Теория вероятностей и математическая статистика. Коллоквиум №4

Lev Khoroshansky

Содержание

Билет №1.

Выборка

Def. *В теории вероятностей используется термин* ***случайная величина****, в то время как в статистике этому же объекту дают название* ***генеральная совокупность****.*

Пусть нам известно распределение $F_\theta(t)$ случайной величины, где θ — неизвестный параметр. Тогда задачей статистики является оценка этого параметра.

Def. *Набор $X_1, \ldots, X_n$ независимых и одинаково распределённых случайных величин называется* ***простой выборкой****. В статистике этот набор зачастую представляется в виде чисел $X_1(w) = x_1, \ldots, X_n(w) = x_n$, где w — конкретный исход.*

Статистика

Def. ***Статистикой*** *$T_N(X_1, \ldots, X_N)$ называется борелевская функция, зависящая от N случайных величин $X_1, \ldots, X_N$ (то есть T_N не зависит от параметра θ).*

Например, $T_N(X_1, \ldots, X_N) = \dfrac{X_1 + \cdots + X_N}{N} = \overline{X}$ — выборочное среднее.

Оценка параметра

Def. *Если статистика $T_N(X_1, \ldots X_N)$ используется для оценки параметра θ, то она называется* ***оценкой*** *и обозначается как $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_N)$.*

Несмещённые и состоятельные оценки

Def. *Оценка $\widehat{\theta}$ является* ***несмещённой****, если $\mathbb{E}\widehat{\theta} = \theta$ для всех значений θ.*

Рассмотрим несколько примеров:

1. Если мы хотим оценить $\theta = \mathbb{E}X_k$, то подойдёт $\widehat{\theta} = \overline{X}$, потому что $\mathbb{E}\widehat{\theta} = \mathbb{E}\overline{X} = \mathbb{E}X_k = \theta$.

2. Несмещённая оценка может не существовать — например, у схемы Бернулли:

$$X_k = \begin{cases} 1, & p \in (0; 1) \\ 0, & q = 1 - p \end{cases}$$

Пусть мы хотим оценить параметр $\theta = \dfrac{1}{p}$ и пусть $\widehat{\theta} = T_N(X_1, \ldots, X_N)$ — наша оценка, тогда

$$\mathbb{E}T_N(X_1, \ldots, X_N) = \sum_{k=0}^{N} \left(\sum_{\#1=k} T_N(\varepsilon_1, \ldots, \varepsilon_N) \right) p^k q^{n-k} = \frac{1}{p}$$

В равенстве выше ε_k принимают значения 0 или 1, а запись $\#1 = k$ означает, что среди всех $\varepsilon_1, \ldots, \varepsilon_N$ было ровно k единиц. Видно, что при $p \to 0$ левая часть стремится к какой-то константе, в то время как правая часть уходит на бесконечность. Следовательно, равенство не выполняется.

3. Несмещённая оценка может не иметь смысла — например, у Пуассоновского распределения:

$$\mathbb{P}(X_1 = k) = \frac{\theta^k}{k!} e^{-\theta}, \text{ где } 0 < \theta < 1.$$

Предположим, что вся выборка состоит из одной случайной величины X_1.
Пусть $\widehat{\theta} = T(X_1)$ — наша оценка, тогда

$$\mathbb{E}T(X_1) = \sum_{k=0}^{+\infty} T(k) \frac{\theta^k}{k!} e^{-\theta} = \theta$$

Умножим на экспоненту:

$$\sum_{k=0}^{+\infty} T(k) \frac{\theta^k}{k!} = \theta e^{\theta}$$

Заметим, что в левой части написано разложение экспоненты по Тейлору, умноженное на θ:

$$\sum_{k=0}^{+\infty} T(k) \frac{\theta^k}{k!} = \theta + \theta^2 + \frac{\theta^3}{2!} + \ldots$$

Откуда можем выразить $T(k)$:

$$\frac{T(k)}{k!} = \frac{1}{(k-1)!} \implies T(k) = k$$

Следовательно, $T(X_1) = X_1$. Но, так как X_1 принимает натуральные значения и $\theta \in (0; 1)$ по условию, данная оценка бессмысленна.

Def. *Оценка $\widehat{\theta} = T_N(X_1, \ldots, X_N)$ называется* ***состоятельной****, если $\widehat{\theta} \xrightarrow[N\to\infty]{\mathbb{P}} \theta$.*

Обычно состоятельность оценки является следствием закона больших чисел. В качестве примера может выступить выборочное среднее при оценке $\theta = \mathbb{E}X_k$:

$$\begin{cases} \overline{X} \xrightarrow[\text{ЗБЧ}]{\mathbb{P}} \mathbb{E}X_k, \\ \widehat{\theta} = \overline{X}, \end{cases} \implies \widehat{\theta} \xrightarrow{\mathbb{P}} \theta$$

Проверка состоятельности с помощью неравенства Чебышёва

St. *Пусть $\widehat{\theta}$ — несмещённая оценка θ и пусть $\mathbb{D}\widehat{\theta} \to 0$, тогда $\widehat{\theta}$ является состоятельной оценкой.*

Доказательство. Применим неравенство Чебышёва:

$$\mathbb{P}\left(|\widehat{\theta} - \theta| \geqslant \delta\right) \leqslant \frac{\mathbb{D}\widehat{\theta}}{\delta^2} \to 0 \implies \widehat{\theta} \xrightarrow{\mathbb{P}} \theta$$

[:|||:]

Билет №2.

Оптимальные оценки

Def. *Несмещённая оценка $\widehat{\theta}$ называется* ***оптимальной****, если $\mathbb{D}\widehat{\theta} \leqslant \mathbb{D}\tilde{\theta}$ для всякой другой несмещённой оценки $\tilde{\theta}$.*

Некорректность задачи о поиске наилучшей в среднеквадратичном оценки среди всех оценок

St. *В общем случае, не существует такой оценки $\widehat{\theta}$, что $\mathbb{E}(\widehat{\theta}-\theta)^2 \leqslant \mathbb{E}(\tilde{\theta}-\theta)^2$ для всех θ и для всякой другой оценки $\tilde{\theta}$.*

Доказательство. Пусть область значений θ состоит хотя бы из двух различных чисел $A \neq B$ (иначе неинтересно). Положим $\tilde{\theta} \equiv A$, тогда

$$\mathbb{E}(\widehat{\theta}-\theta)^2 \leqslant \mathbb{E}(\tilde{\theta}-\theta)^2 \quad \forall\theta$$

Следовательно, это должно быть справедливо и для $\theta = A$:

$$\mathbb{E}(\widehat{\theta}-A)^2 \leqslant \mathbb{E}(A-A)^2 = 0 \implies \widehat{\theta} \overset{\text{a.s.}}{=} A$$

Аналогичным образом получаем, что $\widehat{\theta} \overset{\text{a.s.}}{=} B$ — противоречие. **[:|||:]**

Комментарий и вспомогательное утверждение

Заметим, что при $\theta = A$ случайные величины имеют распределение $F_A(t)$, однако при $\theta = B$ они имеют другое распределение $F_B(t)$. Если же эти распределения не накладываются друг на друга (иными словами, не существует такого события C, что $P_A(C) > 0$ и $P_B(C) > 0$), то противоречия в рассуждениях выше мы не получим. Но и ничего страшного в этом нет: в таком случае два распределения хорошо различимы и относительно каждого из них оценка может быть константой.

St. *Пусть $\widehat{\theta}$ и $\tilde{\theta}$ — оптимальные (и несмещённые) оценки, тогда $\widehat{\theta} \overset{\text{a.s.}}{=} \tilde{\theta}$ относительно распределения $F_\theta(t)$ для всех θ.*

Доказательство. Рассмотрим их среднее арифметическое:

$$\mathbb{E}\left(\frac{\widehat{\theta}+\tilde{\theta}}{2}\right) = \frac{1}{2}\left(\mathbb{E}\widehat{\theta}+\mathbb{E}\tilde{\theta}\right) = \frac{1}{2}(\theta+\theta) = \theta \text{ — несмещённость.}$$

Оптимальность обеих оценок даёт минимальность и совпадение их дисперсий (по определению), следовательно

$$\mathbb{D}\left(\frac{\widehat{\theta}+\tilde{\theta}}{2}\right) \geqslant \mathbb{D}\widehat{\theta} = \mathbb{D}\tilde{\theta}$$

С другой стороны,

$$\mathbb{D}\left(\frac{\widehat{\theta}+\tilde{\theta}}{2}\right) = \frac{1}{4}\left(\mathbb{D}\widehat{\theta} + 2\operatorname{cov}(\widehat{\theta},\tilde{\theta}) + \mathbb{D}\tilde{\theta}\right) \geqslant \mathbb{D}\widehat{\theta}$$

Заменим $\mathbb{D}\tilde{\theta}$ на $\mathbb{D}\widehat{\theta}$ в силу равенства и перенесём всё в одну сторону:

$$\frac{1}{2}\,\mathbb{D}\widehat{\theta} - \frac{1}{2}\,\mathrm{cov}(\widehat{\theta},\tilde{\theta}) \leqslant 0 \quad\Longrightarrow\quad \frac{1}{4}\,\mathbb{D}\widehat{\theta} - \frac{1}{2}\,\mathrm{cov}(\widehat{\theta},\tilde{\theta}) + \frac{1}{4}\,\mathbb{D}\tilde{\theta} \leqslant 0 \quad\Longrightarrow\quad \mathbb{D}(\widehat{\theta} - \tilde{\theta}) \leqslant 0$$

Следовательно, $\widehat{\theta} \overset{\text{a.s.}}{=} \tilde{\theta}$. **[:|||:]**

Необходимое и достаточное условие того, что оценка является оптимальной

Def. *Статистика U называется* ***несмещённой оценкой нуля****, если $\mathbb{E}U = 0$ относительно распределения $F_\theta(t)$ для любого θ.*

St. *Несмещённая оценка $\widehat{\theta}$ является оптимальной оценкой тогда и только тогда, когда $E(\widehat{\theta}\,U) = 0$ для любой несмещённой оценки нуля U.*

Доказательство. В предположении, что оценки имеют конечные дисперсии, докажем, что для оптимальной оценки $\widehat{\theta}$ справедливо равенство $\mathbb{E}(\widehat{\theta}\,U) = 0$. Рассмотрим $\widehat{\theta} + \lambda U$:

$$\mathbb{E}(\widehat{\theta} + \lambda U) = \mathbb{E}\widehat{\theta} + \lambda \cdot \mathbb{E}U = \mathbb{E}\widehat{\theta} + \lambda \cdot 0 = \theta$$

В силу оптимальности,

$$\mathbb{D}\widehat{\theta} \leqslant \mathbb{D}(\widehat{\theta} + \lambda U) = \mathbb{D}\widehat{\theta} + 2\lambda\,\mathrm{cov}(\widehat{\theta}, U) + \lambda^2\mathbb{D}U$$

Сократив $\mathbb{D}\widehat{\theta}$ слева и справа, получим, что для любого λ

$$2\lambda\,\mathrm{cov}(\widehat{\theta}, U) + \lambda^2\mathbb{D}U \geqslant 0 \quad\Longrightarrow\quad \lambda\left(\lambda - \left(-\frac{2\,\mathrm{cov}(\widehat{\theta}, U)}{\mathbb{D}U}\right)\right) \geqslant 0$$

Если $\mathbb{D}U = 0$, то $\mathrm{cov}(\widehat{\theta}, U) = 0$, так как λ любая. В противном случае, мы можем указать такое значение λ, что левая часть будет строго меньше нуля, следовательно, нужно занулить ковариацию. В любом случае получаем, что

$$0 = \mathrm{cov}(\widehat{\theta}, U) = \mathbb{E}\left[(\widehat{\theta} - \theta)(U - 0)\right] = \mathbb{E}(\widehat{\theta}\,U) - \theta \cdot \mathbb{E}U = \mathbb{E}(\widehat{\theta}\,U) = 0$$

В обратную сторону, возьмём другую несмещённую оценку $\tilde{\theta}$. Мы знаем, что

$$\mathbb{D}\tilde{\theta} = \mathbb{E}(\tilde{\theta} - \theta)^2 = \mathbb{E}(\tilde{\theta}\,[-\widehat{\theta} + \widehat{\theta}\,] - \theta)^2 = \mathbb{E}((\tilde{\theta} - \widehat{\theta}) + (\widehat{\theta} - \theta))^2$$

Раскроем квадрат:

$$\mathbb{E}((\tilde{\theta} - \widehat{\theta}) + (\widehat{\theta} - \theta))^2 = \mathbb{E}(\tilde{\theta} - \widehat{\theta})^2 + 2\mathbb{E}\left[(\tilde{\theta} - \widehat{\theta})(\widehat{\theta} - \theta)\right] + \mathbb{E}(\widehat{\theta} - \theta)^2$$

Положим $U = \tilde{\theta} - \widehat{\theta}$, тогда $\mathbb{E}U = \mathbb{E}\tilde{\theta} - \mathbb{E}\widehat{\theta} = 0$. Следовательно, U — несмещённая оценка нуля, откуда

$$2\mathbb{E}\left[(\tilde{\theta} - \widehat{\theta})(\widehat{\theta} - \theta)\right] = 2\mathbb{E}[(\widehat{\theta} - \theta)\,U] = 2\mathbb{E}(\widehat{\theta}\,U) - 2\theta \cdot \mathbb{E}U = 2\mathbb{E}(\widehat{\theta}\,U)$$

Возвращаясь к исходному равенству и помня, что $\mathbb{E}(\widehat{\theta} - \theta)^2 = \mathbb{D}\widehat{\theta}$, получаем

$$\mathbb{D}\tilde{\theta} = \mathbb{E}(\tilde{\theta} - \widehat{\theta})^2 + \mathbb{D}\widehat{\theta} \geqslant \mathbb{D}\widehat{\theta}$$

Следовательно, у любой другой несмещённой оценки дисперсия не меньше дисперсии $\widehat{\theta}$.
По определению, $\widehat{\theta}$ является оптимальной. **[:|||:]**

Пример

Рассмотрим схему Бернулли с выборкой из $X_1, \ldots, X_n$, где

$$X_i = \begin{cases} 1, & \theta \in (0;1) \\ 0, & 1-\theta \end{cases}$$

Мы знаем, что выборочное среднее $\overline{X}$ следующими свойствами:

1. $\mathbb{E}\overline{X} = \mathbb{E}X_k = \theta$ — несмещённость.

2. $\overline{X} \xrightarrow{\mathbb{P}} \theta$ — состоятельность.

3. Оптимальность: покажем, что $\mathbb{E}(\overline{X}\,U) = 0$ для всякой оценки нуля U. По определению,

   $$\mathbb{E}U = 0 \text{ для всех } \theta.$$

   Тогда

   $$\mathbb{E}U = \mathbb{E}U(X_1, \ldots, X_n) = \sum U(\varepsilon_1, \ldots, \varepsilon_n)\,\theta^k(1-\theta)^{n-k},$$

   где ε_k — значение X_k (либо 0, либо 1). Сгруппируем по количеству единиц (#1):

   $$\mathbb{E}U = \sum_{k=0}^{n} \left(\sum_{\#1=k} U(\varepsilon_1, \ldots, \varepsilon_n) \right) \theta^k(1-\theta)^{n-k}$$

   Обозначим $Q_k = \sum\limits_{\#1=k} U(\varepsilon_1, \ldots, \varepsilon_n)$, тогда

   $$\mathbb{E}U = \sum_{k=0}^{n} Q_k\,\theta^k(1-\theta)^{n-k} \quad \longrightarrow \quad \frac{\mathbb{E}U}{(1-\theta)^n} = \sum_{k=0}^{n} Q_k \left(\frac{\theta}{1-\theta}\right)^k$$

   И это должно быть верно для любого $\theta \in (0;1)$. Если опустить Q_k, то псевдомногочлен в правой части пробегает все положительные значения в силу стремления $1-\theta$ к нулю при $\theta \to 1$. Таким образом, тождественность нулю имеет место быть только при условии, что $Q_k = 0$ при всех k. Иными словами,

   $$\mathbb{E}U = 0 \quad \leftrightarrow \quad Q_k = 0 \quad \leftrightarrow \quad \sum_{\#1=k} U(\varepsilon_1, \ldots, \varepsilon_n) = 0$$

   Таким образом, явными претендентами на оптимальность среди всех оценок будут такие оценки, что они зависят от числа единиц:

   $$\mathbb{E}(\overline{X}\,U) = \sum_{k=0}^{n} \left(\frac{k}{n} \left(\sum_{\#1=k} U(\varepsilon_1, \ldots, \varepsilon_n) \right) \theta^k(1-\theta)^{n-k} \right) = \sum_{k=0}^{n} \left(\frac{k}{n} \cdot 0 \cdot \theta^k(1-\theta)^{n-k} \right) = 0$$

   По ранее доказанному утверждению, $\overline{X}$ является оптимальной оценкой.

И ещё один пример

Рассмотрим ситуацию, когда оптимальной оценки не существует.
Для этого возьмём следующее распределение:

X	-1	0	1	2	$\ldots$	k	$\ldots$
$\mathbb{P}$	p	q^2	pq^2	p^2q^2	$\ldots$	p^kq^2	$\ldots$

Пусть мы хотим оценить параметр p, и пусть выборка состоит из одной случайной величины X_1.
Рассмотрим оценку

$$\tilde{\theta}(X_1) = \begin{cases} 1, & X_1 = -1 \\ 0, & \text{иначе} \end{cases}$$

Видно, что $\mathbb{E}\tilde{\theta} = p$ — несмещённость. Мы знаем, что все остальные несмещённые оценки имеют вид $\tilde{\theta} + U$, где U — несмещённая оценка нуля. В этом случае,

$$0 = \mathbb{E}U = U(-1)\,p + \sum_{k=0}^{\infty} U(k)\,p^k q^2 \text{ для всех } p.$$

Помня, что $q = 1 - p$, преобразуем равенство:

$$\frac{U(-1)\,p}{(1-p)^2} = -\sum_{k=0}^{\infty} U(k)\,p^k$$

Распишем левую часть как степенной ряд:

$$[U(-1)\,p]\,\frac{1}{(1-p)^2} = [U(-1)\,p]\,\sum_{k=1}^{\infty} k\,p^{k-1} = \sum_{k=1}^{\infty} U(-1)\,k\,p^k = -\sum_{k=0}^{\infty} U(k)\,p^k$$

Заметим, что $U(0) = 0$, так как в левой части нет p^0, тогда

$$\sum_{k=1}^{\infty} U(-1)\,k\,p^k = -\sum_{k=1}^{\infty} U(k)\,p^k \quad \to \quad U(k) = -k\,U(-1)$$

Обозначим $a = U(-1)$. Таким образом, все несмещённые оценки нуля имеют вид $U(k) = ak$.
Возвращаясь к исходной оценке, получаем, что все несмещённые оценки p имеют вид $\tilde{\theta}(X_1) + a\,X_1$.
Для оптимальности необходима минимальность дисперсии:

$$\mathbb{D}\left(\tilde{\theta}(X_1) + a\,X_1\right) = \mathbb{E}\left(\tilde{\theta}(X_1) + a\,X_1\right)^2 - \left[\mathbb{E}\left(\tilde{\theta}(X_1) + a\,X_1\right)\right]^2 = \mathbb{E}\left(\tilde{\theta}(X_1) + a\,X_1\right)^2 - p^2 \geqslant 0$$

Таким образом, минимальность дисперсии эквивалентна минимальности ожидания квадрата:

$$\mathbb{E}\left(\tilde{\theta}(X_1) + a\,X_1\right)^2 \to \min \text{ для всех } p.$$

Регулировать мы можем только a, поэтому рассмотрим, к примеру, случай, когда $X_1 = -1$:

$$\mathbb{E}\left(\tilde{\theta}(-1) - a\right)^2 = (1-a)^2\,p + \sum_{k=0}^{\infty} a^2k^2q^2p^k$$

Преобразуем к квадратному трёхчлену:

$$a^2 \left(p + \sum_{k=0}^{\infty} k^2 q^2 p^k \right) - 2ap + p \to \min$$

Минимум параболы достигается в точке

$$a = \frac{p}{p + \sum\limits_{k=0}^{\infty} k^2 q^2 p^k}$$

Видно, что a зависит от параметра p, следовательно, и исходная оценка зависит от параметра, но это противоречит определению оценки.

Билет №3.

Введение

В этом билете рассматриваются дискретные распределения — случаи, когда случайные величины принимают конечное(-счётное) число значений.

Рассмотрим случайную величину, имеющую следующее распределение:

X	k_1	$\dots$	k_m
$\mathbb{P}$	$p_\theta(k_1)$	$\dots$	$p_\theta(k_m)$

Пусть мы хотим оценить параметр θ. Необходимо принять во внимание тот факт, что $p_\theta(k) = \mathbb{P}(X = k)$, причём k может быть любым конечным числом (необязательно целым).

Пусть дана простая выборка $X_1, \dots, X_n$ со значениями $k_1, \dots, k_n$ (необязательно различными). Посчитаем вероятность этого события (события, при котором получены именно эти значения):

$$\mathbb{P}_\theta(X_1 = k_1, \dots, X_n = k_n) = p_\theta(k_1) \cdots p_\theta(k_n) \text{ (в силу независимости)}$$

Def. *Функцию выше называют* ***функцией правдоподобия****.*

Def. *Функцию* $L(\theta, k_1, \dots, k_n) = \ln\left(p_\theta(k_1) \cdots p_\theta(k_n)\right)$ *называют* ***логарифмической функцией правдоподобия****.*

Также она может принять вид $L(\theta, X_1, \dots, X_n) = \ln\left(\mathbb{P}_\theta(X_1) \cdots \mathbb{P}_\theta(X_n)\right)$.

Информация Фишера для дискретных распределений

Def. ***Информацией Фишера****, пришедшей от* **всей выборки** *с дискретным распределением, называют ожидание* $\mathbb{E}\left(\dfrac{\partial L(\theta, X_1, \dots, X_n)}{\partial \theta}\right)^2 = \mathbb{E}\left(\dfrac{\mathbb{P}'_\theta(X_1)}{\mathbb{P}_\theta(X_1)} + \dots + \dfrac{\mathbb{P}'_\theta(X_n)}{\mathbb{P}_\theta(X_n)}\right)^2$.

Посчитаем ожидание для одной случайной величины:

$$\mathbb{E}\left(\frac{\mathbb{P}'_\theta(X_1)}{\mathbb{P}_\theta(X_1)}\right) = \sum_i \frac{p'_\theta(k_i)}{p_\theta(k_i)} \cdot p_\theta(k_i) = \sum_i p'_\theta(k_i) = \left(\sum_i p_\theta(k_i)\right)' = 0, \text{ так как } p_\theta(k_i) \text{ – числа.}$$

Так как величины имеют нулевое ожидание и независимы в совокупности, имеем

$$\mathbb{E}\left(\frac{\partial L(\theta, X_1, \dots, X_n)}{\partial \theta}\right)^2 = \mathbb{E}\left(\frac{\mathbb{P}'_\theta(X_1)}{\mathbb{P}_\theta(X_1)} + \dots + \frac{\mathbb{P}'_\theta(X_n)}{\mathbb{P}_\theta(X_n)}\right)^2 = \mathbb{D}\left(\frac{\mathbb{P}'_\theta(X_1)}{\mathbb{P}_\theta(X_1)} + \dots + \frac{\mathbb{P}'_\theta(X_n)}{\mathbb{P}_\theta(X_n)}\right)$$
$$= n\,\mathbb{D}\left(\frac{\mathbb{P}'_\theta(X_1)}{\mathbb{P}_\theta(X_1)}\right) = n\,\mathbb{E}\left(\frac{\mathbb{P}'_\theta(X_1)}{\mathbb{P}_\theta(X_1)}\right)^2 = n\,\mathbb{E}\left(\frac{\partial}{\partial \theta} \ln \mathbb{P}_\theta(X_1)\right)^2$$

Def. ***Информацией Фишера****, пришедшей от* **одной случайной величины** *с дискретным распределением, называют ожидание* $I(\theta) = \mathbb{E}\left(\dfrac{\partial}{\partial \theta} \ln \mathbb{P}_\theta(X_1)\right)^2$.

Неравенство Рао-Крамера

Th. *Пусть $\tilde{\theta}$ — несмещённая оценка, тогда для неё справедливо неравенство*

$$\mathbb{D}\tilde{\theta} \geqslant \frac{1}{n\,I(\theta)},\ \text{где } I(\theta) \text{ — информация Фишера.}$$

Доказательство. Пусть случайные величины принимают какие-либо значения $k_1, \ldots, k_n$. Обозначим $K = (k_1, \ldots, k_m)$, $\tilde{\theta}(K) = \tilde{\theta}(k_1, \ldots, k_n), \mathbb{P}_\theta(K) = p_\theta(k_1) \cdots p_\theta(k_n)$ для удобства.

1. По условию, $\tilde{\theta}$ — несмещённая оценка:

$$\mathbb{E}\tilde{\theta} = \sum_K \tilde{\theta}(K)\,\mathbb{P}_\theta(K) = \theta$$

2. Сумма всех вероятностей равна единице:

$$\sum_K \mathbb{P}_\theta(K) = 1$$

3. Продифференцируем оба равенства по θ:

$$\frac{\partial}{\partial\theta}\sum_K \tilde{\theta}(K)\,\mathbb{P}_\theta(K) = \frac{\partial}{\partial\theta}\,\theta = 1, \qquad \frac{\partial}{\partial\theta}\sum_K \mathbb{P}_\theta(K) = \frac{\partial}{\partial\theta}\,1 = 0$$

4. Умножим второе на θ и вычтем из первого:

$$\sum_K \left(\tilde{\theta}(K) - \theta\right)\,\mathbb{P}'_\theta(K) = 1$$

5. Умножим и поделим на $\mathbb{P}_\theta(K)$:

$$\sum_K \left(\tilde{\theta}(K) - \theta\right)\,\frac{\mathbb{P}'_\theta(K)}{\mathbb{P}_\theta(K)}\,\mathbb{P}_\theta(K) = 1$$

6. Методом пристального взгляда заметим, что

$$\sum_K \left(\tilde{\theta}(K) - \theta\right)\,\frac{\mathbb{P}'_\theta(K)}{\mathbb{P}_\theta(K)}\,\mathbb{P}_\theta(K) = \mathbb{E}\left[\left(\tilde{\theta} - \theta\right)\cdot\frac{\partial}{\partial\theta}L\right] = 1$$

7. Применим неравенство Коши-Буняковского:

$$\sqrt{\mathbb{D}\tilde{\theta}}\cdot\sqrt{\mathbb{E}\left(\frac{\partial}{\partial\theta}L\right)^2} \geqslant 1$$

8. Преобразуем:

$$\mathbb{D}\tilde{\theta} \geqslant \frac{1}{n\,I(\theta)}$$

[:|||:]

Если же выполняется равенство, то есть $\mathbb{D}\tilde{\theta} = \dfrac{1}{n\,I(\theta)}$, то мы можем выразить $\mathbb{P}_\theta(X)$.
Если в неравенстве Коши-Буняковского выполняется равенство, то это означает линейную зависимость между случайными величинами. В нашем случае, ими будут $\left(\tilde{\theta} - \theta\right)$ и $\dfrac{\partial}{\partial\theta}L$:

$$\tilde{\theta}(X) - \theta = C(\theta)\,\frac{\partial}{\partial\theta}\ln p_\theta(X),\ \text{где } C(\theta) \text{ — какая-то функция от } \theta.$$

Поделим на $C(\theta)$ и проинтегрируем по θ:

$$\ln p_\theta(X) = \alpha(\theta)\,\tilde{\theta}(X) + \beta(\theta) + \gamma(X)$$

Применим экспоненту:

$$p_\theta(X) = \exp\left(\alpha(\theta)\,\tilde{\theta}(X) + \beta(\theta)\right) \cdot \exp\left(\gamma(X)\right)$$

Следовательно, если в неравенстве Рао-Крамера достигнуто равенство, распределение устроено конкретным образом:

$$p_\theta(X) = g\left(\theta, \tilde{\theta}\right) \cdot h(X)$$

Билет №4.

Асимптотическая нормальность

Def. *Оценка $\widehat{\theta}$ называется* ***асимптотически нормальной****, если*

$$\frac{\sqrt{n}}{\sigma(\theta)}\,(\widehat{\theta}-\theta) \xrightarrow{d} \xi \sim \mathcal{N}(0;1), \text{ где } \sigma(\theta) \text{ — параметр.}$$

St. *Асимптотически нормальные оценки состоятельны.*

Доказательство. Обозначим $\eta_n = \dfrac{\sqrt{n}}{\sigma(\theta)}\,(\widehat{\theta}-\theta)$, тогда $(\widehat{\theta}-\theta) = \dfrac{\sigma(\theta)}{\sqrt{n}}\,\eta_n$.
Заметим, что $\dfrac{\sigma(\theta)}{\sqrt{n}} \to 0$, а $\eta_n \xrightarrow{d} \xi \sim \mathcal{N}(0;1)$. Следовательно, $(\widehat{\theta}-\theta) \xrightarrow{d} 0\cdot\xi$.
Таким образом, $\widehat{\theta} \xrightarrow{\mathbb{P}} \theta$. **[:|||:]**

Методы стабилизации асимптотической дисперсии

Несколько способов избавиться от зависимости $\sigma(\theta)$ от θ:

1. если $\sigma(\theta)$ непрерывна и не равна 0, можем заменить θ на $\widehat{\theta}$.
2. применим теорему о непрерывности из пятого билета предыдущего коллоквиума:

$$\xi_n = \frac{\sqrt{n}}{\sigma(\theta)}\,(\widehat{\theta}-\theta), \quad \xi \sim \mathcal{N}(0;1), \quad a = \theta, \quad b_n = \frac{\sigma}{\sqrt{n}}$$

Попробуем оценить не θ, а $h(\theta)$, где h — некоторая функция:

$$h\left(\theta + \frac{\sigma}{\sqrt{n}}\,\xi_n\right) = h\left(\widehat{\theta}\right), \quad \frac{\sqrt{n}}{\sigma\, h'(\theta)}\left(h\left(\widehat{\theta}\right) - h(\theta)\right) = \frac{1}{h'(\theta)}\left(\frac{h\left(\theta + \frac{\sigma}{\sqrt{n}}\,\xi_n\right) - h(\theta)}{\frac{\sigma}{\sqrt{n}}}\right) \xrightarrow[\text{Теорема}]{d} \xi$$

Теперь подберём h так, что $\sigma(\theta)\,h'(\theta) = 1$:

$$h(t) = \int \frac{dt}{\sigma(t)}$$

Оценка вероятности успеха в схеме Бернулли

Пусть дана выборка $X_1, \ldots, X_n$:

$$X_k = \begin{cases} 1, & \theta \in (0;1) \\ 0, & 1-\theta \end{cases}$$

В качестве оценки θ возьмём выборочное среднее: $\widehat{\theta} = \overline{X}$, — тогда имеем

1. $\mathbb{E}\widehat{\theta} = \mathbb{E}X_k = \theta$ — несмещённость.
2. Асимптотическую нормальность, согласно ЦПТ (или же теореме Муавра-Лапласа):

$$\frac{\sqrt{n}}{\sqrt{\theta\,(1-\theta)}}\,(\widehat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0;1)$$

   Вместо θ можем

   a. подставить $\widehat{\theta}$.

   b. посчитать и подставить несложный интеграл:

$$h(t) = \int \frac{dt}{\sqrt{t\,(1-t)}} = \arcsin(2t-1)$$

3. Состоятельность, как следствие.
4. Оптимальность, согласно несмещённым оценкам нуля (пример из второго билета).

Билет №5.

Метод моментов

Пусть дана выборка $X_1, \dots, X_n$ с распределением $F_\theta(t)$. Возьмём некоторую функцию $g(x)$ и посчитаем ожидание:

$$\mathbb{E}g(X_1) = f(\theta)$$

Предположим, что определена обратная функция f^{-1} и что она непрерывна.
Воспользовавшись ЗБЧ, имеем

$$\frac{g(X_1) + \dots + g(X_n)}{n} \xrightarrow{\mathbb{P}} f(\theta)$$

Применим обратную функцию и теорему о непрерывности:

$$f^{-1}\left(\frac{g(X_1) + \dots + g(X_n)}{n}\right) \xrightarrow{\mathbb{P}} \theta$$

В качестве оценки θ можем взять $\widehat{\theta} = f^{-1}\left(\dfrac{g(X_1) + \dots + g(X_n)}{n}\right)$.

Состоятельность и асимптотическая нормальность метода моментов

Состоятельность оценки, полученной методом моментов, следует из ЗБЧ.
Покажем асимптотическую нормальность: обозначим $h = f^{-1}$, $\overline{g} = \dfrac{g(X_1) + \dots + g(X_n)}{n}$.
Из ЦПТ следует, что

$$\xi_n = \frac{\sqrt{n}}{\sigma}\left(\overline{g} - f(\theta)\right) \xrightarrow{d} \mathcal{N}(0;1), \text{ где } \sigma = \sqrt{\mathbb{D}g(X_1)}.$$

Далее,

$$h(\overline{g}) = h\left(f(\theta) + \frac{\sigma}{\sqrt{n}}\,\xi_n\right), \qquad h(f(\theta)) = \theta$$

Теперь приведём к виду из теоремы:

$$\frac{\sqrt{n}}{\sigma\, h'(f(\theta))}\left(h(\overline{g}) - \theta\right) = \frac{h\left(f(\theta) + \dfrac{\sigma}{\sqrt{n}}\,\xi_n\right) - h(f(\theta))}{\dfrac{\sigma}{\sqrt{n}}\; h'(f(\theta))} \xrightarrow{d} \mathcal{N}(0;1)$$

В итоге получаем следующее утверждение:

St. *Пусть $f(\theta) = \mathbb{E}g(X_1)$ такие, что f и f^{-1} — непрерывные и дифференцируемые. Пусть также $\mathbb{D}g(X_1) < \infty$, тогда оценка*

$$\widehat{\theta} = f^{-1}\left(\frac{g(X_1) + \dots + g(X_n)}{n}\right)$$

является асимптотически нормальной оценкой θ, причём

$$\frac{\sqrt{n}}{\sqrt{\mathbb{D}g(X_1)}\;\dfrac{1}{f'(\theta)}}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}(0;1)$$

Билет №6.

Энтропия одного распределения относительно другого

Def. ***Информацией по Шеннону*** *или* ***энтропией*** *распределения ρ_1 относительно распределения ρ_0 называют функцию*

$$\widetilde{I} = \int \ln \rho_1(x) \cdot \rho_1(x)\, dx - \int \ln \rho_0(x) \cdot \rho_1(x)\, dx$$

Энтропия оценивает расстояние между двумя распределениями.

Неравенство информации

St. *Пусть ρ_0, ρ_1 — положительные плотности, тогда неравенство*

$$\int \ln \rho_1(x) \cdot \rho_1(x)\, dx \geqslant \int \ln \rho_0(x) \cdot \rho_1(x)\, dx$$

называется ***неравенством информации****, причём равенство достигается тогда и только тогда, когда $\rho_0 = \rho_1$.*

Доказательство. Перенесём всё в правую часть:

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x)\, dx \quad \vee \quad 0$$

Применим неравенство $\ln x \leqslant x - 1$:

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x)\, dx \leqslant \int \left(\frac{\rho_0(x)}{\rho_1(x)} - 1 \right) \rho_1(x) dx = \int (\rho_0(x) - \rho_1(x))\, dx \quad \vee \quad 0$$

Заметим, что

$$\int (\rho_0(x) - \rho_1(x))\, dx = \int \rho_0(x)\, dx - \int \rho_1(x)\, dx = 1 - 1 = 0$$

Тогда

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x)\, dx \leqslant \int \rho_0(x)\, dx - \int \rho_1(x)\, dx = 0$$

Теперь пусть достигнуто равенство, тогда по всё тому же неравенству $x - 1 - \ln x \geqslant 0$ имеем

$$\int \left(\frac{\rho_0(x)}{\rho_1(x)} - 1 - \ln \frac{\rho_0(x)}{\rho_1(x)} \right) \rho_1(x) dx = 0$$

Равенство в $x - 1 - \ln(x - 1) \geqslant 0$ достигается в одной точке $x = 1$, следовательно,

$$\frac{\rho_0(x)}{\rho_1(x)} = 1 \quad \longrightarrow \quad \rho_0 = \rho_1$$

[:|||:]

Метод максимального правдоподобия

Пусть дана выборка $X_1, \dots, X_n$ с распределением ρ_θ. Пусть θ_0 — истинное значение параметра, то есть, на самом деле, X_k распределены с плотностью ρ_{θ_0}. Тогда максимум **истинной функции правдоподобия**

$$W(\theta) = \mathbb{E}_{\theta_0} \ln \rho_\theta(X_1) = \int \ln \rho_\theta(x) \cdot \rho_{\theta_0}(x)\, dx$$

достигается только в точке $\theta = \theta_0$, согласно неравенству информации. Также, величина

$$W(\theta_0) - W(\theta) = \int \ln \frac{\rho_{\theta_0}(x)}{\rho_\theta(x)} \cdot \rho_{\theta_0}(x)\, dx$$

показывает количество информации, которое дало единичное наблюдение.

Состоятельность оценки, полученной методом максимального правдоподобия

Согласно ЗБЧ,

$$\frac{1}{n} \sum_j \ln \rho_\theta(X_j) \xrightarrow{\mathbb{P}} W(\theta), \qquad \text{где } \sum_j \ln \rho_\theta(X_j) = L(\theta, X).$$

Логично предположить, что точки максимумов функции приближаются друг к другу.

St. *Пусть $\theta \in (\alpha; \beta)$ и на этом интервале функция $\theta \longmapsto L(\theta, X)$ имеет единственную точку локального максимума $\widehat{\theta}$, тогда $\widehat{\theta}$ — состоятельная оценка максимума θ_0 функции $W(\theta)$, полученная методом максимального правдоподобия.*

Доказательство. Зафиксируем маленькое $\delta > 0$, тогда

$$W(\theta_0 - \delta) < W(\theta_0) > W(\theta_0 + \delta)$$

$(*)$ При достаточно больших n с вероятностью, близкой к единице, имеем $\dfrac{1}{n} L(\theta, X) \xrightarrow{\mathbb{P}} W(\theta)$ и

$$W(\theta_0 - \delta) \;\xleftarrow{\mathbb{P}}\; \frac{1}{n} L(\theta_0 - \delta, X) \;<\; \frac{1}{n} L(\theta_0, X) \;>\; \frac{1}{n} L(\theta_0 f + \delta, X) \;\xrightarrow{\mathbb{P}}\; W(\theta_0 + \delta)$$

Таким образом, на отрезке $[\theta_0 - \delta; \theta_0 + \delta]$ есть локальный максимум, то есть $|\widehat{\theta} - \theta_0| < \delta$.
Тогда $\mathbb{P}(|\widehat{\theta} - \theta_0| \geqslant \delta) \to 0$, или же $\widehat{\theta}$ состоятельна. **[:|||:]**

Пояснение

В месте, содержащем $(*)$, написана вода, которую нам нужно разгрести — покажем, что

$$\mathbb{P}\left(\frac{1}{n}L(\theta_0-\delta,X)<\frac{1}{n}L(\theta_0,X)\right)\to 1 \quad \text{(случай с } \theta_0+\delta \text{ аналогичен)}$$

Обозначим $W(\theta_0)-W(\theta_0-\delta)=\varepsilon>0$.
Рассмотрим исход, попадающий в событие $A=\left\{\frac{1}{n}L(\theta_0-\delta,X)\geqslant\frac{1}{n}L(\theta_0,X)\right\}$.
Пусть $\frac{1}{n}L(\theta_0-\delta,X)$ отличается от $W(\theta_0-\delta)$ меньше, чем на $\frac{\varepsilon}{10}$,
и пусть $\frac{1}{n}L(\theta_0,X)$ отличается от $W(\theta_0)$ меньше, чем на $\frac{\varepsilon}{10}$, тогда

$$W(\theta_0)\leqslant\frac{1}{n}L(\theta_0,X)+\frac{\varepsilon}{10}\leqslant\frac{1}{n}L(\theta_0-\delta,X)+\frac{\varepsilon}{10}\leqslant W(\theta_0-\delta)+\frac{\varepsilon}{10}+\frac{\varepsilon}{10}\leqslant W(\theta_0-\delta)+\frac{\varepsilon}{5}$$

Таким образом, $\varepsilon=W(\theta_0)-W(\theta_0-\delta)\leqslant\frac{\varepsilon}{5}$ — противоречие. Следовательно, на этом исходе

1. либо $\frac{1}{n}L(\theta_0-\delta,X)$ отличается от $W(\theta_0-\delta)$ больше, чем на $\frac{\varepsilon}{10}$,
2. либо $\frac{1}{n}L(\theta_0,X)$ отличается от $W(\theta_0)$ больше, чем на $\frac{\varepsilon}{10}$.

Также оба высказывания могут быть справедливы одновременно. Переформулировав, получим

$$\mathbb{P}\left(\left|\frac{1}{n}L(\theta_0-\delta,X)-W(\theta_0-\delta)\right|\geqslant\frac{\varepsilon}{10}\right)\xrightarrow[\text{ЗБЧ}]{\mathbb{P}}0, \qquad \mathbb{P}\left(\left|\frac{1}{n}L(\theta_0,X)-W(\theta_0)\right|\geqslant\frac{\varepsilon}{10}\right)\xrightarrow[\text{ЗБЧ}]{\mathbb{P}}0$$

Следовательно, $\mathbb{P}(A)\xrightarrow{\mathbb{P}}0$, что даёт нам $\mathbb{P}\left(\frac{1}{n}L(\theta_0-\delta,X)<\frac{1}{n}L(\theta_0,X)\right)\to 1$.

Билет №7.

Доверительные интервалы

Пусть даны: две статистики θ_1 и θ_2, простая выборка $X_1, \ldots, X_n$ с распределением $F_\theta(t)$ и

$$\mathbb{P}\left(\theta_1 < \theta < \theta_2\right) = 1 - \alpha$$

Def. *Интервал* $(\theta_1; \theta_2)$ *называют* ***доверительным интервалом***.

Def. *Число* $(1-\alpha)$ *называют* ***доверительной вероятностью***.

Построение доверительных интервалов

С помощью центральной функции

Def. *Функция* $g(X_1, \ldots, X_n, \theta)$, *распределение которой не зависит от* θ, *называется* ***центральной функцией***.

Пусть центральная функция непрерывна и строго монотонна, тогда мы можем зажать её в какой-то интервал с заданной вероятностью:

$$\mathbb{P}\left(A < g(X, \theta) < B\right) = 1 - \alpha$$

Отсюда же можем выразить интервал на θ:

$$\mathbb{P}(g^{-1}(A) < \theta < g^{-1}(B)) = 1 - \alpha$$

С помощью асимптотически нормальной оценки

Пусть оценка $\widehat{\theta}$ асимптотически нормальна, то есть

$$\frac{\sqrt{n}}{\sigma}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}(0; 1)$$

Тогда

$$\mathbb{P}\left(\widehat{\theta} - \frac{z_\alpha\,\sigma}{\sqrt{n}} < \theta < \widehat{\theta} + \frac{z_\alpha\,\sigma}{\sqrt{n}}\right) \to \int\limits_{-z_\alpha}^{z_\alpha} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1 - \alpha$$

Доверительные интервалы для параметров нормального распределения

Пусть дана выборка $X_1, \ldots, X_n \sim \mathcal{N}(\mu; \sigma^2)$.

Среднее при известной дисперсии

В данном случае мы можем отнормировать выборочное среднее $\overline{X}$:

$$\frac{\sqrt{n}}{\sigma}\left(\overline{X} - \mu\right) \sim \mathcal{N}(0; 1)$$

Тогда доверительным интервалом будет

$$\left(\overline{X} - \frac{z_\alpha\,\sigma}{\sqrt{n}};\ \overline{X} + \frac{z_\alpha\,\sigma}{\sqrt{n}}\right),$$

а доверительной вероятностью — $(1-\alpha) = \Phi(z_\alpha) - \Phi(-z_\alpha)$.

Среднее при неизвестной дисперсии

Вспомним про выборочную дисперсию и её свойство:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \ \xrightarrow{\mathbb{P}} \sigma^2$$

Далее, мы можем найти *асимптотический* доверительный интервал:

$$\frac{\sqrt{n}}{S} \left(\overline{X} - \mu\right) \xrightarrow{d} \mathcal{N}(0; 1)$$

Для точного доверительного интервала понадобится найти распределение $T_{n-1} = \dfrac{\sqrt{n}}{S} \left(\overline{X} - \mu\right)$.
Поделим числитель и знаменатель на σ и обозначим $Y_i = \dfrac{X_i - \mu}{\sigma} \sim \mathcal{N}(0; 1)$:

$$T_{n-1} = \frac{\sqrt{n}}{S} \cdot \frac{\left(\sum\limits_{i=1}^{n} \dfrac{X_i - \mu}{n\,\sigma}\right)}{\dfrac{1}{\sigma}} = \frac{\sqrt{n}}{S} \cdot \frac{\overline{Y}}{\dfrac{1}{\sigma}}$$

Рассмотрим $\dfrac{S^2}{\sigma^2}$ отдельно:

$$\frac{S^2}{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma}\right)^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i \ \left[-\mu + \mu\right] - \overline{X}}{\sigma}\right)^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$$

Собираем всё в кучу:

$$T_{n-1} = \frac{\sqrt{n} \cdot \overline{Y}}{\sqrt{\dfrac{1}{n-1}} \sqrt{\sum\limits_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}}$$

Выборочное среднее $\overline{Y}$ и выборочная дисперсия $\sum\limits_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$ независимы, причём

$$\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 \sim \chi^2_{n-1} \qquad \text{(имеет хи-квадрат распределение)}$$

Обозначим $Z = \sqrt{n} \cdot \overline{Y}$ и $R = \sum\limits_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$, тогда

$$T_{n-1} = \frac{\sqrt{n-1} \cdot Z}{\sqrt{R}}, \ \text{ где } R \sim \chi^2_{n-1} \text{ и } Z \sim \mathcal{N}(0; 1) \text{ независимы.}$$

Def. *Распределение случайной величины T_{n-1} называется* ***распределением Стьюдента*** *с $(n-1)$ степенью свободы.*

Вычислим плотность χ^2_{n-1} распределения $\left(\sum\limits_{i=1}^{n-1} Z_i^2,\ \text{где } Z_i \sim \mathcal{N}(0;1)\right)$, все константы будем выносить в одну общую:

$$\mathbb{P}\left(Z_1^2 + \ldots Z_{n-1}^2 \leqslant t\right) = \frac{1}{(2\pi)^{(n-1)/2}} \int\limits_{x_1^2+\cdots+x_{n-1}^2 \leqslant t} \exp\left(-\frac{x_1^2+\cdots+x_{n-1}^2}{2}\right) dx_1 \ldots dx_{n-1}$$

$$(\text{сферические координаты}) = C\int\limits_0^{\sqrt{t}} r^{n-2} \exp\left(-\frac{r^2}{2}\right) dr = \left[r^2 = v \implies r = \sqrt{v}\right]$$

$$= C\int\limits_0^{t} v^{\frac{n-3}{2}} \exp\left(-\frac{v}{2}\right) dv$$

Таким образом, χ^2_{n-1} распределение обладает плотностью

$$\rho_{\chi^2_{n-1}}(t) = C\, t^{\frac{n-3}{2}} \exp\left(-\frac{t}{2}\right)$$

Найдём распределение T_{n-1} (без объяснений, это кошмар):

$$F_{T_{n-1}}(t) = \mathbb{P}(T_{n-1} \leqslant t) = \mathbb{P}\left(\frac{\sqrt{n-1}\cdot Z}{\sqrt{R}} \leqslant t\right) = \int\limits_{\frac{\sqrt{n-1}\cdot x}{\sqrt{y}} \leqslant t} \rho_Z(x)\,\rho_R(y)\,dx\,dy$$

$$= C\int\limits_0^{+\infty} dy \int\limits_{-\infty}^{\frac{t\cdot\sqrt{y}}{\sqrt{n-1}}} \exp\left(-\frac{x^2}{2}\right) y^{\frac{n-3}{2}} \exp\left(-\frac{y}{2}\right) dx = \left[x = \frac{\sqrt{y}}{\sqrt{n-1}}\, s\right]$$

$$= C\int\limits_0^{+\infty} dy \int\limits_{-\infty}^{t} y^{\frac{n-2}{2}} \exp\left(-\frac{y}{2}\right) \exp\left(-\frac{ys^2}{2(n-1)}\right) ds$$

$$= C\int\limits_0^{+\infty} dy \int\limits_{-\infty}^{t} y^{\frac{n-2}{2}} \exp\left(-\frac{y}{2}\left(1+\frac{s^2}{n-1}\right)\right) ds = \left[y = \frac{w}{(1+\frac{s^2}{n-1})}\right]$$

$$= C\int\limits_{-\infty}^{t} \left(1+\frac{s^2}{n-1}\right)^{-n/2} ds$$

Получаем, что $T_{n-1} = \dfrac{\sqrt{n}}{S}\,(\overline{X}-\mu)$ обладает плотностью

$$\rho_{T_{n-1}}(t) = C\left(1+\frac{t^2}{n-1}\right)^{-n/2}$$

По итогу, мы можем зажать $\dfrac{\sqrt{n}}{S}\,(\overline{X}-\mu)$ симметрично и найти интервал:

$$\mathbb{P}\left(-t_\alpha < \frac{\sqrt{n}}{S}\,(\overline{X}-\mu) < t_\alpha\right) = 1-\alpha \quad\longrightarrow\quad \mu \in \left(\overline{X} - \frac{t_\alpha\,S}{\sqrt{n}};\ \overline{X} + \frac{t_\alpha\,S}{\sqrt{n}}\right)$$

Дисперсия

Мы знаем, что $\dfrac{(n-1)\,S^2}{\sigma^2}$ имеет распределение χ^2_{n-1}, тогда выберем границы A и B так, что

$$\mathbb{P}\left(A < \frac{(n-1)\,S^2}{\sigma^2} < B\right) = 1 - \alpha \quad \longrightarrow \quad \sigma \in \left(\sqrt{\frac{(n-1)\,S^2}{B}};\ \sqrt{\frac{(n-1)\,S^2}{A}}\right)$$

Билет №8.

Проверка гипотез

Пусть дана выборка $X_1, \dots, X_n$ (необязательно независимые) с распределением $F_\theta(t)$. Предположим, что перед нами стоит выбор между двумя значениями: θ_0 и θ_1. Гипотеза H_0 считается основной и предполагает, что $\theta = \theta_0$. Гипотеза H_1 считается противопоставленной первой и предполагает, что $\theta = \theta_1$ (в простом случае).

Проверка происходит следующим образом: в $\mathbb{R}^n$ выбирается область K (критическая область), после чего смотрят, попала ли выборка в эту область или нет. Если да, то H_0 отклонить, а H_1 принять, иначе — отклонить H_1 и принять H_0.

Уровень значимости критерия

Пусть мы отклонили H_0, но H_0 верна. Эта ситуация называется ошибкой I-го рода. Вероятность этой ошибки считается следующим образом:

$$\mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in K) = \alpha \text{ — } \textbf{уровень значимости}.$$

Мощность критерия

Пусть мы приняли H_0, но H_0 — ложь. Эта ситуация называется ошибкой II-го рода. Вероятность этой ошибки считается следующим образом:

$$\mathbb{P}_{\theta_1}((X_1, \dots, X_n) \in \mathbb{R}^n \setminus K) = \beta$$
$$\mathbb{P}_{\theta_1}((X_1, \dots, X_n) \in K) = 1 - \beta \text{ — } \textbf{мощность критерия}.$$

Теорема Неймана-Пирсона

Пусть нам известны совместные плотности ρ_{θ_0} и ρ_{θ_1} выборки $X_1, \dots, X_n$ с соответствующими распределениями F_{θ_0} и F_{θ_1}. Пусть $t > 0$, тогда рассмотрим область

$$K = \left\{ x \in \mathbb{R}^n \;\middle|\; \frac{\rho_{\theta_1}}{\rho_{\theta_0}} \geqslant t \right\}$$

Выбор t происходит за счёт заданного уровня значимости, то есть $\mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in K) = \alpha$.

Th. *Всякий другой критерий имеет меньшую мощность. Иными словами, для любой такой другой области Q, что $\mathbb{P}_{\theta_0}((X_1, \dots, X_n) \in Q) = \alpha$, справедливо неравенство*

$$\mathbb{P}_{\theta_1}((X_1, \dots, X_n) \in Q) \leqslant \mathbb{P}_{\theta_1}((X_1, \dots, X_n) \in K)$$

Доказательство. Для удобства обозначим $\mathbb{P}_{\theta_1}((X_1, \dots, X_n) \in Q) = \mathbb{P}_{\theta_1}(Q)$.
У множеств в неравенстве вычтем пересечения:

$$\mathbb{P}_{\theta_1}(Q \setminus K) \quad \vee \quad \mathbb{P}_{\theta_1}(K \setminus Q)$$

Найдём вероятности с помощью интегрирования:

$$\mathbb{P}_{\theta_1}(Q \setminus K) = \int\limits_{Q\setminus K} \rho_{\theta_1}(x)\,dx$$

Заметим, что мы находимся вне множества K, где $\dfrac{\rho_{\theta_1}}{\rho_{\theta_0}} \geqslant t$, тогда $\rho_{\theta_1} \leqslant t\,\rho_{\theta_0}$ и

$$\begin{aligned}\mathbb{P}_{\theta_1}(Q \setminus K) &= \int\limits_{Q\setminus K} \rho_{\theta_1}(x)\,dx \leqslant t \int\limits_{Q\setminus K} \rho_{\theta_0}(x)\,dx \\ &= t\left(\int\limits_{Q} \rho_{\theta_0}(x)\,dx - \int\limits_{Q\cap K} \rho_{\theta_0}(x)\,dx\right)\end{aligned}$$

Интеграл по Q считает уровень значимости, который совпадает с уровнем значимости критерия K:

$$\begin{aligned}\ldots &= t\left(\int\limits_{Q} \rho_{\theta_0}(x)\,dx - \int\limits_{Q\cap K} \rho_{\theta_0}(x)\,dx\right) = t\left(\alpha - \int\limits_{Q\cap K} \rho_{\theta_0}(x)\,dx\right) \\ &= t\left(\int\limits_{K} \rho_{\theta_0}(x)\,dx - \int\limits_{Q\cap K} \rho_{\theta_0}(x)\,dx\right) = t\int\limits_{K\setminus Q} \rho_{\theta_0}(x)\,dx\end{aligned}$$

Последний интеграл считается на множестве K, где $\dfrac{\rho_{\theta_1}}{\rho_{\theta_0}} \geqslant t$:

$$\ldots = t\int\limits_{K\setminus Q} \rho_{\theta_0}(x)\,dx \leqslant \int\limits_{K\setminus Q} \rho_{\theta_1}(x)\,dx = \mathbb{P}_{\theta_1}(K \setminus Q)$$

Итого:

$$\mathbb{P}_{\theta_1}(Q \setminus K) \quad \leqslant \quad \mathbb{P}_{\theta_1}(K \setminus Q)$$

[:|||:]

Нижняя оценка суммы вероятностей ошибок двух родов

St. *Для уровня значимости α и мощности критерия $1-\beta$ справедливо неравенство*

$$\alpha + \beta \geqslant 1 - \frac{1}{2}\int\limits_{\mathbb{R}^n} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\,dx$$

Доказательство. Распишем α и β как интегралы:

$$\alpha + \beta = \int\limits_{K} \rho_{\theta_0}(x)\,dx + \int\limits_{\mathbb{R}^n\setminus K} \rho_{\theta_1}(x)\,dx \quad \vee \quad 1 - \frac{1}{2}\int\limits_{\mathbb{R}^n} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\,dx$$

Перепишем интеграл для β как разность:

$$\int\limits_{K} \rho_{\theta_0}(x)\,dx + \left(1 - \int\limits_{K} \rho_{\theta_1}(x)\,dx\right) \quad \vee \quad 1 - \frac{1}{2}\int\limits_{\mathbb{R}^n} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\,dx$$

Сократим единицу и перенесём в другую сторону, согласовывая знаки:

$$\frac{1}{2}\int\limits_{\mathbb{R}^n} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\, dx \quad \vee \quad \int\limits_K \rho_{\theta_1}(x)\, dx - \int\limits_K \rho_{\theta_0}(x)\, dx$$

Немного отойдя в сторону, вспомним несложное неравенство из матанализа:

$$\int\limits_K |\rho_{\theta_1}(x) - \rho_{\theta_0}(x)|\, dx \quad \geqslant \quad \int\limits_K \rho_{\theta_1}(x)\, dx - \int\limits_K \rho_{\theta_0}(x)\, dx \quad (*)$$

И ещё одно:

$$\int\limits_{\mathbb{R}^n\setminus K} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\, dx \quad \geqslant \quad \int\limits_{\mathbb{R}^n\setminus K} \rho_{\theta_0}(x)\, dx - \int\limits_{\mathbb{R}^n\setminus K} \rho_{\theta_1}(x)\, dx$$

Перепишем как разность:

$$\int\limits_{\mathbb{R}^n\setminus K} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\, dx \quad \geqslant \quad \left(1 - \int\limits_K \rho_{\theta_0}(x)\, dx\right) - \left(1 - \int\limits_K \rho_{\theta_1}(x)\, dx\right)$$

Получаем

$$\int\limits_{\mathbb{R}^n\setminus K} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\, dx \quad \geqslant \quad \int\limits_K \rho_{\theta_1}(x)\, dx - \int\limits_K \rho_{\theta_0}(x)\, dx \quad (**)$$

Сложим $(*)$ и $(**)$:

$$\int\limits_{\mathbb{R}^n} |\rho_{\theta_0}(x) - \rho_{\theta_1}(x)|\, dx \quad \geqslant \quad 2\left(\int\limits_K \rho_{\theta_1}(x)\, dx - \int\limits_K \rho_{\theta_0}(x)\, dx\right)$$

[:|||:]

Билет №9.

Эмпирическая функция распределения

Def. *Пусть дана выборка $X_1, \dots, X_n$ с функцией распределения $F(t)$, тогда случайная величина*

$$F_n(t) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{Ind}_{\leqslant t}(X_k)$$

называется ***эмпирической функцией распределения****.*

Другими словами, эмпирическая функция распределения подсчитывает такую долю элементов выборки, что все они не превосходят t.

Свойства

Несмещённость и состоятельность

Обозначим $Y_k = \mathrm{Ind}_{\leqslant t}(X_k)$, тогда

$$\mathbb{E}Y_k = \mathbb{E}\,\mathrm{Ind}_{\leqslant t}(X_k) = \mathbb{P}(X_k \leqslant t) = F(t)$$

Согласно ЗБЧ,

$$F_n(t) = \frac{Y_1 + \dots + Y_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}Y_k = F(t)$$

Таким образом, $F_n(t)$ является несмещённой и состоятельной оценки $F(t)$.

Равномерная сходимость по вероятности

St. *Эмпирическая функция сходится по вероятности к функции распределения равномерно:*

$$\sup_t |F_n(t) - F(t)| \xrightarrow{\mathbb{P}} 0$$

Доказательство. Пусть F непрерывна. Разобьём прямую на конечное число n отрезков $\Delta_i = [t_{i-1}; t_i)$ (крайние два — лучи) так, что на каждом из них $F(t_i) - F(t_{i-1}) < \varepsilon$ для любого i и любого наперёд заданного $\varepsilon > 0$ (это можно сделать за счёт непрерывности).
Пусть $t \in [t_{i-1}; t_i]$, тогда

$$F_n(t_{i-1}) - F(t_{i-1}) - \varepsilon \leqslant F_n(t_{i-1}) - F(t_i) \leqslant F_n(t) - F(t) \leqslant F_n(t_i) - F(t_{i-1}) \leqslant F_n(t_i) - F(t_i) + \varepsilon$$

Это справедливо для любого i, поэтому

$$|F_n(t) - F(t)| \leqslant \max_i |F_n(t_i) - F(t_i)| + \varepsilon$$

Теперь перейдём к супремуму:

$$\sup_t |F_n(t) - F(t)| \leqslant \max_i |F_n(t_i) - F(t_i)| + \varepsilon$$

Проверим сходимость по определению:

$$\mathbb{P}\left(\sup_t |F_n(t) - F(t)| \geqslant 2\varepsilon\right) \leqslant \sum_{i=1}^{n} \mathbb{P}\left(|F_n(t_i) - F(t_i)| \geqslant \varepsilon\right) \to 0 \quad \text{по условию.}$$

[:|||:]

Th. *(Гливенко-Кантелли)*

$$\sup_t |F_n(t) - F(t)| \xrightarrow{a.\ s.} 0$$

Независимость распределения супремума разности

St. *Пусть F непрерывна, тогда распределение $D_n = \sup\limits_t |F_n(t) - F(t)|$ не зависит от конкретного распределения X_k.*

Доказательство. Докажем для строго монотонной функции:

$$\begin{aligned} D_n &= \sup_t \left| \frac{1}{n} \sum_{k=1}^{n} \mathrm{Ind}_{\leqslant t}(X_k) - F(t) \right| = \begin{bmatrix} t = F^{-1}(s) \\ s \in [0;1] \end{bmatrix} \\ &= \sup_s \left| \frac{1}{n} \sum_{k=1}^{n} \mathrm{Ind}_{\leqslant F^{-1}(s)}(X_k) - F(F^{-1}(s)) \right| \end{aligned}$$

Рассмотрим индикатор отдельно, обозначив $Y_k = F(X_k)$:

$$\begin{aligned} \mathrm{Ind}_{\leqslant F^{-1}(s)}(X_k) = \mathrm{Ind}_{\leqslant s}(F(X_k)) = \mathrm{Ind}_{\leqslant s}(Y_k) \\ \mathbb{P}(Y_k \leqslant s) = \mathbb{P}(F(X_k) \leqslant s) = \mathbb{P}(X_k \leqslant F_{-1}(s)) = s \end{aligned}$$

Следовательно, $Y_k \sim U[0;1]$, тогда

$$D_n = \sup_s \left| F_n^{Y_k}(s) - s \right|, \quad \text{где } F_n^{Y_k} \text{ — эмпирическая функция для } Y_k.$$

Иными словами, можно считать, что $X_k \sim U[0;1]$. **[:|||:]**

Th. *(Колмогоров)*

$$\mathbb{P}\left(\sqrt{n}\, D_n \leqslant t\right) \to K_n(t) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp\left(-2k^2 t^2\right)$$

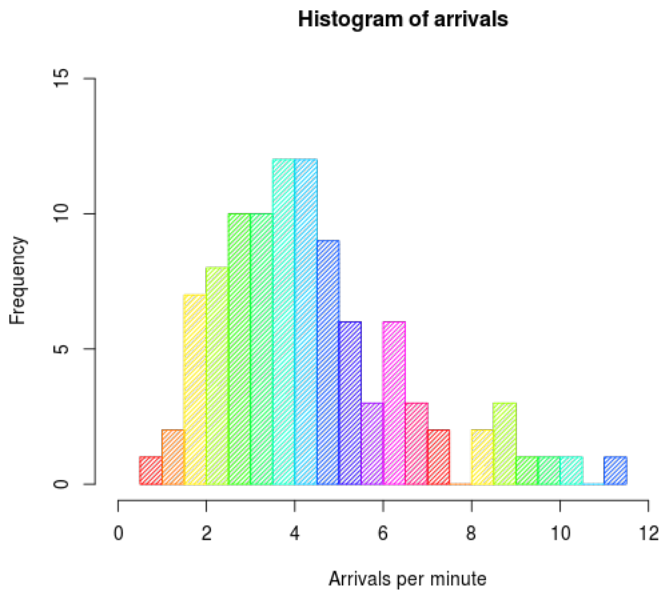
Критерий согласия Колмогорова

Имеется одна гипотеза H_0: распределение выборки согласуется с F. Выбирается такое z_α, что $1 - K(z_\alpha) = \alpha$. Если в этом случае выходит, что $D_n > \dfrac{z_\alpha}{\sqrt{n}}$, то H_0 отклонить.

Гистограмма

Берётся прямая или отрезок, разбивается на конечное число промежутков Δ_i, после чего на каждом из них рисуется столбик высоты ν, который отражает отношение количества значений n_i выборки, попавших на этот промежуток, к её объёму n, умноженному на $|\Delta_i|$:

$$\nu = \frac{n_i}{n \cdot |\Delta_i|}$$



Пример гистограммы.

Мы знаем, что частота (или же доля значений, выпавших на Δ_i) стремится к вероятности:

$$\frac{n_i}{n} \to \mathbb{P}(X_1 \in \Delta_i)$$

Что примерно равно интегралу:

$$\frac{n_i}{n} \to \mathbb{P}(X_1 \in \Delta_i) \simeq \int\limits_{\Delta_i} \rho(x)\,dx = \rho(c)\,|\Delta_i| \quad \text{(по теореме о среднем)}$$

Таким образом, $v = \rho(c)$, то есть на каждом промежутке плотность заменяется на её значение в какой-то точке, принадлежащей этому промежутку. То есть гистограмма является ступенчатым приближением плотности (при учёте, что выборка достаточно большая).

χ^2 критерий Пирсона

Составляется следующая таблица:

	$\ldots$	Δ_i	$\ldots$
Н	$\ldots$	n_i	$\ldots$
О	$\ldots$	$n\,p_i$	$\ldots$

где Δ_i — наблюдаемый промежуток, n_i — количество значений, выпавшее на этот промежуток (наблюдаемое), а $n\,p_i$ — ожидание этого количества при условии, что вероятность попасть на Δ_i равна p_i (ожидаемое). Далее, формируется статистика

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - n\,p_i)^2}{n\,p_i}$$

Оказывается, что с ростом выборки, $\chi^2 \xrightarrow{d} \chi^2_{N-1}$.

Сам критерий Пирсона заключается в следующем: для гипотезы H_0, которая утверждает, что гистограмма согласуется с плотностью, выбирается такое z_α, что вероятность попасть на площадь графика χ^2 распределения на луче $[z_\alpha; +\infty)$ равна α, после чего проверяется неравенство $\chi^2 > z_\alpha$. Если оно верно, то гипотеза отклоняется.

Обоснование состоит в следующем: обозначим $Y_i^k = \mathrm{Ind}_{\Delta_i}(X_k)$, тогда

$$\overline{Y_i} = \frac{Y_i^1 + \cdots + Y_i^n}{n} = \frac{n_i}{n}$$

Посмотрим на ожидание и дисперсию:

$$\mathbb{E}Y_i^1 = p_i, \qquad \mathbb{D}Y_i^1 = p_i\,q_i$$

Тогда, согласно ЦПТ, имеем

$$\frac{\sqrt{n}}{\sqrt{p_i\,q_i}} \left(\overline{Y_i} - p_i\right) \xrightarrow{d} \xi \sim \mathcal{N}(0;1)$$

Перепишем:

$$\frac{1}{\sqrt{q_i}} \frac{n_i - n\,p_i}{\sqrt{n\,p_i}} \xrightarrow{d} \xi \sim \mathcal{N}(0;1)$$

И посмотрим на такую сумму:

$$\sum_{i=1}^{N} \left(\frac{n_i - n\,p_i}{\sqrt{n\,p_i}}\right)^2 = \sum_{i=1}^{N} q_i\,\xi_i^2 \sim \chi^2_{N-1}$$