Теория вероятностей и математическая статистика—2

Винер Даниил @danya_vin
Рейтман Яна

Версия от 10 марта 2025 г.

Содержание

1 Закон больших чисел. Центральная предельная теорема

1.1 Закон больших чисел в форме Бернулли

Пусть имеются некоторые случайные величины $\xi_i = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$, где p — вероятность, что какое-то событие произошло. Тогда $\mathbb{E}[\xi] = p$, $\mathbb{D}[\xi] = p(1-p) \leqslant \frac{1}{4}$

Теорема. Пусть $\hat{p} = \frac{1}{n}\sum\limits_{i=1}^{n} \xi_i$ — доля успехов в n испытаниях Бернулли, тогда $\hat{p} \xrightarrow{p} p$

Доказательство. Распишем по неравенству Чебышёва:

$$\mathbb{P}(|\hat{p} - p| \geqslant \varepsilon) \leqslant \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0$$

Пример

Пусть 87% новорожденных доживают до 50 лет. Тогда $p = 0,87$ — вероятность дожить до 50. Рассмотрим $n = 1000$ новорожденных

Опредедлим с какой вероятностью данная случайная величина отклонится от своего математического ожидания не более, чем на $0,04$ — $\mathbb{P}(|\hat{p} - 0,87| \leqslant 0,04)$. По Чебышёву:

$$\mathbb{P}(|\hat{p} - p| \leqslant 0,04) \geqslant 1 - \frac{\mathbb{D}[\hat{p}]}{(0,04)^2} = 1 - \frac{0,87 \cdot 0,13}{0,0016 \cdot 1000} = 0,929$$

1.2 Центральная предельная теорема

Рассмотрим сумму независимых одинаково распределенных случайных величин:

$$S_n = \xi_1 + \ldots + \xi_n,$$

при этом существует $\mathbb{D}[\xi_i] \leqslant c$, $\mathbb{E}[\xi_i] = \mu$, $\mathbb{D}[\xi_n] = \sigma^2$

Тогда, $Z_n = \dfrac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z$, где $Z \sim \mathcal{N}(0;1)$ — имеет стандартное нормальное распределение

Функция плотности:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

1.3 Теорема Муавра-Лапласа

Теорема. Имеется $\xi_i = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$. $S_n = \sum \xi_i$ — число успехов в n испытаниях. Тогда

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z \sim \mathcal{N}(0;1)$$

Пример

Проходит суд над Бенджамином Споком. Из 300 человек 90 — женщины, которые симпатизируют Споку, при этом 12 присяжных будут судить Спока. Требуется определить мог ли отбор присяжных быть случайным.

Число успехов в данном случае — число женщин среди 300 присяжных. Будем считать, что $p = 0.5$, то есть половина женщин.

$$\mathbb{P}\left(\frac{S_{300} - 150}{\sqrt{0.5 \cdot 0.5 \cdot 300}} \leqslant \frac{90 - 150}{\sqrt{75}}\right) \simeq \Phi(-6.93) \simeq 2.3 \cdot 10^{-12}$$

Значит, практически невозможно случайным образом выбрать 90 или меньше женщин среди 300 присяжных при справедливом распределении, то есть отбор был предвзятым

1.4 Неравенство Берри-Эссена

$$|F_n - \Phi| \leqslant \frac{C_0 \cdot \mathbb{E}\left[|\xi_1 - \mu|^3\right]}{\sigma^3 \sqrt{n}}, \text{ где } \begin{cases} F_n \text{ — функция распределения стандартизированной СВ} \\ C_0 \text{ — константа} \\ \mathbb{E}\left[|\xi_1 - \mu|^3\right] \text{ — третий абсолютный центральный момент} \end{cases}$$

Пример

Пусть имеется $n = 1000$ заключенных договоров страхования с 1 января на 1 год. С вероятностью $p = 0.05$ произойдет страховой случай, выплаты по каждому договору — 2000 у.е. R — резерв страховой компании

Требуется определить какой должен быть размер резерва, чтобы страховая компания выполнила свои обязательства с вероятностью 0.99

$S_n = 2000(\xi_1 + \ldots + \xi_n)$, $\xi_i \sim Bi(p = 0.05)$

$$\mathbb{P}\left(S_n \leqslant R\right) = \mathbb{P}\left(\frac{\sum \xi_i - 0.05 \cdot 1000}{\sqrt{1000 \cdot 0.05 \cdot 0.95}} \leqslant \frac{\frac{R}{2000} - 0.05 \cdot 1000}{\sqrt{1000 \cdot 0.05 \cdot 0.95}}\right) \geqslant 0,99$$

Значит, требуется найти квантиль уровня 0.99. Он равен 2.33, тогда

$$\frac{\frac{R}{2000} - 0.05 \cdot 1000}{\sqrt{1000 \cdot 0.05 \cdot 0.95}} = 2.33 \Longrightarrow R = 132117$$

То есть, для покрытия 99% страховых случаев у страховой компании резерв должен быть размером 132117 у.е. Напротив, для покрытия всех случаев $R = 2000000$

2 Многомерное нормальное распределение—1

2.1 Одномерное нормальное распределение

Определение. Случайная величина имеет нормальное распределение $X \sim \mathcal{N}(\mu, \sigma^2)$, если функция плотности равна

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

2.2 Многомерное нормальное распределение

Определение. Пусть случайные величины $z_1, \ldots, z_n$ независимы и $\sim N(0, 1)$. Тогда $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ имеет многомерное нормальное распределение $N(0, I)$, где I — единичная матрица

Функция плотности:

$$f_Z(z) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} z_i^2} = \frac{1}{(\sqrt{2\pi})^n} e^{-0.5 Z^T Z}$$

Примечание. Пусть $Z \sim N(0, I)$, $A \in \mathrm{Mat}_{k \times n}$ — матрица полного ранга и $k < n$, то есть $\mathrm{rank} A = k$. Тогда

$$Y = AZ + b \sim N(b, AA^T)$$

$$\begin{aligned} f_Y(y) &= \frac{1}{|\det A|} f_Z(A^{-1}(y-b)) \\ &= \frac{1}{(\sqrt{2\pi})^n} \frac{1}{|\det A|} e^{-0.5(y-b)^T (A^{-1})^T A^{-1} (y-b)} \\ &\text{пусть } AA^T = C \\ &= \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\sqrt{|C|}} e^{-0.5(y-b)^T C^{-1} (y-b)} \end{aligned}$$

Определение. Случайная величина $Y \sim N(b, C)$, если

$$f_Y(y) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\sqrt{|C|}} e^{-0.5(y-b)^T C^{-1} (y-b)}$$

Определение. Случайный вектор $Y \sim N(0, C)$, если $\forall a_1, \ldots, a_n$

$$a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n$$

либо $N(0, \dot{})$ либо const

2.3 Свойства многомерного нормального распределения

Пусть $Y \sim N(b, C)$

1. $\mathbb{E}[Y] = b, cov(Y) = C$

   Доказательство. $Y = AZ + b$, $Z \sim N(0, I)$

   $$cov(Y) = \mathbb{E}\left[(AZ + b - \mathbb{E}[AZ + b])(AZ + b - AEZ - b)^T\right] = AcovZA^T = AA^T = C$$

2. Любое линейное невырожденное преобразование многомерного нормаьного дает многомерный нормальный вектор

   $\forall B, a : BY + a \sim N(Bb + a, BCB^T)$

3. $\forall$ подвектор нормального вектора нормален

4. Если $Y \sim N(b, D)$, то его компоненты независимы

   Примечание. Некоррелированность = независимость

   Доказательство.

$$
\begin{aligned}
f_Y(y) &= \frac{1}{(\sqrt{2\pi}^n)} e^{-0.5(y-b)^T D^{-1}(y-b)} \\
&= \frac{1}{(\sqrt{2\pi}^n)} e^{-0.5 \sum \left(\frac{y_i - b_i}{\sigma_i}\right)^2} \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-0.5\left(\frac{y_i-b_i}{\sigma_i}\right)^2}
\end{aligned}
$$

Пример. $Y_1 \sim N(0,1),\ \lambda = \begin{cases} 1, & p = 0.5 \\ -1, & p = 0.5 \end{cases},\ Y_2 = 2Y_1$

$$
\begin{aligned}
\mathbb{P}(Y_2 \leqslant y) &= \mathbb{P}(Y_1 \leqslant y | \alpha = 1) \cdot \mathbb{P}(\alpha = 1) + \mathbb{P}(-Y_1 \leqslant y | \alpha = -1) \cdot \frac{1}{2} \\
&= \Phi(y)
\end{aligned}
$$

$cov(Y_1, Y_2) = cov(Y_1, 2Y_1) = \mathbb{E}\left[\alpha Y_1^2\right] - \mathbb{E}[Y]\,\mathbb{E}[\alpha Y_1] = 0$. То есть они не коррелированы

2.4 Условное нормальное распределение

Имеется случайный вектор $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, пишут $\Phi_2(z_1, z_2; \rho)$

Допустим, что z_1 фиксирован, тогда $z_2 | z_1 = z \sim N(\rho z, 1 - \rho^2)$

$z_2 = \rho z_1 + u$, где z_1 и u независимы и $u \sim N(.,.)$

3 Многомерное нормальное распределение—2

3.1 Условное нормальное распределение

$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, пишут $\Phi_2(z_1, z_2; \rho)$

Допустим, что z_1 фиксирован, тогда $z_2|z_1 = z \sim N(\rho z, 1-\rho^2)$

Утверждение. $z_2 = \rho z_1 + u$, где z_1 и u независимы и $(u, z_1) \sim N(.,.)$

Доказательство. $u = z_1 - \rho z_1 \Longrightarrow (z_1, u) = (z_1, z_2 - \rho z_1) = \begin{pmatrix} 1 & 0 \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1-\rho^2 \end{pmatrix}\right)$

$cov(z_1, u) = A \cdot cov(z_1, z_2) \cdot A^T = \begin{pmatrix} 1 & 0 \\ 0 & 1-\rho^2 \end{pmatrix}$ $\square$

Доказательство свойства. $\mathbb{E}[z_2] = \rho\mathbb{E}[z_1] + \mathbb{E}[u] = 0$, $\mathbb{D}[z_2] = \rho^2\mathbb{D}[z_1] + \mathbb{D}[u] = \rho^2 + 1 - \rho^2 = 1$

$(z_2|z_1 = z) = \rho z + u \Longrightarrow \begin{cases} \mathbb{E}[z_2|z_1 = z] = \rho z + \mathbb{E}[u] = \rho z \\ \mathbb{D}[z_2|z_1 = z] = \mathbb{D}[u] = 1 - \rho^2 \end{cases}$ $\square$

Примечание. Пусть вектор Y такой, что $AY \sim N(.,.)$ (многомерное нормальное), меньшей размерности, чем Y, тогда говорят, что Y имеет обобщенное нормальное распределение

Примечание. Двумерная Гауссова копула представима в виде $\Phi_2(\Phi^{-1}(F_1(u_1)), \Phi^{-1}(F_2(u_2)); \rho)$

3.2 Многомерная центральная предельная теорема

Теорема. Пусть $\xi_1^{(1)}, \ldots, \xi_n^{(n)}$ — последовательность независимых одинаково распределенных случайных векторов, у каждого из которых $\mathbb{E}\left[\xi^{(k)}\right] = b\ \forall k$, $cov(\xi^{(k)}) = c$, $\det C > 0$.

Обозначим $S_n = \xi_1^{(1)} + \ldots + \xi_n^{(n)}$ — вектор частичных сумм. Тогда, при $n \to \infty$ последовательность $\eta^{(n)}$, где $\eta^{(n)} = \frac{S_n - nb}{\sqrt{n}}$ сходится по распределению к вектору $\eta \sim N\left(\overrightarrow{0}, C\right)$

4 TBA

5 Введение в математическую статистику

- Имеется n независимых случайных величин $X_1, \ldots, X_n$, которые имеют одинаковые функции распределения: $F_{X_1}(x) = \ldots = F_{X_n}(x) = F(x)$
- Пусть функция распределения $F(x)$ зависит от некоторого вектора неизвестных параметров $\theta = (\theta_1, \ldots, \theta_r)$
- $F(x) = F(x; \theta)$, где x — переменная, а θ — вектор неизвестных параметров
- Θ — множество допустимых значений вектора θ

  Пример. Если $X_i \sim N(\mu, \sigma^2)$, то $\theta = (\mu, \sigma^2) \in (-\infty, +\infty) \times (0; +\infty)$

Определение. Случайной выборкой объема n наблюдений из распределения с функицей распределения $F(x; \theta)$ называется случайный вектор $X = (X_1, \ldots, X_n)$, компоненты которого удовлетворяют следующим условиям

- случайные величины $X_1, \ldots, X_n$ — независимы
- случайные величины $X_1, \ldots, X_n$ имеют одну и ту же функцию распределения $F(x; \theta)$:

$$F_{X_1}(x; \theta) = \ldots = F_{X_n}(x; \theta) = F(x; \theta)$$

Примечание. Продифференцировав эти равенства, получаем, что все функции плотностей распределения равны

Примечание. Если все величины $X_1, \ldots, X_n$ дискретны, то они должны иметь одинаковые таблицы распределения

Примечание. При $i \neq j:\ cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] = 0$, так как X_i и X_j независимы

Имеются случайные величины $X_1(\omega), \ldots, X_n(\omega)$. Пусть произошел вероятностый эксперимент, в результате которого реализовался исход $\omega_0 \in \Omega$. То есть

$$X_1(\omega_0), \ldots, X_n(\omega_0)$$

Тогда, вектор $x = (X_1(\omega_0), \ldots, X_n(\omega_0))$ называется *реализацией случайной выборки*

Пример. $\Omega = \{a, b, c, d, e, f, g, h\},\ \mathcal{F} =$ все подмножества Ω

ω	a	b	c	d	e	f	g	h
$\mathbb{P}(\omega)$	p^3	p^2q	p^2q	pq^2	p^2q	pq^2	pq^2	q^3
$X_1(\omega)$	1	1	1	1	0	0	0	0
$X_2(\omega)$	1	1	0	0	1	1	0	0
$X_3(\omega)$	1	0	1	0	1	0	1	0

- $\mathbb{P}(X_1 = 1) = p^3 + p^2q + p^2q + pq^2 = p(p^2 + pq + pq + q^2) = p \Longrightarrow X_1 \sim Be(p)$
- $\mathbb{P}(X_2 = 1) = p^3 + p^2q + p^2q + pq^2 = p(p^2 + pq + pq + q^2) = p \Longrightarrow X_2 \sim Be(p)$
- $\mathbb{P}(X_3 = 1) = p^3 + p^2q + p^2q + pq^2 = p(p^2 + pq + pq + q^2) = p \Longrightarrow X_3 \sim Be(p)$

$$\mathbb{P}(\{X_1 = 1\} \cap \{X_2 = 1\} \cap \{X_3 = 1\}) = p^3$$
$$\mathbb{P}(X_1 = 1) \cdot \mathbb{P}(X_2 = 1) \cdot \mathbb{P}(X_3 = 1) = p^3$$

Рассуждая аналогично, перебираем оставшиеся 7 случаев и получаем, что X_1, X_2, X_3 — независимы

Пусть $\omega_0 = c$, тогда $(X_1(\omega_0), X_2(\omega_0), X_3(\omega_0)) = (1, 0, 1)$

- Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка
- Для каждого $\omega \in \Omega$ расположим числа $X_1(\omega), \ldots, X_n(\omega)$ в порядке возрастания
- Получим набор чисел $X_{(1)}(\omega) \leqslant \ldots \leqslant X_{(n)}(\omega)$, где (i) означает уже отсортированный номер

  При этом $X_{(1)}(\omega) = \min(X_1(\omega), \ldots, X_n(\omega))$, а $X_{(n)}(\omega) = \max(X_1(\omega), \ldots, X_n(\omega))$

Определение. Набор случайных величин $X_{(1)}, \ldots, X_{(n)}$ называется вариационным рядом

Определение. $\overline{X} := \dfrac{X_1 + \ldots + X_n}{n}$ — выборочное среднее

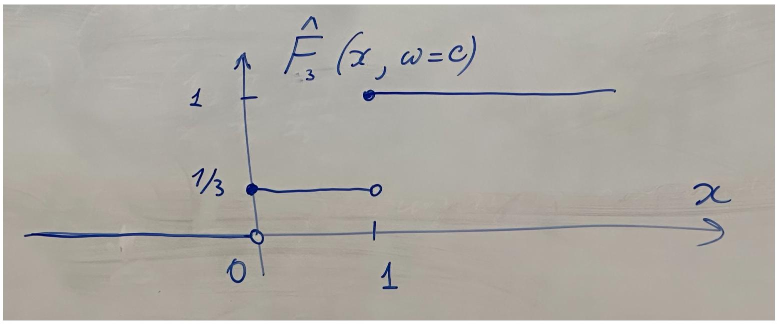
Определение. $s^2 := \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — выборочная дисперсия

Определение. $\widehat{\sigma}^2 := \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — исправленная выброчная дисперсия

Определение. Выборочной функцией распределения называется

$$\begin{aligned}\hat{F}_n(x) &= \frac{\text{число элементов случайной выборки, которые нестрого меньше } x}{n} \\ &= \frac{\#\{i \in \{1, \ldots, n\}:\ X_i \leqslant x\}}{n}\end{aligned}$$

Пример. Рассмотрим $\hat{F}_3(x; \omega = c)$ и выборку $(1, 0, 1)$. Тогда график будет выглядеть так



Утверждение. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка, компоненты которой имеют конечное матожидание. Тогда

$$\mathbb{E}\left[\overline{X}\right] = \mathbb{E}\left[X_i\right]$$

Утверждение. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка, компоненты которой имеют конечные дисперсии. Тогда

$$\mathbb{D}\left[\overline{X}\right] = \frac{\mathbb{D}\left[X_i\right]}{n}$$

Доказательство.

$$\begin{aligned}\mathbb{D}\left[\overline{X}\right] &= \mathbb{D}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n^2}\mathbb{D}\left[\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}\left[X_i\right] \\ &= \frac{1}{n^2} n\mathbb{D}\left[X_i\right] \\ &= \frac{\mathbb{D}\left[X_i\right]}{n}\end{aligned}$$

Определение. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из распределения с функцией распределения $F(x; \theta)$, где θ — неизвестный параметр

Оценкой параметра θ называется случайная величина $\hat{\theta}$, которая является произвольной борелевской функцией от элементов случайной выборки, то есть

$$\hat{\theta} = g(X_1, \ldots, X_n),$$

где $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ — произвольная борелевская функция

Пример. $X = (X_1, \ldots, X_n)$ — случайная выборка из распределения Бернулли с параметром $p \in (0; 1)$

$\hat{p} = \dfrac{X_1 + \ldots + X_n}{n}$ — оценка параметра p

Определение. Оценка $\hat{\theta}$ неизвестного параметра $\theta \in \Theta$ называется *несмещенной*, если

$$\mathbb{E}\left[\hat{\theta}\right] = \theta \ \forall \theta \in \Theta$$

Пример. $X = (X_1, \ldots, X_n)$ — случайная выборка из распределения Бернулли с параметром $p \in (0; 1)$. Определим, является ли $\hat{p} = \frac{X_1 + \ldots + X_n}{n}$ несмещенной оценкой для параметра p

$$\begin{aligned}
\mathbb{E}\left[\hat{p}\right] &= \frac{\mathbb{E}\left[X_1 + \ldots + X_n\right]}{n} \\
&= \frac{\mathbb{E}\left[X_1\right] + \ldots + \mathbb{E}\left[X_n\right]}{n} \\
&= \frac{np}{p} \\
&= p
\end{aligned}$$

То есть $\hat{p}$ — несмещенная оценка

Определение. Говорят, что последовательность случайных величин $(X_n)_{n=1}^{\infty}$ сходится по вероятности к случайной величине X, если

$$\forall \varepsilon > 0 \ \lim_{n \to \infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0$$

Обозначение: $X_n \xrightarrow{\mathbb{P}} X$ при $n \to \infty$ ИЛИ $p \lim\limits_{n \to \infty} X_n = x$

Определение. Последовательность оценок $\hat{\theta}_n$ называется *состоятельной оценкой* неизвестного параметра $\theta \in \Theta$, если

$$\forall \theta \in \Theta \ \hat{\theta}_n \xrightarrow{\mathbb{P}} \theta \text{ при } n \to \infty$$

Пример. Проверим, что $\hat{p} = \dfrac{X_1 + \ldots + X_n}{n}$ является состоятельной. По ЗБЧ:

$$\hat{p} = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\left[X_i\right] = p$$

Значит, $\hat{p}$ — состоятельная оценка

6 Оптимальные оценки, характеристики выборок

6.1 Свойства оптимальных оценок

Определение. Оценкой для параметра будет некоторая функция $\hat{\Theta} = T(X_1, \ldots, X_n)$

Определение. Оценка *состоятельна*, если $\hat{\Theta} \xrightarrow[n\to\infty]{\mathbb{P}} \Theta$

Определение. Оценка *желательно* должна обладать свойством *несмещенности*, то есть $\mathbb{E}\left[\hat{\Theta}\right] = \Theta$

Пример. $X_1, \ldots, X_n$. Построим две оценки

1. $\hat{\mu_1} = \frac{1}{4}(X_1, \ldots, X_4)$
2. $\hat{\mu_2} = \frac{1}{8}X_1 + \frac{1}{8}X_2 + \frac{1}{4}X_3 + \frac{1}{2}X_4$

Первая оценка лучше как минимум потому, что в ней веса каждого наблюдения равны

Определение. Среднеквадратичная ошибка оценки — это

$$\mathbb{E}\left[\hat{\Theta} - \Theta\right]^2$$

$$\begin{aligned}
\mathbb{E}\left[\hat{\Theta} - \Theta\right]^2 &= \mathbb{E}\left[\left(\hat{\Theta} - \mathbb{E}\left[\hat{\Theta}\right]\right) + \underbrace{\left(\mathbb{E}\left[\hat{\Theta}\right] - \Theta\right)}_{b(\Theta)\text{— смещение}}\right]^2 \\
&= \underbrace{\mathbb{E}\left[\hat{\Theta} - \mathbb{E}\left[\hat{\Theta}\right]\right]^2}_{\mathbb{D}[\hat{\Theta}]} + b^2(\Theta) + 2\underbrace{\mathbb{E}\left[\hat{\Theta} - \mathbb{E}\left[\hat{\Theta}\right]\right]}_{=0} \cdot b(\Theta) \\
&= \underbrace{\mathbb{E}\left[\hat{\Theta} - \mathbb{E}\left[\hat{\Theta}\right]\right]^2}_{\mathbb{D}[\hat{\Theta}]} + b^2(\Theta)
\end{aligned}$$

Определение. $\hat{\Theta}_1$ эффективнее, чем $\hat{\Theta}_2$, если

$$\mathbb{E}\left[\hat{\Theta_1} - \Theta\right]^2 \leqslant \mathbb{E}\left[\hat{\Theta_2} - \Theta\right]^2 \text{причем } \exists \Theta_0 :$$
$$\mathbb{E}\left[\hat{\Theta_1} - \Theta_0\right]^2 < \mathbb{E}\left[\hat{\Theta_2} - \Theta_0\right]^2$$

Определение. Оценка $\hat{\Theta}$ называется эффективной (=оптимальной, *the best*) в классе K, если $\forall\, \overline{\Theta} \in K$

$$\mathbb{E}\left[\hat{\Theta} - \Theta\right]^2 \leqslant \mathbb{E}\left[\overline{\Theta} - \Theta\right]^2 \text{причем } \exists \Theta_0 :$$
$$\mathbb{E}\left[\hat{\Theta} - \Theta_0\right]^2 < \mathbb{E}\left[\overline{\Theta} - \Theta_0\right]^2$$

Класс K обладает смещением $b(\Theta)$

Определение. Оценка $\hat{\Theta}$ называется эффективной в классе несмещенных оценок, если $\forall\, \overline{\Theta} \in K$

$$\mathbb{D}\left[\hat{\Theta}\right] \leqslant \mathbb{D}\left[\overline{\Theta}\right], \text{причем } \exists \Theta_0 :$$
$$\mathbb{D}\left[\hat{\Theta}\right] < \mathbb{D}\left[\overline{\Theta}\right] \text{ для } \Theta_0$$

6.2 Способы организации выборки

Определение. Генеральная совокупность — выборка, которая в теории может быть доступна, но на практике вряд ли ее можно собрать. Например, точное население РФ

1. Простой случайный отбор — практически нереализуемо
2. Отбор с помощью механистической процедуры, не влияющий на случайность

   Например, на автоматическом производстве надо посчитать процентр брака. Тогда, для выборки можно ограничиться промежутком времени 14:20-14:30
3. Стратифицированные выборки — выборка производится из страт, то есть берут людей из всех социальных слоев населения
4. Комбинирование

Таким образом, важно следить, чтобы выборка обладали всеми или большинством свойств генеральной совокупности

6.3 Характеристики выборки

1. Вариационный ряд — наблюдения, упорядоченные по возрастанию

   $$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)} \text{— случайные величины}$$

   При этом, $X_{(1)}$ и $X_{(n)}$ — экстремальные статистики
2. Гистограмма и выборочная функция распределения

   Определение. Выборочная (эмпирическая) функция распределения — это

   $$\hat{F}_n(x) = \frac{\sum\limits_{i=1}^{n} I\{X_i \leqslant x\}}{n},$$

   где $I\{X_i \leqslant x\}$ — индикаторная функция, показывающая сколько наблюдений меньше либо равны какого-то x

   Определение. $\hat{p}_i = \dfrac{m(\Delta i)}{|\Delta i| \cdot n}$, где $m(\Delta i)$ — число наблюдений, попавших в интервал Δi, $|\delta i|$ — длина интервала, n — число наблюдений

   Гистограмма — тоже своего рода функция плотности

7 Lecture 24.02.2025

7.1 Характеристики выборок

Определение. Элементы вариационного ряда называются порядковыми статистиками

Определение. Выборочная функция распределения:

$$\widehat{F_n} := \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leqslant x\}$$

Определение. Выборочные (эмпирические) моменты

$$\widehat{\alpha}_k := \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Определение. Выборочное среднее — $\overline{X} = \alpha_1 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

Определение. Выборочная дисперсия — $s^2 := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$

Определение. $\widehat{\sigma}^2 := \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$ — несмещенная (скорректированная) дисперсия

Определение. Выборочный центральный момент k-го порядка

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

Определение. Выборочная ковариация

$$\widehat{cov}(X, Y) := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{X})$$

Определение. Выборочная ковариация

$$\widehat{corr}(X, Y) := \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}}$$

7.2 Свойства эмпиричесой (выборочной) функции распределения

Теорема. Пусть $X_1, \ldots, X_n \sim F(x;\theta) \equiv F(x)$, $\widehat{F}_n = \dfrac{1}{n}\sum I\{X_i \leqslant x\}$, где

$$I\{X_i \leqslant x\} = \begin{cases} 1, & X_i \leqslant x \\ 0, & \text{иначе} \end{cases}$$

Тогда, для $\forall x \in \mathbb{R}$ $\widehat{F}_n(x) \xrightarrow[n\to\infty]{\mathbb{P}} F(x)$

Доказательство.

$$\mathbb{E}\left[\widehat{F}_n(x)\right] = \frac{1}{n}\sum_{i=1}^{n} F(x) = F(x)$$

$$\mathbb{D}\left[\widehat{F}_n\right] = \frac{F(x)(1 - F(x))}{n}$$

$$\mathbb{P}\left(\left|\widehat{F}_n - F(x)\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{D}\left[\widehat{F}_n(x)\right]}{\varepsilon^2}$$

то есть $\widehat{F}_n \xrightarrow[n\to\infty]{\mathbb{P}} F(x)$ $\square$

Теорема. (Гливенко-Кантелли)

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

При этом, *почти наверное* тоже верно

Пример. Чему должно быть равно n, чтобы $\mathbb{P}\left(\left|\widehat{F}_n(x) - F(x)\right| \geqslant 0.01\right) \leqslant 0.01$?

Способ 1. Используем Чебышёва

$$\mathbb{P}(|\widehat{F}_n(x) - F(x)| \geqslant 0.01) \leqslant \frac{F(x)(1-F(x))}{n\cdot(0.01)^2} \leqslant \frac{1}{4n\cdot 0.0001} \leqslant 0.01$$

Получается, что $n \geqslant 250000$

Способ 2. По теореме Муавра-Лапласа

$$\frac{\widehat{F}_n(x) - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0;1)$$

$$\mathbb{P}(|\widehat{F}_n(x) - F(x)| \geqslant 0.01) = \mathbb{P}\left(\frac{|\widehat{F}_n(x) - F(x)|}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \geqslant \frac{0.01}{\sqrt{\frac{F(x)(1-F(x))}{n}}}\right) \leqslant 0.01$$

$$\begin{aligned}\frac{0.01}{\sqrt{\frac{F(x)(1-F(x))}{n}}} &\geqslant 2.58\\ \sqrt{n}\cdot 0.01 &\geqslant 2.58\sqrt{F(x)(1-F(x))}\\ \sqrt{n} &\geqslant \frac{2.58\cdot 0.5}{0.01} \approx 129\\ n &\geqslant 16641\end{aligned}$$

7.3 Теорема Колмогорова

Теорема. Пусть $X_1,\ldots,X_n \sim F(x;\theta) \equiv F(x)$, $\widehat{F}_n = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leqslant x\}$, есть случайная величина D_n:

$$\sqrt{n}\cdot\sup_x |\widehat{F}_n(x) - F(x)| \xrightarrow[n\to\infty]{d} \mu \sim K(x),$$

где $K(x)$ — распределение Колмогорова, определенная как

$$\mathbb{P}(D_n \leqslant x) = K(x) = \sum_{j=-\infty}^{\infty} (-1)^j \exp\left(-2j^2x^2\right)$$

7.4 Свойства выборочной функции распределения

1. состоятельная оценка
2. несмещенная оценка $F(x)$: $\mathbb{E}\left[\widehat{F}_n(x)\right] = F()x$
3. $\mathbb{D}\left[\widehat{F}_n(x)\right] = \frac{F(x)(1-F(x))}{n} \xrightarrow[n\to\infty]{} 0$
4. $\frac{\widehat{F}_n(x) - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0;1) \Longrightarrow \widehat{F}_n(x)$ — асимптотически (в пределе) нормаьная оценка $F(x)$
5. $n\cdot\widehat{F}_n(x) \sim Bi(n, F(x))$

7.5 Свойства выборочных моментов

Теорема.

1. Если $\exists \mathbb{E}\left[|X_1|^k\right] < \infty$, тогда $\mathbb{E}\left[\widehat{\alpha}_k\right] = \alpha_k$ (несмещенность)

2. $\widehat{\alpha}_k \xrightarrow[n\to\infty]{P} \alpha_k$ (состоятельность)

3. Если $\mathbb{D}\left[X_1^k\right] \neq 0, < \infty$, то $\frac{\widehat{\alpha}_k - \alpha_k}{\sqrt{\frac{\mathbb{D}\left[X_1^k\right]}{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0;1)$

8 Информации Фишера

8.1 Количество информации Фишера

Имеется выборка с неизвестным параметром — $X_1, \dots, X_n \sim F(x;\theta)$

Определение. Информацией Фишера называется

$$I(\theta; X) = \mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right],$$

где $\mathcal{L}$ — функция правдоподобия

Примечание. Определние применимо для регулярного случая, то есть область значений X не зависит от θ

Определение. Также информацией Фишера называется

$$I(\theta; X) = -\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right]$$

$$\ln \mathcal{L} = \sum_{i=1}^{n} \ln f_i, \quad f_i = f(x_i; \theta)$$

Также существует малая информация Фишера

$$i(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln f}{\partial \theta}\right)^2\right]$$

8.2 Эквивалентность определений

Докажем, что определения эквивалентны, то есть

$$-\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right] = \mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right],$$

$$\frac{\partial \ln f}{\partial \theta} = \frac{1}{f}\frac{\partial f}{\partial \theta} \Longrightarrow \frac{\partial f}{\partial \theta} = f\frac{\partial \ln f}{\partial \theta}$$

Покажем, что $\mathbb{E}\left[\frac{\partial \ln f}{\partial \theta}\right] = 0$

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial \ln f}{\partial \theta}\right] &= \int\limits_{-\infty}^{+\infty} \frac{\partial \ln f}{\partial \theta} \cdot f \cdot \mathrm{d}x \\
&= \int\limits_{-\infty}^{+\infty} \frac{\partial f}{\partial \theta} \mathrm{d}x \\
&= \frac{\partial}{\partial \theta} \int\limits_{\infty}^{+\infty} f \mathrm{d}x \\
&= 0
\end{aligned}$$

Теперь покажем эквивалентность определений

$$\int_{-\infty}^{+\infty} \frac{\partial \ln f}{\partial \theta} \cdot f \cdot \mathrm{d}x = 0$$

$$\int_{-\infty}^{+\infty} \left(\frac{\partial^2 \ln f}{\partial \theta^2} f + \frac{\partial \ln f}{\partial \theta} \frac{\partial f}{\partial \theta} \right) \mathrm{d}x = 0$$

$$\int_{-\infty}^{+\infty} \frac{\partial^2 \ln f}{\partial \theta^2} f \mathrm{d}x + \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f}{\partial \theta} \right)^2 f \mathrm{d}x = 0$$

$$\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial \theta^2} \right] + \mathbb{E}\left[\left(\frac{\partial \ln f}{\partial \theta} \right)^2 \right] = 0$$

Значит, $\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial \theta^2} \right] = -\mathbb{E}\left[\left(\frac{\partial \ln f}{\partial \theta} \right)^2 \right]$ □

Примечание. Если $X = \frac{\partial \ln f}{\partial \theta}$ — случайная величина, то

$$\mathbb{E}[X] = 0$$
$$\mathbb{D}[X] = i(\theta)$$

Доказательство. Покажем, что $I(\theta) = n \cdot i(\theta)$

$$\frac{\partial \ln f}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f_i}{\partial \theta} \text{ — сумма независимых сл. величин}$$

Отсюда, очевидно следует, что $I(\theta) = n \cdot i(\theta)$ □

8.3 Пример

Пусть имеется выборка $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, причем дисперсия известна

$$\begin{aligned}
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right) \\
\ln f &= -\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \\
\frac{\partial \ln f}{\partial \mu} &= \frac{x-\mu}{\sigma} \cdot \frac{1}{\sigma} \\
i(\theta) &= \mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \mu} \right)^2 \right] \\
&= \frac{1}{\sigma^4} \mathbb{E}\left[(x-\mu)^2 \right] \\
&= \frac{1}{\sigma^2}
\end{aligned}$$

Теперь воспользуемся второй формулой

$$\frac{\partial^2 \ln f}{\partial \mu^2} = -\frac{1}{\sigma^2} \Longrightarrow i(\mu) = -\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial \theta^2} \right] = \frac{1}{\sigma^2}$$

8.4 Неравенство Рао-Крамера-Фреше

Определение. Считаем, что выполняются все условия гладкости и регулярности

$$\mathbb{D}\left[\widehat{\theta} \right] \geqslant \frac{1}{I(\theta)} \left(1 + \frac{\partial b(\theta)}{\partial \theta} \right)^2,$$

где $b(\theta) = \mathbb{E}\left[\widehat{\theta} - \theta\right]$

Для несмещенных оценок:

$$\mathbb{D}\left[\widehat{\theta}\right] \geqslant \frac{1}{I(\theta)}$$

Доказательство. $\mathbb{E}\left[\widehat{\theta}\right] = \theta, \mathbb{E}\left[\widehat{\theta} - \theta\right] = 0$

$$\int \widehat{\theta} f \mathrm{d}x = 0$$

$$\int \widehat{\theta} \frac{\partial f}{\partial \theta} \mathrm{d}x = 1$$

$$\int \widehat{\theta} \frac{\partial \ln f}{\partial \theta} f \mathrm{d}x = 1$$

$$\mathbb{E}\left[\widehat{\theta} \frac{\partial \ln f}{\partial \theta}\right] = 1$$

$$E\left(\widehat{\theta} - \theta\right)\left(\frac{\partial \ln f}{\partial \theta} - 0\right) = 1 \Longrightarrow cov\left(\widehat{\theta}, \frac{\partial \ln f}{\partial \theta}\right) = 1$$

Получается, что

$$1 \leqslant \mathbb{D}\left[\widehat{\theta}\right] \underbrace{\mathbb{D}\left[\frac{\partial \ln \mathcal{L}}{\partial \theta}\right]}_{I(\theta)}$$

8.5 Критерии эффективности

Если $\mathbb{D}\left[\widehat{\theta}\right] = \frac{1}{I(\theta)} \left(1 + \frac{\partial b(\theta)}{\partial \theta}\right)^2$, то оценка эффективна в классе оценок со смещением $b(\theta)$

Определение. $\widehat{\theta}$ асимптотически нормальна, если $\exists \sigma^2(\theta)$, что

$$\sqrt{\pi}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$$

Определение. Если $\widehat{\theta}$ — асимптотически нормальная, то $\frac{\sigma^2(\theta)}{n}$ называется асимптотической дисперсией

Определение. Оценка $\widehat{\theta}$ — асимптотически эффективна, если ее асимптотическая дисперсия

$$\frac{\sigma^2(\theta)}{n} = \frac{1}{I(\theta)} \text{ для несмещенных}$$