Введение в математическую статистку

Содержание

1 Введение в математическую статистику. Первичная обработка данных

1.1 Вместо введения

Математическая статистика является направлением в математике, исследующим способы обработки, представления, систематизации данных для получения практических выводов.

Математическая статистика тесно связана с теорией вероятности, в виду того, что использует аналогичные методы и приемы рассуждений. В то же время, задачи математической статистики в некотором смысле являются двойственными к задачам теории вероятностей.

В теории вероятностей с использованием понятий случайных событий и случайных величин определяется модель явления и исследуются его свойства. В математической статистике известны реализации некоторых случайных величин, которые называются *статистическими данными*, и определяются как совокупность числовых (как правило) или качественных характеристик, порожденных некоторым источником.

Математическую статистику можно определить как раздел математики, в котором для математических моделей процессов порождения статистических данных изучаются методы систематизации, обработки, методы использования результатов для научных и (или) практических выводов.

Данный курс представляет собой лишь введение в математическую статистику и рассматривает следующие задачи:

- оценивание неизвестных параметров распределений;
- проверка статистических гипотез о виде распределения;
- прогнозирование.

Введем основные понятия математической статистики: наблюдение, выборка, реализация выборки.

Определение 1.1. *Наблюдение — это значение, которое приняла случайная величина.*

Определение 1.2. *Выборкой будем называть случайный вектор* $X = (X_1, \ldots, X_n)$*, компонентами которого является наблюдаемые случайные величины.*

Будем говорить, что выборка X из распределения ξ (или $\mathscr{L}(\xi)$), если компоненты вектора X есть независимые, одинаково распределенные случайные величины, имеющие такое же распределение, как и случайная величина ξ.

Определение 1.3. *Реализацией выборки будем называть набор из n наблюдений* $x = (x_1, \ldots, x_n)$*.*

Можно рассматривать реализацию выборки как наблюдение за выборкой.

Определение 1.4. *Выборочным пространством будем называть множество всех возможных значений реализаций выборки вместе с* σ*-алгеброй измеримых подмножеств этого пространства.*

В теории вероятностей каждая случайная величина $\eta : \Omega \to B$, определенная на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ порождает на пространстве $(B, \mathcal{B})$ вероятностные меры $\mathsf{P}_\eta = \mathsf{P}\,(\eta \in C)$, $C \in \mathcal{B}$, определяемые функцией распределения F_η. В математической статистике в большинстве случаев на выборочном пространстве определяется не одна мера, а конечное или бесконечное семейство вероятностных мер.

Определение 1.5. *Семейством вероятностных мер (или параметрическим семейством), заданных на пространстве $(B, \mathcal{B})$ будем называть множество $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$, где Θ — некоторое множество параметров.*

Наверное самым часто используемым понятием (после выборки) в курсе будет понятие статистики.

Определение 1.6. *Статистикой будем называть произвольное отображение*

$$T(X) = (T_1(X), \ldots, T_l(X)) : \Omega^n \to \mathbb{R}^l,$$

где $T_i(X) : \Omega^n \to \mathbb{R}$ — произвольное измеримое отображение.

1.2 Вариационный ряд

Рассмотрим выборку $X = (X_1, \ldots, X_n)$ из распределения $\mathscr{L}(\xi)$, а $x = (x_1, \ldots, x_n)$ — реализация этой выборки. Вектору x можно поставить в соответствие упорядоченную последовательность:

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}, \text{ где}$$

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

$$x_{(i)} = \min\{x_1, x_2, \ldots, x_n\} \setminus \{x_{(1)}, \ldots, x_{(i-1)}\},$$

$$x_{(n)} = \max\{x_1, x_2, \ldots, x_n\}.$$

Обозначим через $X_{(i)}$ случайную величину, которая для каждой реализации выборки x принимает значение $x_{(i)}$.

Определение 1.7. *Упорядоченную последовательность случайных величин $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, полученную по выборке X будем называть вариационным рядом выборки, а сами значения $X_{(i)}$ — порядковыми статистиками. Значения $X_{(1)}$ и $X_{(n)}$ будем называть минимальными и максимальными значениями выборки, а величину $\rho = X_{(n)} - X_{(1)}$ будем называть размахом выборки.*

Определение 1.8. *Медианой выборки будем называть величину, равную $X_{(k+1)}$, если $n = 2k + 1$ и $\frac{1}{2}\left(X_{(k)} + X_{(k+1)}\right)$, если $n = 2k$.*

1.3 Эмпирическая функция распределения

Пусть имеется выборка $X = (X_1, \ldots, X_n)$ из распределения $\mathscr{L}(\xi)$. Функцию распределения (теоретическую функцию распределения) случайной величины ξ будем обозначать $F(y)$.

Определение 1.9. *Для каждого $y \in \mathbb{R}$ рассмотрим случайную величину*

$$\mu_n(y) = \sum_{i=1}^{n} \mathrm{Ind}(X_i \leq y),$$

равную числу элементов выборки X меньших или равных y. Функцию $\widehat{F}_n(y) = \dfrac{\mu_n(y)}{n}$ будем называть эмпирической функцией распределения (э.ф.р.), соответствующей выборке X.

Заметим, что эмпирическая функция распределения является случайной величиной и принимает следующие значения:

$$\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\right\}.$$

При этом, используя равенство $\mathsf{P}\left(X_i \leqslant y\right) = F(y)$, легко показать, что $\widehat{F}_n(x)$ принимает значение равное k/n с вероятностью:

$$\mathsf{P}\left(\widehat{F}_n(x) = \frac{k}{n}\right) = \binom{n}{k} F^k(x) \left(1 - F(x)\right)^{n-k}.$$

Заметим, что для каждой реализации выборки x функция $\widehat{F}_n(x)$ однозначно определена и очевидным образом обладает следующими свойствами, присущими теоретическим функциям распределения:

- $0 \leqslant \widehat{F}_n(x) \leqslant 1$,
- $\widehat{F}_n(x)$ является не убывающей функцией,
- $\widehat{F}_n(x)$ непрерывна справа.

Докажем теорему, показывающую, что для произвольного фиксированного $y \in \mathbb{R}$ э.ф.р. $\widehat{F}_n(y)$ с увеличением объема выборки n стремиться к значению функции распределения $F(y)$.

Теорема 1.10. *Для $\forall x \in \mathbb{R}$ и для $\forall \varepsilon > 0$ при $n \to \infty$*

$$\mathsf{P}\left(\left|\widehat{F}_n(x) - F(x)\right| < \varepsilon\right) \to 1$$

(то есть $\widehat{F}_n(x) \xrightarrow{\mathsf{P}} F(x)$*).*

Доказательство. Введем обозначение для последовательности независимых, одинаково распределенных случайных величин $\eta_j = \mathrm{Ind}(X_j \le x)$, где $j = \overline{1,n}$. Тогда, $\mathsf{M}\,\eta_j = \mathsf{P}\,(\eta_j \le x) = F(x)$, а $\mathsf{D}\,\eta_j = F(x)(1 - F(x))$. Заметим, что

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\eta_i$$

Тогда, согласно закону больших чисел:

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\eta_i \xrightarrow{\mathsf{P}} F(x)$$

□

Оценим скорость сближения э.ф.р. и теоретической функции распределения. Найдем математическое ожидание и дисперсию $\widehat{F}_n(y)$

$$\mathsf{M}\,\widehat{F}_n(y) = \mathsf{M}\,\frac{1}{n}\sum_{i=1}^{n}\eta_i = F(x),$$

$$\mathsf{D}\,\widehat{F}_n(x) = \mathsf{D}\,\frac{1}{n}\sum_{i=1}^{n}\eta_i = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{D}\,\eta_i = \frac{1}{n}F(x)(1 - F(x)).$$

Используя неравенство Чебышева:

$$\mathsf{P}(|\xi - \mathsf{M}\,\xi| > \varepsilon) \le \frac{\mathsf{D}\,\xi}{\varepsilon^2}$$

получаем следующую оценку:

$$\mathsf{P}\left(\sqrt{n}\left|\widehat{F}_n(x) - F(x)\right| > t\right) \le \frac{F(x)(1 - F(x))}{t^2}.$$

Приведем без доказательства теорему, которую можно использовать для оценивания скорости сходимости выборочной функции распределения к её теоретическому аналогу.

Теорема 1.11 (Теорема Колмогорова)**.** *Рассмотрим функцию*

$$D_n = D_n\,(X) = \sup\left|\widehat{F}_n(x) - F(x)\right|.$$

Если $F(x)$ – непрерывная функция распределения, то $\forall t > 0$:

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

Статистику $K(t)$ называют распределением Колмогорова, которое не зависит от вида функции распределения.

Теорема Колмогорова дает асимптотическую оценку и использование статистики Колмогорова возможно при объеме выборки $n \geqslant 20$. Помимо предельного результата Колмогоров в работе [?] предложены рекуррентные соотношения для конечных n. На русском языке доказательство теоремы можно найти, например в [?].

Следующая теорема, которая также дается без доказательства, также дает важное представление о свойствах э.ф.р.

Теорема 1.12 (Теорема Смирнова)**.** *Пусть $\widehat{F}_{1n}(x), \widehat{F}_{2m}(x)$ – две эмпирические функции распределения построенные на двух независимых выборках n и m соответственно. И $D_{n,m} = \sup\limits_{x\in\mathbb{R}} \left|\widehat{F}_{1n}(x) - \widehat{F}_{2m}(x)\right|$, а $F(x)$ – непрерывная функция распределения. $\forall t > 0$:*

$$\lim_{n,m\to\infty} P\left(\sqrt{\frac{n\cdot m}{n+m}} D_{n,m} \leq t\right) = K(t).$$

2 Введение в теорию оценивания

2.1 Выборочные характеристики

Пусть имеется выборка $X = (X_1, \ldots, X_n)$ из распределения $\mathscr{L}(\xi)$.

Определение 2.1. *Выборочными характеристиками будем называть измеримые функционалы от эмпирического распределения (то есть функции от выборки).*

Определение 2.2. *Выборочным моментом порядка k называется величина*

$$\widehat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k = \int\limits_{-\infty}^{\infty} t^k d\widehat{F}_n(t),$$

Иногда, для краткости, выборочный момент 1 порядка обозначают $\overline{X}$.

Определение 2.3. *Выборочным центральным моментом порядка k называется величина, определенная формулой:*

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\alpha}_1)^k = \int\limits_{-\infty}^{\infty}(t - \widehat{\alpha}_1)^k d\widehat{F}_n(t)$$

Определение 2.4. $\widehat{\alpha}_1$ *будем называть выборочным средним,* $\widehat{\mu}_2$ *будем называть выборочной дисперсий*

2.2 Понятие оценки

Одним из направлений математической статистики является разработка «рациональных» методов оценивания неизвестных истинных значений характеристик наблюдаемых случайных величин. Понятие статистической оценки использовалось ранее (хоть и не называлось), когда говорилось, что значение э.ф.р. $\widehat{F}_n(x)$ в каждой точке x можно рассматривать в качестве приближенного значения (оценки) для теоретической функции распределения (см. теорему 1.10). При этом, для выборок большого объема значительная разница между значениями выборочной и теоретической характеристиками маловероятна (см. теорему 1.11).

Оценка истинного значения g_0 некоторой характеристики $g = g(\xi)$ случайной величины ξ по выборке $X = (X_1, X_2, \ldots, X_n)$ из параметрического семейства $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ подразумевает построение такой функции $T_n = T_n(X)$, значение которой t при наблюдавшейся в эксперименте реализации выборки является приближением (в некотором смысле) для g_0. Нестрого можно написать, что $g_0 \approx t$. В этом случае говорят, что статистика T_n оценивает g или что T_n есть оценка для g и обозначают: $\widehat{g} = T_n(X)$. Такая оценка называется *точечной оценкой.*

Приведем пример построения оценок. Пусть $x = (x_1, \ldots x_n)$ — реализация выборки из равномерного распределения $R(0, \theta]$. Очевидно, что следующие функции являются оценками на неизвестный параметр θ.

$$\widehat{\theta}_1 = \left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) \cdot 2,$$

$$\widehat{\theta}_2 = x_{(n)} = \max\{x_1, \ldots, x_n\}.$$

Эти оценки являются функциями от выборки и не зависят от θ. При $n \to \infty$ обе оценки «должны» должны приближаться к истинному значению θ.

2.3 Несмещенные оценки

Любая оценка $T = T(X)$, где T — измеримая функция, является случайной величиной. Для произвольной случайной величины определены понятия математического ожидания и дисперсии.

Рассмотрим параметрическую модель

$$\mathcal{F} = \{F_\theta,\, \theta \in \Theta\}$$

которое задает распределение вероятностей на выборочном пространстве.

Нижним индексом θ при символах математического ожидания будем обозначать то, что соответствующие величины вычисляются для распределения F_θ. Обозначим $\mathsf{M}_\theta T(X)$ и $\mathsf{D}_\theta T(X)$ соответственно математическое ожидание и дисперсию статистики T в случае, когда функция распределения реализации выборки $\overline{x} = (x_1, x_2, \ldots, x_n)$ есть

$$F_{X,\theta}(X) = F_{\xi,\theta}(x_1) \cdot \ldots \cdot F_{\xi,\theta}(x_n).$$

Математическое ожидание выписывается по определению:

$$\mathsf{M}_\theta(T(x)) = \int\limits_B T(t)dF_\theta(t),$$

что в непрерывном случае соответствует формуле:

$$\int\limits_{\mathbb{R}^n} T(t)f_\theta(x_1) \cdot \ldots \cdot f_\theta(x_n)dx_1 \ldots dx_n,$$

где $f(x)$ — плотность случайной величины ξ, и в дискретном случае соответствует:

$$\sum_{x \in B} T(t)p_\theta(x_1) \cdot \ldots \cdot p_\theta(x_n),$$

где $p(x)$ соответствует значению вероятности $\mathsf{P}\left(\xi = x\right)$.

Определение 2.5. *Статистика $T = T(X)$ называется несмещенной в среднем (или просто несмещенная) оценкой для заданной параметрической* *функции $\tau(\theta)$, если она удовлетворяет условию*

$$\mathsf{M}_\theta T(x) = \tau(\theta), \forall\theta \in \Theta,$$

которое называется уравнением несмещенности.

Определение 2.6. *Смещением оценки $T(X)$ неизвестного параметра $\tau(\theta)$ будем называть величину*

$$b(\theta) = \mathsf{M}_\theta T(X) - \tau(\theta).$$

Если смещение равно 0, то оценка несмещенная.

Найдем математическое ожидание выборочного среднего и выборочной дисперсии.

$$\mathsf{M}_\theta \widehat{\alpha}_1 = \mathsf{M}_\theta \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} \mathsf{M}_\theta X_i = \mathsf{M}_\theta X_1 = \mathsf{M}_\theta \xi.$$

Получили, что выборочное среднее является несмещенной оценкой для математического ожидания случайной величины ξ.

Введем обозначение $Y_i = X_i - \alpha_1$, где $\alpha_1 = \mathsf{M}\,\xi$, и найдем математическое ожидание выборочной дисперсии. Для начала преобразуем выражение для выборочной дисперсии:

$$\widehat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i \pm \alpha_1 - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2.$$

Так как $\mathsf{M}_\theta Y_i = 0$, $\mathsf{M}_\theta Y_i^2 = \mu_2$, $\mathsf{M}_\theta Y_i Y_j = \mathsf{M}_\theta Y_i \, \mathsf{M}_\theta Y_j = 0$, где $\mu_2 = \mathsf{D}\,\xi$. Если $i \neq j$, верно следующее:

$$\mathsf{M}\,\overline{Y}^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} \mathsf{M}_\theta Y_i Y_j = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{M}_\theta Y_i^2 = \frac{\mu_2}{n},$$

Так как

$$\frac{n\mu_2}{n} - \frac{\mu_2}{n} = \frac{n-1}{n}\mu_2,$$

при $n \to \infty$ значение выборочной дисперси хоть и сходится к значению дисперсии случайной величины ξ, но является смещенной оценкой.

Можно построить несмещенную оценку дисперсии:

$$\widehat{\mu}_2' = \frac{n}{n-1}\widehat{\mu}_2.$$

Пример 2.7. Несмещенные оценки не всегда существуют. Пусть $\xi \sim \mathrm{Pois}(\theta)$. Покажем, что ни для одной функции $T = T(X)$ не будет выполнено равенство $\mathsf{M}_\theta T(X) = \theta^{-1}$. Предположим обратное:

$$\frac{1}{\theta} = \mathsf{M}_\theta T(X) = \sum_{k=0}^{\infty} T(k) \cdot \frac{\theta^k}{k!} \cdot e^{-\theta}.$$

Тогда будет верно равенство:

$$\sum_{k=0}^{\infty} T(k) \cdot \frac{\theta^{k+1}}{k!} = e^{\theta} = \sum_{r=0}^{\infty} \frac{\theta^r}{r!}, \forall \theta > 0.$$

Такой функции $T(x)$ не существует (можно показать устремив θ к 0).

2.4 Состоятельность оценки

Определение 2.8. *Статистика $T_n = T_n(X)$ называется состоятельной оценкой для неизвестного параметра $\tau(\theta)$ случайной величины ξ, если при $n \to \infty$:*

$$T_n(X) \xrightarrow[n\to\infty]{\mathsf{P}} \tau(\theta).$$

Утверждение 2.9. *Пусть элементы выборки $X = (X_1, \ldots X_n)$ являются независимыми одинаково распределенными случайными величинами и $\mathsf{M}\,|X_1| < \infty$. Тогда при $n \to \infty$ выборочные моменты порядка k сходятся к k-ым моментам случайной величины X_1:*

$$\widehat{\alpha}_k \xrightarrow{\mathsf{P}} \mathsf{M}\, X_1^k.$$

Доказательство. Рассмотрим статистику

$$S = S(X) = \int g(x) d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(x_i).$$

Статистика S представляет собой сумму независимых одинаково распределенных случайных величин, при этом $\mathsf{M}\, g(X_1) = \int g(t)dF(t)$. По закону больших чисел:

$$S \xrightarrow{P} \mathsf{M}\, g(x_i),$$

что и заканчивает доказательство утверждения. $\square$

Верно и более сильное утверждение, что любая непрерывная функция от выборочных моментов сходится по вероятности к значению этой функции от соответствующих моментов случайной величины.

Теорема 2.10. *Пусть случайные величины $\eta_1(n), \ldots, \eta_k(n)$ сходятся по вероятности при $n \to \infty$ к некоторым константам $c_1, \ldots, c_k$. Тогда для любой непрерывной функции $\varphi(x_1, \ldots, x_k)$ случайная величина $\varphi(\eta_1(n), \ldots, \eta_k(n))$ сходится по вероятности к $\varphi(c_1, \ldots, c_k)$.*

Доказательство. По условию теоремы φ — непрерывная функция. Тогда $\forall \varepsilon > 0\ \exists \delta = \delta(\varepsilon)$:

$$|\varphi(x_1, \ldots, x_k) - \varphi(c_1, \ldots, c_k)| < \varepsilon \text{ при } |x_i - c_i| < \delta,\ i = \overline{1,k}.$$

Пусть $B_i = \{|\eta_i(n) - c_i| < \delta\}$. Рассмотрим событие $B = B_1 \cdot \ldots \cdot B_k$. Выполнение события B влечет за собой выполнение события

$$C = \{|\varphi\left(\eta_1(n), \ldots, \eta_k(n)\right) - \varphi(c_1, \ldots, c_k)| < \varepsilon\}.$$

Рассмотрим следующую последовательность неравенств:

$$\mathsf{P}(C) \geqslant \mathsf{P}(B) = 1 - \mathsf{P}(\overline{B}) = 1 - \mathsf{P}(\overline{B}_1 \cup \cdots \cup \overline{B}_k) \geqslant 1 - \sum_{i=1}^{k} \mathsf{P}(\overline{B}_i).$$

Так как $\eta_i(n) \xrightarrow{\mathsf{P}} c_i$, то для фиксированного δ для произвольного $\gamma > 0$ существует номер $n_i = n_i(\gamma)$ такой, что для всех $n \geqslant n_i$ выполняется

$$\mathsf{P}\{|\eta_i(n) - c_i| \geq \delta\} \leq \frac{\gamma}{k}.$$

Пусть $n_0 = \max\{n_1, \ldots, n_k\}$. Тогда для любого $n \geqslant n_0$ верно неравенство: $\mathsf{P}(\overline{B}_i) < \frac{\gamma}{k}$ и

$$\mathsf{P}(C) \geqslant 1 - \gamma.$$

Отсюда и от произвольности выбора γ следует доказательство теоремы. $\square$

Заметим, что для проверки состоятельности несмещенной оценки T_n для неизвестного параметра $\tau(\theta)$ достаточно убедиться, что ее дисперсия стремиться к 0 при $n \to \infty$. Действительно, по неравенству Чебышева

$$\mathsf{P}\left(|T_n(X) - \tau(\theta)| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}_\theta\, T_n(X)}{\varepsilon} \to 0,$$

что говорит о сходимости по вероятности. Таким образом, легко показать, что значение э.ф.р. в произвольной точке y есть состоятельная оценка для функции распределения в точке y.

3 Некоторые способы построения оценок

3.1 Метод моментов

Пусть имеется выборка $X = (X_1, \ldots, X_n)$ из распределения $\mathscr{L}(\xi)$, $\mathscr{L}(\xi) \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$, где $\theta = (\theta_1, \ldots, \theta_r) \in \mathbb{R}^r$.

Рассмотрим один из простейших способов построения состоятельных оценок. Пусть у случайной величины ξ имеются первые r моментов, т.е. $\alpha_k = \mathsf{M}_\theta\left(\xi^k\right) < \infty$, являющиеся функциями от неизвестного θ: $\alpha_k = \alpha_k(\theta)$, $k = \overline{1, r}$.

Рассмотрим следующую систему:

$$\begin{cases} \alpha_k(\theta) = \widehat{\alpha}_k,\ k = \overline{1,r} \end{cases}$$

в которой ровно r неизвестных $\theta_1, \ldots, \theta_r$.

Пусть эта система однозначно разрешима и ее решением являются $\widehat{\theta}_1, \ldots, \widehat{\theta}_r$, $\widehat{\theta}_i = \phi_i(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_k)$, ϕ_i — некоторая функция.

Оценки $\widehat{\theta}_i = \phi_i(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_k)$ будем называть оценками, построенными по методу моментов. Заметим, что если функция ϕ_i является непрерывной функцией, то по теореме 2.10 оценка $\widehat{\theta}_i$ является состоятельной.

Заметим, что метод напрямую неприменим, когда выборочный момент не существует — $\mathsf{M}_\theta\left(\xi^k\right) = \alpha^k = \infty$.

Пример 3.1. Рассмотрим выборку из распределения $\mathscr{L}(\xi) \sim N(\theta_1, \theta_2^2)$. Известно, что $\mathsf{M}(\xi) = \theta_1$, и $\mathsf{D}(\xi) = \theta_2 = \mathsf{M}(\xi^2) - (\mathsf{M}(\xi))^2$. При этом $\mathsf{M}(\xi^2) = \alpha_2$, $(\mathsf{M}(\xi))^2 = (\alpha_1)^2$.

Тогда легко получить следующие оценки методом моментов $\widehat{\theta}_1 = \widehat{\alpha}_1$, $\widehat{\theta}_2 = \sqrt{\widehat{\alpha}_2 - (\widehat{\alpha}_1)^2}$. При этом, данные оценки будут состоятельными.

3.2 Метод максимального правдоподобия

Рассмотрим еще один метод построения оценок неизвестных параметров. Оценки полученные методом, рассмотренным в этом разделе принято называть *оценками максимального правдоподобия.*

Пусть есть выборка $X = (X_1, \ldots, X_n)$ из распределения $\mathscr{L}(\xi)$. Пусть $f_\theta(x)$ — функция плотности случайной величины ξ, которая известна с точностью до параметра из распределения; (в дискретном случае вместо функции плотности берем функцию вероятности $\mathsf{P}_\theta(\xi = x)$). Пусть $\overline{x}$ — реализация выборки, $\overline{x} = (x_1, \ldots, x_n)$.

Определение 3.2. *Функцию, заданную равенством $L(\overline{x};\theta) = \prod\limits_{i=1}^{n} f_\theta(x_i)$ будем называть функцией правдоподобия.*

Фактически функция правдоподобия совпадает с плотностью $\overline{x}$ как вектора.

Если для двух значений параметра $\theta_1, \theta_2 \in \Theta$ $L(\overline{x};\theta_1) > L(\overline{x};\theta_2)$, то говорят, что θ_1 более правдоподобна, чем θ_2.

Определение 3.3. *Оценкой максимального правдоподобия называется построенная по реализации выборки $\overline{x}$ значение $\widehat{\theta} = \arg\max\limits_{\theta\in\Theta} L(\overline{x};\theta)$.*

То есть, чтобы найти оценку максимального правдоподобия (о.м.п.), надо найти такое значение θ, при котором функция правдоподобия принимает максимальное значение.

Если для каждого $\overline{x}$ из выборочного пространства максимум $L\left(\overline{x};\theta\right)$ достигается в некоторой внутренней точке и $L\left(\overline{x};\theta\right)$ дифференцируема по θ, то $\widehat{\theta}\left(\overline{x}\right)$ удовлетворяет условию

$$\frac{\partial L(\overline{x};\theta)}{\partial \theta} = 0.$$

В случае векторного параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$ условие на максимум функции правдоподобия представляется системой

$$\left\{ \frac{\partial L\left(\overline{x};\theta\right)}{\partial \theta_i} = 0, i = \overline{1,r}. \right.$$

Вместо функции правдоподобия для простоты часто рассматривают следующую функцию $\ln L\left(\overline{x};\theta\right) = \sum\limits_{i=1}^{n} \ln f_\theta(x_i)$.

Пример 3.4. Рассмотрим выборку из распределения $\mathscr{L}(\xi) \sim \mathrm{N}(\theta_1,\, \theta_2^2)$, $\overline{x}$ — реализация выборки.

$$f_\theta(x_i) = \frac{1}{\theta_2 \cdot \sqrt{2\pi}} \cdot \exp\left(\frac{-\frac{1}{2}\cdot(x_i - \theta_1)^2}{\theta_2^2}\right).$$

Выпишем функцию правдоподобия:

$$L(\overline{x};\theta) = \frac{1}{\theta_2^n \cdot (2\pi)^{n/2}} \cdot \prod_{i=1}^{n} \exp\left(\frac{-\frac{1}{2}\cdot(x_i - \theta_1)^2}{\theta_2^2}\right)$$

и прологарифмируем ее:

$$\ln L(\overline{x};\theta) = -n \cdot \ln\theta_2 - \frac{n}{2}\cdot\ln[2\pi] - \frac{1}{2}\cdot\sum_{i=1}^{n}\frac{(x_i-\theta_1)^2}{\theta_2^2}.$$

Продифференцировав по θ_1 и θ_2 получаем систему уравнений:

$$\begin{cases} \dfrac{\partial(\ln[L(\overline{x};\theta)])}{\partial\theta_1} = & \dfrac{1}{\theta_2^2}\cdot\sum\limits_{i=1}^{n}(x_i - \theta_1) = 0 \\ \dfrac{\partial(\ln[L(\overline{x};\theta)])}{\partial\theta_2} = & -\dfrac{n}{\theta_2} + \dfrac{1}{\theta_2^3}\cdot\sum\limits_{i=1}^{n}(x_i-\theta_1)^2 = 0 \end{cases}$$

$$\begin{cases} \widehat{\theta}_1 = \frac{1}{n}\cdot\sum\limits_{i=1}^{n} x_i = \widehat{\alpha}_1, \\ \widehat{\theta}_2^2 = \frac{1}{n}\cdot\sum\limits_{i=1}^{n}(x_i - \theta_1)^2; \end{cases}$$

$$\begin{cases} \widehat{\theta}_1 = \widehat{\alpha}_1, \\ \widehat{\theta}_2 = \sqrt{\widehat{\alpha}_2 - 2\widehat{\alpha}_1^2 + \widehat{\alpha}_1^2} = \sqrt{\widehat{\alpha}_2 - \widehat{\alpha}_1^2}; \end{cases}$$

Заметим, что система имеет одно решение, поскольку точка экстремума единственная и при $\theta_1, \theta_2 \to \infty$ получаем, что $L(\overline{x};\theta) \to 0$, то есть на концах отрезка функция правдоподобия убывает.

Оценка, полученная методом моментов, совпала с оценкой максимального правдоподобия, однако это не всегда так.

Важным свойством оценок максимального правдоподобия является их инвариантность относительно преобразований параметра. Это означает, что если $q = q(\theta)$ — произвольная функция, взаимно однозначно отображающая параметрическое множество Θ рассматриваемой модели в некоторое множество Q, то о.м.п. в Q будет $\widehat{q} = q\left(\widehat{\theta}\right)$, где $\widehat{\theta}$ — о.м.п. в параметрическом семействе Θ. Принцип инвариантности позволяет в каждой конкретной задаче выбирать наиболее удобную параметризацию, а о.м.п. получать затем с помощью соответствующих преобразований.

4 Сравнения различных оценок

В предыдущем разделе были рассмотрены различные подходы к построению статистических оценок. В данном разделе ответим на вопрос как сравнивать различные оценки.

4.1 Среднеквадратичный подход к сравнению оценок

Пусть заданно параметрическое семейство распределений: $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$, $X = (X_1, X_2, \ldots, X_n)$ — выборка. Пусть получены две оценки T_1, T_2 неизвестного параметра $\tau = \tau(\theta)$.

Определение 4.1. *Будем говорить что оценка T_1, лучше оценки T_2 в среднеквадратическом смысле, если*

$$\mathsf{M}_\theta \left(T_1 - \tau\right)^2 < \mathsf{M}_\theta \left(T_2 - \tau\right)^2.$$

Определение 4.2. *Функцию $r = \sqrt{\mathsf{M}_\theta (T - \tau)^2}$ называют среднеквадратическим отклонением оценки T.*

Определение 4.3. *Среднеквадратической ошибкой оценки T параметра $\tau(\theta)$ будем называть величину $\Delta T = \mathsf{M}_\theta (T - \tau)^2$*

Заметим, что ΔT порождает критерий оптимальности оценок. Класс несмещенных оценок параметра τ будем обозначать $\mathcal{T}_\tau$.

Определение 4.4. *Оценку минимизирующую среднеквадратическое отклонение в классе несмещенных оценок будем называть оптимальной оценкой $T^* = \arg\min\limits_{T \in \mathcal{T}_\tau} \Delta T$.*

Утверждение 4.5. *Для несмещенной оценок среднеквадратическая ошибка совпадает с ее дисперсией, а для смещенной оценки больше ее дисперсии.*

Доказательство.

$$\begin{aligned}\mathsf{M}_\theta(T-\tau)^2 = \mathsf{M}_\theta(T - \mathsf{M}_\theta T + \mathsf{M}_\theta T - \tau)^2 = \\ = \underbrace{\mathsf{M}_\theta(T - \mathsf{M}_\theta T)^2}_{\mathsf{D}_\theta T} + \mathsf{M}_\theta(\underbrace{\mathsf{M}_\theta T - \tau}_{b})^2 + 2\underbrace{\mathsf{M}_\theta(T - \mathsf{M}_\theta T)(\mathsf{M}_\theta T - \tau)}_{0} = \\ = \mathsf{D}_\theta T + b^2.\end{aligned}$$

□

Теорема 4.6. *Пусть T_1 и T_2 две оптимальные оценки для неизвестного параметра $\tau = \tau(\theta)$. Тогда $T_1 = T_2$.*

Доказательство. Пусть $D = \mathsf{D}_\theta T_1 = \mathsf{D}_\theta T_2$. Обозначим $\Delta_k = T_k - \tau$, $k \in \overline{1,2}$. Рассмотрим оценку

$$\widehat{T} = \frac{T_1 + T_2}{2}.$$

Очевидно, что $\widehat{T}$ также будет несмещенной оценкой.

$$\left(\frac{\Delta_1 + \Delta_2}{2}\right)^2 + \left(\frac{\Delta_1 - \Delta_2}{2}\right)^2 = \frac{\Delta_1^2}{2} + \frac{\Delta_2^2}{2}.$$

Рассмотрим подробнее выражения в скобках:

$$\frac{\Delta_1 + \Delta_2}{2} = \frac{T_1 - \tau + T_2 - \tau}{2} = \widehat{T} - \tau;$$

$$\frac{\Delta_1 - \Delta_2}{2} = \frac{T_1 - \tau - T_2 + \tau}{2} = \frac{T_1 - T_2}{2}.$$

После подстановки и вычисления математического ожидания от левой и правой части получаем:

$$\mathsf{M}_\theta\left(\widehat{T} - \tau\right)^2 + \frac{1}{4}\mathsf{M}_\theta\left(T_1 - T_2\right)^2 = D.$$

Так как $\mathsf{M}_\theta\left(\widehat{T} - \tau\right)^2 \geqslant D$, то получаем, что $\mathsf{D}\left(T_1 - T_2\right) = 0$, что доказывает теорему. □

Пример 4.7. Рассмотрим выборку $X = (X_1, \ldots, X_n)$ из $\mathscr{L}(\xi)$, $\xi \sim Bi(1,\theta)$, $\theta \in (0,1)$. Хотим оценить параметр θ. Напомним, что $\mathsf{M}\, X_i = \theta$, $\mathsf{D}\, X_i = \theta(1-\theta)$.

Множество несмещенных оценок не пусто. Действительно, статистика $\frac{1}{n}\sum\limits_{i=1}^{n} x_i = \widehat{\alpha}$ как известно является несмещенной оценкой параметра θ:

$$\mathsf{M}\, \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} \mathsf{M}\, X_i = \frac{n}{n}\theta = \theta.$$

Эта оценка является состоятельной:

$$\mathsf{D}\, \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{D}\, X_i = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n} \to 0, n \to \infty.$$

Заметим, что таких оценок бесконечно много. Пусть $b_1 + \ldots + b_n = n$. Рассмотрим статистику T вида

$$T(X) = \frac{1}{n}\sum_{i=1}^{n} b_i X_i,$$

которая является как несмещенной так и состоятельной:

$$\mathsf{D}\, T(X) = \frac{1}{n^2}\sum_{i=1}^{n} b_i^2 \theta(1-\theta) \leq \frac{b^2}{n^2} n\theta(1-\theta) \leq \frac{b^2}{n} \to 0, n \to \infty,$$

где $b = \max(b_1, \ldots, b_n)$.

Покажем, что для любой оценки

$$\mathsf{D}_\theta \geqslant \frac{\theta(1-\theta)}{n} \ \forall \theta \in (0,1),$$

то есть $\widehat{\alpha}_1$ является оптимальной и несмещенной оценкой.

Для начала выпишем функцию правдоподобия:

$$L(\overline{x};\theta) = \theta^{\sum\limits_{i=1}^{n} x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i},$$

где $\overline{x}$ — реализация выборки. Очевидно, что

$$\sum_{\overline{x}} L(\overline{x};\theta) = 1.$$

Пусть T — произвольная несмещенная оценка

$$\theta = \mathsf{M}_\theta\, T(X) = \sum_{\overline{x}} T(\overline{x}) L(\overline{x};\theta).$$

Продифференцируем первое равенство по θ:

$$0 = \sum_{\overline{x}} \frac{\partial L(\overline{x};\theta)}{\partial\theta} = \sum_{\overline{x}} \frac{\partial \ln L(\overline{x};\theta)}{\partial\theta} L(\overline{x};\theta) = \mathsf{M}_\theta \left(\sum_{\overline{x}} \frac{\partial \ln L(\overline{x};\theta)}{\partial\theta} \right), \qquad (1)$$

пользуясь следующими равенствами:

$$\frac{\partial \ln f(x)}{\partial x} = \frac{1}{f(x)} \frac{\partial f(x)}{\partial x};$$

$$\frac{\partial f(x)}{\partial x} = \frac{\partial \ln f(x)}{\partial x} f(x).$$

Аналогично продифференцируем второе равенство:

$$1 = \sum_{\overline{x}} T(\overline{x}) \frac{\partial L(\overline{x};\theta)}{\partial\theta} = \mathsf{M}_\theta \left(T(\overline{x}) \frac{\partial \ln L\,(\overline{x};\theta)}{\partial\theta} \right). \qquad (2)$$

Так как $\mathsf{M}(a \cdot b) \leq \sqrt{\mathsf{M}\,a^2 \cdot \mathsf{M}\,b^2}$, то вычтя из (2) умноженное на θ значение (1) получим:

$$1 = \mathsf{M}_\theta \left((T(X) - \theta) \frac{\ln(L(X;\theta))}{\partial\theta} \right) \leq \sqrt{\mathsf{M}_\theta (T(X)-\theta)^2\, \mathsf{M}_\theta \left(\frac{\partial \ln(L(X;\theta))}{\partial\theta} \right)^2}.$$

И после возведения левой и правой части в квадрат:

$$1 \leq \mathsf{M}_\theta (T(X)-\theta)^2 \cdot \mathsf{M}_\theta \left(\frac{\partial \ln(L(X;\theta))}{\partial\theta} \right)^2$$

Так как $\mathsf{M}_\theta (T(X)-\theta)^2 = \mathsf{D}_\theta\, T(X)$, то

$$\mathsf{D}_\theta\, T(X) \geq \frac{1}{\mathsf{M}_\theta \left(\frac{\partial \ln(L(X;\theta))}{\partial\theta} \right)^2} \quad \forall\theta \in (0;1).$$

Осталось найти значение математического ожидания в знаменателе. Рассмотрим величину:

$$\frac{\partial \ln(L(\bar{x};\theta))}{\partial\theta} = \frac{\partial}{\partial\theta}\left(\sum_{i=1}^{n} x_i \ln\theta + (n - \sum_{i=1}^{n} x_i)\ln(1-\theta)\right) =$$

$$= \frac{\sum\limits_{i=1}^{n} x_i}{\theta} - \frac{n - \sum\limits_{i=1}^{n} x_i}{1-\theta} = \frac{(1-\theta)\sum\limits_{i=1}^{n} x_i - \theta_n + \theta\sum\limits_{i=1}^{n} x_i}{\theta(1-\theta)} = \frac{1}{\theta(1-\theta)}\sum_{i=1}^{n}(x_i - \theta).$$

$$\mathsf{M}_\theta\left(\frac{\partial \ln(L(X;\theta))}{\partial\theta}\right)^2 = \frac{1}{\theta^2(1-\theta)^2}\cdot\mathsf{M}_\theta\left(\sum_{i=1}^{n}(X_i-\theta)\right)^2 = \frac{n\cdot\mathsf{D}_\theta X_i}{\theta^2(1-\theta)^2} = \frac{n}{\theta(1-\theta)}.$$

Теорема 4.8. *Относительная частота произвольного события является оптимальной несмещенной оценкой для вероятности этого события.*

Следствие 4.9. *Значение эмпирической функции $\widehat{F}_n(x)$ в каждой точке x является оптимальной несмещенной оценкой для значения $F(x)$.*

4.2 Эффективные оценки. Неравенство Рао-Крамера.

Пусть заданно параметрическое семейство распределений: $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$, $X = (X_1, X_2, \ldots, X_n)$ — выборка, $\overline{x} = (x_1, \ldots, x_n)$ — реализация выборки.

Определение 4.10. *Параметрическое семейство $\mathcal{F} = \{F_\theta,\ \theta \in \Theta\}$ называется регулярным (по Рао-Крамеру), если выполнены следующие условия, которые в дальнейшем будем называть условиями регулярности:*

1. *$L(\bar{x};\theta) > 0$ для всех значений $\bar{x} \in B$ из выборочного пространства и дифференцируема по θ, $\forall\theta \in \Theta$ (Θ — параметрическое множество).*
2. *Случайная величина $V(X;\theta)$, называемая функцией вклада выборки и определенная равенством*

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

имеет ограниченную дисперсию:

$$0 < \mathsf{M}_\theta V^2(X;\theta) < \infty.$$

При этом значение $\dfrac{\partial \ln(f_\theta(X_i))}{\partial\theta}$ *будем называть вкладом* i*-го наблюдения выборки.*

3. $\forall\theta\in\Theta$ $\forall$ *статистики* $T(X)$ *верно равенство:*

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} T(\bar{x})L(\bar{x};\theta)d\bar{x} = \int\limits_{\mathbb{R}^n} T(\bar{x})\frac{\partial L(\bar{x};\theta)}{\partial\theta}d\bar{x}.$$

Рассмотрим некоторые свойства вклада выборки $V(\bar{x};\theta)$. Воспользуемся известным равенством:

$$\int\limits_{\mathbb{R}^n} L(\bar{x};\theta)d\bar{x} = 1.$$

Откуда продифференцировав по θ получаем:

$$0 = \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} L(\bar{x};\theta)d\bar{x} = \int\limits_{\mathbb{R}^n} \frac{\partial L(\bar{x};\theta)}{\partial\theta}d\bar{x} = \int\limits_{\mathbb{R}^n} \frac{\partial \ln(L(\bar{x};\theta)}{\partial\theta}L(\bar{x};\theta)d\bar{x} =$$
$$= \mathsf{M}_\theta \frac{\partial \ln(L(X);\theta)}{\partial\theta} = \mathsf{M}_\theta V(X;\theta).$$

Для регулярных моделей определена $\mathsf{D}_\theta V(X)$ корректно и равна $\mathsf{M}_\theta V^2(X)$.

Определение 4.11. *Функцию* $i_n(\theta)$*, определенную равенством:*

$$i_n(\theta) = \mathsf{D}_\theta V(X;\theta) = \mathsf{M}_\theta V^2(X;\theta)$$

будем называть информацией Фишера.

Заметим, что верно равенство:

$$i_n(\theta) = \mathsf{D}_\theta V(X;\theta) = \sum_{i=1}^{n} \mathsf{D}\, f(x_i;\theta) = n\cdot i_1(\theta) = n\cdot i(\theta)$$

Теорема 4.12. *Неравенство Рао-Крамера.*

Пусть $\tau(\theta)$ *— неизвестный параметр в регулярной модели* $\mathcal{F} = \{F_\theta, \theta\in\Theta\}$ *и* θ *— скалярный параметр. Пусть также* $\tau(\theta)$ *— дифференцируема по* θ*.*

Тогда для любой оценки $T = T(X)$ *параметр* $\tau(\theta), T\in\mathcal{T}_\tau$ *справедливо неравенство:*

$$\mathsf{D}_\theta T(X) \geq \frac{[\tau'(\theta)]^2}{n\cdot i(\theta)}.$$

При этом равенство выполняется тогда и только тогда, когда:

$$T(X) - \tau(\theta) = a(\theta) \cdot V(X; \theta),$$

где $a(\theta)$ — некоторая функция.

Доказательство.

$$\mathsf{M}_\theta T(X) = \int\limits_{\mathbb{R}^n} T(\bar{x}) L(\bar{x}; \theta) d\bar{x} = \tau(\theta)$$

так как оценка несмещенная. В виду регулярности модели:

$$\tau'(\theta) = \int\limits_{\mathbb{R}^n} T(\bar{x}) \frac{\partial \ln(L(\bar{x}; \theta))}{\partial \theta} L(\bar{x}; \theta) d\bar{x} = \mathsf{M}_\theta \left(T(\bar{x}) \cdot \frac{\partial \ln(L(\bar{x}; \theta))}{\partial \theta} \right) =$$
$$= \mathsf{M}_\theta (T(X) V(X; \theta)).$$

$$\begin{aligned} \tau'(\theta) = \mathsf{M}_\theta (T(X) V(X; \theta)) &= \mathsf{M}_\theta ((T(X) + \tau(\theta) - \tau(\theta)) \cdot V(X; \theta)) = \\ &= \mathsf{M}_\theta ((T(X) - \tau(\theta)) V(X; \theta)) + \tau(\theta) \mathsf{M}_\theta V(X; \theta) = \\ &= \mathsf{M}_\theta (T(X) - \tau(\theta)) \cdot (V(X; \theta) - \mathsf{M}_\theta V(X; \theta)) = \mathsf{cov}_\theta (T(X); V(X; \theta)). \end{aligned}$$

Используя неравенство Коши-Буняковского $\mathsf{cov}(\xi, \eta)^2 \leqslant \mathsf{D}\,\xi\, \mathsf{D}\,\eta$:

$$\tau'(\theta)^2 \leq \mathsf{D}_\theta T(X) \cdot \mathsf{D}_\theta V(X, \theta) = \mathsf{D}_\theta T(X) \cdot i_n(\theta)$$

Напомним, что неравенство Коши-Буняковского превращается в равенство, если ξ, η линейно связаны: $\xi = a\eta + b$. В нашем случае:

$$T(X) = aV(X; \theta) + b.$$

Вычислим математическое ожидание от левой и правой части:

$$b = \mathsf{M}_\theta T(X) - a \, \mathsf{M}_\theta V(X; \theta)$$

Так как $\mathsf{M}_\theta T(X) = \tau(\theta)$ и $\mathsf{M}_\theta V(X; \theta) = 0$, то:

$$T(X) = aV(X; \theta) + \tau(\theta).$$

Следовательно для любого значения θ существует такое $a = a(\theta)$, что завершает доказательство теоремы. $\square$

Следствие 4.13. *Если* $\mathsf{D}_\theta T(X) = \frac{[\tau'(\theta)]^2}{n \cdot i(\theta)}$, *то в этом случае* $\mathsf{D}_\theta T(X) = a(\theta) \cdot \tau'(\theta)$

Доказательство.

$$\mathsf{cov}_\theta(T(X); V(X;\theta)) = \tau'(\theta)$$

$$\mathsf{cov}_\theta(T(X); V(X;\theta)) = \mathsf{M}_\theta \left(T(X) \cdot V(X;\theta)\right) - \mathsf{M}_\theta T(X) \cdot \mathsf{M}_\theta V(X;\theta)$$

Так как $\mathsf{M}_\theta T(X) \cdot \mathsf{M}_\theta V(X;\theta) = 0$, то

$$\mathsf{M}_\theta \left(T(X) \cdot V(X;\theta)\right) = \mathsf{M}_\theta \left(T(X) \cdot \frac{T(X) - \tau(\theta)}{a(\theta)}\right) = \frac{1}{a(\theta)} \mathsf{M}_\theta \left(T^2(X) - T(X)\right),$$

откуда получаем:

$$\tau'(\theta) = \frac{1}{a(\theta)} \mathsf{D}\, T(X).$$

$\square$

Замечание 4.14. *Если* $T(X)$ *- смещенная оценка,* $\mathsf{M}_\theta T(X) = \tau(\theta) + b(\theta)$, *то*

$$\mathsf{D}_\theta T(X) \leq \frac{[\tau'(\theta) + b'(\theta)]^2}{n \cdot i(\theta)}.$$

Следствие 4.15. *Если* $T(X)$ *является несмещнной оценкой для* θ, *то:*

$$\mathsf{D}_\theta T(X) \leq \frac{1}{n \cdot i(\theta)}.$$

Определение 4.16. *Эффективностью оценки* $T = T(X)$ *параметра* $\tau(\theta)$ *будем называть величину*

$$e(T) = \frac{(\tau'(\theta))^2}{i_n(\theta)\, \mathsf{D}_\theta T},\ 0 \leq e(T) \leq 1.$$

Если выполняется равенство $e(t) = 1$, *то в этом случае оценка* $T(X)$ *называется эффективной.*

Заметим, что ранее, в частности было показано, что

$$\frac{\theta(1-\theta)}{n} V(X;\theta) = \overline{X} - \theta.$$

Тогда положив $\tau(\theta) = \theta$ и $a(\theta) = \theta(1-\theta)/n$ получим, что статистика $\overline{X}$ является эффективной оценкой параметра в модели $Bi(1,\theta)$. Рассмотрим еще примеры эффективных статистик.

Пример 4.17. Рассмотрим выборку $X = (X_1, \dots X_n)$ из $\mathscr{L}(\xi), \xi \sim N(\theta,\sigma^2)$. Выпишем вклад оного наблюдения:

$$V(X_1,\theta) = \frac{\partial \ln f_\theta(X_1)}{\partial\theta} = \frac{\partial}{\partial\theta} \ln\left(\left(\frac{1}{\sqrt{2\pi}\sigma}\right)\exp\left(-\frac{(X_1-\theta)^2}{2\sigma^2}\right)\right) = \frac{X_1-\theta}{\sigma^2}.$$

Заметим, что $f_\theta(X_1)$ — дважды дифференцируема и $\mathsf{M}_\theta(V(X,\theta)) = 0$, откуда:

$$\frac{\partial}{\partial\theta}\mathsf{M}_\theta\left(V(X,\theta)\right) = \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} \frac{\partial \ln f_\theta(\overline{x})}{\partial\theta} f_\theta(\overline{x})d\overline{x} =$$
$$= \int\limits_{\mathbb{R}^n} \frac{\partial^2 \ln f_\theta(\overline{x})}{\partial^2\theta} f_\theta(\overline{x})d\overline{x} + \int\limits_{\mathbb{R}^n}\left(\frac{\partial \ln f_\theta(\overline{x})}{\partial\theta}\right)^2 f_\theta(\overline{x})d\overline{x},$$

откуда получаем, что $\mathsf{M}_\theta \frac{\partial^2 \ln f_\theta(X_1)}{\partial^2} = -i_1(\theta)$ — информация Фишера. Таким образом доказано следующее

Утверждение 4.18. *Если плотность случайной величины ξ дважды дифференцируема, то*

$$-\mathsf{M}_\theta \frac{\partial^2 \ln f_\theta(X_1)}{\partial^2} = i_1(\theta).$$

Вернемся к примеру и воспользуемся доказанным утверждением:

$$\frac{\partial^2 \ln f_\theta(X_1)}{\partial\theta^2} = -\frac{1}{\sigma^2}.$$

Откуда получаем, что $i(\theta) = \frac{1}{\sigma^2}$.

Рассмотрим статистику

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и покажем, что она эффективная оценка параметра θ. Действительно,

$$\frac{1}{ni(\theta)} = \frac{\sigma^2}{n} = \mathsf{D}_\theta\,\overline{X}.$$

Также можно заметить, что

$$\frac{\sigma^2}{n} V(X;\theta) = \overline{X} - \theta.$$

4.3 О доказательстве регулярности модели

Для доказательства регулярности модели необходимо показать возможность перестановки знаков дифференцирования и интегрирования. Напомним следующую теорему из курса математической статистики, которая в большинстве случаев помогает доказать свойство регулярности:

Теорема 4.19. *Пусть функция $f(x,\theta)$, определенная в прямоугольнике $[a,b;c,d]$, будет непрерывна по x в $[a,b]$ при любом постоянном θ в $[c,d]$. Предположим, что во всей области существует частная производная $\frac{\partial f(x,\theta)}{\partial \theta}$, непрерывная как функция двух переменных. Тогда при любом θ из $[c,d]$ имеет место формула*

$$\frac{\partial}{\partial \theta} \int\limits_a^b f(x,\theta) dx = \int\limits_a^b \frac{\partial f(x,\theta)}{\partial \theta} dx.$$

Аналогично для несобственного интеграла:

Теорема 4.20. *Если:*

- *функции $f(x,\theta)$, $\frac{\partial f(x,\theta)}{\partial \theta}$ непрерывны на множестве $\left\{ (x,\theta) \in \mathbb{R}^2 \middle| \, x \in [a,\omega), \theta \in [c,d] \right\}$,*
- *интеграл*

$$\Phi(\theta) = \int\limits_a^\omega \frac{\partial f(x,\theta)}{\partial \theta} dx$$

  сходится равномерно на множестве $Y = [c,d]$,
- *интеграл*

$$F(\theta) = \int\limits_a^\omega f(x,\theta) dx$$

  сходится хотя бы при одном значении $\theta_0 \in Y$, *то он сходится и даже равномерно на всем множестве Y, при этом функция $F(\theta)$ оказывается дифференцируемой и справедливо равенство*

$$\frac{\partial F(\theta)}{\partial \theta} = \int\limits_a^\omega \frac{\partial f(x,\theta)}{\partial \theta} dx.$$

В эквивалентной форме условия регулярности можно переписать следующим образом:

- Существует такое множество C, что $\mathsf{P}_\theta(\xi) = 1$, что при каждом $y \in C$ функция $\sqrt{f_\theta(y)}$ непрерывно дифференцируема по θ всюду в области Θ,
- информация Фишера $i_n(\theta)$ существует, положительна и непрерывна по θ во всех точках $\theta \in \Theta$.

4.4 Экспоненциальное семейство

Определение 4.21. *Параметрическое семейство $\mathcal{F} = \{F_\theta,\, \theta \in \Theta\}$ называется экспоненциальным, если плотность $f_\theta(x)$ имеет следующий вид:*

$$f_\theta(x) = \exp\left\{A(\theta) \cdot B(x) + C(\theta) + D(x)\right\}.$$

Заметим, что следующие модели очевидно являются экспоненциальными: $N(\theta, \sigma^2)$, $N(\mu, \theta^2)$, $\Gamma(\theta, \lambda)$, $Bi(k,\theta)$, $\overline{Bi}(r, \theta)$, $\mathrm{Pois}(\theta)$, $R(0,\theta)$.

Надем вклад выборки для экспоненциальной модели:

$$V(X;\theta) = \frac{\partial \ln \prod\limits_{i=1}^{n} f_\theta(X_i)}{\partial \theta} = A'(\theta) \sum_{i=1}^{n} B(X_i) + nC'(\theta) = \\ = nA'(\theta) \left(\frac{1}{n} \sum_{i=1}^{n} B(X_i) + \frac{C'(\theta)}{A'(\theta)} \right).$$

Рассмотрим статистику

$$T(X) = \frac{1}{n} \sum_{i=1}^{n} B(X_i)$$

и покажем, что $T(X)$ является эффективной оценкой параметра

$$\tau(\theta) = -\frac{C'(\theta)}{A'(\theta)}.$$

Действительно, пользуясь неравенством Рао-Крамера получаем, что

$$T(x) - \tau(\theta) = \underbrace{\frac{1}{nA'(\theta)}}_{a(\theta)} \cdot V(X;\theta).$$

Таким образом *в случае регулярности параметрического семейства* статистика $T(X)$ является эффективной оценкой для параметрической функции $\tau(\theta)$. Более того, согласно следствию 4.13 к теореме Рао-Крамера получаем:

$$\mathsf{D}_\theta T = \frac{\tau'(\theta)}{nA'(\theta)}.$$

Пример 4.22. Заметим, что условие регулярности здесь играет очень важную роль. Действительно, легко показать, что параметрическое семейство $R(0,\theta)$ является экспоненциальным, однако оно не является регулярным. Покажем, что семейство $R(0,\theta)$ не является регулярным. Воспользуемся тождеством

$$\int\limits_0^\theta \frac{dx}{\theta} = 1.$$

Если модель регулярна, то должно выполнятся следующее равенство:

$$\int\limits_0^\theta \frac{\partial}{\partial\theta}\frac{dx}{\theta} = 0,$$

что неверно.

На самом деле верно и обратное утверждение: если существует эффективная оценка для некоторой функции $\tau(\theta)$, то семейство является регулярным. Чтобы доказать это воспользуемся равенством из теоремы Рао-Крамера:

$$(T(X) - \tau(\theta)) \cdot a^{-1}(\theta) = V(X;\theta)$$

и проинтегрируем его по θ:

$$A_1(\theta)T(X) + C_1(\theta) + D_1(X) = \ln L(X;\theta).$$

Отсюда однозначно определяется вид плотности распределения.

Для регулярных экспоненциальных семейств функцию информации можно вычислить по следующей формуле:

$$i(\theta) = \frac{(\tau'(\theta))^2}{\mathsf{D}_\theta T(X)} = \frac{(\tau'(\theta))^2}{a(\theta)\cdot\tau'(\theta)} = \tau'(\theta)\cdot A'(\theta) = \left(-\frac{C'(\theta)}{A'(\theta)}\right)'\cdot A'(\theta) = \\ = \frac{C'(\theta)A''(\theta)}{A'(\theta)} - C''(\theta).$$

5 Условное математическое ожидание

Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathsf{P})$ и заданную на нем случайную величину ξ. В курсе теории вероятности определялось понятие условной вероятности события $C \in \mathcal{F}$ при условии, что происходит некоторое событие $B \in \mathcal{F}$:

$$\mathsf{P}\left(C\,|B\right) = \frac{\mathsf{P}(CB)}{\mathsf{P}(B)}.$$

Если событие B фиксировано, то совокупность условных вероятностей $\mathsf{P}\left(C\,|B\right)$, $C \in \mathcal{F}$, задает на измеримом пространстве $(\Omega, \mathcal{F})$ новую вероятностную меру P_B, $\mathsf{P}_B(C) = \mathsf{P}\left(B|C\right)$, сосредоточенную на множестве B:

$$\mathsf{P}_B(B) = 1, \mathsf{P}_B(\Omega\backslash B) = 0.$$

Пусть ξ — случайная величина: заданная на $(\Omega, \mathcal{F}, \mathsf{P})$ и принимающая значения из конечного или счетного множества $X = \{x_1, x_2, \ldots\}$. Этой случайной величине соответствует разбиение $\{A_1, A_2, \ldots\}$ пространства Ω на непересекающиеся $\mathcal{F}$-измеримые множества

$$A_k = \{\omega \in \Omega\colon \xi(\omega) = x_k\} \in \mathcal{F},\ k \geqslant 1,$$

порождающие σ-алгебру $\sigma(\xi)$. В этом случае математическое ожидание ξ можно вычислить как относительно основного разбиения P:

$$\mathsf{M}\,\xi = \sum_{\omega\in\Omega}\xi(\omega)\mathsf{P}(\omega) = \sum_{k\geqslant 1} x_k \mathsf{P}(A_k),$$

так и относительно условного распределения P_B:

$$\mathsf{M}\left(\xi|B\right) = \sum_{\omega\in\Omega}\xi(\omega)P(\omega|B) = \sum_{k\geqslant 1} x_k P\left(A_k\,|B\right).$$

Пусть $(\Omega, \mathcal{F}, \mathsf{P})$ — вероятностное пространство и $\mathcal{B} = \{B_1, B_2, \ldots\}$ — измеримое разбиение Ω на множества положительной меры. Заметим, что здесь и далее все разбиения будем считать измеримыми, не оговаривая это специально.

Рассмотрим произвольное событие $C \in \mathcal{F}$. Для данного события определены условные вероятности $\mathsf{P}\left(C \,|B_k\right)$, $k \geqslant 1$. Введем случайную величину, опредлив измеримую функцию на Ω, которая на множестве B_k принимает значение $P\left(C \,|B_k\right)$:

$$\xi_{\mathcal{B}} = \xi_{\mathcal{B}}(\omega) = \sum_{k \geqslant 1} \mathsf{P}\left(C \,|B_k\right) \operatorname{Ind}(\omega \in B_k) = \mathsf{P}\left(C \,|\mathcal{B}\right).$$

Случайная величина $\xi_{\mathcal{B}}(\omega)$ называется *условной вероятностью события C относительно разбиения $\mathcal{B}$*.

Такое определение позволяет сравнивать условные вероятности относительно разных разбиений, поскольку эти условные вероятности являются функциями с общей областью определения Ω. Например, пусть ξ и η — независимые случайные величины, принимающие значения $\{x_1, \ldots, x_n\}$ и $\{y_1, \ldots, y_m\}$ и порождающие разбиения $\mathcal{A} = \{A_1, \ldots, A_n\}$ и $\mathcal{B} = \{B_1, \ldots, B_m\}$,

$$\mathsf{P}(\xi = x_i) = p_i,\ i \in \overline{1,n},$$

$$\mathsf{P}(\eta = y_i) = q_j,\ j \in \overline{1,m}.$$

Тогда при любом $i \in \overline{1,n}$

$$\mathsf{P}\left(A_i \,|\mathcal{B}\right) = \mathsf{P}\left(\xi = x_i \,|\mathcal{B}\right) = \mathsf{P}\left(\xi = x_i\right) = p_i.$$

В тоже время

$$\mathsf{P}\left(A_i \,|\mathcal{A}\right) = \mathsf{P}\left(\xi = x_i \,|\mathcal{A}\right) = \operatorname{Ind}(x \in A_i),$$

то есть условные вероятности относительно разбиения $\mathcal{B}$ являются константами со значениями $p_1, \ldots, p_n$ (что соответствует независимости событий), а относительно разбиения $\mathcal{A}$ — ступенчатыми функциями, принимающими только два значения: 0 и 1.

В терминах условных вероятностей стандартная формула полной вероятности

$$\mathsf{P}(C) = \sum_{k \geqslant 1} \mathsf{P}\left(C \,|B_k\right) \mathsf{P}(B_k),\ B_i \cap B_j = \emptyset,\ i \neq j,\ \bigcup_{k \geqslant 1} B_k = \Omega,$$

принимает другой вид. Заменим в ней $\mathsf{P}(B_k)$ на $\mathsf{M}\,\mathrm{Ind}(\omega \in B_k)$ и воспользуемся аддитивностью математического ожидания:

$$\mathsf{P}(C) = \sum_{k \geqslant 1} \mathsf{P}\left(C \,|B_k\right) \mathsf{P}(B_k) = \sum_{k \geqslant 1} \mathsf{P}\left(C \,|B_k\right) \mathsf{M}\,\mathrm{Ind}(\omega \in B_k) =$$
$$= \mathsf{M} \sum_{k \geqslant 1} \mathsf{P}\left(C \,|B_k\right) \mathrm{Ind}(\omega \in B_k) = \mathsf{M}\,\mathsf{P}\left(C \,|\mathcal{B}\right),$$

где $\mathcal{B} = \{B_1, B_2, \ldots\}$ — разбиение, или

$$\mathsf{P}(C) = \mathsf{M}\,\mathsf{P}\left(C \,|\mathcal{B}\right).$$

Если разбиение $\mathcal{B}$ порождается случайной величиной η, то говорят, об *условной вероятности относительно случайной величины* η или относительно порожденной ею σ-алгебры $\sigma(\eta)$:

$$\mathsf{P}\left(C \,|\eta\right) = \mathsf{P}\left(C \,|\sigma(\eta)\right) = \mathsf{P}\left(C \,|\eta\right)(\omega).$$

Замечание 5.1. *Так как случайная величина η на каждом элементе порожденного ей разбиения постоянна, то условная вероятность $\mathsf{P}\left(C \,|\eta\right)$ есть суть функция от C и $\eta(\omega)$.*

Рассмотрим теперь аналогичную конструкцию для случайных величин. Пусть на вероятностном пространстве $(\Omega, \mathcal{F}, \mathsf{P})$ заданы:

- разбиение $\mathcal{B} = \{B_1, B_2, \ldots\}$;
- случайная величина ξ с конечным или счетным множеством значений $x_1, x_2, \ldots \in \mathbb{R}$;
- порожденное случайной величиной разбиение $\mathcal{A} = \{A_1, A_2, \ldots\}$ пространства Ω на непересекающиеся множества

$$A_k = \{\omega \in \Omega \colon \xi(\omega) = x_k\} \in \mathcal{F},\ k \geq 1.$$

Условные вероятности

$$\mathsf{P}\left(A_k\,|\mathcal{B}\right) = \mathsf{P}\left(A_k\,|\mathcal{B}\right)(\omega) = \mathsf{P}\left(\xi = x_k\,|\mathcal{B}\right)(\omega),\ \omega \in \Omega,$$

определены для всех $k \geqslant 1$, и для каждого $\omega \in \Omega$ они образуют вероятностное распределение на множестве $\{x_1, x_2, \ldots\}$ значений ξ, зависящее от ω. Определим *условное математическое ожидание* ξ *относительно разбиения* $\mathcal{B}$ как функцию, отображающую Ω в $\mathbb{R}$ (т.е. как случайную величину):

$$\mathsf{M}\left(\xi|\mathcal{B}\right)(\omega) = \sum_{k\geqslant 1} x_k \mathsf{P}\left(A_k\,|\mathcal{B}\right)(\omega) = \sum_{k\geqslant 1} x_k \mathsf{P}\left(\xi = x_k\,|\mathcal{B}\right)(\omega).$$

Преобразуем последнюю формулу:

$$\mathsf{M}\left(\xi|\mathcal{B}\right)(\omega) = \sum_{k\geqslant 1} x_k \mathsf{P}\left(A_k\,|\mathcal{B}\right)(\omega) = \sum_{k\geqslant 1} x_k \sum_{j\geqslant 1} \mathsf{P}\left(A_k\,|B_j\right) \mathrm{Ind}(\omega \in B_j) =$$
$$= \sum_{j\geqslant 1} \mathrm{Ind}(\omega \in B_j) \sum_{k\geqslant 1} x_k \mathsf{P}\left(A_k\,|B_j\right) = \sum_{j\geqslant 1} \mathrm{Ind}(\omega \in B_j)\, \mathsf{M}\left(\xi|B_j\right).$$

Таким образом, условное математическое ожидание ξ относительно разбиения $\mathcal{B}$ есть случайная величина, принимающая на каждом элементе B_j этого разбиения постоянное значения равное $\mathsf{M}\left(\xi|B_j\right)(\omega)$. Отсюда следует, что $\mathsf{M}\left(\xi|\mathcal{B}\right)(\omega)$ — функция измеримая относительно разбиения $\mathcal{B}$. Действительно, для любого $\mathcal{B}$-измеримого множества D событие $\{\omega\colon\ \mathsf{M}\left(\xi|\mathcal{B}\right) \in D\}$ является объединением множеств B_j, то есть принадлежит σ-алгебре, порожденной разбиением $\mathcal{B}$.

Пример 5.2. Пусть $\Omega = [0,1)$, $\xi(\omega) = \omega$ и σ-алгебра $\mathcal{B}$ порождается множествами $\left[0, \frac{1}{5}\right)$, $\left[\frac{1}{5}, \frac{2}{5}\right)$, $\left[\frac{2}{5}, \frac{3}{5}\right)$, $\left[\frac{3}{5}, \frac{4}{5}\right)$, $\left[\frac{4}{5}, 1\right)$. Тогда случайную величину $\mathsf{M}(\xi|\mathcal{B})$ можно описать таблицей

B_j	$\left[0, \frac{1}{5}\right)$	$\left[\frac{1}{5}, \frac{2}{5}\right)$	$\left[\frac{2}{5}, \frac{3}{5}\right)$	$\left[\frac{3}{5}, \frac{4}{5}\right)$	$\left[\frac{4}{5}, 1\right)$
$\mathsf{M}\left(\xi|\mathcal{B}\right)(\omega)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{7}{10}$	$\frac{9}{10}$

Утверждение 5.3. *Для любой случайной величины* ξ *и для любого события* C*, измеримого относительно разбиения* $\mathcal{B} = \{B_1, B_2, \ldots\}$*, справедлива формула*

$$\mathsf{M}\,\mathrm{Ind}\left(\omega \in C\right) \cdot \xi(\omega) = \mathsf{M}\left(\mathrm{Ind}\left(\omega \in C\right) \cdot \mathsf{M}\left(\xi|\mathcal{B}\right)\right).$$

Доказательство. Действительно, так как C измеримо относительно $\mathcal{B}$, то

$$C = \bigcup_{j\colon B_j \in C} B_j$$

и

$$\begin{gathered}\mathsf{M}\,\mathrm{Ind}\,(\omega \in C)\cdot \xi(\omega) = \sum_{j\colon B_j\in C}\sum_{\omega\in B_j}\xi(\omega)\mathsf{P}(\omega) = \\ = \sum_{j\colon B_j\in C}\sum_{\omega\in B_j}\xi(\omega)\mathsf{P}\left(\omega\,|B_j\right)\mathsf{P}(B_j) = \sum_{j\colon B_j\in C}\mathsf{M}\left(\xi|B_j\right)\mathsf{P}(B_j) = \\ = \sum_{j\colon B_j\in C}\mathsf{M}\left(\xi|B_j\right)\sum_{\omega\in B_j}\mathsf{P}(\omega) = \sum_{\omega\in C}\mathsf{M}\left(\xi|\mathcal{B}\right)(\omega)\mathsf{P}(\omega) = \mathsf{M}\left(\mathrm{Ind}\,(\omega\in C)\cdot\mathsf{M}\left(\xi|\mathcal{B}\right)\right).\end{gathered}$$

□

Следствие 5.4. *Для любой случайной величины ξ и любого разбиения $\mathcal{B}$ справедлива формула полного математического ожидания:*

$$\mathsf{M}\,\xi = \mathsf{M}\,\mathsf{M}(\xi|\mathcal{B}).$$

Аналогично (хоть и использованием много более сложного математического аппарата) определяется условное математическое ожидание, когда разбиение $\mathcal{B}$ порождаются величинами, имеющими непрерывные распределения. В этом случае условное математическое ожидание случайной величины ξ относительно σ-алгебры $\mathcal{B}$ называется случайная величина $\mathsf{M}(\xi|\mathcal{B})$, измеримая относительно $\mathcal{B}$ и удовлетворяющая тождеству:

$$\int_C \xi d\mathsf{P} = \int_C \mathsf{M}(\xi|\mathcal{B})d\mathsf{P}$$

для любого события $C \in \mathcal{B}$.

Это равенство аналогично определению плотности абсолютно непрерывного распределения — определяется с точностью до множеств, имеющих меру 0.

Для условных математических ожиданий можно доказать следующие свойства:

1. Линейность: для любых случайных величин ξ, η с конечными математическими ожиданиями и чисел a,b справедливо равенство:

$$\mathsf{M}\,(a\xi + b\eta|\mathcal{B}) = a\,\mathsf{M}\,(\xi|\mathcal{B}) + b\,\mathsf{M}\,(\eta|\mathcal{B})\,.$$

2. Если существует $\mathsf{M}\,\xi$, то $\mathsf{M}\,(\xi|\{\emptyset,\Omega\}) = \mathsf{M}\,\xi$.
3. Если распределение случайной величины ξ вырождено, т.е. $\mathsf{P}\,(\xi = C = \mathrm{const}) = 1$, то $\mathsf{M}(\xi|\mathcal{B}) = C$.
4. $\mathsf{M}(\mathrm{Ind}(A)|\mathcal{B}) = \mathsf{P}(A|\mathcal{B})$ для любого $A \in \mathcal{F}$. Действительно, если $\mathcal{B}$ порождается разбиением $\{B_1, B_2, \ldots\}$, то в силу свойства (1):

$$\mathsf{M}(\mathrm{Ind}(A)|\mathcal{B}) = \mathsf{M}\left(\sum_{i\geqslant 1} \mathrm{Ind}(AB_i)\middle|\,\mathcal{B}\right) =$$

$$= \sum_{i\geqslant 1} \mathsf{M}\,(\mathrm{Ind}(AB_i)|\,B_i)\,\mathrm{Ind}(B_i) = \sum_{i\geqslant 1} \mathsf{P}\,(A\,|B_i)\,\mathrm{Ind}(B_i) = \mathsf{P}\,(A\,|\mathcal{B})\,.$$

5. Если $\mathsf{P}(\xi \leqslant \eta) = 1$, то $\mathsf{M}(\xi|\mathcal{B}) \leqslant \mathsf{M}(\eta|\mathcal{B})$. Действительно, для любого $A \in \mathcal{B}$:

$$\int\limits_A \mathsf{M}(\xi|\mathcal{B})d\mathsf{P} = \mathsf{M}\,\mathsf{M}(\xi|\mathcal{B})\,\mathrm{Ind}(A) = \mathsf{M}\,\xi\,\mathrm{Ind}(A) \leq$$

$$\leqslant \mathsf{M}\,\eta\,\mathrm{Ind}(A) = \int\limits_A \mathsf{M}(\eta|\mathcal{B})d\mathsf{P},$$

поэтому $\mathsf{P}\,(\mathsf{M}(\eta|\mathcal{B}) - \mathsf{M}(\xi|\mathcal{B}) \geqslant 0) = 1$.
6. Для условных математических ожиданий верны теоремы сходимости, справедливые для обычных математических ожиданий, в частности, теорема о монотонной сходимости и теорема о мажорируемой сходимости.

6 Способы построения оценок

6.1 Достаточные статистики

Определение 6.1. *Статистика $T = T(X)$ называется достаточной для параметра $\tau(\theta)$ (в модели $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$), если для любого события $A \subset \mathcal{B}$ условная вероятность $\mathsf{P}_\theta(X \in A|T(X) = t)$ не зависит от θ.*

Это свойство означает, что статистика T содержит всю информацию о параметре θ, содержащуюся в выборке. Иными словами: статистика T достаточна, если условная плотность $L(\overline{x}|t;\theta)$ выборки $X = (X_1, \ldots, X_n)$ при условии $T(X) = t$ не зависит от параметра θ. Очевидно, что сама выборка X, очевидно, является достаточной статистикой так как $\mathsf{P}_\theta(X \in A|X = \overline{x})$ при любом θ есть либо 1, если $\overline{x} \in A$, либо 0 в противном случае.

Обычно стараются найти достаточную статистику наименьшей размерности, представляющую данные в наиболее сжатом виде. При наличии достаточной статистики все статистические выводы об исследуемой модели в конечном итоге формулируются в терминах этой статистики.

Пример 6.2. Рассмотрим выборку $X = (X_1, X_2, \ldots, X_n)$ из распределения $\mathscr{L}(\xi), \xi \sim \mathrm{Pois}(\lambda)$.

Покажем, что $T(X) = \sum\limits_{i=1}^{n} X_i$ является достаточной статистикой. Выпишем условную вероятность вектора:

$$\mathsf{P}(X = \overline{x}|T(X) = t) = \begin{cases} \mathsf{P}(X = \overline{x}), & \sum\limits_{i=1}^{n} x_i = t \\ 0, & \text{в противном случае.} \end{cases}$$

где $\overline{x} = (x_1, x_2, \ldots, x_n)$.

$$\mathsf{P}(X = \overline{x}|T(X) = t) = \\ = \frac{\mathsf{P}(X = \overline{x}, T(X) = t)}{\mathsf{P}(T(X) = t)} = \frac{\mathsf{P}(X = \overline{x})}{\sum\limits_{\overline{x}:T(\overline{x})=t} \mathsf{P}(\overline{x})} = \frac{\mathsf{P}(X = \overline{x})}{\mathsf{P}\left(\sum\limits_{i=1}^{n} x_i = t\right)}.$$

Так как $\sum\limits_{i=1}^{n} X_i \sim \mathrm{Pois}(\lambda n)$ получаем, что

$$\mathsf{P}(X = \overline{x}|T(X) = t) = \frac{\prod\limits_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}}{\frac{(n\lambda)^t}{t!} e^{-n\lambda}}$$

Так как $\prod\limits_{i=1}^{n} \lambda^{a_i} e^{-\lambda} = \lambda^t e^{-n\lambda}$, то все выражения с параметром λ сокращаются. Таким образом искомая условная вероятность не зависит от λ и статистика $T(X)$ является достаточной.

Теорема 6.3 (Критерий факторизации). *Для того, чтобы статистика $T = T(X)$ была достаточной, необходимо и достаточно, чтобы функция правдоподобия $L(\overline{x};\theta)$ имела следующий вид:*

$$L(\overline{x};\theta) = g\left(T(\overline{x});\theta\right) \cdot h(\overline{x}),$$

где g и h — неотрицательные функции, и функция h не зависит от θ.

Доказательство. Если статистика T достаточна, то для любого t из области значений $T(X)$ функция $L(\overline{x}|t;\theta)$ не зависит от θ и ее можно записать в виде $h\left(\overline{x};t\right)$.

Пусть $\mathsf{P}_\theta(T(X) = t) = g(t;\theta)$ и $T\left(\overline{x}\right) = t$. Тогда событие $\{X = \overline{x}\} \subseteq \{T(X) = t\}$, поэтому

$$\begin{aligned} L\left(\overline{x};\theta\right) &= \mathsf{P}_\theta\left(X = \overline{x}\right) = \mathsf{P}_\theta(X = \overline{x}, T(x) = t) = \\ &= \mathsf{P}_\theta(T(x) = t) \cdot \mathsf{P}_\theta(X = \overline{x}|T(X) = t) = g(t;\theta) \cdot L(\overline{x}|t;\theta) = g(t;\theta) \cdot h(\overline{x};t). \end{aligned}$$

Доказательство в обратную сторону проведем только для дискретной модели. Пусть имеет место разложение (факторизация). Тогда для любого $\overline{x}$ такого, что $T\left(\overline{x}\right) = t$ верно:

$$\begin{aligned} L(\overline{x}|t;\theta) = \mathsf{P}_\theta(X = \overline{x}|T(X) = t) &= \frac{\mathsf{P}_\theta(X = \overline{x}, T(X) = t)}{\mathsf{P}_\theta(T(X) = t)} = \\ &= \frac{L(\overline{x};\theta)}{\sum\limits_{\overline{x}:T(\overline{x})=t} L(\overline{x};\theta)} = \frac{g(T(\overline{x});\theta) \cdot h(\overline{x})}{\sum\limits_{\overline{x}} g(T(\overline{x});\theta) \cdot h(\overline{x})} = \frac{h(\overline{x})}{\sum\limits_{\overline{x}} h(\overline{x})} \end{aligned}$$

— не зависит от θ. □

Следствие 6.4. *Эффективная оценка является достаточной.*

Доказательство. Пользуясь доказанной теоремой Рао-Крамера:

$$T(\overline{x}) - \tau(\theta) = a(\theta)\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}.$$

Так как:

$$\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta} = \frac{1}{L(\overline{x};\theta)} \cdot \frac{\partial L(\overline{x};\theta)}{\partial\theta},$$

то

$$L(\overline{x};\theta) = \frac{a(\theta)\frac{\partial L(\overline{x};\theta)}{\partial\theta}}{T(x) - \tau(\theta)}.$$

Для завершения доказательства осталось ввести следующие обозначения:

$$\frac{a(\theta)}{T(\overline{x}) - \tau(\theta)} = g(T(\overline{x}); \theta),$$

$$\frac{\partial \ln L(\overline{x}; \theta)}{\partial \theta} = h(\overline{x}).$$

□

Следствие 6.5. *Если $T = T(X)$ — достаточная статистика параметра $\tau(\theta)$, то о.м.п. является функцией от T.*

Доказательство.

$$\widehat{Q}(x) = \arg\max_{\theta} L(\overline{x}; \theta) = \arg\max_{\theta} g(T(\overline{x}); \theta) h(x) = \arg\max_{\theta} g(T(\overline{x}); \theta).$$

□

Пример 6.6 (Равномерная модель и достаточная статистика для нее)**.** Пусть $X = (X_1, \ldots, X_n)$ — выборка из равномерного распределения $R(0,\theta)$. Выпишем плотность равномерного распределения:

$$f_\theta(x) = \theta^{-1} \operatorname{Ind}(0 \leqslant x \leqslant \theta)$$

и найдем плотность X как случайного вектора:

$$L(\overline{x}; \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \theta^{-n} \prod_{i=1}^{n} \operatorname{Ind}(0 \leqslant x_i \leqslant \theta) = \theta^{-n} \operatorname{Ind}(x_{(n)} \leqslant \theta) \operatorname{Ind}(x_{(1)} \geqslant 0).$$

Отсюда следует, что в данном случае достаточной статистикой является максимальное значение выборки $X_{(n)}$.

Теорема 6.7 (Рао-Блекуэлла-Колмогорова)**.** *Оптимальная оценка, если существует, является функцией от достаточной статистики.*

Доказательство. Пусть $T = T(X)$ - достаточная статистика, T_1 - произвольная несмещенная оценка заданной параметрической функции (она существует по условию теоремы).

Рассмотрим функцию

$$H(t) = \mathsf{M}_\theta(T_1 | T = t) = \int_{\mathcal{B}} T_1(\overline{x}) \cdot L(\overline{x} | t; \theta) d\overline{x}.$$

Эта функция не зависит от θ, так как условная плотность $L(\overline{x}|t;\theta)$ не зависит от параметра θ. Докажем, что $H(T(X))$ будет являться несмещнной оценкой для $\tau(\theta)$. Действительно, пользуясь следствием 5.3:

$$\mathsf{M}(H(T(X))) = \mathsf{M}(\mathsf{M}(T|T_1)) = \mathsf{M}(T_1) = \tau(\theta).$$

Покажем, что

$$\mathsf{D}_\theta(H(T)) \leqslant \mathsf{D}_\theta(T_1),\ \forall\theta,$$

причем равенство имеет место в том и только в том случае, когда $T_1 = H(T)$. Для доказательства заметим, что

$$\begin{aligned}&\mathsf{D}_\theta(T_1) = \mathsf{M}_\theta(T_1 - \tau(\theta) + H(T) - H(T))^2 = \\ &= \mathsf{M}_\theta[(T_1 - H(T)]^2 + 2\,\mathsf{M}_\theta[(T_1 - H(T))(H(T) - \tau(\theta))] + \mathsf{M}_\theta[-\tau(\theta) - H(T))]^2 = \\ &\qquad = \mathsf{M}_\theta[T_1 - H(T)]^2 + D(H(T)) \geqslant \mathsf{D}(H(T)),\end{aligned}$$

поскольку в силу следствия 5.3

$$\begin{aligned}&\mathsf{M}_\theta[(T_1 - H(T))(H(T) - \tau(\theta))] = \mathsf{M}_\theta[H(T)\cdot(T_1 - H(T))] = \\ &\quad = \mathsf{M}_\theta\,\mathsf{M}_\theta[H(T)\cdot(T_1 - H(T))|T] = \mathsf{M}_\theta[H(T)\cdot\mathsf{M}_\theta(T_1 - H(T)|T)] = \\ &\qquad = \mathsf{M}_\theta[H(T)\cdot(\mathsf{M}_\theta(T_1|T) - H(T))] = 0\end{aligned}$$

Таким образом, для любой несмещенной оценки функции $\tau(\theta)$, не являющейся функцией от достаточной статистики, всегда можно указать несмещенную оценку, которая зависит от достаточной статистики и имеет дисперсию меньшую, чем исходная оценка. Следовательно, оптимальную оценку надо искать среди функций от достаточной статистики.

□

Определение 6.8. *Достаточная статистика $T = T(X)$ называется полной, если для любой функции φ из того, что*

$$\mathsf{M}_\theta\,\varphi(T) = 0$$

следует, что $\varphi \equiv 0$ на всем множестве значений статистики T.

Теорема 6.9. *Если существует полная достаточная статистика, то произвольная функция от нее будет являться оптимальной оценкой своего математического ожидания.*

Доказательство. Пусть $T = T(X)$ — полная достаточная статистика и $H(T)$ — произвольная функция от T. Обозначим

$$\mathsf{M}_\theta(H(T)) = \tau(\theta),$$

и покажем, что $H(T)$ — единственная функция от T с математическим ожиданием $\tau(\theta)$.

Предположим противное — пусть $H_1(T)$ — другая такая функция, тогда

$$\mathsf{M}(H_1(T) - H(T)) = 0, \forall\theta,$$

т.к. T - полная статистика. Значит $H_1(T) = H(T)$.

По теореме 6.7 оптимальную оценку для $\tau(\theta)$ необходимо искать в классе функций, зависящих от T. Но $H(T)$ — единственная функция от T, несмещенно оценивающая $\tau(\theta)$ и она является искомой оптимальной оценкой. $\square$

Пример 6.10. Пусть $X = (X_1, \ldots, X_n)$ — выборка из равномерного распределения $R(0,\theta)$. Из примера 6.6 известно, что $T(X) = X_{(n)}$ является достаточной статистикой. Покажем, что $T(X)$ является полной.

Пусть φ такова, что

$$\mathsf{M}_\theta\, \varphi(T(X)) \equiv 0,\ \forall\theta \in (0, \infty).$$

Выпишем плотность статистики $T(x)$:

$$f_{X_{(n)}}(y) = \frac{\partial}{\partial y} F_{X_{(n)}}(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & y \in [0,\theta] \\ 0, & \text{иначе.} \end{cases}$$

Тогда:

$$\mathsf{M}_\theta\, \varphi\left(X_{(n)}\right) = \int\limits_0^\theta \varphi(y) \frac{ny^{n-1}}{\theta^n} dy \equiv 0 \Rightarrow \int\limits_0^\theta \varphi(y) y^{n-1} dy \equiv 0.$$

Продифференцируем последнее равенство по θ:

$$\varphi(\theta)\theta^{n-1} \equiv 0,$$

отсюда следует, что

$$\varphi(\theta) \equiv 0 \ \forall\theta \in (0, \infty).$$

Доказали полноту.

Согласно теореме 6.9 она является оптимальной статистикой для своего математического ожидания:

$$\mathsf{M}_\theta\, X_{(n)} = \int\limits_{\mathbb{R}} y f_{X_{(n)}}(y)dy = \frac{n}{n+1}\theta.$$

Отсюда следует, что $X_{(n)}$ является смещенной оценкой θ. Найдем оптимальную оценку для θ. По теореме 6.9 необходимо найти такую функцию H, что $\mathsf{M}_\theta\, H(T) = \theta$. Очевидно, что такая функция

$$H(x) = \frac{n+1}{n}x.$$

Таким образом, оценка

$$\widehat{\theta} = \frac{n+1}{n}X_{(n)}$$

является оптимальной оценкой параметра θ.

Найдем дисперсию $X_{(n)}$:

$$\mathsf{M}_\theta\, X_{(n)}^2 = \int\limits_{\mathbb{R}} y^2 f_{X_{(n)}}(y)dy = \frac{n\theta^2}{n+2}.$$

Откуда

$$\mathsf{D}\, X_{(n)} = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

А

$$\mathsf{D}\, \widehat{\theta} = \frac{\theta^2}{(n+2)n}.$$

Если бы выполнялись условия регулярности (см. пример 4.22) и $\widehat{\theta}$ являлась бы регулярной оценкой, то ее дисперсия была бы ограничена снизу величиной $(ni(\theta))^{-1}$, которая имеет порядок малости n^{-1}. Полученная выше оценка имеет порядок малости n^{-2}. Такие оценки иногда называют *сверхэффективными.*

Задача 6.11. Показать, что в случае $X = (X_1, \ldots, X_n)$ — выборка из равномерного распределения $R(0,\theta)$, оценка $2\overline{X}$ является несмещенной и найти условия, когда она является оптимальной оценкой.

Резюмируем все вышесказанное. Пусть в рассматриваемом семействе $\mathcal{F} = \{F_\theta,\, \theta \in \Theta\}$ существует полная и достаточная статистика T и требуется оценить параметрическую функцию $\tau(\theta)$. Тогда:

- если существует какая-то несмещенная оценка $\tau(\theta)$, то существует и несмещенная оценка, являющаяся функцией от T (в противном случае класс несмещенных оценок пуст);
- оптимальная оценка, когда она существует, всегда является функцией от T и определяется уравнением $\mathsf{M}_\theta(H(T)) = \tau(\theta)$;
- оптимальную оценку можно искать по формуле

  $$H(T) = \mathsf{M}_\theta(T_1|t),$$

  исходя из любой несмещенной оценки T_1 функции $\tau(\theta)$. (Однако это используется редко в виду значительных аналитических сложностей вычисления условного математического ожидания).

Пример 6.12. Ранее, в примере 6.2 было показано, что статистика $T(X) = \sum\limits_{i=1}^{n} X_i$ является достаточной статистикой в случае, когда распределение из $\mathscr{L}(\xi)$, $\xi \sim \mathrm{Pois}(\theta)$. Покажем, что она является полной. Напомним, что $T(X) \sim \mathrm{Pois}(n\theta)$. Пусть

$$\mathsf{M}_\theta\, \varphi(T(X)) = \sum_{k=0}^{\infty} \varphi(k) \exp\{-n\theta\} \frac{(n\theta)^k}{k!} = 0\ \forall \theta \in \Theta.$$

Рассмотрим последнее выражение при $\theta = 0$. Так как $\mathsf{M}_\theta\, \varphi(T(X)) = \varphi(0)$, то $\varphi(0) = 0$.

Умножим последнее выражение на θ^{-1} и устремим θ к 0 и получим, что $\varphi(1) = 0$. Аналогично получим, что $\varphi(x) = 0$, $x \geqslant 2$. Таким образом показали, что $\varphi \equiv 0$.

Приведем без доказательства следующую теорему.

Теорема 6.13 (О полноте экспоненциальных семейств). *Пусть $\mathcal{F} = \{F_\theta,\ \theta \in \Theta\}$ — экспоненциальное семейство, определенное в 4.21, и функция $A(\theta)$ и параметрическое пространство Θ таковы, что $A(\theta)$ содержит некоторый отрезок, когда θ пробегает множество Θ. Тогда $T(X) = B(X)$ является полной и достаточной статистикой.*

Пример 6.14. Пусть $X = (X_1, \ldots, X_n)$ — выборка из гамма-распределения $\Gamma\left(\frac{1}{\theta},1\right)$. Покажем, что

$$T(X) = \frac{n}{n+1}\overline{X}^2$$

является оптимальной оценкой θ^2.

Выпишем функцию правдоподобия:

$$L(X;\theta) = \frac{1}{\theta^n}\exp\{-\frac{n}{\theta}\overline{X}\}.$$

По теореме о полноте экспоненциальных семейств $\overline{X}$ — полная и достаточная статистика, поэтому любая измеримая функция от $\overline{X}$ является оптимальной оценкой своего математического ожидания. В частности, $T(X)$ оптимальная оценка для

$$\mathsf{M}_\theta(T(X)) = \frac{n}{n+1}\,\mathsf{M}_\theta\,\overline{X}^2 = \frac{n}{n+1}\left(\mathsf{D}_\theta\,\overline{X} + \left(\mathsf{M}_\theta\,\overline{X}\right)^2\right) =$$
$$= \frac{n}{n+1}\left(\frac{1}{n}\,\mathsf{D}_\theta\,X_1 + (\mathsf{M}_\theta\,X_1)^2\right) = \frac{n}{n+1}\left(\frac{\theta^2}{n} + \theta^2\right) = \theta^2.$$

6.2 Модели с выборочным пространством, зависящем от параметра θ

Пусть $a \in \mathbb{R}$, $X = (X_1, \ldots, X_n)$ из параметрического семейства $\mathcal{F} = \{F_\theta, \theta \in \mathbb{R}, \theta > a\}$, при этом плотность случайных величин из рассматриваемых семейств равна:

$$f_\theta(x) = \begin{cases} Q_1(\theta)M_1, & x \in [a,\theta) \\ 0, & x \notin [a,\theta) \end{cases}$$

Где функция $Q_1 \geq 0$ для всех $\theta > a$ и дифференцируема.

Найдем ограничение на вид функции M_1:

$$1 = \int\limits_a^\theta f_\theta(x)dx = \int\limits_a^\theta Q_1(\theta)M_1(x)dx = Q_1(\theta)\int\limits_a^\theta M_1(x)dx \Rightarrow$$

$$\Rightarrow \frac{1}{Q_1(\theta)} = \int\limits_a^\theta M_1(x)dx. \quad (3)$$

Возьмем производную от обоих частей (3):

$$-\frac{Q_1'(\theta)}{Q_1^2(\theta)} = M_1(\theta). \quad (4)$$

Найдем достаточную статистику. Пусть $\overline{x}$ — реализация выборки нашего распределения, тогда функция правдоподобия равна:

$$L(\overline{x};\theta) = \prod_{i=1}^n f_\theta(x) = Q_1^n(\theta)\prod_{i=1}^n (M_1(x_i)\operatorname{Ind}(a \le x_i \le \theta)) =$$

$$= Q_1^n(\theta)\prod_{i=1}^n (M_1(x_i)\operatorname{Ind}(x_{(1)\ge a})\operatorname{Ind}(x_{(n)})).$$

Воспользуемся критерием факторизации для того, чтобы найти достаточную статистику:

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

То есть $T(x) = x_{(n)}$ — достаточная статистика.

Воспользуемся (3) и найдем ее функцию распределения:

$$F_{x_{(n)}}(y) = P(x_{(n)} \le y) =$$

$$= \prod_{i=1}^n (P(x_i \le y)) = P^n(x_1 \le y) = \left(\int_a^y Q_1(\theta)M_1(x)\right)^n =$$

$$= Q_1(\theta)^n \left(\frac{1}{Q_1(y)}\right)^n = \left(\frac{Q_1(\theta)}{Q_1(y)}\right)^n.$$

Откуда с использованием (4) получим выражение для плотности:

$$f_{x_{(n)}}(y) = \frac{-nQ_1(\theta)^n Q_1'(\theta)}{Q_1(y)^{n-1}} = Q_1^n(\theta)nQ_1(y)^{(-n-1)}M_1(y)Q_1^2(y).$$

Зная плотность, можем проверить полноту. Для этого необходимо доказать, что из

$$\int\limits_a^\theta \varphi(y) f_{x_{(n)}}(y) dy = 0$$

следует, что $\varphi(y) \equiv 0$. Действительно:

$$\int\limits_a^\theta \varphi(y) Q_1^n(\theta) n Q_1(y)^{(-n-1)} M_1(y) Q_1^2(y) dy = 0,$$

откуда:

$$\int\limits_a^\theta \varphi(y) Q_1(y)^{(-n+1)} M_1(y) dy = 0$$

Возьмем производную от левой и правой части:

$$\varphi(\theta) Q_1(\theta)^{(-n+1)} M_1(\theta)$$

Значит $\varphi(\theta) \equiv 0$ и статистика $T(x) = X_{(n)}$ полная. Тогда для любой функции H статистика $H(x_{(n)})$ будет оптимальной статистикой своего математического ожидания:

$$\int_a^\theta H(t) g(t,\theta) dt = \tau(\theta).$$

Предполагая дифференцируемость функции τ найдем выражение для $H(x)$:

$$\int_a^\theta H(y) n Q_1(\theta)^n Q_1(y)^{-n+1} M_1(y) dy = \tau(\theta) \Rightarrow$$

$$\Rightarrow n \int_a^\theta H(y Q_1(y)^{-n+1} M_1(y) dy = \frac{\tau(\theta)}{Q_1(\theta)^n} \Rightarrow$$

$$\Rightarrow n H(\theta) Q_1(\theta)^{-n+1} M_1(\theta) = \frac{\tau'(\theta) Q_1(\theta)^n - n Q_1(\theta)^{n-1} \tau(\theta)}{Q_1(\theta)^{2n}}.$$

Воспользуемся (4):

$$H(\theta) = \frac{\tau(\theta)' Q(\theta) - n\tau(\theta) Q_1'(\theta)}{n Q_1(\theta)^{n+1} Q_1(\theta)^{-n+1} M_1(\theta)}$$

$$H(\theta) = \frac{\tau(\theta)' Q(\theta) - n\tau(\theta) Q_1'(\theta)}{n Q_1(\theta)^2} M_1(\theta) = \tau(\theta) + \frac{\tau'(\theta)}{n Q_1(\theta) M_1(\theta)}.$$

Или:
$$H(\theta) = \tau(\theta) + \frac{\tau'(\theta)}{nf_\theta(\theta)}$$

Пусть дана выборка из распределения $\mathcal{R}[0;\theta)$. Найдем оптимальную оценку для параметрической функции $\tau(\theta) = \theta^r$. В этом случае $Q_1(\theta) = \theta^{-1}$, $M_1(x) = 1$ и оптимальной оценкой для θ^r является:

$$X^r_{(n)} + \frac{r}{n} X^r_{(n)}.$$

Заметим, что для случая $r = 1$ эта оценка ранее была получена нами.

Используя аналогичные рассуждения можно показать, что в случае, когда

$$f_\theta(x) = Q_2(\theta) \cdot M_2(x),\ \theta \leqslant x \leqslant b,$$

для некоторого фиксированного b. В этом случае оптимальной оценкой произвольной дифференцируемой функции $\tau(\theta)$ является статистика:

$$\tau\left(X_{(1)} - \frac{\tau'\left(X_{(1)}\right)}{n \cdot f_{X_{(1)}}\left(X_{(1)}\right)}\right).$$

7 Асимптотические свойства оценки максимального правдоподобия.

Пусть здесь и далее $X = (X_1, X_2, \ldots, X_n)$ — некоторая выборка из распределения $\mathscr{L}(X_i) \sim \xi$, $\overline{x} = (x_1, x_2, \ldots, x_n)$ — реализация выборки.

Определение 7.1. *Пусть имеется последовательность случайных величин $\{\xi_n\}_{n\geqslant 1}$, и при $n \to \infty$ $\xi_n \to \xi$ по распределению, $\xi \sim \mathcal{N}(\mu, \sigma^2)$. Тогда будем говорить, что случайная величина ξ_n асимптотически нормальна с параметрами μ_n, σ_n^2.*

Определение 7.2. *Пусть $T_n(X)$ является оценкой параметра $\tau(\theta)$, пусть также T_n — асимптотически нормальная с параметрами $\tau(\theta)$, $\frac{\sigma^2(\theta)}{n}$ (иными словами $L(T_n;\theta) \to \mathcal{N}\left(\tau(\theta), \frac{\sigma^2(\theta)}{n}\right)$).*

Асимптотической эффективностью такой оценки будем называть величину
$$\varepsilon_0(T_n, \theta) = \frac{[\tau'(\theta)]^2}{i(\theta)\sigma^2(\theta)}.$$

Оценку будем называть асимптотически эффективной, если $\varepsilon_0(T_n;\theta) = 1$.

Теорема 7.3. *Пусть выполнены следующие условия:*

1. $\theta \in \Theta, \Theta$ *— невырожденный замкнутый интервал на* $\mathbb{R}$.
2. *Существуют производные:*

$$\frac{\partial \ln f_\theta(x)}{\partial\theta}, \frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2}, \frac{\partial^3 \ln f_\theta(x)}{\partial\theta^3}\ \forall\theta\in\Theta.$$

3. *Для всех* $\theta \in \Theta$ *существуют интегрируемые на* $\mathbb{R}$ *функции* H_1, H_2 *и функция* H, *такая, что:*

$$\mathsf{M}_\theta H = \int\limits_{\mathbb{R}} H(x) f_\theta(x) dx < M,$$

где не зависит от θ *и*

$$\left|\frac{\partial \ln f_\theta(x),\theta)}{\partial\theta}\right| \leqslant H_1(x),$$

$$\left|\frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2}\right| \leqslant H_2(x),$$

$$\left|\frac{\partial^3 \ln f_\theta(x)}{\partial\theta^3}\right| \leqslant H(x).$$

То есть, все три производные ограничены по x.

4. $0 < i(\theta) = \int\limits_{\mathbb{R}} \left(\frac{\partial \ln L(\overline{x},\theta)}{\partial\theta}\right)^2 L(\overline{x},\theta)dx < \infty.$

Тогда оценка максимального правдоподобия обладает следующими свойствами:

1. *Такая оценка состоятельна;*
2. *Такая оценка асимптотически нормальна;*
3. *Такая оценка асимптотически эффективна.*

Условие теоремы содержит в себе условие регулярности. Таким образом, везде далее считаем, что рассматриваемая модель регулярна.

Лема 7.4. *Пусть*

$$B_0 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right)\Bigg|_{\theta=\theta_0},$$

$$B_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2 \ln f_\theta(x_i)}{\partial\theta^2}\right)\Bigg|_{\theta=\theta_0},$$

$$B_2 = \frac{1}{n}\sum_{i=1}^{n} H(x_i).$$

Тогда

1. $B_0 \xrightarrow[n\to\infty]{\mathsf{P}} 0$,
2. $B_1 \xrightarrow[n\to\infty]{\mathsf{P}} -k^2$, *где* $k^2 = i(\theta)$,
3. $B_2 \xrightarrow[n\to\infty]{\mathsf{P}} \mathsf{M}(H(x_1)) < M$.

Доказательство. Все выражения B_j является суммой независимых одинаково распределенных случайных величин с конечным математическим ожиданием. Воспользуемся законом больших чисел:

$$\frac{1}{n}\sum \xi_i \xrightarrow[n\to\infty]{\mathsf{P}} \mathsf{M}\,\xi$$

Поэтому, в частности, очевидно, что $B_2 \xrightarrow[n\to\infty]{\mathsf{P}} \mathsf{M}(H(x_1)) < M$.

Докажем лемму для B_0. В условиях регулярности можем менять пределы интегрирования и дифференцирования:

$$\mathsf{M}\,\frac{\partial \ln f_\theta(\xi)}{\partial\theta} = \int\limits_{\mathbb{R}} \frac{\partial \ln f_\theta(x)}{\partial\theta} f_\theta(x)dx =$$

$$= \int\limits_{\mathbb{R}} \frac{1}{f_\theta(x)}\frac{\partial f_\theta(x)}{\partial\theta} f_\theta(x)dx = \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}} f_\theta(x)dx = \frac{\partial}{\partial\theta}1 = 0.$$

Докажем лемму для B_1:

$$\mathsf{M}\,\frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2} = \int\limits_{\mathbb{R}} \frac{\partial}{\partial\theta}\left(\frac{\partial \ln f_\theta(x)}{\partial\theta}\right) f_\theta(x)dx = \int\limits_{\mathbb{R}} \frac{\partial}{\partial\theta}\left(\frac{1}{f_\theta(x)}\frac{\partial f_\theta(x)}{\partial\theta}\right) f_\theta(x)dx =$$

$$= \int\limits_{\mathbb{R}} \frac{(-1)}{f_\theta^2(x)}\left(\frac{\partial f_\theta(x)}{\partial\theta}\right)^2 f_\theta(x)dx + \int\limits_{\mathbb{R}} \frac{\partial^2 f_\theta(x)}{\partial\theta^2}\frac{1}{f_\theta(x)} f_\theta(x)dx =$$

$$= -\int\limits_{\mathbb{R}} \frac{1}{f_\theta(x)}\left(\frac{\partial f_\theta(x)}{\partial\theta}\right)^2 dx + \underbrace{\frac{\partial^2}{\partial\theta^2}\int\limits_{\mathbb{R}} f_\theta(x)dx}_{=0} =$$

$$= -\int\limits_{\mathbb{R}} \left(\frac{\partial \ln f_\theta(x)}{\partial} f_\theta(x)\right)^2 \frac{1}{f_\theta(x)} =$$

$$= -\int\limits_{\mathbb{R}} \left(\frac{\partial \ln f_\theta(x)}{\partial\theta}\right)^2 f_\theta(x)dx = -i(\theta) = -k^2.$$

$\square$

Лема 7.5 (без доказательства). *Пусть* X_n,Y_n, Z_n *— последовательности случайных величин такие, что:*

1. $X_n \xrightarrow[n\to\infty]{d} X$, $Y_n \xrightarrow[n\to\infty]{\mathsf{P}} 0$, *тогда* $X_n + Y_n \xrightarrow[n\to\infty]{d} X$,
2. $X_n \xrightarrow[n\to\infty]{d} X, Z_n \xrightarrow[n\to\infty]{\mathsf{P}} 1$, *тогда* $X_n \cdot Z_n \xrightarrow[n\to\infty]{d} X$; $X_n/Z_n \xrightarrow[n\to\infty]{d} X$.

Доказательство. Пользуясь данными леммами, докажем теорему. Разложим в ряд Тейлора с остаточным членом в форме Лагранжа плотность распределения:

$$\frac{\partial \ln f_\theta(x)}{\partial\theta} = \left.\frac{\partial \ln f_\theta(x)}{\partial\theta}\right|_{\theta=\theta_0} + \left.\frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2}\right|_{\theta=\theta_0}(\theta-\theta_0) + \frac{1}{2}\left.\frac{\partial^3 \ln f_\theta(x)}{\partial\theta^3}\right|_{\theta=\theta^\star}(\theta-\theta_0)^2,$$

где $\theta^\star$ лежит между θ_0 и θ).

Можно показать, что существует $0 \le \tau \le 1$ такое, что

$$\frac{1}{2}\left.\frac{\partial^3 \ln f_\theta(x)}{\partial\theta^3}\right|_{\theta=\theta^\star}(\theta-\theta_0)^2 = \frac{1}{2}\tau H(\xi)(\theta-\theta_0)^2.$$

Пусть
$$A(\theta)=\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right).$$
Тогда $A(\theta)=B_0+(\theta-\theta_0)B_1+\frac{1}{2}\tau(\theta-\theta_0)^2B_2$.

Для нахождения оценки максимального правдоподобия необходимо решить уравнение
$$\frac{1}{n}\frac{\partial \ln L(\overline{x};\theta)}{\partial\theta}=0$$
относительно θ.

Фиксируем $\varepsilon>0$, $\delta>0$ — некоторые достаточно малые числа и выберем $n>n_0(\epsilon;\delta)$ так, что выполнены следующие неравенства (по условию леммы 7.4 это возможно):

- $\mathsf{P}(|B_0|\geqslant\delta^2)<\frac{\varepsilon}{3}$,
- $\mathsf{P}(|B_1|\geqslant-\frac{k^2}{2})<\frac{\varepsilon}{3}$,
- $\mathsf{P}(|B_2|\geqslant 2M)<\frac{\varepsilon}{3}$.

Рассмотрим событие:
$$S=\left\{x\colon\ |B_0|<\delta^2, B_1<-\frac{k^2}{2}, |B_2|<2M\right\}.$$
Тогда, при $n>n_0:\mathsf{P}(S)>1-\varepsilon$.

Пусть $x\in S$. Рассмотрим $\theta\in[\theta_0-\delta;\theta_0+\delta]$: Оценим $A(\theta)$:
$$\left|B_0+\frac{1}{2}\tau\delta^2B_2\right|\leqslant\delta^2\left(\frac{1}{2}M+1\right).$$

Тогда так как $\theta\in[\theta_0-\delta;\theta_0+\delta]$:
$$A(\theta)=B_0+\delta B_1+\frac{1}{2}\tau\delta^2B_2.$$
В этом случае знак $A(\theta)$ определяется слагаемым $B_1\delta$.

Тогда $A(\theta)\in(\theta_0-\delta;\theta_0+\delta)$ и точка максимума θ^* с вероятностью не меньшей $1-\varepsilon$ будет лежать в интервале $(\theta_0-\delta;\theta_0+\delta)$, а это как раз и показывает состоятельность оценки.

Покажем асимптотическую нормальность. $A\left(\theta^*\right)=0$, тогда:
$$\theta^*-\theta_0=\frac{B_0}{-B_1-\frac{1}{2}\tau(\theta^*-\theta_0)B_2},$$

откуда

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

Введем случайную величину: $Y_i = \frac{\partial \ln f(X_i)}{\partial \theta}$, при этом

$$\mathsf{M}\, Y_i = 0, \mathsf{D}\, Y_i - MY_i = k^2.$$

Используя центральную предельную теорему:

$$\frac{1}{\sqrt{nk^2}} \sum \frac{\partial \ln f(X_i)}{\partial \theta} \underset{n\to\infty}{\xrightarrow{d}} \mathcal{N}(0,1)$$

Пользуясь леммой 7.4:

$$B_1 \underset{n\to\infty}{\xrightarrow{\mathsf{P}}} -k^2,$$

$$B_2 \underset{n\to\infty}{\xrightarrow{\mathsf{P}}} \mathsf{M}(H(\xi_1)) < M \Rightarrow \frac{1}{2}\tau \frac{\theta^* - \theta_0}{k^2} B_2 \leqslant \frac{1}{2}\frac{M}{k^2}(\theta^* - \theta_0) \to 0, (\theta^* - \theta_0) < \delta \to 0,$$

получаем:

$$\frac{B_1}{-k^2} + \frac{1}{2}\tau \frac{\theta^* - \theta_0}{k^2} B_2 \underset{n\to\infty}{\xrightarrow{\mathsf{P}}} 1, \text{ при } \delta \to 0.$$

Используя лемму 7.5, получаем доказательство асимптотической нормальности.

Отсюда следует, что $\sqrt{nk^2}(\theta^* - \theta_0)$ асимптотически нормальна с $\mathcal{N}(0,1)$, а θ^* — с параметрами $\mathcal{N}\left(\theta_0, \frac{1}{nk^2}\right)$.

Дисперсия $nk^2 = i_n(\theta)$, что доказывает эффективность оценки.

□

8 Асимптотически доверительный интервал.

Пусть заданы $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$. Ранее мы занимались построением точечных оценок неизвестного параметра $\tau(\theta)$. Такие оценки могут быть не очень интересны с практической точки зрения (например в непрерывных распределениях, когда вероятность конкретной точки всегда равна 0). В данном разделе рассмотрим иной подход, основная идея которого заключается в построении доверительного множества, в котором с повышенной вероятностью лежит оцениваемый параметр.

Определение 8.1. *Будем называть γ-доверительным интервалом параметрической функции $\tau(\theta)$ такой случайный интервал $(T_1(x), T_2(x))$, $T_1(x) < T_2(x)$, который удовлетворяет условию:*

$$\mathsf{P}_\theta\left(T_1(x) < \tau(\theta) < T_2(x)\right) \geq \gamma \ \forall \theta \in \Theta.$$

В данном курсе рассмотрим только так называемые асимптотические доверительные интервалы.

Пусть имеется асимптотически нормальная оценка $T_n = T_n(X)$ для параметра $\tau(\theta)$. То есть иными словами: пусть при $n \to \infty$ имеет место соотношение:

$$\mathcal{L}_\theta\left(\sqrt{n}(T_n - \tau(\theta))\right) \to N\left(0, \sigma_n^2(\theta)\right) \ \forall \theta \in \Theta,$$

причем $\sigma_n^2(\theta)$ непрерывна по θ. Тогда

$$\mathcal{L}_\theta\left(\frac{\sqrt{n}(T_n - \tau(\theta))}{\sigma_n(\theta)}\right) \to N\left(0, 1\right).$$

Для стандартного нормального распределения легко показать, что для заданной вероятности интервал минимальной дляны будет иметь вид $(-t,t)$. Построим такой интервал:

$$\mathsf{P}\left(\sqrt{n} \cdot \left|\frac{T_n - \tau(\theta)}{\sigma_n(\theta)}\right| < c_\gamma\right) = \mathsf{P}\left(\tau(\theta) \in \left(T - \frac{c_\gamma \sigma_n(\theta)}{\sqrt{n}}, T + \frac{c_\gamma \sigma_n(\theta)}{\sqrt{n}}\right)\right) =$$
$$= \Phi(c_\gamma) - \Phi(-c_\gamma) = 2\Phi(c_\gamma) - 1 = \gamma,$$

где Φ - функция распределения случайной величины, имеющей стандартное нормальное распределение.

!!!!!!!!!!!Добавить классный пример про доверительный интервал для параметра бернулиевской модели (стр 285)

9 Проверка статистических гипотез

9.1 Основные понятия

Одним из основных направлений математической статистики является теория проверки статистических гипотез. Неформально статистическую гипотезу можно понимать как некоторое предположение о виде или параметрах распределения. Необходимо уметь подтверждать или опровергать гипотезы о виде распределения, о зависимости данных, об однородности выборок и многие другие.

Пусть как и ранее X — выборка, $\mathfrak{X}$ — выборочное пространство, $\mathcal{F}$ — совокупность априори доступных распределений выборки X.

Задачу проверки статистических гипотез можно сформулировать следующим образом. Дана выборка X из неизвестного распределения $F_X \in \mathcal{F}$. Выделим подмножество $\mathcal{F}_0 \subset \mathcal{F}$, где $\mathcal{F}_0$. По реализации выборки X необходимо проверить справедливо ли утверждение:

$$F_X \in \mathcal{F}_0$$

или ложно, то есть верно, что

$$F_X \in \mathcal{F}_1, \, \mathcal{F}_1 = \mathcal{F} \backslash \mathcal{F}_0.$$

При этом говорят о проверке гипотез.

- гипотеза H_0 — основная гипотеза (нулевая гипотеза), заключается в том, что $F_X \in \mathcal{F}_0$;
- гипотеза H_1 — альтернативная гипотеза, заключается в том, что $F_X \in \mathcal{F}_1$.

Если определены гипотезы H_0 и H_1, то говорят о *проверке гипотезы H_0 против альтернативы H_1*.

Определение 9.1. *Если множество $\mathcal{F}_0$ состоит из одного распределения, то говорят, что H_0 — простая гипотеза, иначе H_0 - сложная гипотеза.*

Определение 9.2. *Если множество $\mathcal{F}_1$ состоит из одного распределения, то говорят, что H_1 - простая гипотеза, иначе H_1 - сложная гипотеза.*

Пример 9.3. Пусть $\mathcal{F} \in \left\{\mathcal{N}(\mu,\sigma^2), \mu, \sigma \in \mathbb{R}\right\}$. В качестве примера простой основной гипотезы H_0: $F_X = \mathcal{N}(0,1)$, и сложной альтернативы H_1: $F_X \neq \mathcal{N}(0,1)$.

Пример 9.4. Пример простой основной и альтернативной гипотезы: $\mathcal{F} = \{\mathcal{N}(0,1), \mathcal{N}(1,1)\}$, то $H_0 : F_X = N(0,1)$ и $H_1 : F_X = N(1,1)$.

Определение 9.5. *Статистический критерий - это правило, по которому каждой реализации выборки ставится в соответствие решение: принимаем гипотезу H_0 или отвергаем ее (то есть принимаем гипотезу H_1).*

Определение 9.6. *Иногда альтернативу H_1 не конкретизируют, то есть задана только гипотеза H_0 (множество распределений $\mathcal{F}_0$). В этом случае говорят о согласии данных X с нулевой гипотезой. Такой критерий называется критерием согласия.*

Пример 9.7. Примеры критериев согласия:

- выборка X из распределения Пуассона: $H_0\colon F_X \in \{\mathrm{Pois}(\lambda), \lambda > 0\}$,
- выборка X из распределения Пуассона с параметром λ_0: $H_0\colon F_X = \mathrm{Pois}(\lambda_0)$.

Зачастую множество распределений $\mathcal{F}$ есть некоторое параметрическое семейство:

$$\mathcal{F} = \{F_\theta, \theta \in \Theta\}.$$

Тогда основную и альтернативную гипотезы можно определить следующим образом:

- гипотеза $H_0\colon \theta \in \Theta_0 \subset \Theta$;
- гипотеза $H_1\colon \theta \in \Theta_1$, $\Theta_1 = \Theta \backslash \Theta_0$.

В данном случае говорят *о параметрических гипотезах* и *параметрических критериях.*

Говоря о критерии, подразумеваем некоторое правило, согласно которому по каждому $X \in \mathfrak{X}$ говорим верна ли гипотеза H_0. Таким образом, статистический критерий может задаваться двумя множествами:

- $\mathfrak{X}_0 \subset \mathfrak{X}$ — принимается H_0, $\mathfrak{X}_0$ называется областю принятия гипотезы;
- $\mathfrak{X}_1 = \mathfrak{X} \backslash \mathfrak{X}_0$ — принимается H_1, $\mathfrak{X}_1$ называется критической областю.

Тогда $\mathfrak{X}_0 \sqcup \mathfrak{X}_1 = \mathfrak{X}$ и задача построения статистического критерия равносильна построению критического множества $\mathfrak{X}_1$. В дальнейшем будем отождествлять понятие статистического критерия и критического множества.

Определение 9.8. *Если $x \in \mathfrak{X}_0$, то говорят, что H_0 соглашается с данными или данные не противоречат гипотезе H_0*

Определение 9.9. *Если $x \in \mathfrak{X}_1$, то говорят, что данные противоречат гипотезе H_0.*

Рассмотрим возможные варианты:

	$\mathfrak{X}_0$	$\mathfrak{X}_1$
верна H_0	определили правильно	отвергаем истину
верна H_1	принимаем ложь за истину	определили правильно

Очевидно ходим, чтобы вероятность $\mathsf{P}(x \in \mathfrak{X}_1 | H_0)$ была как можно меньше. При построении критерия фиксируют некоторе малое значение α, которое называется уровнем значимости критерия, так, чтобы

$$\mathsf{P}(x \in \mathfrak{X}_1 | H_0) \leqslant \alpha.$$

Если это неравенство выполняется, то говорят, что критерий $\mathfrak{X}_1$ имеет уровень значимости α и обозначается $\mathfrak{X}_{1,\alpha}$. При этом, очевидно, существуют различные критические области с одинаковым уровнем значимости.

В случае, когда H_0 и H_1 — простые гипотезы, вводятся понятия ошибок критерия. Критерий описывается вероятностями:

1. $\mathsf{P}(X \in \mathfrak{X}_1 | H_0) = \alpha$ — *ошибка 1 рода*;
2. $\mathsf{P}(X \in \mathfrak{X}_0 | H_1) = \beta$ — *ошибка 2 рода.*

Определение 9.10. *Функцией мощности критерия W назовем функционал на множестве допустимых распределений $\mathcal{F}$ и выборке X.*

$$W\left(F_X\right) = W\left(F_X; \mathfrak{X}_{1,\alpha}\right) = \mathsf{P}\left(X \in \mathfrak{X}_{1,\alpha} | F_X\right),$$

где $\mathsf{P}\left(x \in \mathfrak{X}_{1,\alpha} | F_X\right)$ - вероятность попасть в $\mathfrak{X}_{1,\alpha}$, если F_X - истинное распределение.

Через функцию мощности критерия легко можно выразить вероятности ошибок первого и второго рода:

$$\alpha = W(F_X),\, \beta = 1 - W(F_X).$$

Аналогично уровень значимости определяется:

$$\alpha = \sup_{F \in \mathcal{F}_0} W(F).$$

Среди всех критериев с уровнем значимости α нас интересует наиболее мощный критерий, то есть тот, который минимизирует величину:

$$\beta = \sup_{F \in \mathcal{F}_1} \left(1 - W(F)\right).$$

Пример 9.11. Пусть $\mathcal{F}$ состоит из двух распределений:

1. случайная величина принимает значение 0 с вероятностью p_0 и значение 1 с вероятностью p_1;
2. случайная величина принимает значение 1 с вероятностью q_1 и значение 2 с вероятностью q_2.

При этом $p_1, q_1 \neq 0$, $p_1, q_1 \neq 1$, $p_0 + p_1 = 1 = q_1 + q_2$. В качестве основной гипотезы H_0 рассмотрим распределение Бернулли с вероятностью успеха p_1.

Критическая область задается следующим образом:

$$\mathfrak{X}_1 = \left\{X_i = 1, \forall i = \overline{1,n}\right\}.$$

Иными словами: если есть хотя бы один 0, то выбираем гипотезу H_0, иначе H_1.

Тогда

$$\alpha = P\{H_1|H_0\} = p_1^n,$$

$$\beta = P\{H_0|H_1\} = 0.$$

Если же задать критическую область иначе:

$$\mathfrak{X}_1 = \{\exists\ i \colon X_i = 2\},$$

или принимаем $\mathfrak{X}_1$, если есть хотя бы одна 2. В этом случае:

$$\alpha = P\{H_1|H_0\} = 0,$$

$$\beta = P\{H_0|H_1\} = q_1^n.$$

9.2 Проверка гипотезы о виде распределения

Рассмотрим задачу проверки гипотезы о виде распределения. Пусть дана выборка $X = (X_1 \dots X_n)$ из распределения $\mathscr{L}(\xi)$ и F_ξ - неизвестное распределение. Рассмотрим основную гипотезу $H_0 : F_\xi = F(x)$, при этом никак не конкретизируем альтернативную гипотезу (H_1 — сложная гипотеза).

9.2.1 Критерий согласия Колмогорова

Критерий Колмогорова основан на теореме Колмогорова (см. 1.10). Статистика критерия определяется формулой:

$$D_n = D_n(X) = \sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right|,$$

где D_n — это отклонение эмпирической функции распределения от теоретической функции распределения.

Мы знаем, что $\widehat{F}_n$ является оптимальной, несмещенной и состоятельная оценкой для $F(x)$. Отсюда следует, что D_n не должно «сильно» отклоняться от 0.

По теореме Колмогорова для непрерывных функций распределения F и при $n \geqslant 20$:

$$\mathsf{P}\left(\sqrt{n}D_n \geq \lambda_\alpha | H_0\right) = 1 - K(\lambda_\alpha) = \alpha.$$

При этом по значению α возможно однозначно определить величину λ_α.

Критерий формулируется следующим образом. Проверяем, выполняется ли неравенство: $\sqrt{n}D_n \geq \lambda_\alpha$, если да, то отвергаем гипотезу H_0. Данному критерию соответсвует критическая область

$$\mathfrak{X}_1 = \left\{\overline{x} : D_n(\overline{x})\sqrt{n} \geq \lambda_\alpha\right\}.$$

Вместо статистики $\sqrt{n}D_n \geq \lambda_\alpha$ при малых значениях n ($n \geqslant 20$) рекомендуется использовать статистику

$$S_n = \frac{6nD_n + 1}{6\sqrt{n}}$$

(см. напр. [?]), которая также сходится к распределению Колмогорова.

Опишем способ вычисления значения $D_n = \sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right|$. Вычисление супремума функции, вообще говоря, не является тривиальной задачей. Однако в виду того, что $\widehat{F}_n(x)$ принимает конечное число значений: $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\right\}$, задача нахождения супремума функции сильно упрощается.

Пусть имеется вариационный ряд реализации выборки: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. Определим следующие две функции:

$$D_n^+ = \max_{1\leq k\leq n} \left|\frac{k}{n} - F\left(x_{(k)}\right)\right|,$$

$$D_n^- = \max_{1 \le k \le n} \left| F(x_{(k)}) - \frac{k-1}{n} \right|.$$

Тогда вычислить значение D_n можно следующим образом:

$$D_n = \max\{D_n^-, D_n^+\}.$$

Однако критерий Колмогорова обладает рядом минусов:

1. Функция $D_n = \sup\limits_{x \in R} \left| \widehat{F}_n(x) - F(x) \right|$ не зависит от вида функции распределения $F(x)$, только в случае, когда $F(x)$ — непрерывная функция. Встает вопрос, что делать если $F(x)$ имеет точки разрыва.

   Утверждение 9.12. *Пусть $Y_1, \ldots, Y_n$ — независимые, одинаково распределенные случайные величины, $Y_i \sim \mathcal{R}[0,1]$, $X_1, \ldots, X_n$ — выборка из некоторого распределения, функция которого имеет точки разрыва. Построим следующую случайную величину:*

   $$U_i = F(X_i-) + Y_i[F(X_i - F(X_i-)],$$

   где $F(x_i-) = \lim\limits_{z \downarrow 0} F(x_i - z)$. Тогда случайная величина $U_i \sim \mathcal{R}[0,1]$.

2. В случае сложных гипотез распределение $\mathsf{D}_n(\theta)$, зависит как от вида априорных распределений, так и от способа получения оценок, размера выборки n, вида Θ. Существуют два подхода к проверке гипотезы о виде распределения в этом случае:
   - Для проверки гипотезы $H_0 : F_\xi(x) \in \mathcal{F}_0 = \{F_\theta(x), \theta \in \Theta\}$ сначала вычисляется оценка $\widehat{\theta}$ неизвестных параметров. Имея оценку вычислить значение статистики Колмогорова K_n (или S_n). Критическую область определить по предвычисленным таблицам в зависимости от распределения, количества и вида параметров.
   - В случае достаточно большой выборки, можно разбить ее на две части: по одной получить оценки на неизвестные параметры, по второй проверить гипотезу о виде распределения.

Определение 9.13. *Пусть случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ имеют стандартное нормальное распределение, тогда случайная величина:*

$$\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$$

имеет распределение, которое называется распределением хи-квадрат с n степенями свободы

Аналогичным образом определяется критерий Смирнова. Пусть дана выборка $X_1, \ldots, X_n$.

$$D^{+}_{Sm} = \sup_{x \in R} \left| \widehat{F}_n(x) - F(x) \right|,$$

$$D^{-}_{Sm} = -\inf_{x \in R} \left| \widehat{F}_n(x) - F(x) \right|.$$

$$Sm_n = \frac{(6nD^{+}_{S_n} + 1)^2}{9n}.$$

Известно, что эта статистика имеет следующее распределение:

$$\mathcal{L}(S_m) = \chi^2_2.$$

Рассмотрим вопрос об оценке вероятности ошибки второго рода при применении критерия согласия Колмогорова. Обозначим функцию распределения, соответствующую гипотезе H_0 – $F_0(x)$, а функцию распределения, соответствующую альтернативной гипотезе, $F_1(x)$.

$$\beta = \mathsf{P}\left(\sqrt{n} D_n \leqslant \lambda_\alpha \middle| H_1\right) = \mathsf{P}\left(\sqrt{n} \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F_0(x) \right| \leqslant \lambda_\alpha \middle| H_1\right)$$

Выберем некоторых x_0. Для данного значения верно, что событие $\left\{ \sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F_0(x) \right| \leqslant a \right\} \subset \left\{ \left| \widehat{F}_n(x_0) - F(x_0) \right| \leqslant a \right\}$, соответственно верно

$$\beta \leqslant \mathsf{P}\left(\sqrt{n} \, \left| \widehat{F}_n(x_0) - F(x_0) \right| \leqslant \lambda_\alpha \middle| H_1\right) =$$

$$= \mathsf{P}\left(\sqrt{n} \, \left| \widehat{F}_{1n}(x_0) - F_0(x_0) \right| \leqslant \lambda_\alpha\right) =$$

$$= \mathsf{P}\left(\sqrt{n} \frac{\left| \widehat{F}_{1n}(x_0) - F_0(x_0) - F_1(x_0) + F_1(x_0) \right|}{\sqrt{F_1(x_0)\left(1 - F_1(x_0)\right)}} \leqslant \frac{\lambda_\alpha}{\sqrt{F_1(x_0)\left(1 - F_1(x_0)\right)}}\right) =$$

$$= \mathsf{P}\left(\frac{-\lambda_\alpha + \sqrt{n}\left(F_0(x_0) - F_1(x_0)\right)}{\sqrt{F_1(x_0)\left(1 - F_1(x_0)\right)}} \leqslant\right.$$

$$\left. \leqslant \sqrt{n} \frac{\widehat{F}_{1n}(x_0) - F_1(x_0)}{\sqrt{F_1(x_0)\left(1 - F_1(x_0)\right)}} \leqslant \frac{\lambda_\alpha + \sqrt{n}\left(F_0(x_0) - F_1(x_0)\right)}{\sqrt{F_1(x_0)\left(1 - F_1(x_0)\right)}}\right) \approx$$

$$\approx^{1} \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{-\lambda_\alpha+\sqrt{n}(F_0(x_0)-F_1(x_0))}{\sqrt{F_1(x_0)(1-F_1(x_0))}}}^{\frac{\lambda_\alpha+\sqrt{n}(F_0(x_0)-F_1(x_0))}{\sqrt{F_1(x_0)(1-F_1(x_0))}}} e^{-\frac{t^2}{2}}\,dt$$

Несложно видеть, что с ростом n пределы интегрирования будут одновременно стремится к бесконечности (к положительной или отрицательной зависит от знака разности $F_0(x_0)-F_1(x_0)$). То есть с ростом n значение интеграла будет стремиться к нулю, что свидетельствует о состоятельности критерия.

9.2.2 Критерий согласия хи-квадрат

Утверждение 9.14. *Пусть* $\overline{\xi}$ *- случайный вектор* $\overline{\xi} = (\xi_1, \ldots, \xi_n)$, *и* $\xi_i \sim \mathcal{N}(0,1)$. *И, вектор* $\overline{\xi}$ *имеет единичную матрицу ковариаций. Пусть также* $\overline{c} = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$, *такой что* $|\overline{c}| = 1$.

Рассмотрим проекцию $\overline{\xi}^{(c)}$ *на гиперплоскость* $L_{\overline{c}} = \{x \in \mathbb{R}^n : (\overline{x},\overline{c}) = 0\}$, *которая ортогональна вектору* $\overline{c}$. *Тогда*

- *вектор* $\overline{\xi}^{(c)}$ *имеет математическое ожидание равное* $\theta = (0, \ldots, 0)$ *и матрицу ковариации*

$$C(\overline{\xi}^{(c)}) = E- \parallel c_i c_j \parallel_{i,j=1}^{n},$$

- *квадрат длины вектора* $\overline{\xi}^{c}$ *имеет расспределение* χ^2_{n-1} *(хи-квадрат с* $n-1$ *степенью свободы)*

Доказательство. Выпишем представление вектора $\overline{\xi}$ в исходном базисе:

$$\overline{\xi} = \overline{e}_1\xi_1 + \ldots + \overline{e}_{n-1}\xi_{n-1} + \overline{e}_n\xi_n.$$

Так как $|\overline{c}| = 1$ мы можем рассмотреть ортонормированный базис $\overline{e'}_1 \ldots, \overline{e'}_{n-1}, \overline{e'}_n$, где $\overline{e'}_n = \overline{c}$:

$$\overline{\xi} = \overline{e'}_1\xi'_1 + \ldots + \overline{e'}_{n-1}\xi'_{n-1} + \overline{c}\xi'_n.$$

Из-за перехода от одного ортонормированного базиса к другому $\xi'_i \sim \mathcal{N}(0,1)$. Действительно, пусть матрица перехода от одного базиса к другому равна

[1] Интегральная теорема Муавра-Лапласа для последовательности бернуллевских случайных величин.

$A = \|a_{i,j}\|_{i,j=1}^{n}$. Тогда

$$\mathsf{M}\xi_i' = \mathsf{M}\left(a_{1,i}\xi_1 + \ldots + a_{n,i}\xi_n\right) = 0.$$

Так как оба базиса ортонормированны, то $A \cdot A^T = E$. Найдем $\mathsf{D}\,\xi_i'$:

$$\mathsf{D}\,\xi_i' = D\left(a_{1,i}\xi_1 + \ldots + a_{n,i}\right) = \sum_{j=1}^{n} a_{j,i}^2 = 1.$$

Аналогично доказывается $\mathsf{cov}(\xi_i', \xi_j') = 0$, $i \neq j$.

Выпишем проекцию вектора ξ на плоскость $L_{\overline{c}}$:

$$\overline{\xi}^c = \overline{e'}_1\xi_1' + \ldots + \overline{e'}_{n-1}\xi_{n-1}'\overline{e'}_1\xi_1' + \ldots + \overline{e'}_{n-1}\xi_{n-1}'$$

и рассмотрим квадрат длины проекции $\overline{\xi}^c$. Так как базис $\overline{e'}_1 \ldots, \overline{e'}_{n-1}$ ортонормированный, получим следующее:

$$\left|\overline{\xi}^{(c)}\right|^2 = \left(\xi_1'\right)^2 + \ldots + \left(\xi_{n-1}'\right)^2,$$

то есть $\left|\overline{\xi}^{(c)}\right|^2$ имеет распределение хи-квадрат с $n-1$ степенью свободы.

Очевидно, что

$$\mathsf{M}\left(\overline{\xi}^{(c)}\right) = \overline{0}.$$

Найдем ковариационную матрицу:

$$C\left(\overline{\xi}\right) = C\left(\overline{\xi}^{(c)} + \xi_n' e_n'\right) = C\left(\overline{\xi}^{(c)}\right) + C\left(\xi_n'\overline{e'}_n\right).$$

Найдем $C\left(\xi_n'\overline{e'}_n\right)$:

$$C\left(\xi_n'\overline{e'}_n\right) = C\left(\xi_n'\overline{c}\right) = \left\|\mathsf{cov}\left(\xi_n' c_i, \xi_n' c_j\right)\right\|.$$

Учитывая то $\mathsf{cov}(\xi_i' c_i, \xi_j' c_j) = c_i c_j$ получим:

$$C\left(\overline{\xi}^{(c)}\right) = E - \parallel c_i c_j \parallel_{i,j=1}^{n}.$$

□

Пусть $\xi_1, \ldots \xi_n$ — независимые одинаково распределенные случайные величины, которые принимают значения $1, \ldots, N$ с вероятностью $p_1, \ldots, p_n$.

Введем случайную величину

$$\nu_k^{(n)} = \sum_{i=1}^{n} \mathrm{Ind}(\xi_i = k).$$

Величину $\nu_k^{(n)}$ называют частотой встречаемости значения k. Определим случайный вектор частот, имеющий полиномиальное распределение:

$$\overline{\nu^{(n)}} = \left(\nu_1^{(n)}, \ldots, \nu_N^{(n)}\right),$$

$$\mathsf{P}\left(\nu_i^{(n)} = m_i, i = \overline{1,N}\right) = \frac{n!}{m_1! \ldots m_N!} p_1^{m_1} \ldots p_N^{m_N}.$$

Рассмотрим следующую статистику:

$$X_N^2 = \sum_{i=1}^{N} \frac{\left(\nu_i^{(n)} - np_i\right)^2}{np_i} = \sum_{i=1}^{N} \frac{\left(\nu_i^{(n)}\right)^2}{np_i} - n,$$

которую будем называть статистикой Пирсона или статистика хи-квадрат.

Заметим, что $\dfrac{\nu_k^{(n)}}{n} \to p_k$ и значение статистики Пирсона должно стремится к 0 при выборке из рассматриваемого распределения. Тогда критерий можно сформулировать следующим образом:

$$\mathfrak{X}_{1\alpha} = \left\{X_N^2 \geq t_\alpha\right\}.$$

Остается вопрос о том, как правильно выбрать t_α.

Теорема 9.15 (Предельное распределение статистики Пирсона)**.** *Пусть случайный вектор частот* $\overline{\nu^{(n)}} = \left(\nu_1^{(n)}, \ldots, \nu_N^{(n)}\right)$ *имеет полиномиальное распределение с параметрами* n *и* $\overline{p} = (p_1, \ldots, p_N)$*. Если вектор* $\overline{p}$ *фиксирован, а* $n \to \infty$*, то распределение статистики Пирсона сходится к распределению* χ^2 *с* $N-1$ *степенью свободы.*

Доказательство. Вектор частот $\overline{\nu^{(n)}} = \left(\nu_1^{(n)}, \ldots, \nu_N^{(n)}\right)$ можно представить в виде суммы n независимых одинаково распределенных случайных векторов

$$\overline{\nu^{(n)}} = \sum_{t=1}^{N} \left(\mathrm{Ind}\{\xi_t = 1\}, \ldots, \mathrm{Ind}\{\xi_t = n\}\right).$$

Математическое ожидание очевидно равно $(p_1, \ldots, p_n)$, а матрица ковариаций равна

$$C\left(\overline{\nu^{(n)}}\right) = \operatorname{diag}(p_1, \ldots, p_N) - \parallel p_i p_j \parallel_{i,j=1}^{N}.$$

По этому согласно многомерной центральной предельной теореме распределение вектора

$$\left(\frac{\nu_1^{(n)} - np_1}{\sqrt{n}}, \ldots, \frac{\nu_N^{(n)} - np_N}{\sqrt{n}}\right) \tag{5}$$

при $n \to \infty$ сходится к нормальному распределению с нулевым средним и матрицей ковариации $c\left(\overline{\nu^{(n)}}\right)$. Статистику Пирсона можно интерпретировать как квадрат длины вектора

$$\left(\frac{\nu_1^{(n)} - np_1}{\sqrt{np_1}}, \ldots, \frac{\nu_N^{(n)} - np_N}{\sqrt{np_N}}\right), \tag{6}$$

который получается из (5) линейным преобразованием: делением j-ой координаты на $\sqrt{p_j}$. При этом математическое ожидание останется нулевым, а в матрице ковариаций элемент i-ой строки и j-го столбца делится на $\sqrt{p_i p_j}$. Тогда распределение векторов (6) сходится к нормальному распределению с нулевым вектором средних и матрицей ковариаций

$$E - \left\| \sqrt{p_i p_j} \right\|.$$

Заметим, что для $\overline{c} = \left(\sqrt{p_1}, \ldots, \sqrt{p_N}\right)$ верно равенство:

$$\sum_{j=1}^{N} \frac{\nu_j^{(n)} - np_j}{\sqrt{np_j}} \cdot \sqrt{p_j} = 0.$$

Тогда распределение квадрата длины такого случайного вектора имеет распределение хи-квадрат с $N-1$ степенью свободы. Поскольку квадрат длины — непрерывная функция, то распределение ее значения от допредельных векторов (6) сходится к распределению ее значений от случайного вектора. $\square$

На практике критерий хи-квадрат можно использовать для расчетов с хорошим приближением при $n \geqslant 50$ и $\nu_j \geqslant 5$, $j \in \overline{1,N}$.

Таким образом: гипотеза H_0 отвергается тогда и только тогда, когда $X_n^2 > \chi^2_{1-\alpha, N-1}$, где α — заданный уровень значимости.

Рассмотрим случай сложной гипотезы. Пусть гипотеза для полиномиального распределения имеет следующий вид:

$$H_0\colon \overline{p} = \overline{p}(\theta),\, \theta = (\theta_1, \ldots, \theta_r), \theta \in \Theta, r < N - 1.$$

В этом случае при гипотезе H_0 вероятности исходов являются некоторыми функциями от параметра θ. Если выписать статистику Пирсона, то она будет зависеть от параметра θ. При фиксации некоторого θ мы можем вычислить статистику критери хи-квадрат. Верна следующая

Теорема 9.16. *Пусть $p_j(\theta)$, $j = \overline{1,N}$, $\theta = (\theta_1, \ldots, \theta_r)$, при этом:*

1. $\sum_{j=1}^{N} p_i(\theta) = 1$, $\forall \theta \in \Theta$,
2. $p_j(\theta) \geq c > 0, \forall j$
3. *существуют непрерывные производные*

$$\frac{\partial p_j(\theta)}{\partial \theta_k},\ k = \overline{1,r},$$

$$\frac{\partial^2 p_j(\theta)}{\partial \theta_k \theta_l},\ k,l = \overline{1,r},$$

4. *Матрица размера $N \times r$ $\left\| \frac{\partial^2 p_j(\theta)}{\partial \theta_k} \right\|$ имеет ранг r для всех $\theta \in \Theta$.*

Пусть также

$$\widehat{\theta} = \arg\max_{\theta} \prod_{j=1}^{N} \left(p_j(\theta)\right)^{\nu_j},$$

Тогда статистика

$$\widehat{X}_n^2 = X_n^2\left(\widehat{\theta}\right) = \sum_{j=1}^{N} \frac{\left(\nu_j - np_j(\theta)\right)^2}{np_j(\theta)}$$

имеет распределение хи-квадрат с $N - 1 - r$ степенями свободы.

В этом случае гипотеза H_0 отвергается тогда и только тогда, когда $\widehat{X}_n^2 > \chi^2_{1-\alpha,N-1-r}$, где α — заданный уровень значимости.

!!!!!!!!!!Добавить про состоятельность критерия хи-квадрат и близкие альтернативы

9.3 Гипотеза и критерии однородности

Пусть $X = (X_1 \dots X_n)$ из распределения $\mathcal{L}(\xi)$ с неизвестной функцией распределения $F_1(x)$ и $Y = (Y_1 \dots Y_n)$ из распределения $\mathcal{L}(\eta)$ также с неизвестной функцией распределения $F_2(x)$. Гипотеза однородности формулируется следующим образом $H_0 : F_1(x) = F_2(x)$ и заключается в проверке гипотезы о том, что рассматриваются две выборки из одного и того же распределения.

9.3.1 Критерий однородности Смирнова

Пусть X, Y - две выборки объема n и m соответственно $\widehat{F}_{1n}$ - э.ф.р., построенная по выборке X, $\widehat{F}_{2m}$ - э.ф.р., построенная по выборке Y. Рассмотрим статистику:

$$D_{n,m} = \sup_{x \in \mathbb{R}} \left| \widehat{F}_{1n}(x) - \widehat{F}_{2n}(x) \right|$$

В случае если F_1 и F_2 непрерывные функции распределения, то по теореме Смирнова статистика

$$\sqrt{\frac{n \cdot m}{n+m}} D_{n,m}$$

имеет распределение Колмогорова. Тогда критерий проверки гипотезы однородности можно сформулировать следующим образом: если $D_{n,m} > t_\alpha(n,m)$, то гипотезу H_0 отвергаем, где

$$t_\alpha(n,m) = \sqrt{\frac{1}{n} + \frac{1}{m}} t_\alpha,\ K(t_\alpha) = 1 - \alpha.$$

При этом:

$$\mathsf{P}\left(D_{n,m} > \sqrt{\frac{1}{n} + \frac{1}{m}} t_\alpha \middle| H_0 \right) = \mathsf{P}\left(\sqrt{\frac{mn}{n+m}} D_{n,m} > t_\alpha \middle| H_0 \right) \underset{\substack{m\to\infty \\ n\to\infty}}{=} 1 - K(\lambda_\alpha) = \alpha.$$

Данный критерий имеет ряд преимуществ:

1. Можем использовать статистику даже если не знаем вид распределения, кроме того, что оно **непрерывное**
2. $D_{n,m}$ считается легко

$$D^+_{n,m} = \max_{1 \le r \le m} \left| \frac{r}{m} - \widehat{F}_{1n}\left(Y_{(r)}\right) \right| = \max_{1 \le r \le n} \left| \widehat{F}_{2m}\left(X_{(r)}\right) - \frac{r-1}{n} \right|$$

$$D_{n,m}^{-} = \max_{1\le r\le m}\left|\widehat{F}_{1n}\left(Y_{(r)}\right) - \frac{r-1}{m}\right| = \max_{1\le r\le n}\left|\frac{r}{n} - \widehat{F}_{2m}\left(X_{(r)}\right)\right|$$

$$D_{n,m} = \max\left\{D_{n,m}^{-}, D_{n,m}^{+}\right\}.$$

Однако применять мы его можем только в случае непрерывных распределений.

Стоит отметить, что сходимость к распределению Колмогорова достаточно медленная. В случае небольших значений n и m можно пользоваться таблицами, приведенными, например в [?].

9.3.2 Критерий однородности хи-квадрат

Рассмотрим критерий, который можно применить для проверки гипотезы однородности в случае, когда наблюдается некоторый переменный признак, принимающий конечное число $N \geqslant 2$ различных значений.

Как и для критерия однородности Пирсона изначально рассмотрим случай полиномиально распределенных случайных величин. Очевидным образом критерий обобщается на случай произвольных случайных величин. Очевидным плюсом данного критерия является тот факт, что с его помощью можно проверять однородность произвольного количества выборок. Пусть k — количество серий наблюдений $\left(X^{(1)}, X^{(2)}, \ldots, X^{(k)}\right)$. Каждая серия $X^{(i)}$ порождает вектор частот $\nu_i = (\nu_{i,1}, \ldots, \nu_{i,N})$, $i = \overline{1,k}$. При этом для каждой выборки имеется свое вероятностное распределения $(p_{i1}, \ldots, p_{iN}) = \overline{p}_i$, которое называется вероятностью появления *признаков*.

Гипотеза однородности H_0 заключается в том, что

$$\overline{p}_0 = \overline{p}_1 = \cdots = \overline{p}_k = \overline{p},$$

где $\overline{p} = (p_1, \ldots, p_N)$, $p_i > 0, i = \overline{1,N}$, $\sum\limits_{i=1}^{N} p_i = 1$.

В случае верности гипотезы H_0 верно равенство: $\mathsf{M}(\nu_{ij}|H_0) = n_i p_j$. Рассмотрим статистику

$$\sum_{i=1}^{k} X_k^2 = \sum_{i=1}^{k}\sum_{j=1}^{N}\frac{(\nu_{ij} - n_i p_j)^2}{n_i p_j}.$$

Как и в случае использования критерия хи-квадрат для случая сложной гипотезы заменим значения p_i их оценкой максимального правдоподобия $\widehat{p}_j$, построенная по всем выборкам:

$$\widehat{p} = (\widehat{p}_1, \ldots, \widehat{p}_N) = \arg\max_{\overline{p}} \prod_{i,j} p_j^{\nu_{i,j}} = \arg\max_{\overline{p}} \prod_{j} p_j^{\nu_{\cdot,j}},$$

где $\nu_{\cdot,j} = \sum\limits_{i=1}^{N} \nu_{i,j}$. Отсюда легко можно заметить, что

$$\widehat{p}_j = \frac{\nu_{\cdot,j}}{n},$$

где $n = n_1 + \ldots + n_k$. Получаем следующий вид критерия проверки однородности хи-квадрат:

$$\widehat{X}^2_{n_1,\ldots,n_k} = X^2_{n_1,\ldots,n_k}(\widehat{p}) = n \sum_{i=1}^{k} \sum_{j=1}^{N} \frac{1}{n_i \nu_{\cdot j}} \left(\nu_{i,j} - \frac{n_i \nu_{\cdot j}}{n} \right)^2.$$

H_0 отвергаем тогда и только тогда, когда $X^2_{n_1,\ldots,n_k} > t_\alpha$.

Осталось узнать как считать величину t_α. Можно доказать, что $\mathcal{L}(X^2_{n_1,\ldots,n_k} | H_0) \underset{n_i \to \infty}{\longrightarrow} \chi^2_{(k-1)(N-1)}$. Что дает нам следующий критерий проверки однородности: гипотезу H_0 отвергаем тогда и только тогда, когда $\widehat{X}^2_{n_1,\ldots,n_k} > \chi^2_{1-\alpha,(k-1)(N-1)}$, где α — заданный уровень значимости.

9.4 Выбор из двух простых гипотез. Критерий Неймана-Пирсона

Рассмотрим случай двух простых гипотез H_0 и H_1. Иными словами, считаем, что допустимыми распределениями случайной величины ξ являются лишь два заранее заданных распределения $F_0(x)$ и $F_1(x)$. Необходимо по выборке $X = (X_1, \ldots, X_n)$ из $\mathcal{L}(\xi)$ проверить гипотезу H_0 против альтернативы H_1.

Выбор из двух простых гипотез можно представить в виде параметрической гипотезы. Действительно, пусть $\Theta = \{0,1\}$, и $F_\theta(x) = (1-\theta)F_0(x) + \theta F_1(x)$.

В случае параметрических гипотез функцию мощности критерия можно переписать в виде:

$$W(\theta) = W(\theta; \mathfrak{X}_{1,\alpha}) = \mathsf{P}_\theta \left(X \in \mathfrak{X}_{1,\alpha} \right).$$

Сначала рассмотрим случай абсолютно непрерывных распределений. Для реализации выборки $\overline{x}$ с помощью функции плотности можем выписать правдоподобие данных:

$$L(\overline{x}, \theta_i) = f_i(x_1) \cdot \ldots \cdot f_i(x_n).$$

Как и ранее рассмотрим задачу нахождения такого статистического критерия, что при заданной ошибке 1 рода

$$\alpha = W(\theta_0; \mathfrak{X}_{1,\alpha}) = \int\limits_{\mathfrak{X}_{1,\alpha}} L\left(\overline{x}; \theta_0\right) d\overline{x}$$

максимизировать функцию мощности

$$W(\theta_1; \mathfrak{X}_{1,\alpha}) = \int\limits_{\mathfrak{X}_{1,\alpha}} L\left(\overline{x}; \theta_1\right) d\overline{x} = 1 - \beta \to \max,$$

где β — ошибка 2 рода. Параметрический критерий, минимизирующий ошибку 2 рода при заданной ошибке 1 рода называется *наиболее мощным критерием с уровнем значимости* α. Заметим, что функции $W(\theta_i; \mathfrak{X}_{1,\alpha})$ есть суть вероятности того, что выборка попадет в i-ю критическую область. Критическую область можно построить следующим образом: множество $\mathfrak{X}_{1,\alpha}$ состоит из таких $\overline{x}$, для которых правдоподобие $L(\overline{x}, \theta_1)$ будет больше правдоподобия $L(\overline{x}, \theta_0)$.

Определение 9.17. *Функция, имеющая вид:*

$$l(\overline{x}) = \frac{L(\overline{x},\theta_1)}{L(\overline{x}, \theta_0)} = \frac{\prod\limits_{i=1}^{n} f_1(x_i)}{\prod\limits_{i=1}^{n} f_0(x_i)}$$

называется функцией отношения правдоподобия.

Выберем некоторую границу c. Если $l(\overline{x}) \geq c$, то принимаем H_1, иначе — H_0. Далее возьмем все $x \in X$ и упорядочим их по значению $l(\overline{x})$ и зададим такую c, чтобы выполнялось условие ошибка первго рода была в точности α.

Определение 9.18. *Назовем $T(X)$ тестовой статистикой, если выполняется равенство $T(\overline{x}) = l(\overline{x})$,и $l(\overline{x})$ — отношение правдоподобия.*

Определение 9.19. *Критическим множеством критерия Неймана-Пирсона называется множество $\mathfrak{X}^*_{1,\alpha}$ имеющее вид:*

$$\mathfrak{X}^*_{1,\alpha} = \{\overline{x} \in \mathfrak{X} : l(\overline{x}) \geq c_\alpha)\},$$

где c_α такое, что ошибка 1 рода равна α.

Предположим далее, что $f_i(x) > 0$, так как из $f_i(x) = 0$ следует $L(x_i,\theta_0) = 0$, что говорит о верности другой гипотезы. Здесь и далее для сокращения записи вместо P_{θ_i} будем писать P_i. Определим вспомогательную функцию $\phi(c)$:

$$\phi(c) = \mathsf{P}_0\left(l(\overline{x}) \geq c\right) = \int\limits_{\overline{x}:l(\overline{x})\geq c} L(\overline{x}, \theta_0) d\overline{x}.$$

Заметим, что при $c = 0$ функция принимает значение равное $\phi(0) = 1$. Покажем, что чем больше c, тем меньше значение $\phi(c)$:

$$1 \geq \mathsf{P}_1(l(\overline{x}) \geq c) = \int\limits_{\overline{x}:l(\overline{x})\geq c} L(\overline{x},\theta_1) d\overline{x} \geq c \cdot \int\limits_{\overline{x}:l(\overline{x})\geq c} L(\overline{x},\theta_0) d\overline{x} = c \cdot \phi(c).$$

Следовательно:

$$\phi(c) \leq \frac{1}{c}.$$

Отсюда следует, что при $c \to \infty$ значение $\phi(c) \to 0$. Предположим, что $\phi(c)$ — непрерывная функция. Тогда для любого $\alpha \in (0; 1)$ можно найти такое $c_\alpha : \phi(c_\alpha) = \alpha$. Соответственно для множества $\mathfrak{X}^*_{1,\alpha} = \{\overline{x} : l(\overline{x}) \geq c_\alpha\}$ верно равенство:

$$W(\theta_0, \mathfrak{X}^*_{1,\alpha}) = \mathsf{P}_0\left(\mathfrak{X}^*_{1,\alpha}\right) = \phi(c_\alpha) = \alpha.$$

Теорема 9.20 (Лемма Неймана-Пирсона)**.** *Пусть для фиксированного α существует такое c_α, что $\phi(c_\alpha) = \alpha$. Тогда критическая область $\mathfrak{X}^*_{1,\alpha} = \{\overline{x} : l(\overline{x}) \geq c_\alpha\}$ задает наиболее мощный критерий для гипотезы $H_0\colon F_\xi = F_0$ относительно альтернативы $H_1\colon F_\xi = F_1$ среди всех критериев с уровнем значимости α.*

Доказательство. Пусть $\mathfrak{X}_{1,\alpha}$ - другая критическая область с уровнем значимости α:

$$W(\theta_0; \mathfrak{X}_{1,\alpha}) = \mathsf{P}_0(\overline{x} \in \mathfrak{X}_{1,\alpha}) = \int\limits_{\mathfrak{X}_{1,\alpha}} L(\overline{x},\theta_0) d\overline{x} = \alpha.$$

Тогда

$$\mathsf{P}_0(\mathfrak{X}_{1,\alpha}\backslash\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}) = \mathsf{P}_0(\mathfrak{X}_{1,\alpha})-\mathsf{P}_0(\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}) = \mathsf{P}_0(\mathfrak{X}^*_{1,\alpha})-\mathsf{P}_0(\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}) = \\ = \mathsf{P}_0(\mathfrak{X}^*_{1,\alpha}\backslash\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha})$$

Согласно определению множества $\mathfrak{X}^*_{1,\alpha}$, для любого $\overline{x}$ не лежащего в нем, выполняется неравенство: $l(\overline{x}) < c_\alpha$ или

$$c_\alpha \cdot L(\overline{x};\theta_0) > L(\overline{x};\theta_1).$$

Следовательно,

$$\mathsf{P}_1(\mathfrak{X}^*_{1,\alpha}\backslash\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}) \geq c_\alpha\cdot\mathsf{P}_0(\mathfrak{X}^*_{1,\alpha}\backslash\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}) = c_\alpha\cdot\mathsf{P}_0(\mathfrak{X}_{1,\alpha}\backslash\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}) > \\ > \mathsf{P}_1(\mathfrak{X}_{1,\alpha}\backslash\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha}).$$

Прибавив к обоим частям неравенства $\mathsf{P}_1(\mathfrak{X}_{1,\alpha}\cap\mathfrak{X}^*_{1,\alpha})$ получим:

$$W(\theta_1;\mathfrak{X}^*_{1,\alpha}) = \mathsf{P}_1(\mathfrak{X}^*_{1,\alpha}) > \mathsf{P}_1(\mathfrak{X}_{1,\alpha}) = W(\theta_1,\mathfrak{X}_{1,\alpha}).$$

□

Определение 9.21. *Будем говорить, что статистический критерий является несмещенным, если $W(\theta) \geqslant \alpha$ для всех $\alpha \in \Theta_1$.*

Утверждение 9.22. *Критерий Неймана-Пирсона является несмещенным критерием.*

Доказательство. 1) Рассмотрим случай $c_\alpha \geq 1$. Тогда

$$W(\theta_1,\mathfrak{X}^*_{1,\alpha}) = \int\limits_{\mathfrak{X}^*_{1,\alpha}} L(\overline{x},\theta_1)d\overline{x} \geq c_\alpha\cdot\int\limits_{\mathfrak{X}^*_{1,\alpha}} L(\overline{x},\theta_0)d\overline{x} \geq \alpha\cdot c_\alpha > \alpha$$

2) Рассмотрим случай $c_\alpha < 1$:

$$W(\theta_1,\mathfrak{X}^*_{1,\alpha}) = 1 - P(\overline{\mathfrak{X}^*_{1,\alpha}}) > 1 - c_\alpha P_0(\overline{\mathfrak{X}_{0,\alpha}}) \geq 1 - P_0(\overline{\mathfrak{X}_{1,\alpha}}) = \alpha$$

□

Пример 9.23. Пусть имеется выборка $X = (X_1,\ldots,X_n)$ из распределения $\mathcal{L}(\xi)$. Рассмотрим две гипотезы:

$$H_0 : \xi \sim N(\theta_0, \sigma^2)$$
$$H_1 : \xi \sim N(\theta_1, \sigma^2)$$

Найдем функцию отношения правдоподобия:

$$l(\overline{x}) = \frac{\prod\limits_{i=1}^{n} f_1(x_i)}{\prod\limits_{i=1}^{n} f_2(x_i)} = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-\theta_0)^2-(x_i-\theta_1)^2\right]\right\} =$$
$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2+\theta_0^2-2x_i\theta_0-x_i^2-\theta_1^2+2x_i\theta_1\right\} =$$
$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \theta_0^2-\theta_1^2+2x_i\theta_1-2x_i\theta_0\right\} =$$
$$= \exp\left\{-\frac{n}{2\sigma^2}(\theta_0^2-\theta_1^2)-\frac{n}{\sigma^2}(\theta_1-\theta_0)\frac{1}{n}\sum_{i=1}^{n} x_i\right\} =$$
$$= \exp\left\{\frac{n}{\sigma^2}(\theta_0-\theta_1)\overline{X}-\frac{n}{2\sigma^2}(\theta_0^2-\theta_1^2)\right\} \geq c$$

Прологарифмируем:

$$\frac{n}{\sigma^2}(\theta_0-\theta_1)\overline{X} \geq \ln c + \frac{n}{2\sigma^2}(\theta_0^2-\theta_1^2),$$

$$\overline{X} \geq \frac{\sigma^2}{n\cdot(\theta_0-\theta_1)}\ln c + \frac{\theta_0-\theta_1}{2}.$$

Получаем следующее эквивалентное равенство:

$$\mathsf{P}_i(l(\overline{x}) \geq c_\alpha) = \mathsf{P}_i\left(\overline{X} \geq t_\alpha\right).$$

Тогда:

$$\phi(c_\alpha) = \mathsf{P}_0\left(\overline{X} \geq t_\alpha\right) = \mathsf{P}_0\left(\frac{\sqrt{n}}{\sigma}\left(\overline{X}-\theta_0\right) \geq \frac{(t_\alpha-\theta_0)\sqrt{n}}{\sigma}\right) = \Phi\left(-\frac{(t-\theta_0)\sqrt{n}}{\sigma}\right),$$

где Φ - функия стандартного нормального распределения.

Так как $\Phi(c)$ — непрерывная функция, то всегда найдем такое t_α. Тогда,

$$\mathsf{P}_1\left(\overline{X} \geq t_\alpha\right) = \mathsf{P}_1\left(\frac{\sqrt{n}}{\sigma}(\overline{X}-\theta_1) \geq \frac{\sqrt{n}(t_\alpha-\theta_1)}{\sigma}\right) = \Phi\left(-\frac{\sqrt{n}(t_\alpha-\theta_1)}{\sigma}\right).$$

Точное вычисление параметров критерия возможен в том случае, когда распределение статистики $l(X)$ (или эквивалентной ей как в примере выше) известно как при гипотезе H_0, так и при альтернативе H_1. Это не всегда возможно. В случае большого объема выборки ($n \to \infty$) возможно рассмотреть асимптотический подход, основанный на центральной предельной теореме. При логарифмировании функции отношения правдоподобия мы получаем сумму одинаково распределенных независимых величин вида

$$Z_i = \ln \frac{f_1(X_i)}{f_0(X_i)},\ i \in \overline{1,n}.$$

Из $l(\overline{X}) \geq c_\alpha$ следует, что $S_n \geq \ln c_\alpha$, где $S_n = \sum\limits_{i=1}^{n} Z_i$. Воспользуемся центральной предельной теоремой (если это возможно) и получим предельное распределение статистики $\ln l(\overline{X})$:

$$\begin{gathered}\text{Если } S_n \sim N\left(\mu_0, \sigma_0^2\right), \text{ то принимаем гипотезу } H_0;\\ \text{Если } S_n \sim N\left(\mu_1, \sigma_1^2\right), \text{ то принимаем гипотезу } H_1,\end{gathered}$$

где

$$\mu_i = \mathsf{M}_{\theta_i} Z_1,\ \sigma_i^2 = \mathsf{D}_{\theta_i} Z_1,\ i = \overline{1,2}.$$

В этом случае ошибку первого рода можно вычислить по формуле:

$$\alpha = \mathsf{P}_0\left(S_n \geqslant \ln c_\alpha\right) = \mathsf{P}_0\left(\frac{S_n - n\mu_0}{\sqrt{n}\sigma_0} \geqslant \frac{\ln c_\alpha - n\mu_0}{\sqrt{n}\sigma_0}\right),$$

а ошибку второго рода:

$$\beta = \mathsf{P}_1\left(S_n < \ln c_\alpha\right) = \mathsf{P}_1\left(\frac{S_n - n\mu_1}{\sqrt{n}\sigma_1} < \frac{\ln c_\alpha - n\mu_1}{\sqrt{n}\sigma_1}\right).$$

Рассмотрим вопрос применения Критерия Неймана-Пирсона в случае дискретных распределений.

Можно провести аналогичные рассуждения, как и в случае с непрерывными распределениями. Пусть как и ранее $f_j(z_k) > 0$, для всех z_k — возможных значений случайной величины ξ, и для $j = \overline{0,1}$. То есть для рассматриваемых гипотез вероятность любого из возможных значений случайной величины ξ, распределение которой мы хотим установить, ненулевая.

Общий принцип остаётся прежним: рассматриваем статистику отношения правдоподобия $l(\overline{x})$. Все $x \in \mathfrak{X}$ «упорядочиваем» в соответствии с величиной значения функции отношения правдоподобия:

$$l(\overline{x}) = \frac{L(\overline{x}, \theta_1)}{L(\overline{x}, \theta_0)}.$$

В критическое множество $\mathfrak{X}_1$ включим максимальное число этих x, так, чтобы выполнялось неравенство:

$$\sum_{\overline{x} \in \mathfrak{X}_1} L(\overline{x}; \theta_0) \leqslant \alpha.$$

Однако в отличие от непрерывного случая здесь не всегда можно получить равенство величине α в силу дискретности распределения.

Пусть $\ldots l_k < l_{k+1} < \ldots$ — возможные значения статистики $l(\overline{x})$. Могут возникнуть две ситуации:

1. Существует $l_k \in \mathbb{N}$ такой, что верно равенство:

$$\sum_{\overline{x}: l(\overline{x}) \geqslant l_k} L(\overline{x}; \theta_0) = \alpha.$$

   В этом случае критерий не отличается от непрерывного случая и применяется аналогично, то есть: H_0 отвергается тогда и только тогда, когда $l(\overline{x}) \in \mathfrak{X}^*_{1\alpha}$, $\mathfrak{X}^*_{1\alpha} = \{\overline{x} \colon l(\overline{x} \geqslant\}$ — наиболее мощный критерий (н.м.к.) при альтернативе H_1 среди всех критериев уровня значимости α.

2. При заданном уровне α можно определить такое $k = k(\alpha)$, что

$$\sum_{\overline{x}: l(\overline{x}) \geqslant l_{k+1}} L(\overline{x}; \theta_0) < \alpha < \sum_{\overline{x}: l(\overline{x}) \geqslant l_k} L(\overline{x}; \theta_0).$$

   Иначе говоря, обозначив

$$\sum_{\overline{x}: l(\overline{x}) \geqslant l_{k+1}} L(x; \theta_j) = \alpha_j,\ j = \overline{0{,}1},$$

$$p_j = \mathsf{P}(l(\overline{x}) = l_k) = \sum_{\overline{x}\colon\, l(\overline{x}) = l_k} L(\overline{x}; \theta_j),\ j = \overline{0{,}1}$$

   получаем:

$$\alpha_0 < \alpha < \alpha_0 + p_0.$$

Рассмотрим более подробно второй случай. Как в этом случае построить критерий с уровнем значимости α? Можно предложить два решения:

- Перейти от α_0 к α_0+p_0 и получить 1 случай, для которого все доказано. Но в этом случае точность уровня значимости может быть низкой.
- Рассмотрим случайную величину

$$\xi \sim \mathrm{Bin}\left(1, \frac{\alpha-\alpha_0}{p_0}\right).$$

Предложим следующее правило принятия решения:

$$\varphi^*(\overline{x}) = \begin{cases} 1, & \text{если } l(\overline{x}) > l_k \\ \xi, & \text{если } l(\overline{x}) = l_k \\ 0, & \text{если } l(\overline{x}) < l_k \end{cases}$$

То есть если $l(\overline{x}) > l_k$ — отвергаем гипотезу H_0, $l(\overline{x}) < l_k$ — принимаем гипотезу H_0. Если же $l(\overline{x}) = l_k$ бросаем монетку с вероятностью 1 (орел) равной $\frac{\alpha-\alpha_0}{p_0}$. Если выпала единица, то отвергаем H_0, иначе — принимаем. Такое правило называют *рандомизированным критерием.*

Покажем, что рандомизированный критерий является критерием с уровнем значимости α. Действительно,

$$\mathsf{P}(H_1|H_0) = \mathsf{P}_0\left(\varphi^*(X) = 1\right) = \\ = \mathsf{P}_0\left(l\left(X\right) > l_k\right) + \frac{\alpha-\alpha_0}{p_0}\mathsf{P}_0\left(l\left(X\right) = l_k\right) = \alpha_0 + \frac{\alpha-\alpha_0}{p_0}p_0 = \alpha.$$

Его мощность вычисляется аналогично:

$$W(\theta_1;\varphi^*) = \mathsf{P}_1\left(l\left(X\right) > l_k\right) + \frac{\alpha-\alpha_0}{p_0}\mathsf{P}_1\left(l\left(X\right) = l_k\right) = \alpha_1 + \frac{\alpha-\alpha_0}{p_0}p_1.$$

Покажем, что φ^* - наиболее мощный критерий с уровнем значимости α. Рассмотрим другой (произвольный) критерий φ с уровнем значимости α:

$$\varphi\colon \mathsf{P}_0\left(\varphi\left(X\right) = 1\right) = \alpha.$$

Представим выборочное пространство в следующем виде:

$$\mathfrak{X} = \mathfrak{X}^+ \bigcup \mathfrak{X}^0 \bigcup \mathfrak{X}^-,$$

где

- $\mathfrak{X}^+ = \{\overline{x} : \varphi^*(\overline{x}) - \varphi(\overline{x}) > 0\}$,
- $\mathfrak{X}^- = \{\overline{x} : \varphi^*(\overline{x}) - \varphi(\overline{x}) < 0\}$,
- $\mathfrak{X}^0 = \{\overline{x} : \varphi^*(\overline{x}) = \varphi(\overline{x}) = 0\}$.

Если $\overline{x} \in \mathfrak{X}^+$, то $\varphi^*(x) > 0$. Тогда

$$\frac{L(\overline{x};\theta_1)}{L(\overline{x};\theta_0)} \geqslant l_k$$

и $L(\overline{x};\theta_1) \geqslant l_k L(\overline{x};\theta_0)$. Аналогично, если $\overline{x} \in \mathfrak{X}^-$, то $\varphi_1^*(\overline{x}) < 1$ и $L(\overline{x};\theta_1) \leqslant l_k L(\overline{x};\theta_0)$.

Таким образом:

$$\sum_{\overline{x}} (\varphi^*(\overline{x}) - \varphi(\overline{x}))(L(\overline{x};\theta_1) - l_k L(x;\theta_0)) = \sum_{\overline{x}\in\mathfrak{X}^+} + \sum_{\overline{x}\in\mathfrak{X}^-} \geqslant 0.$$

Отсюда для разности мощностей получаем:

$$\begin{aligned} W(\theta_1;\varphi^*) - W(\theta_1;\varphi) = \mathsf{P}_1(\varphi^*(X) = 1) - \mathsf{P}_1(\varphi(X) = 1) = \\ = \sum_{\overline{x}} (\varphi^*(x) - \varphi(\overline{x})) L(\overline{x};\theta_1) \geqslant \\ \geqslant l_k \sum_{\overline{x}} (\varphi^*(x) - \varphi(x)) L(\overline{x};\theta_0) = l_k(\mathsf{P}_0(\varphi^*(X) = 1) - \mathsf{P}_0(\varphi(X) = 1)) \geqslant 0. \end{aligned}$$

То есть, $W(\theta_1;\varphi^*) \geqslant W(\theta_1;\varphi)$. Из этого следует, что φ^* — наиболее мощный критерий.

Пример 9.24. Пусть о неизвестной вероятности успеха в бернулиевской модели $\mathrm{Bin}(1,\theta)$ имеются две простые гипотезы:

- $H_0\colon \theta = \theta_0$;
- $H_1\colon \theta = \theta_1$, $\theta_1 \geq \theta_0$.

Такие критерии зачастую называют односторонними и обозначают H_1^+.

Выпишем функцию правдоподобия:

$$L(\overline{x};\theta_j) = (\theta_j)^r (1-\theta_j)^{n-r},$$

где $r = r(\overline{x}) = \sum\limits_{i=1}^{n} x_i$ — наблюдавшееся число «успехов».

Рассмотрим функцию отношения правдоподобия:

$$l(\overline{x}) = \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^r \left(\frac{(1-\theta_1)}{(1-\theta_0)}\right)^n.$$

Функция $f(x) = x/(1-x)$ возрастает на интервале (0,1), поэтому, при $\theta_1 > \theta_0$ $f(\theta_1)/f(\theta_0) > 1$ и неравенство $l(\overline{x}) \geq c$ эквивалентно неравенству

$$r(\overline{x}) \geq \frac{\ln c - n\rho_1}{\rho},$$

где

$$\rho_1 = \ln\left(\frac{1-\theta_1}{1-\theta_0}\right), \rho = \ln\left(\frac{\theta_0(1-\theta_1)}{\theta_1(1-\theta_0)}\right).$$

Следовательно критическая область критерия Неймана-Пирсона выражается через статистику $r(\overline{x})$ и имеет вид:

$$\mathfrak{X}^*_{1\alpha} = \{\overline{x} \colon r(\overline{x}) \geq r_\alpha\}.$$

Так как статистика $r(\overline{x})$ при верности гипотезы H_j имеет распределение $\mathrm{Bin}(n, \theta_j)$, то легко вычисляем ошибки первого и второго рода. При этом ошибка первого рода не зависит от альтернативы. От нее зависит только ошибка второго рода.

Заметим, что в случае $n \to \infty$ можем воспользоваться ЦПТ и рассчитать ошибки критерия пользуясь нормальным приближением.

10 Сложные гипотезы

Случай, когда основная и альтернативная гипотезы являются простыми — достаточно редки. Рассмотрим пример, в котором основная гипотеза H_0 : $F_\theta, \theta \in \Theta$ — простая гипотеза (один параметр), а гипотеза $H_1 : F_{\theta_2}, \theta_2 \in \Theta \backslash \theta$ (альтернативная) — сложная гипотеза.

Определение 10.1. *Семейство распределений $\{F_\theta, \theta \in \Theta\}$ называется семейством с монотонным отношением правдоподобия, если существует такая достаточная статистика $T(\overline{x})$ такая, что:*

$$l(\overline{x}) = \frac{f_{\theta_1}(\overline{x})}{f_{\theta_0}(\overline{x})}$$

является монотонной функцией от $T(\overline{x})$

Пользуясь критерием факторизации получим, что если $T(\overline{x})$ — достаточная статистика, то

$$L(\overline{x};\theta) = g(T(\overline{x});\theta) \cdot h(\overline{x})$$

и

$$l(\overline{x}) = \frac{g(T(\overline{x}),\theta_1)h(\overline{x})}{g(T(\overline{x}),\theta_0)h(\overline{x})}.$$

Функция $l(\overline{x})$ должна монотонно возрастать (убывать) по $T(\overline{x})$.

Теорема 10.2. *Пусть семейство распределений $\{F_\theta, \theta \in \Theta\}$ имеет монотонное отношение правдоподобия. Тогда в классе всех критериев проверки простой гипотезы $H_0 = \theta$ против сложной альтернативы $H_1 = \theta_1 > \theta$ с уровнем значимости α существует равномерно наиболее мощный критерий, задаваемый функцией:*

$$\varphi^*(x) = \begin{cases} 1 & \textit{если } T(\overline{x}) > c \\ \xi & \textit{если } T(\overline{x}) = c \\ 0 & \textit{если } T(\overline{x}) < c \end{cases}$$

где с и р определяются по формуле:

$$\mathsf{P}_0(T(\overline{x}) > c) + p\mathsf{P}_0(T(\overline{x} \geqslant c) = \alpha.$$

Доказательство. Пусть отношение

$$\frac{g(T(\overline{x});\theta_1)}{g(T(\overline{x});\theta_0)}$$

возрастает по T. Построим критерий Неймана-Пирсона в задаче (H_0, H_1). В данном случае:

$$l(\overline{x}) \geqslant c \Leftrightarrow T(\overline{x}) \geqslant c_\alpha^+$$

причем граница c_α^+ определяется лишь распределением

$$F(x;\theta_0) = \mathsf{P}_{\theta_0}\{T(\overline{x}) \geqslant c_\alpha^+\} = \alpha.$$

Таким образом, критерий $\mathfrak{X}^{*+}_{1\alpha} = \{T(\overline{x}) \geqslant c^{+}_{\alpha}\}$ не зависит от конкретной альтернативы $\theta_1 > \theta_0$, следовательно, он есть р.н.м.к. для задачи $(\theta_0; \theta_1)$.

Если же
$$\frac{g(T(\overline{x}); \theta_1)}{g(T(\overline{x}); \theta_0)}$$
убывает по T, то неравенство для T заменяется на противоположное. Аналогично рассматривается и случай для левосторонней H^{-}_{1}. □

Пример 10.3. Рассмотрим экспоненциальное семейство:

$$f(x; \theta) = \exp\{A(\theta)B(x) + C(\theta) + D(x)\}.$$

Достаточной статистикой в этом случае является: $T(\overline{x}) = \sum\limits_{i=1}^{n} B(X_i)$. Действительно:
$$l(\overline{x}) = \exp\{(A(\theta) - A(\theta_0)) \sum B(\overline{x}) + n(c(\theta) - c(\theta_0))\}.$$

Если A — монотонная функция, то и $l(\overline{x})$ — монотонная функция. В этом случае существует Р.Н.М.К.