Математическая статистика

Михайлов Максим

26 сентября 2021 г.

Оглавление

Лекция 1

6 сентября

1 Организационные вопросы

Большая часть баллов зарабатывается индивидуальными заданиями, выполняемыми в Excel — 30 баллов. Тест с большим числом вопросов — 20 или 25 баллов.

2 Введение

Теория вероятности состоит в следующем: исследуется случайная величина с заданным распределением. Математическая статистика занимается обратным — даны данные, нужно приближенно найти числовые характеристики случайной величины и с некоторой уверенностью найти вид распределения. Матстатистика также исследует связанность случайных величин, их корреляцию. В идеале есть цель построить модель, которая по значениям одних случайных величин предсказывает другие.

Пусть проводится некоторое количество экспериментов, в ходе которых появились некоторые данные.

Определение. **Генеральная совокупность** — набор всех исходов проведенных экспериментов.

В реальности наблюдается некоторая выборка генеральной совокупности, ибо рассматривать всю совокупность нецелесообразно.

Определение. **Выборочная совокупность** — исходы наблюдаемых экспериментов.

Определение. Выборка называется **репрезентативной**, если её распределение совпадает с распределением генеральной совокупности.

Выборка может быть нерепрезентативной, как в примере с ошибкой выжившего. Мы считаем, что таких ошибок у нас нет и все выборки репрезентативны, ибо исправление этих ошибок — задача конкретной области, в которой используется матстатистика.

Определение (после опыта). Пусть проведено n наблюдаемых независимых экспериментов, в которых случайная величина приняла значение $X_1, X_2 \dots X_n$. Набор[1] этих данных называется **выборкой объема** n.

Определение (до опыта). **Выборкой объема** n называется набор из n независимых одинаково распределенных случайных величин.

Пусть имеется выборка в смысле "после опыта" объема n. Её можно интерпретировать как следующую дискретную случайную величину:

X_i	X_1	X_2	$\dots$	X_n	$\sum$
p_i	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$	1

Средневыборочное:

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

Выборочная дисперсия:

$$S^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

2.1 Выборочная функция распределения

$$F_n^*(z) \coloneqq \frac{1}{n} \sum_{i=1}^{n} I(X_i < z) = \frac{\text{число } X_i \in (-\infty, z)}{n}$$

Примечание. I — индикатор:

$$I(X_i < z) = \begin{cases} 1, & X_i < z \\ 0, & X_i \geq z \end{cases}$$

Теорема 1.

$$\forall x \in \mathbb{R} \quad F_n^*(z) \xrightarrow[n \to \infty]{P} F(z)$$

Доказательство. Заметим, что

$$\mathbb{E} I(X_1 < z) = 1 \cdot P(X_1 < z) + 0 \cdot P(X_1 \geq z) = P(X_1 < z) = F(z)$$

[1] Или вектор.

, где $F(z)$ — функция распределения X_1. Заметим, что $F(z) \le 1 < \infty$, следовательно применим ЗБЧ Хинчина:

$$F_n^*(z) = \frac{\sum_{i=1}^n I(X_i < z)}{n} \xrightarrow{P} \mathbb{E}I(X_1 < z) = F(z)$$

$\square$

Примечание. На самом деле имеется даже равномерная сходимость по вероятности — это теорема Гливенко-Кантелли:

$$\sup_{z \in \mathbb{R}} |F_n^*(z) - F(z)| \xrightarrow[n \to \infty]{P} 0$$

3 Первоначальная обработка статданных

Если отсортировать данные, то получим **вариационный ряд**: $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$. Если учесть повторяющиеся экземпляры, то получим **частотный вариационный ряд**:

$X_{(i)}$	$X_{(1)}$	$X_{(2)}$	$\dots$	$X_{(k)}$	$\sum$
n_i	n_1	n_2	$\dots$	n_k	n
p_i^*	$\dfrac{n_1}{n}$	$\dfrac{n_2}{n}$	$\dots$	$\dfrac{n_k}{n}$	1

Определение. $h \coloneqq X_{\max} - X_{\min}$ — **размах выборки**

Допустим, что разбили интервал $(X_{\min}, X_{\max})$ на k интервалов, чаще всего одинаковой длины.[2] Тогда $l_i = \frac{h}{k}$ — длина каждого интервала и интервальный ряд можно заменить интервальным вариационным рядом.

i	l_1	l_2	$\dots$	l_k	$\sum$
m_i	m_1	m_2	$\dots$	m_k	n
$\dfrac{m_i}{n}$	$\dfrac{m_1}{n}$	$\dfrac{m_2}{n}$	$\dots$	$\dfrac{m_k}{n}$	1

m_i — число попавших в i-тый интервал данных.

По такой таблице можно построить **гистограмму**. На координатной плоскости построим прямоугольники с основаниями l_i и высотами $\frac{m_i}{nl_i}$. В результате получаем ступенчатую фигуру площади 1, которая и называется гистограммой.

Теорема 2. При $n \to \infty, k(n) \to \infty$, причем $\frac{k(n)}{n} \to 0$, гистограмма будет стремиться к плотности распределения:

$$\frac{m_i}{n} \xrightarrow{P} P(X_i \in l_i) = \int_{l_i} f(x)dx$$

[2] Применяются и другие разбиения, например равнонаполненное.

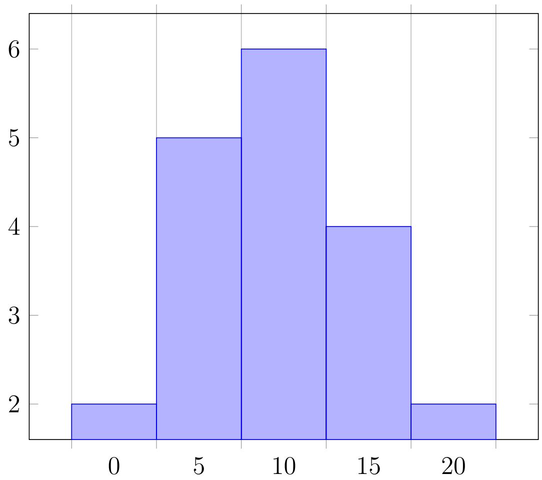

Рис. 1.1: Пример гистограммы

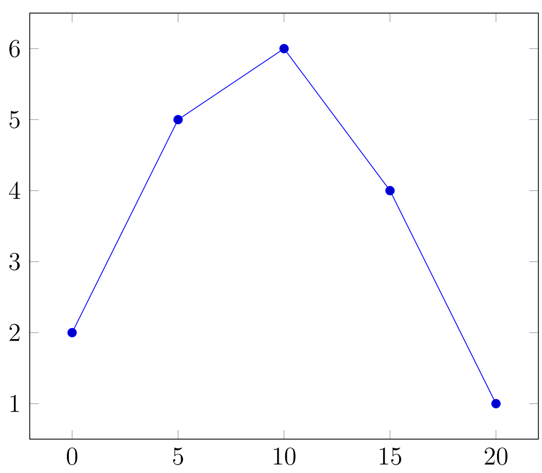

Рис. 1.2: Пример полигона

Чаще всего число интервалов берется по формуле Стёрджесса: $k \approx 1 + \log_2 n$. Иногда $k \approx \sqrt[3]{n}$.

Примечание. Иногда выборка изображается в виде **полигона**: отображаются точки, соответствующие серединам интервалов и ставим точки на высоте $\frac{m_i}{n}$.

Лекция 2

13 сентября

4 Точечные оценки

Пусть имеется выборка объема n: $X = \begin{pmatrix} X_1 & \dots & X_n \end{pmatrix}$

Определение. **Статистикой** называется измеримая функция $\theta^* = \theta^*(X_1, \dots, X_n)$.

Пусть требуется найти значение параметра θ случайной величины X по данной выборке. Оценку будем считать с помощью некоторой статистики θ^*.

4.1 Свойства статистических оценок

4.1.1 Состоятельность

Определение. Статистика $\theta^* = \theta^*(X_1, \dots, X_n)$ называется **состоятельной оценкой** параметра θ, если:

$$\theta^* \xrightarrow[n\to\infty]{P} \theta$$

4.1.2 Несмещённость

Определение. Статистика $\theta^* = \theta^*(X_1, \dots, X_n)$ называется **несмещенной оценкой** параметра θ, если

$$\mathbb{E}\theta^* = \theta$$

Примечание. То есть с равной вероятностью можем ошибиться как в меньшую, так и в большую сторону. Нет систематической ошибки.

Определение. Статистика $\theta^* = \theta^*(X_1, \dots, X_n)$ называется **асимптотически несмещенной оценкой** параметра θ, если

$$\mathbb{E}\theta^* \xrightarrow[n\to\infty]{} \theta$$

Примечание. То есть при достаточно большом объеме выборки ошибка исчезает, но при малом она может существовать.

4.1.3 Эффективность

Определение. Оценка θ_1^* **не хуже** оценки θ_2^*, если

$$\mathbb{E}(\theta_1^* - \theta)^2 \leq \mathbb{E}(\theta_2^* - \theta)^2$$

или, если оценки несмещенные,

$$\mathbb{D}\theta_1^* \leq \mathbb{D}\theta_2^*$$

Определение. Оценка θ^* называется **эффективной**, если она не хуже всех остальных оценок.

Теорема 3. Не существует эффективной оценки в классе всех возможных оценок.

Теорема 4. В классе несмещённых оценок существует эффективная оценка.

4.2 Точечные оценки моментов

Определение. **Выборочным средним** $\overline{X_B}$ называется величина

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

Определение. **Выборочной дисперсией** $\mathbb{D}_B$ называется величина

$$\boxed{\mathbb{D}_B = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_B})^2}$$

Определение. **Исправленной выборочной дисперсией** S^2 называется величина

$$\boxed{S^2 = \frac{n}{n-1}\mathbb{D}_B}$$

или

$$\boxed{S^2 = \frac{1}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_B})^2}$$

Определение. **Выборочным средним квадратическим отклонением** называется величина

$$\boxed{\sigma_B = \sqrt{\mathbb{D}_B}}$$

Определение. **Исправленным выборочным средним квадратическим отклонением** называется величина

$$\boxed{S = \sqrt{S^2}}$$

Определение. **Выборочным k-тым моментом** называется величина

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

Определение. **Модой** M_0^* вариационного ряда называется варианта с наибольшей частотой:

$$\boxed{M_0^* = X_i : n_i = \max_{1 \le j < n} n_j}$$

Определение. **Медианой** M_e^* вариационного ряда называется значение варианты в середине ряда:

1. Если $n = 2k - 1$, то $M_e^* = X_k$
2. Если $n = 2k$, то $M_e^* = \frac{X_k + X_{k+1}}{2}$

Величина	Команда в Excel	
	Русский	Английский
$\overline{X_B}$	СРЗНАЧ	AVERAGE
$\mathbb{D}_B$	ДИСПР	VARP
S^2	ДИСП	VAR
σ_n	СТАНДОТКЛОНП	STDEVP
S	СТАНДОТКЛОН	STDEV
M_0^*	МОДА	MODE
M_e^*	МЕДИАНА	MEDIAN

Теорема 5. Выборочное среднее $\overline{X_B}$ является несмещенной состоятельной оценкой для математического ожидания, то есть:

1. $\mathbb{E}\overline{X_B} = \mathbb{E}X = a$ — несмещенность
2. $\overline{X_B} \xrightarrow[n\to\infty]{P} \mathbb{E}X$ — состоятельность

Доказательство.

1. $$\mathbb{E}\overline{X} = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i = \frac{1}{n}\cdot n\mathbb{E}X_i = \mathbb{E}X$$

2.
$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{P} \mathbb{E}X$$

Это верно по закону больших чисел.

□

Теорема 6. Выборочный k-тый момент является несмещенной состоятельной оценкой для теоретического k-того момента, то есть:

1. $\mathbb{E}\overline{X^k} = X^k$
2. $\overline{X^k} \xrightarrow{P} \mathbb{E}X^k$

Доказательство. Следует из предыдущей теоремы, если в качестве случайной величины взять X^k. □

Теорема 7.

- $\mathbb{D}_B$ — смещённая состоятельная оценка дисперсии
- S^2 — несмещённая состоятельная оценка дисперсии

Доказательство.
$$\mathbb{D}_B = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \overline{X^2} - (\overline{X})^2$$

$$\begin{aligned}
\mathbb{E}\mathbb{D}_B &= \\
\mathbb{E}(\overline{X^2} - (\overline{X})^2) &= \\
\mathbb{E}\overline{X^2} - \mathbb{E}(\overline{X})^2 &= \\
\mathbb{E}X^2 - \mathbb{E}(\overline{X})^2 &= \\
\mathbb{D}\overline{X} &= \\
\mathbb{E}(\overline{X})^2 - (\mathbb{E}\overline{X})^2 &= \\
\mathbb{E}X^2 - (\mathbb{D}\overline{X} + (\mathbb{E}\overline{X})^2) &= \\
\mathbb{E}X^2 - (\mathbb{E}X)^2 - \mathbb{D}\overline{X} &= \\
(\mathbb{E}X^2 - (\mathbb{E}X)^2) - \mathbb{D}\overline{X} &= \\
\mathbb{D}X - \mathbb{D}\overline{X} &= \\
\mathbb{D}X - \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) &= \\
\mathbb{D}X - \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}X_i &=
\end{aligned}$$

$$\mathbb{D}X - \frac{1}{n^2} \cdot n\mathbb{D}X =$$
$$\mathbb{D}X - \frac{1}{n}\mathbb{D}X =$$
$$\frac{n-1}{n}\mathbb{D}X \neq \mathbb{D}X$$

$$\mathbb{E}S^2 = \mathbb{E}\left(\frac{n}{n-1}\mathbb{D}_B\right) = \frac{n}{n-1} \cdot \frac{n-1}{n}\mathbb{D}X = \mathbb{D}X$$
$$\mathbb{D}_B = \overline{X^2} - (\overline{X})^2 \xrightarrow{P} \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{D}X$$
$$S^2 = \frac{n}{n-1}\mathbb{D}_B \xrightarrow{P} \underbrace{\frac{n}{n-1}}_{\to 1}\mathbb{D}X$$

□

Примечание. $\mathbb{D}_B$ — асимптотически несмещённая оценка, т.к. при $n \to \infty$, $\frac{n-1}{n} \to 1$. Таким образом, при большой[1] выборке можно игнорировать смещённость.

4.3 Метод моментов

Изобретен Карлом Пирсоном.

Пусть имеется выборка $(X_1 \dots X_n)$ неизвестного распределения, при этом известен тип[2] распределения. Пусть этот тип определяется k неизвестными параметрами $\theta_1 \dots \theta_k$. Теоретическое распределение задает теоретические k-тые моменты. Например, если распределение непрерывное, то оно задается плотностью $f(X, \theta_1 \dots \theta_k)$ и $m_k = \int_{-\infty}^{+\infty} X^k f(x, \theta_1 \dots \theta_k)dx = h_k(\theta_1 \dots \theta_k)$. Метод моментов состоит в следующем: вычисляем выборочные моменты и подставляем их в эти равенства вместо теоретических. В результате получаем систему уравнений:

$$\begin{cases} \overline{X} = h_1(\theta_1 \dots \theta_k) \\ \overline{X^2} = h_2(\theta_1 \dots \theta_k) \\ \vdots \\ \overline{X^k} = h_k(\theta_1 \dots \theta_k) \end{cases}$$

Решив эту систему, мы получим оценки на $\theta_1 \dots \theta_k$. Эти оценки будут состоятельными[3], но смещёнными.

Пример. Пусть $X \in U(a, b), a < b$. Обработав статданные, получили оценки первого и второго момента: $\overline{X} = 2.25; \overline{X^2} = 6.75$

[1] $n \geq 100$, например.
[2] Нормальное, показательное и т.д.
[3] Если не придумывать специально плохие примеры

Решение. Плотность $f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \le x \le b \\ 0, & x > b \end{cases}$

$$\mathbb{E}X = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a} = \frac{1}{b-a} \cdot \frac{x^2}{2}\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \boxed{\frac{a+b}{2}}$$

$$\mathbb{E}X^2 = \int_a^b x^2 \cdot \frac{1}{b-a}dx = \frac{1}{b-a} \cdot \frac{x^3}{3}\Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \boxed{\frac{a^2+ab+b^2}{3}}$$

$$\begin{cases} 2.25 = \frac{a+b}{2} \\ 6.75 = \frac{a^2+ab+b^2}{3} \end{cases}$$

$$\begin{cases} a+b = 4.5 \\ a^2+ab+b^2 = 20.25 \end{cases}$$

$$\begin{cases} a+b = 4.5 \\ ab = 0 \end{cases}$$

$$\begin{cases} a = 0 \\ b = 4.5 \end{cases}$$

□

Лекция 3

20 сентября

4.4 Метод максимального правдоподобия

Метод максимального правдоподобия состоит в том, чтобы подобрать параметры таким образом, чтобы вероятность получения данной выборки была наибольшей. Если распределение дискретное, то вероятность выборки

$$P_\theta(X_1 = x_1, X_2 = x_2 \dots X_n = x_n) = P_\theta(X_1 = x_1)P_\theta(X_2 = x_2)\dots P_\theta(X_n = x_n)$$

Для непрерывной величины аналогично.

Поэтому исследуем такую функцию:

Определение. **Функцией правдоподобия** называется функция $L(\overline{X}, \theta)$, зависящая от выборки и неизвестных параметров, равная:

- В случае дискретного распределения:

$$P_\theta(X_1 = x_1)P_\theta(X_2 = x_2)\dots P_\theta(X_n = x_n)$$

- В случае абсолютно непрерывного распределения:

$$f_\theta(x_1)f_\theta(x_2)\dots f_\theta(x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

Эту функцию неудобно исследовать, поэтому мы используем следующую функцию:

Определение. **Логарифмическая функция правдоподобия**:

$$M(\overline{X}, \theta) = \ln L(\overline{X}, \theta)$$

Т.к. логарифм — строго возрастающая функция, экстремумы обычной и логарифмической функций правдоподобия совпадают.

Определение. Оценкой максимального правдоподобия $\hat{\theta}$ называется значение θ, при котором функция правдоподобия достигает наибольшего значения.

Пример. Пусть $X_1 \dots X_n$ — выборка неизвестного распределения Пуассона с параметром λ: $X \in \Pi_\lambda, \lambda > 0$

$$P(X = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$L(\overline{X}, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{n \cdot \overline{X}}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

$$\ln L(\overline{X}, \lambda) = n \cdot \overline{X} \cdot \ln \lambda - \ln \prod_{i=1}^{n} x_i! - n\lambda$$

$$\frac{\partial \ln L(\overline{X}, \lambda)}{\partial \lambda} = \frac{n\overline{X}}{\lambda} - n$$

Приравняем производную к нулю, чтобы найти точки экстремума:

$$\frac{n\overline{X}}{\lambda} - n = 0 \Rightarrow \lambda = \overline{X}$$

Таким образом $\hat{\theta} = \overline{X}$ — ОМП.

Пример. Пусть $X_1 \dots X_n$ — выборка неизвестного нормального распределения: $X \in N(a, \sigma^2), a \in \mathbb{R}, \sigma > 0$

$$f_{a,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

$$L(\overline{X}, a, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-a)^2}{2\sigma^2}} = \frac{1}{\sigma^n \sqrt{2\pi}^n} e^{-\frac{\sum(x_i-a)^2}{2\sigma^2}}$$

$$\ln L(\overline{X}, a, \sigma^2) = n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum (x_i - a)^2$$

Не дописано

Пример. Пусть $X_1 \dots X_n$ — выборка равномерного распределения вида $U(0, \theta)$

1. Метод моментов.

$$\mathbb{E} = \frac{a+b}{2} = \frac{\theta}{2} \Rightarrow \theta = 2\overline{X}$$

2. Метод максимального правдоподобия.

$$f_\theta(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{\theta}, & x \in [0, \theta] \\ 0, & x > \theta \end{cases}$$

$$L(\overline{X},\theta)=\prod_{i=1}^{n}f_\theta(x_i)=\begin{cases}0, & \theta<\max x_i=X_{(n)}\\ \frac{1}{\theta^n}, & \theta\geq X_{(n)}\end{cases}$$

L достигает наибольшего значения при $\theta=X_{(n)}$.

Сравним полученные оценки.

1. $\theta^*=2\overline{X}$ — несмещённая оценка, т.к. $\mathbb{E}\theta^*=\mathbb{E}2\overline{X}=2\mathbb{E}\overline{X}=\theta$

$$\mathbb{E}(\theta^*-\theta)=\mathbb{D}(\theta^*)=\mathbb{D}2\overline{X}=4\frac{1}{n}\mathbb{D}X=\frac{4}{n}\frac{\theta^2}{12}=\frac{\theta^2}{3n}$$

2. Изучим случайную величину $X_{(n)}$. Её функция распределения это

$$F_{X_{(n)}}(x)=P(X_{(n)}<x)=P(X_1<x)\dots P(X_n<x)=(F_X(x))^n$$

$$F_X(x)=\begin{cases}0, & x<0\\ \frac{x}{\theta}, & x>0\end{cases}\Rightarrow F_{X_{(n)}}(x)=\begin{cases}0, & x<0\\ x^{\frac{n}{\theta^n}}, & x\geq 0\end{cases}\Rightarrow f_{X_{(n)}}(x)=\begin{cases}0, & x<0\\ \frac{nx^{n-1}}{\theta^n}, & x\geq 0\end{cases}$$

$$\mathbb{E}X_{(n)}=\int_0^\theta x\frac{nx^{n-1}}{\theta^n}dx=\frac{n}{\theta^n}\int_0^\theta x^n dx=\frac{n}{\theta^n}\frac{x^{n+1}}{n+1}\Big|_0^\theta=\frac{n\theta}{n+1}$$

Таким образом, оценка смещённая, но асимптотически несмещенная.

Заменим эту оценку на несмещённую оценку $\tilde{\theta}=\frac{n+1}{n}\hat{\theta}=\frac{n+1}{n}X_{(n)}$ — сходятся к θ с одинаковой скоростью.

$$\begin{aligned}\mathbb{E}\tilde{\theta}^2 &=\\ \mathbb{E}\left(\frac{n+1}{n}X_{(n)}\right)^2 &=\\ \frac{(n+1)^2}{n^2}\mathbb{E}X_{(n)}^2 &=\\ \frac{(n+1)^2}{n^2}\int_0^\theta x^2\frac{nx^{n-1}}{\theta^n}dx &=\\ \frac{(n+1)^2}{n^2}\frac{n}{\theta^n}\frac{x^{n+2}}{n+2}\Big|_0^\theta &=\frac{(n+1)^2}{n(n+2)}\theta^2\end{aligned}$$

$$\mathbb{D}\tilde{\theta}^2=\mathbb{E}\tilde{\theta}^2-\mathbb{E}^2\tilde{\theta}=\frac{(n+1)^2}{n(n+2)}\theta^2-\theta^2=\theta^2\left(\frac{n^2+2n+1-n^2-2n}{n^2+2n}\right)=\frac{\theta^2}{n(n+2)}$$

Итак, сравним оценки.

$$\mathbb{D}\tilde{\theta}=\frac{\theta^2}{n(n+2)}<\frac{\theta^2}{3n}=\mathbb{D}\theta^*$$

Таким образом, оценка с помощью метода максимального правдоподобия лучше, её дисперсия стремится к нулю со скоростью $\frac{1}{n^2}$, а дисперсия первой оценки — со скоростью $\frac{1}{n}$. $\tilde{\theta} \to \theta$ со скоростью $\frac{1}{n}$, а $\theta^* \to \theta$ со скоростью $\frac{1}{\sqrt{n}}$

Следствие 7.1. Оценка математического ожидания $\overline{X} = 2\theta$ не будет эффективной оценкой, т.к. можно показать, что в данном случае эффективной оценкой будет

$$\boxed{\mathbb{E}X = \frac{n+1}{n} \cdot \max\{X_1 \dots X_n\}}$$

Примечание. ОМП состоятельны, часто эффективны, но могут быть смещенными.

5 Неравенство Рао-Крамера

Пусть известно, что случайная величина $X \in \mathcal{F}_\theta$ — семейству распределений с θ.

Определение. **Носителем** семейства распределений $\mathcal{F}_\theta$ называется множество $C \subset \mathbb{R}$, такое что $\forall\theta \;\; P(X \in C) = 1$.

Обозначение.

$$f_\theta(x) = \begin{cases} f_\theta(x), & \text{если распределение абсолютно непрерывное} \\ P_\theta(X = x), & \text{если распределение дискретное} \end{cases}$$

Определение. **Информацией Фишера** называется величина

$$I(\theta) = \mathbb{E}\left(\frac{\partial \ln f_\theta(x)}{\partial\theta}\right)^2$$

, если она существует.

Определение. Семейство распределений $\mathcal{F}_\theta$ называется **регулярным**, если:

1. Существует носитель C семейства $\mathcal{F}_\theta$, такой что $\forall x \in C$ функция $\ln f_\theta(x)$ непрерывно дифференцируема по θ.
2. $I(\theta)$ существует и непрерывна по θ.

Теорема 8 (неравенство Рао-Крамера). Пусть $X_1 \dots X_n$ — выборка объема n из регулярного семейства распределений $\mathcal{F}_\theta$, θ^* — несмещенная оценка параметра θ, дисперсия которой ограничена на любом компакте в области θ.

Тогда

$$\boxed{\mathbb{D}\theta^* \geq \frac{1}{nI(\theta)}}$$

Следствие 8.1. Если при условиях выше $\mathbb{D}\theta^* = \frac{1}{nI(\theta)}$, то θ^* — эффективная оценка. Это не всегда достижимо.

Пример. Пусть $X_1 \dots X_n$ — выборка нормального распределения $N(a, \sigma^2), a \in \mathbb{R}, \sigma^2 > 0$. Проверим эффективность оценки $a^* = \overline{X}$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Рассмотрим носитель $C = \mathbb{R}$.

$$\ln f(x) = -\ln\sigma - \frac{1}{2}\ln(2\pi) - \frac{(x-a)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x)}{\partial a} = \frac{1}{2\sigma^2} 2(x-a) = \frac{x-a}{\sigma}$$

Производная непрерывна по a $\forall a \in \mathbb{R}$

$$I(a) = \mathbb{E}\left(\frac{\partial \ln f(x)}{\partial a}\right)^2 = \mathbb{E}\left(\frac{x-a}{\sigma}\right)^2 = \frac{1}{\sigma^4}\mathbb{E}(X - \mathbb{E}X) = \frac{1}{\sigma^4}\mathbb{D}X = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$$

Сравним обе части неравенства Рао-Крамера:

$$\mathbb{D}a^* = \mathbb{D}\overline{X} = \frac{1}{n}\mathbb{D}X = \frac{1}{n}\sigma^2 = \frac{\sigma^2}{n}$$

$$\mathbb{D}a^* = \frac{\sigma^2}{n} \stackrel{?}{=} \frac{1}{nI(a)} = \frac{1}{n\frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

Таким образом, оценка эффективна.

Примечание. Исправленная дисперсия S^2 также является эффективной оценкой.

Определение. BLUE[1]-оценка — лучшая оценка из оценок вида $\theta^* = \alpha_1 X_1 + \dots + \alpha_n X_n$.

[1] Best linear unbiased estimate.