R}$ — параметр. Гипотеза H_0 состоит в однородности выборок (т.е. $\theta = 0$); $H_1 = \neg H_0$.

Перейдём к разностям $z_i = x_i - y_i$ (это значения случайных величин $\eta_i = \xi_i - \xi'_i$; $\eta_1, \ldots, \eta_n$ независимы в совокупности). Пусть на самом деле $\xi'_i = \xi + \theta + \varepsilon_i$, где ε_i - случайная величина с абсолютно непрерывным распределением и $\mathsf{P}(\varepsilon_i) > 0$. Тогда гипотеза H_0 равносильна гипотезе H'_0: $\forall i\, \mathsf{P}(\eta_i > 0) = \mathsf{P}(\eta_i < 0) = \frac{1}{2}$. Судим о справедливости гипотезы H_0 по соотношению знаков „+” и „−” среди z_i. По сути имеется n испытаний Бернулли, где элементарными событиями являются $A_i = \{\eta_i > 0\}$, $\overline{A_i} = \{\eta_i < 0\}$. Гипотеза H'_0 равносильна гипотезе H''_0: „в этой схеме Бернулли $p = \frac{1}{2}$” ($p = \mathsf{P}(A_i)$ — параметр схемы Бернулли). Строим биномиальный критерий для случайной величины

$$\mu_n^+ = \sum_{i=1}^n \varphi_i, \qquad \varphi_i = \begin{cases} 1, & \eta_i > 0 \\ 0, & \eta_i < 0 \end{cases}$$

с основной гипотезой $H_0''': p = \frac{1}{2}$.

Практический совет на случай, когда среди z_i встречаются нули: если нулей много, то критерий неприменим (потому что в этом критерии считается, что $\mathsf{P}(\eta_i = 0) = 0$). Если же нулей мало, то просто выкидываем те испытания, в которых $z_i = 0$.

7.8.4 Полиномиальный критерий

Обобщим наш критерий. Будем проверять гипотезу о том, что данная выборка $x_1, \ldots, x_n$ имеет полиномиальное распределение с s исходами и заданными вероятностями: $H_0 : p_1 = p_1^0, \ldots, p_s = p_s^0$, $H_1 = \neg H_0$. Обозначим за m_k количество исходов типа k и составим статистику а là $\widehat{X}_1^2$.

Статистика Пирсона:

$$\widehat{X}_{s-1}^2 = \sum_{k=1}^{s} \frac{(m_k - np_k^0)^2}{np_k^0}$$

Далее мы покажем, что в условии H_0 и $n \to \infty$ $\widehat{X}_{s-1}^2$ имеет асимптотическое распределение χ_{s-1}^2 (теорема Пирсона).

Критерий Пирсона (χ^2) о данном распределении в полиномиальной модели. Пусть $g_{1-\alpha}$ — $(1-\alpha)$-квантиль χ_{s-1}^2. Если $\widehat{X}_{s-1}^2 > g_{1-\alpha}(s-1)$, то H_0 отклоняется, иначе принимается. Асимптотическая ошибка первого рода, как будет следовать из теоремы, которую мы сейчас докажем, будет равна α.

7.8.5 Теорема Пирсона

Пусть $\xi_1, \ldots, \xi_n$ взаимно независимы и одинаково распределены: $\xi_i = (\xi_{i_1}, \ldots, \xi_{i_s})^T$ — случайные векторы, причем $\mathsf{P}(\xi_{i_1} = a_1, \ldots, \xi_{i_s} = a_s) = p_1^{a_1} \cdot \ldots \cdot p_s^{a_s}$ для всех i, $p_i > 0$, $\sum_i p_i = 1$; $a_i \in \{0, 1\}$, $\sum_i a_i = 1$ (ξ_i — вектор из 0 и 1 с ровно одной единицей). Очевидно, что $\mathsf{M}\xi_{i_k} = p_k$, $\mathsf{D}\xi_{i_k} = p_k(1-p_k)$, $\mathsf{M}\xi_i = (p_1, \ldots, p_s)^T = \vec{p}$. Элементы ковариационной матрицы $\Sigma = \Sigma_\xi = \mathsf{M}(\xi - \mathsf{M}\xi)(\xi - \mathsf{M}\xi)^T = (\sigma_{jl})$ имеют вид $\sigma_{jl} = \mathsf{M}(\xi_{i_j} - p_j)(\xi_{i_l} - p_l) = -p_j p_l$ при $j \neq l$ (в общем случае $\sigma_{jl} = p_j \delta_{jl} - p_j p_l$). Рассмотрим случайный вектор $S_n = \xi_1 + \ldots + \xi_n = (\mu_1, \ldots, \mu_s)^T$, где $\mu_k = \sum_{i-1}^n \xi_{i_k}$ – сумма независимых случайных величин. $\mathsf{M}\mu_k = np_k$, $\mathsf{D}\mu_k = np_k(1-p_k)$, $\mathsf{M}S_n = (np_1, \ldots, np_s)^t = n\vec{p}$; $\Sigma_{S_n} = n\Sigma_\xi$ (докажите это!).

Распределение S_n называется *полиномиальным* распределением. Найдём распределение статистики

$$\widehat{X}_{s-1}^2 = \sum_{k=1}^{s} \frac{(\mu_k - np_k)^2}{np_k}, \quad \mathsf{M}\widehat{X}_{s-1}^2 = s-1$$

Теорема 7.4 (Пирсон). $\widehat{X}_{s-1}^2 \xrightarrow{\mathrm{d}} \chi_{s-1}^2$

□ Перейдём к $S_n^* = \frac{S_n - n\vec{p}}{\sqrt{n}}, \mathsf{M}S_n^* = 0, \Sigma_{S_n^*} = \Sigma_\xi = \Sigma$. Доказательство будет основано на одном из вариантов многомерной ЦПТ, а именно $\boxed{S_n^* \xrightarrow{\mathrm{d}} Z \sim \mathcal{N}(0, \Sigma)}$ [Севастьянов, гл. 11, § 46, теорема 7]

Положим $\Delta_k = \dfrac{\mu_k - np_k}{\sqrt{np_k}}$, $k = 1, \ldots, s$; $\Delta = (\Delta_1, \ldots, \Delta_s)^T$. $\sum\limits_{k=1}^{s} \sqrt{p_k}\Delta_k = \dfrac{(n-n)}{\sqrt{n}} = 0$, поэтому s-мерное распределение Δ — вырожденное.

$S_n^* = B\Delta$, $B = \mathrm{diag}(\sqrt{p_1}, \ldots, \sqrt{p_s})$. $\Delta = B^{-1}S_n^*$, поэтому $\mathsf{M}\Delta = 0$. Ковариационная матрица $\Sigma_\Delta = \mathsf{M}(\Delta\Delta^T) = B^{-1}\mathsf{M}(S_n^* S_n^{*T})B^{-1} = B^{-1}\Sigma B^{-1}$. $(\Sigma_\Delta)_{ij} = \delta_i^j - \sqrt{p_i p_j}$ (на главной диагонали $1 - p_1, \ldots, 1 - p_k$, вне неё: $-\sqrt{p_k p_j}$).

Положим $\eta = C\Delta$, где C — ортогональная матрица с заданной первой строкой: $c_{1k} = \sqrt{p_k}$. $\eta = (\eta_1, \ldots, \eta_s)^T$, $\mathsf{M}\eta = 0$; $\eta_1 = \sum_{k=1}^s c_{1k}\Delta_k = 0$. Докажем, что $(\Sigma_\eta)_{ij} = \delta_i^j$, если $i, j \geqslant 2$, и 0 иначе. При $j, l \geqslant 2$ имеем:

$$\mathsf{M}(\eta_j\eta_l) = \mathsf{M}\left(\sum_k c_{jk}\Delta_k \cdot \sum_k c_{lk}\Delta_k\right) = \mathsf{M}\left(\sum_{s,t} c_{js}c_{lt}\Delta_s\Delta_t\right) = \sum_{s=1}^{m} c_{js}c_{ls}\mathsf{M}\Delta_s^2 + \sum_{s\neq t} c_{js}c_{lt}\mathsf{M}(\Delta_s\Delta_t) =$$
$$= \sum_{s=1}^{m} c_{js}c_{ls}(1-p_s) - \sum_{s\neq t} c_{js}c_{lt}\sqrt{p_s p_t} = \sum_s c_{js}c_{ls} - \left(\sum_s c_{js}\sqrt{p_s}\right)\left(\sum_t c_{lt}\sqrt{p_t}\right) = \sum_s c_{js}c_{ls} = \delta_{jl}$$

Последнее равенство верно в силу ортогональности строк матрицы C.

$$\widehat{X}_{s-1}^2 = \sum_{k=1}^{s} \Delta_k^2 = \sum_{k=1}^{s} \eta_k^2 = \eta_2^2 + \ldots + \eta_s^2 \quad (\eta_1 = 0)$$

Мы доказали, что при $i > 2$ η_i не коррелируют. Если докажем, что $\eta_i \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ (и к тому же они независимы, а не просто не коррелируют — тут-то нам и потребуется многомерная ЦПТ!), то получим χ^2_{s-1}. При $j \geqslant 2$

$$\eta_j = \sum_{k=1}^{s} c_{jk}\Delta_k = \sum_{i=1}^{n}\sum_{k=1}^{s} \frac{c_{jk}}{\sqrt{np_k}}(\xi_{i_k} - p_k) = \sum_{i=1}^{n} \eta_{ij}.$$

При фиксированном j и разных i η_{ij} независимы и одинаково распределены, $\mathsf{M}\eta_{ij} = 0, \quad \mathsf{D}\eta_j = 1.$

$$\eta_j = \sum_i \eta_{ij}, \quad \mathsf{M}\eta_{ij} = 0, \quad \mathsf{D}\eta_{ij} = \frac{1}{n} \quad \text{докажите это!}$$

По одномерной ЦПТ $\eta_j \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$ для всех $j \geqslant 2$.

Памятуя, что $\eta = (CB^{-1})S_n^*$, а S_n^* сходится по распределению к $\mathcal{N}(0,\Sigma)$ (по формуле в рамочке), получаем, что $\eta \xrightarrow{\mathrm{d}} \mathcal{N}(0,E')$, где E' — единичная матрица без верхней левой единицы ($\eta_1 = 0$), причём компоненты его некоррелированы, а потому независимы (это свойство многомерного нормального распределения). Поэтому $\widehat{X}^2_{s-1} = \eta_2^2 + \ldots + \eta_s^2 \xrightarrow{\mathrm{d}} \chi^2_{s-1}$, что и требовалось. ■

7.8.6 Критерий χ^2

Пусть $x_1, \ldots, x_n$ — повторная выборка, ξ_i одинаково распределены с ф.р. $\mathcal{F}(x)$.
$H_0 : \mathcal{F}(x) = F_0(x), \quad H_1 = \neg H_0$.

Разобьем $\mathbb{R}$ на s интервалов $\Delta_k = (t_{k-1}, t_k]$ произвольным образом. Тогда наша задача сведется к проверке гипотезы о данном распределении вероятностей успеха в полиномиальной модели, где m_k — число x_i, попавших в Δ_k, p_k^0 — вероятностная мера Δ_k (при условии H_0), и $\widehat{X}^2_{s-1}$, определенное ранее, будет иметь асимптотическое распределение χ^2_{s-1}

Замечания. Задачи, решаемые с помощью χ^2:

- $F_0(x)$ может быть задана с точностью до r параметров; тогда статистика Пирсона, в которой p_k^0 вычислены с подстановкой вместо неизвестных параметров статистик для них, будет иметь асимптотическое распределение χ^2_{s-1-r}.

- Критерий χ^2 можно использовать для проверки гипотезы однородности. Пусть есть две полиномиальные схемы с n и l испытаниями, с результатами испытаний $(m_1, \ldots, m_s)$, вероятностями $(p_1, \ldots, p_s)$ и $(m_1', \ldots, m_s')$ и $(p_1', \ldots, p_s')$ соответственно.

  $H_0 : p_i' = p_i, \quad H_1 = \neg H_0$. При условии H_0 выборки объединяются и оцениваются общие значения $p_i = p_i'$: $\hat{p}_i = \frac{m_i + m_i'}{n+l}$ (по методу МП). Тогда

  $$\widehat{X}^2_{s-1} = \sum_{k=1}^{s} \frac{(m_k - n\hat{p}_k)^2}{n\hat{p}_k} + \sum_{k=1}^{s} \frac{(m_k' - l\hat{p}_k)^2}{l\hat{p}_k} \Rightarrow \chi^2_{s-1}$$

  (в общем случае, имея r полиномиальных схем, получим $\chi^2_{(s-1)(r-1)}$).

- Пусть μ – медиана распределения ξ_i. $H_0 : \mu = 0$, $H_1 : \mu \neq 0$. При гипотезе μ_n^+ из критерия знаков имеет биномиальное распределение с $p = 1/2$.

7.9 Критерий Колмогорова

Рассмотрим две гипотезы о функции распределения $F(x)$: $H_0\colon F(x) = F_0(x)$ (нулевая гипотеза), где $F_0(x)$ — заданная непрерывная функция распределения; $H_1\colon F(x) \neq F_0(x)$ (альтернативная гипотеза).

Статистика Колмогорова позволяет сформулировать критерий, согласно которому выбирается одна из этих двух гипотез. А именно:

Критерий Колмогорова. *Если $\sqrt{n}D_n > y_*$, то H_0 отклоняем (H_1 принимаем), если же $\sqrt{n}D_n \leqslant y_*$, то H_0 принимаем (H_1 отклоняем). Здесь число y_* называется критическим значением и равно $y_* = y_{1-\alpha}$ — квантиль уровня $(1-\alpha)$ функции Колмогорова $K(y)$ (т. е. решение уравнения $K(y) = 1-\alpha$).*

На практике для заданного α квантиль $y_{1-\alpha}$ находится по таблице квантилей функции Колмогорова.

Действительно, по теореме Колмогорова $\mathsf{P}_{H_0}(\sqrt{n}D_n > y_*) = 1 - \mathsf{P}_{H_0}(\sqrt{n}D_n \leqslant y_{1-\alpha}) \xrightarrow[n\to\infty]{} 1 - K(y_{1-\alpha}) = 1 - (1-\alpha) = \alpha$, т. е. вероятность ошибки I рода приближенно равна α (если n достаточно велико).

8 Регрессия

Рассмотрим пару случайных величин (ξ_1, ξ_2). Задача *регрессии* состоит в том, чтобы найти такую измеримую функцию $f(\xi_1)$, что $\mathsf{M}(\xi_2 - f(\xi_1))^2$ минимально ($\mathsf{M}\xi_2^2 < \infty$). Как мы знаем, что такой минимум достигается в $m(\xi_1) := \mathsf{M}(\xi_2 \mid \xi_1)$. Функция $m(x) = \mathsf{M}(\xi_2 \mid \xi_1 = x)$ называется *функцией регрессии* ξ_2 по ξ_1. График $y = m(x)$ называется *линией регрессии.*

Регрессия является оценкой $\widehat{\xi_2}$ случайной величины ξ_2. Эта оценка обладает следующими свойствами (по свойствам УМО):

1. $\mathsf{M}\widehat{\xi_2} = \mathsf{M}\xi_2$;
2. $\mathsf{M}(\xi_2 - \widehat{\xi_2})^2$ – минимально.

То есть регрессия предсказывает ξ_2 по ξ_1 с *наименьшей среднеквадратичной ошибкой.*

8.1 Разложение дисперсии случайной величины по отношению к регрессии.

Рассмотрим условную дисперсию, определенную как

$$\mathsf{D}(\xi_2 \mid \xi_1) = \mathsf{M}((\xi_2 - \mathsf{M}(\xi_2 \mid \xi_1))^2 \mid \xi_1)$$

Тогда по свойствам УМО

$$\mathsf{MD}(\xi_2 \mid \xi_1) = \mathsf{M}\{\mathsf{M}((\xi_2 - \mathsf{M}(\xi_2 \mid \xi_1))^2 \mid \xi_1)\} = \mathsf{M}(\xi_2 - \mathsf{M}(\xi_2 \mid \xi_1))^2;$$

$$\begin{gathered}\mathsf{MD}(\xi_2 \mid \xi_1) + \mathsf{DM}(\xi_2 \mid \xi_1) = \mathsf{E}\{\xi_2^2 - 2\xi_2\mathsf{E}(\xi_2 \mid \xi_1) + \mathsf{E}^2(\xi_2 \mid \xi_1) + \mathsf{E}^2(\xi_2 \mid \xi_1) - 2\mathsf{E}\xi_2\mathsf{E}(\xi_2 \mid \xi_1) + \mathsf{E}^2\xi_2\} = \\ \mathsf{E}\xi_2^2 + \mathsf{E}^2\xi_2 - 2\underbrace{\mathsf{E}\{\mathsf{E}\xi_2\mathsf{E}(\xi_2 \mid \xi_1)\}}_{=\mathsf{E}\xi_2\mathsf{E}\{\mathsf{E}(\xi_2 \mid \xi_1)\} = \mathsf{E}^2\xi_2} + 2\mathsf{E}\{\mathsf{E}^2(\xi_2 \mid \xi_1) - \xi_2\mathsf{E}(\xi_2 \mid \xi_1)\} = \\ \mathsf{E}\xi_2^2 - \mathsf{E}^2\xi_2 + 2\mathsf{E}\{\mathsf{E}(\mathsf{E}(\xi_2 \mid \xi_1)(\mathsf{E}(\xi_2 \mid \xi_1) - \xi_2)) \mid \xi_1\} = \mathsf{D}\xi_2 + 2\mathsf{E}\{\mathsf{E}(\xi_2 \mid \xi_1)(\mathsf{E}(\xi_2 \mid \xi_1) - \mathsf{E}(\xi_2 \mid \xi_1))\} = \mathsf{D}\xi_2.\end{gathered}$$

Определение. $\mathsf{MD}(\xi_2 \mid \xi_1)$ – *остаточная регрессия.*

Разберём крайние случаи.

1. Если ξ_1 и ξ_2 – независимы, то $\mathsf{M}(\xi_2 \mid \xi_1) = \mathsf{M}\xi_2$ п.н., поэтому $\mathsf{DM}(\xi_2 \mid \xi_1) = 0$, $\mathsf{MD}(\xi_2 \mid \xi_1) = \mathsf{D}\xi_2$.
2. Если $\xi_1 = \xi_2$ п.н., то $\mathsf{M}(\xi_2 \mid \xi_1) = \xi_2$ п.н., поэтому $\mathsf{MD}(\xi_2 \mid \xi_1) = 0$, $\mathsf{DM}(\xi_2 \mid \xi_1) = \mathsf{D}\xi_2$.

8.2 Связь регрессии и корреляции

Введём понятие *корреляционного отношения*

$$\eta^2 = \eta^2(\xi_2 \mid \xi_1) := 1 - \frac{\mathsf{MD}(\xi_2 \mid \xi_1)}{\mathsf{D}\xi_2}$$

Свойства корреляционного отношения

1. $0 \leqslant \eta^2 \leqslant 1$;
2. $\eta^2 = 0$ тогда и только тогда, когда $\mathsf{MD}(\xi_2 \mid \xi_1) = \mathsf{D}\xi_2$ (ξ_2 и ξ_1 не коррелированы);
3. $\eta^2 = 1$ тогда и только тогда, когда $\mathsf{MD}(\xi_2 \mid \xi_1) = 0$;
4. Если зависимость – линейная, то $\eta^2 = r^2$, где r – коэффициент корреляции Пирсона;
5. $\eta^2 \geqslant r^2$.

8.3 Регрессия для двумерной нормальной выборки

Рассмотрим частный случай. Пусть существует совместная плотность $p(x, y)$ случайных величин (ξ, η). Тогда $p(x) = \int_X p(x, y)dy$. Рассмотрим

$$q(y \mid x) = \frac{p(x, y)}{p(x)}$$

плотность условного распределения ξ_2 при условии $\xi_1 = x$.

$$\int_Y yq(y \mid x)dy = \frac{\int_Y yp(x, y)dy}{\int_X p(x, y)dy} = \mathsf{M}(\xi_2 \mid \xi_1 = x).$$

$m(\xi_1) = \mathsf{M}(\xi_2 \mid \xi_1)$ – регрессия ξ_2 по ξ_1. Линейная регрессия $m(x) = kx + b$.

Пример 3.1.

$$p(x,y) = \frac{1}{2\pi\sqrt{(1-r^2)\sigma_1\sigma_2}} \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{(x-a_1)^2}{\sigma_1^2} - 2r\frac{(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right)\right\}$$

Если $r = 0$,то получаем

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{(x-a_1)^2}{\sigma_1^2} + \frac{(y-a_2)^2}{\sigma_2^2}\right)\right\}$$

Для ξ_1

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-a_1)^2/2}$$

Тогда для $a_1 = a_2 = 0$

$$q(y \mid x) = \frac{p(x,y)}{p(x)} = \frac{1}{\sqrt{2\pi(1-r^2)\sigma_2}} \exp\left\{-\frac{(y - r\frac{\sigma_2}{\sigma_1}x)^2}{2(1-r^2)\sigma_2^2}\right\}$$

Можно заметить, что $q(y \mid x)$ – плотность $\mathcal{N}(r\sigma_2 x/\sigma_1, (1-r^2)\sigma_2^2)$. Тогда

$$\mathsf{M}(\xi_2 \mid \xi_1) = r\frac{\sigma_2}{\sigma_1}\xi_1,\ \mathsf{MD}(\xi_2 \mid \xi_1) = (1-r^2)\sigma_2^2$$

То есть для нормального двумерного распределения (ξ_1, ξ_2) регрессия линейная. В общем случае

$$\widehat{\xi_2} = r\frac{\sigma_2}{\sigma_1}(\xi_1 - a_1) + a_2.$$

8.4 Эмпирическая регрессия

Пусть даны выборочные значения $(x_1, y_1), \ldots, (x_n, y_n)$. Это реализация пары случайных величин (ξ_1, ξ_2). Мы можем нанести эти точки на график Этот график называется *диаграммой рассеяния*. Предположим, что $y = m(x) = kx + b$. Будем искать k и b такие, чтобы

$$\sum_{i=1}^{n}(y_i - (kx_i + b))^2$$

было бы минимально. Такой метод называется *методом наименьших квадратов (МНК)*. Дифференцируя и приравнивая производные нулю получим, что оптимальные параметры

$$k = \widehat{r}\frac{S_2}{S_1},\ b = \bar{y} - k\bar{x},$$

где

$$\widehat{r} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

Но что, если линейного приближения недостаточно? Можно тогда использовать выборочный коэффициент корреляции. Тогда нам нужно проверить гипотезу $\nu^2 = r^2$.

В нормальном случае $\widehat{r}$ имеем асимптотическое распределение $\mathcal{N}(r, (1-r^2)/n)$. При гипотезе

$$\widehat{t}_{n-2} = \frac{\widehat{r}\sqrt{n-2}}{\sqrt{1 - \widehat{r}r^2}}.$$

Тогда для заданного критического уровня α составляем неравенство и проверяем его.

9 Вопросы к экзамену

1. Эмпирическая функция распределения: её свойства как функции распределения и как оценки теоретической функции распределения.
2. Теорема Гливенко-Кантелли о сходимости эмпирической функции распределения.
3. Асимптотические свойства эмпирических моментов и функций от них.
4. Теорема Колмогорова с доказательством независимости распределения статистики Колмогорова от вида непрерывной функции распределения.
5. Вариационный ряд выборки и порядковые статистики. Распределение порядковых статистик. Оценка выборочных квантилей.
6. Информация Фишера и её свойства.
7. Условные математические ожидания и условные распределения относительно сигма-алгебр и случайных величин. Условная плотность распределения одной случайной величины относительно другой.
8. Свойства условных математических ожиданий.
9. Достаточные статистики. Теорема Неймана-Фишера (критерий достаточности).
10. Сравнение точечных статистических оценок по их свойствам. Асимптотические свойства оценок. Примеры состоятельных и асимптотически нормальных оценок.
11. Эффективные оценки в регулярном случае. Неравенство Крамера-Рао.
12. Многомерное обобщение неравенства Крамера-Рао. Информация Фишера в двумерном случае, пример двумерной нормальной выборки.
13. Улучшение оценок с помощью достаточных статистик. Теорема Колмогорова-Блекуэла-Рао.
14. Полные достаточные статистики и теорема о несмещённых оценках с минимальной дисперсией.
15. Свойства частоты как оценки вероятности «удачи» в схеме Бернулли. Сравнение с другими оценками.
16. Метод моментов оценивания параметров. Теорема о состоятельности оценок метода моментов.
17. Метод максимального правдоподобия. Свойства оценок максимального правдоподобия. Примеры оценок максимального правдоподобия.
18. Метод максимального правдоподобия. Теорема об асимптотической нормальности оценок максимального правдоподобия.
19. Байесовский метод. Теорема о байесовской оценке при квадратичной функции риска. Априорное и апостериорное распределение. Априорный и апостериорный риск.
20. Байесовские оценки параметров биномиального и нормального распределений.
21. Свойства байесовских оценок: байесовские оценки и достаточные статистики; минимаксные оценки как байесовские с постоянным риском.
22. Многомерное нормальное распределение: эквивалентные определения и основные характеристики. Свойства многомерного нормального распределения.
23. Лемма Фишера о независимости среднего арифметического и среднего квадратического для независимых одинаково нормально распределённых случайных величин.
24. Распределение хи-квадрат, Стьюдента и Фишера-Снедекора. Вывод формулы плотности распределения.
25. Следствие из леммы Фишера о распределениях хи-квадрат и Стьюдента как распределениях статистик для нормальных выборок.
26. Интервальные оценки и их характеристики. Два метода построения точных доверительных интервалов.
27. Построение точного доверительного интервала для параметра биномиального распределения.

28. Доверительные интервалы для параметров нормального распределения (для среднего и дисперсии).

29. Асимптотические доверительные интервалы. Построение асимптотического доверительного интервала на основе асимптотической нормальности подходящей статистики.

30. Асимптотические доверительные интервалы для параметров биномиального распределения.

31. Теорема Неймана-Пирсона. Критерий отношения правдоподобий для проверки двух простых гипотез, как наиболее мощный и несмещённый критерий.

32. Равномерно наиболее мощный критерий. Теорема о существовании РНМ-критерия при условии свойства монотонности отношения правдоподобий.

33. Критерий отношения правдоподобий для проверки двух гипотез о среднем значении нормального распределения.

34. Проверка гипотез о параметрах нормального распределения (о среднем и дисперсии).

35. Проверка гипотез о параметрах нормального распределения с помощью доверительных интервалов.

36. Критерий Стьюдента равенства средних значений двух независимых нормальных выборок.

37. Критерий Фишера равенства дисперсий двух независимых нормальных выборок.

38. Оценка наименьшего размера выборки, необходимого для проверки двух гипотез о параметре биномиального распределения с заданными вероятностями ошибок первого и второго рода.

39. Дисперсионный анализ однофакторной модели для нормальных выборок.

40. Множественное сравнение параметров однофакторной модели с помощью доверительных интервалов.

41. Критерий проверки гипотез о значении параметра биномиального распределения. Критерий знаков.

42. Полиномиальное распределение, как многомерное распределение. Статистика Пирсона для проверки гипотезы о значениях параметров полиномиального распределения. Теорема об асимптотическом хи-квадрат распределении статистики Пирсона.

43. Критерий хи-квадрат для проверки гипотезы о данном полиномиальном распределении.

44. Критерий Колмогорова для проверки гипотезы о данном непрерывном распределении.

45. Решение задачи регрессии. Регрессия одной случайной величины по другой случайной величине. Разложение дисперсии случайной величины по отношению к регрессии. Остаточная дисперсия.

46. Сравнение байесовских оценок и оценок максимального правдоподобия для параметра биномиального распределения.[11]

47. Линейная регрессия. Роль коэффициента корреляции при оценивании одной случайной величины по другой. Регрессия в случае двумерного нормального распределения. Статистическое оценивание коэффициентов регрессии для нормальных выборок методом наименьших квадратов.

[11] Не знаю что лектор рассказывал по теме данного вопроса, но приведённый пример показывает разницу между ОМП и байесовскими, минимаксными оценками — *примеч. В. Х.*