

Математическая статистика

Михайлов Максим

28 сентября 2021 г.

Оглавление

Лекция 1	6 сентября	3
1	Организационные вопросы	3
2	Введение	3
2.1	Выборочная функция распределения	4
3	Первоначальная обработка статданных	5
Лекция 2	13 сентября	7
4	Точечные оценки	7
4.1	Свойства статистических оценок	7
4.1.1	Состоятельность	7
4.1.2	Несмещённость	7
4.1.3	Эффективность	8
4.2	Точечные оценки моментов	8
4.3	Метод моментов	11
Лекция 3	20 сентября	13
4.4	Метод максимального правдоподобия	13
5	Неравенство Рао-Крамера	16
Лекция 4	18 сентября	18
6	Распределения в матстатистике	18
6.0	Нормальное распределение	18
6.1	Гамма-распределение	18
6.2	Распределение “хи-квадрат”	19
6.3	Распределение Стьюдента	19
6.4	Распределение Фишера-Снедекора	19
7	Линейные преобразования нормальных выборок	20

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 $X_1, X_2 \dots X_n$. Набор¹ 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:

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

Выборочная дисперсия:

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

2.1 Выборочная функция распределения

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

Примечание. I — индикатор:

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

Теорема 1.

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

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 $F(z) \leq 1 < \infty$, следовательно применим ЗБЧ Хинчина:

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

□

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

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

3 Первоначальная обработка статданных

Если отсортировать данные, то получим **вариационный ряд**: $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. 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^*	$\frac{n_1}{n}$	$\frac{n_2}{n}$	\dots	$\frac{n_k}{n}$	1

Определение. $h := X_{\max} - X_{\min}$ — **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
$\frac{m_i}{n}$	$\frac{m_1}{n}$	$\frac{m_2}{n}$	\dots	$\frac{m_k}{n}$	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 $n \rightarrow \infty, k(n) \rightarrow \infty$, причем $\frac{k(n)}{n} \rightarrow 0$, 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.

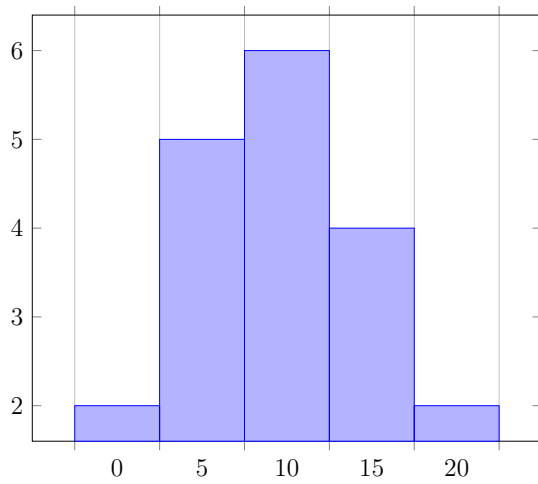


Рис. 1.1: Пример
гистограммы

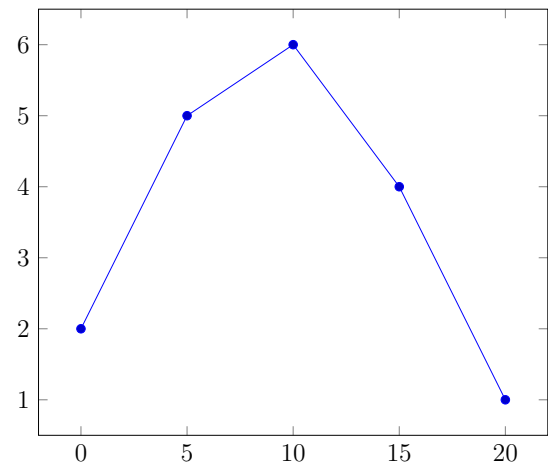


Рис. 1.2: Пример
полигона

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 : $X = (X_1 \dots X_n)$

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:

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

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

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

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

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

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

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

Определение. Исправленной выборочной дисперсией S^2 называется величина

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

или

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

Определение. Выборочным средним квадратическим отклонением называется величина

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

Определение. Исправленным выборочным средним квадратическим отклонением называется величина

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

Определение. Выборочным k -тым моментом называется величина

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

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

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

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. $\overline{X_B} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}X$ — 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.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}X$$

Это верно по закону больших чисел.

□

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

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

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.

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

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

$$\begin{aligned}
\mathbb{D}X - \frac{1}{n^2} \cdot n\mathbb{D}X &= \\
\mathbb{D}X - \frac{1}{n}\mathbb{D}X &= \\
\frac{n-1}{n}\mathbb{D}X &\neq \mathbb{D}X \\
\mathbb{E}S^2 = \mathbb{E}\left(\frac{n}{n-1}\mathbb{D}_B\right) &= \frac{n}{n-1} \cdot \frac{n-1}{n}\mathbb{D}X = \mathbb{D}X \\
\mathbb{D}_B = \overline{X^2} - (\overline{X})^2 &\xrightarrow{P} \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{D}X \\
S^2 = \frac{n}{n-1}\mathbb{D}_B &\xrightarrow{P} \underbrace{\frac{n}{n-1}}_{\rightarrow 1} \mathbb{D}X
\end{aligned}$$

□

Примечание. \mathbb{D}_B — асимптотически несмещённая оценка, т.к. при $n \rightarrow \infty$, $\frac{n-1}{n} \rightarrow 1$. Таким образом, при большой¹ выборке можно игнорировать смещённость.

4.3 Метод моментов

Изобретен Карлом Пирсоном.

Пусть имеется выборка $(X_1 \dots X_n)$ неизвестного распределения, при этом известен тип² 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 состоятельными³, 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 $f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x > b \end{cases}$

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

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

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

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

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

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

□

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 $L(\bar{X}, \theta)$, 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:

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

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 : $X \in \Pi_\lambda, \lambda > 0$

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

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

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

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

Приравняем производную к нулю, чтобы найти точки экстремума:

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

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

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

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

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

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

Не дописано

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

1. Метод моментов.

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

2. Метод максимального правдоподобия.

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

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

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

Сравним полученные оценки.

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

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

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

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

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

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

Таким образом, оценка смещённая, но асимптотически несмещённая.

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.

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

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 $\frac{1}{n}$. $\tilde{\theta} \rightarrow \theta$ со скоростью $\frac{1}{n}$, а $\theta^* \rightarrow \theta$ со скоростью $\frac{1}{\sqrt{n}}$

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

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

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}_θ 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}_θ называется **регулярным**, если:

1. Существует носитель C семейства \mathcal{F}_θ , 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}_θ , θ^* — несмещенная оценка параметра θ , дисперсия которой ограничена на любом компакте в области θ .

Тогда

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

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 $a^* = \bar{X}$.

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

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 $a \ \forall a \in \mathbb{R}$

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

Сравним обе части неравенства Рао-Крамера:

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

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

Таким образом, оценка эффективна.

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

Определение. BLUE¹-оценка — лучшая оценка из оценок вида $\theta^* = \alpha_1 X_1 + \dots + \alpha_n X_n$.

¹ Best linear unbiased estimate.

Лекция 4

18 сентября

6 Распределения в матстатистике

6.0 Нормальное распределение

$X \in N(a, \sigma^2)$:

$$\mathbb{E}X = a, \mathbb{D}X = \sigma$$

$N(0, 1)$ — стандартное нормальное распределение

6.1 Гамма-распределение

$X \in \Gamma_{\alpha, \lambda}$, если её плотность равна:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x > 0 \end{cases}$$

Свойства.

1. $\mathbb{E}\xi = \frac{\lambda}{\alpha}, \mathbb{D}\xi = \frac{\lambda}{\alpha^2}$
2. Если $\xi_1 \in \Gamma_{\alpha, \lambda_1}, \xi_2 \in \Gamma_{\alpha, \lambda_2}$, то $\xi_1 + \xi_2 \in \Gamma_{\alpha, \lambda_1 + \lambda_2}$
3. $\Gamma_{\alpha, 1} = E_\alpha$ — показательное распределение.
4. Если $X_i \in E_\alpha$, то $\sum_{i=1}^n X_i \in \Gamma_{\alpha, n}$
5. Если $X \in N(0, 1)$, то $X^2 \in \Gamma_{\frac{1}{2}, \frac{1}{2}}$

Примечание. Гамма-распределение возникает в матстатистике как распределение квадрата стандартно нормально распределенной величины. Обобщим эту идею:

6.2 Распределение “хи-квадрат”

Определение. Распределение хи-квадрат с k степенями свободы называется распределение суммы k квадратов независимых стандартных нормальных величин.

$$\chi_k^2 = X_1^2 + X_2^2 + \dots + X_k^2, \quad X_i \in N(0, 1)$$

Обозначение. $\chi^2 \in H_k$

Свойства.

1. $\chi_k^2 \in \Gamma_{\frac{1}{2}, \frac{k}{2}}$
2. $\chi_n^2 + \chi_m^2 = \chi_{n+m}^2$ — по определению
3. $\mathbb{E}\chi_k^2 = \frac{\lambda}{\alpha} = \frac{\frac{k}{2}}{\frac{1}{2}} = k, \mathbb{D}\chi_k^2 = \frac{\lambda}{\alpha^2} = \frac{\frac{k}{2}}{(\frac{1}{2})^2} = 2k$

6.3 Распределение Стьюдента

Определение. Пусть случайные величины $X_0, X_1 \dots X_k$ — независимы и имеют стандартное нормальное распределение. Распределением Стьюдента с k степеней свободы называется распределение случайной величины

$$t_k = \frac{X_0}{\sqrt{\frac{1}{k}(X_1^2 + \dots + X_k^2)}} = \frac{X_0}{\sqrt{\frac{1}{k}\chi_k^2}}$$

Свойства.

1. $\mathbb{E}t_k = 0$
2. $\mathbb{D}t_k = \frac{k}{k-2}$

6.4 Распределение Фишера-Снедекора

Определение. Распределение $F_{m,n}$ называется распределением Фишера-Снедекора (или ***F-распределением***) со степенями свободы m и n называется распределение случайной величины

$$f_{m,n} = \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}}$$

, где χ_n^2 и χ_m^2 — независимые случайные величины с распределением χ^2 .

Свойства.

1. $\mathbb{E}f_{m,n} = \frac{n}{n-2}$
2. $\mathbb{D}f_{m,n} = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

$$3. F_{m,n}(x) = P(f_{m,n} < X) = P\left(\frac{1}{f_{m,n}} > \frac{1}{X}\right) = P\left(f_{m,n} > \frac{1}{X}\right) = 1 - F_{n,m}\left(\frac{1}{X}\right)$$

При $n, k, m \rightarrow \infty$ эти распределения слабо сходятся к нормальному. При $n > 30$ они достаточно близки.

7 Линейные преобразования нормальных выборок

Пусть $\vec{X} = (X_1 \dots X_n)$, где $X_i \in N(0, 1)$ и независимы. Будем рассматривать линейные комбинации этого вектора. Пусть A — невырожденная матрица размера $n \times n$. Рассмотрим случайный вектор $\vec{Y} = A\vec{X}$, где координаты случайного вектора $Y_i = a_{i1}X_1 + \dots + a_{in}X_n$. Будем исследовать, что из себя представляют Y_i и их совместное распределение.

Примечание. Если $\eta = a\xi + b$, то $f_\eta(\xi) = \frac{1}{|a|} f_\xi\left(\frac{\xi-b}{a}\right)$

Теорема 9. Пусть случайный вектор \vec{X} имеет плотность распределения $f_{\vec{X}}(\vec{x})$ и A невырожденная матрица.

Тогда случайный вектор $\vec{Y} = A\vec{X} + \vec{b}$ имеет плотность

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{|\det A|} \cdot f_{\vec{X}}(A^{-1}(\vec{y} - \vec{b}))$$

Примечание. $f_{\vec{X}}(\vec{x})$ — плотность \vec{X} , если $P(\vec{x} \in B) = \int \dots \int_B f_{\vec{X}}(\vec{x}) d\vec{x}$

Доказательство.

$$\begin{aligned} P(\vec{y} \in B) &= P(A\vec{x} + \vec{b} \in B) \\ &= P(\vec{x} \in A^{-1}(\vec{y} - \vec{b})) \\ &= \int \dots \int_{A^{-1}(B - \vec{b})} f_{\vec{x}}(x) d\vec{x} \end{aligned}$$

Сделаем замену $\vec{y} = A\vec{x} + \vec{b}$. Тогда $A^{-1}(B - \vec{b})$ перейдёт в B , \vec{x} перейдёт в $A^{-1}(\vec{y} - \vec{b})$, $\vec{y} \in B$, $d\vec{x}$ перейдёт $|J|d\vec{y}$, где $J = |A^{-1}| = |A|^{-1}$

Итого:

$$= \int \dots \int_B f(A^{-1}(\vec{y} - \vec{b})) \cdot \frac{1}{|\det A|} d\vec{y} \Rightarrow f_{\vec{Y}}(\vec{y}) = \frac{1}{|\det A|} f_{\vec{X}}(A^{-1}(\vec{y} - \vec{b}))$$

□

Определение. $A = C$ — ортогональна, т.е. $C^T = C^{-1}$, $|\det C| = 1$

Теорема 10. Пусть дан случайный вектор $\vec{X} = (X_1 \dots X_n)$, где $\forall i \ X_i \in N(0, 1)$ и X_i независимы, а C — ортогональная матрица.

Тогда координаты случайного вектора $\vec{Y} = C\vec{X}$ независимы и также имеют стандартное нормальное распределение.

Доказательство. Т.к. координаты $X_i \in N(0, 1)$ и независимы, то плотность \vec{X} :

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(X_1^2 + X_2^2 + \dots + X_n^2)} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\|\vec{X}\|^2}$$

По предыдущей теореме:

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(C^T \vec{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\|C^T \vec{y}\|^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\|\vec{y}\|^2} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2} = \prod_{i=1}^n f_i(y_i)$$

Следовательно, $Y_i \in N(0, 1)$ и независимы. \square

Лемма 1 (Фишера). Пусть случайный вектор \vec{X} состоит из независимых стандартных нормальных случайных величин, $\vec{Y} = C\vec{X}$, где C — ортогональная матрица. Тогда $\forall k : 1 \leq k \leq n-1$ случайная величина

$$T(\vec{X}) = \left(\sum_{i=1}^n X_i^2 \right) - Y_1^2 - \dots - Y_k^2$$

не зависит от случайного вектора $Y_1 \dots Y_k$ и имеет распределение H_{n-k}

Доказательство. Т.к. C ортогональна:

$$\|\vec{Y}\|^2 = \|C\vec{X}\|^2 = \|\vec{X}\|^2 = X_1^2 + \dots + X_n^2 = Y_1^2 + \dots + Y_n^2$$

Отсюда

$$T(\vec{X}) = \sum_{i=1}^n Y_i^2 - Y_1^2 - \dots - Y_k^2 = Y_{k+1}^2 + \dots + Y_n^2$$

$Y_{k+1} \dots Y_n$ — независимы, имеют стандартное нормальное распределение и $T(\vec{X}) \in H_{n-k}$

$T(\vec{X})$ не зависит от $Y_1 \dots Y_k$, т.к. $Y_{k+1} \dots Y_n$ по предыдущей лемме от них не зависит. \square

Теорема 11 (основная).

- $X_1 \dots X_k$ независимы и имеют нормальное распределение с параметрами a и σ^2
- \bar{X} — выборочное среднее

- S^2 — исправленное выборочное среднее

Тогда имеют место следующие распределения:

1.

$$\sqrt{n} \cdot \frac{\bar{X} - a}{\sigma} \in N(0, 1)$$

2.

$$\sum_{i=1}^n \left(\frac{X_i - a}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} \in H_n$$

3.

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} \in H_{n-1}$$

4.

$$\sqrt{n} \cdot \frac{\bar{X} - a}{S} \in T_{n-1}$$

5. \bar{X} и S^2 — независимые случайные величины

Доказательство.

1.

$$X_i \in N(a, \sigma^2) \Rightarrow \sum_{i=1}^n X_i \in N(na, n\sigma^2) \Rightarrow \bar{X} \in N\left(a, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\sqrt{n}}{\sigma}(\bar{X} - a) \in N(0, 1)$$

2. Верно, т.к. $\frac{X_i - a}{\sigma} \in N(0, 1)$

3.

$$\begin{aligned} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 &= \sum_{i=1}^n \left(\frac{X_i - a}{\sigma} - \frac{\bar{X} - a}{\sigma} \right)^2 \\ &= \sum_{i=1}^n (z_i - \bar{z})^2 \end{aligned}$$

, где

$$z_i = \frac{X_i - a}{\sigma} \in N(0, 1), \bar{z} = \frac{\sum_{i=1}^n z_i}{n} = \frac{\sum x_i - na}{\sigma n} = \frac{\bar{X} - a}{\sigma}$$

Поэтому можем считать, что $X_i \in N(0, 1)$. Применим лемму Фишера.

$$T(\vec{X}) = \sum_{i=1}^n (X_i - \bar{X})^2 = n(\bar{X}^2 - (\bar{X})^2) = \sum_{i=1}^n X_i^2 - n(\bar{X})^2 = \sum_{i=1}^n X_i^2 - Y_1$$

, где

$$Y_1 = n(\bar{X})^2 = \sqrt{n}\bar{X} = \frac{1}{\sqrt{n}}X_1 + \dots + \frac{1}{\sqrt{n}}X_n$$

Так как длина строки $\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}$ равна 1, поэтому эту строку можем дополнить до ортогональной матрицы C . Тогда Y_1 — первая координата случайного вектора $\vec{Y} = C\vec{X}$ и по лемме Фишера $T(\vec{X}) \in H_{n-1}$

5. $T(\vec{X}) = \frac{(n-1)S^2}{\sigma^2}$ не зависит от $Y_1 = \sqrt{n}\bar{X} \Rightarrow S^2$ и \bar{X} независимы.

4.

$$\sqrt{n}\frac{\bar{X} - a}{S} = \sqrt{n}\frac{\bar{X} - a}{\sigma} \cdot \frac{1}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \cdot \frac{1}{n-1}}} = \frac{X_0}{\sqrt{\chi_{n-1}^2(n-1)}} \in T_{n-1}, \text{ т.к.:}$$

$X_0 \in N(0, 1)$ по пункту 1, $\frac{(n-1)S^2}{\sigma^2} \in H_{n-1}$ по пункту 3 и X_0 не зависит от $\frac{(n-1)S^2}{\sigma^2}$ по пункту 5.

□