Лекция 2 МС Баженов

1. Состоятельность:

Опр. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ наз-ся <u>сост-ой</u> оценкой парам-а θ, если $\theta^* \xrightarrow{P} \theta$ при $n \to \infty$

2. Несмещенность

Опр. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ наз-ся <u>несм-ой</u> оценкой параметра θ, если $E\theta^* = \theta$

Определение означает:

Это значит, что если оценка несмещ-ая то мы можем ошибиться как в меньшую так и в большую сторону с одинаковой вероятность, а если смещённая то появляется систематическая ошибка. Несмещённость достигается калибровкой прибора

Note: Оценка θ^* наз-ся асимптотически несмещённой, если

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

Означает, что если объём выборки достаточно большой то смещ-ем можем пренебречь

3 Эффективность

Опр. Оценка θ_1^* не хуже θ_2^*, если $E(\theta_1^* - \theta)^2 \le E(\theta_2^* - \theta)^2$ и если θ_1^* и θ_2^* несмещенные оценки

$$D\theta_1^* \le D\theta_2^*$$

same as variance

!

Опр:

Оценка θ^* называется эффективной если она не хуже всех ост-х оценок

Замечание: в классе всех оценок не существует эффективной оценки

Теор. В классе несмещённых оценок эффективная оценка существует и она единственна.

Доказывать через проекцию всех несмещённых оценок.

Опр. Оценка наз-ся асимптотически нормальной если она сходится по нормальному закону
(в частности скорость сходимости $1/\sqrt{n}$)
чтобы получить более точную оценку необходимо увеличить объём выборки

Точечные оценки моменты

Пусть дана выборка

$$\exists\ \overline{X} = (X_1, \dots, X_n)$$

выборка это набор независимых, одинаково распред. случайных величин.

выборку можем оценивать как случайную величину

Опр. Выборочным средним $\overline{X}$ наз-ся величина $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Опр. Выборочной дисперсией D_X^* наз-ся вел-на

Variance

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

Опр. Исправленной выбор-ой дисперсией S^2 наз-ся величина

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

Опр. Выбор-м средним квад. откл-ем наз-ся

$$\sigma_X^* = \sqrt{D_X^*}$$

Опр. Исправленным средним квад. откл-ем наз-ся $S = \sqrt{S^2}$

Опр. Выборочным „k-м" моментом $\overline{X^k}$ наз-ся вел-на

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

Опр. Выборочной модой M_0^* наз-я варианта с наибольшей частотой.

$M_0^* = X_i$, т.ч. $n_i = \max(n_1, \dots, n_k)$

Опр. Медианой M_e^* наз-я значение варианты в середине ряда

> Варианта - значение вариационного ряда

скобки когда ряд выстроили по возрастанию

$$M_e^* = \begin{cases} X_{(k)} & \text{если } n = 2k-1 \text{ - нечётный} \\ \dfrac{X_{(k)} + X_{(k+1)}}{2} & \text{если } n = 2k \text{ чётное} \end{cases}$$

Теорема 1: Выборочное среднее $\overline{X}$ является несмещённой состоят-ой оценкой для мат-го ожидания

1) Несм.: $E\overline{X} = EX = a$

2) $\overline{X} \xrightarrow{P} EX = a$

по вероятности

Док-во:

рассматриваем как нез-ые, одинаково распред. случайные величины.

1) $E\overline{X} = E\left(\frac{\overbrace{X_1 + \dots + X_n}}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{1}{n} \cdot nEX_1$

первый экземпляр

$= EX_1 = a$

2) $\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow[n \to \infty]{P} EX$

закон больших чисел

LLN $\overline{X}_n = \frac{\sum^n X_i}{n}$

$P(\mu - \varepsilon < \overline{X}_n < \mu + \varepsilon) \to 1$

Теорема 2: $\overline{X^k}$ является несмещённой состоятельной оценкой для теорет-го k-го момента X^k

1) $E\overline{X^k} = EX^k \xrightarrow[n \to \infty]{P} m_k$

2) $E\overline{X^k} \xrightarrow[n \to \infty]{P} m_k$

следует из теоремы 1 если в качестве случайной величины взять X^k.

Теорема 3: выборочной дисперсии D^* и S^2 являются состоятельными оценками для дисперсии, при этом D^* смещённая оценка (есть систем-я ошибка вниз), а S^2 несм-ая оценка.

Док-во: Заметим что $D^* = \frac{1}{n}\sum(X_i - \overline{X})^2 = \overline{X^2} - \overline{X}^2$

Посчитаем мат. ожидание оценки
и покажем смещённость выборочной оценки

$$E D^* = E(\overline{X^2} - \overline{X}^2) = E\overline{X^2} - E\overline{X}^2 =$$

$$= EX^2 - E\overline{X}^2 = \quad D\overline{X} = E\overline{X}^2 - (E\overline{X})^2 \Rightarrow$$

$$E\overline{X}^2 = D\overline{X} + (E\overline{X})^2 =$$

$$= EX^2 - (D\overline{X} + (E\overline{X})^2) = (EX^2 - (E\overline{X})^2) - D\overline{X} =$$

$$= DX - D\overline{X} = DX - D\left(\frac{X_1 + \ldots + X_n}{n}\right) =$$

$$= DX - \frac{1}{n^2} n DX_i = DX - \frac{1}{n} DX =$$

help note about Variance (дисперсия)

$$\mathrm{Var}\, X = E\left[(X - EX)^2\right] = E\left[X^2 - 2X\,EX + (EX)^2\right] =$$

$$= E\left[X^2\right] - (EX)^2$$

$\frac{n-1}{n} DX$

т.к.

$ES^2 = E\left(\frac{n}{n-1} D^*\right) = \frac{n}{n-1} \frac{n-1}{n} DX$

несмещенная оценка

Continue from the next page

Theorem: sample variance s^2 is an unbiased estimate of σ^2

$$\sum (X_i - \bar{X})^2 = \sum [(X_i - m) - (\bar{X} - m)]^2 =$$

$$= \sum (X_i - m)^2 - 2(\bar{X} - m) \sum (X_i - m) + n(\bar{X} - m)^2$$

$$= \sum (X_i - m)^2 - 2n(\bar{X} - m)^2 + n(\bar{X} - m)^2 =$$

$$= \sum (X_i - m)^2 - n(\bar{X} - m)^2$$

$$E\left[\sum (X_i - \bar{X})^2\right] = E\left[\sum (X_i - m)^2\right] - nE\left[(\bar{X} - m)^2\right] =$$

$$= nE\left[(X - m)^2\right] - nE\left[(\bar{X} - m)^2\right] =$$

$$= nV(X) - nV(\bar{X}) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

by Theorem 4 page 119

Therefore we need to devide by $(n-1)$ to get an unbiased estimate

2) Состоятельность

$$D^* = \overline{X^2} - \overline{X}^2 \xrightarrow{P} EX^2 - (EX)^2 = DX,$$

$$S^2 = \frac{n}{n-1} D^* \xrightarrow{P} DX$$

оценки сост

Замечание

т.к. $\frac{n}{n-1} \xrightarrow[n\to\infty]{} 1$, то $ED^* \xrightarrow[n\to\infty]{P} DX$

след-но выборочная дисперсия явл-ся ассимптотически несмещённой оценкой. Это значит на практике если n большое ($n \geq 100$) можно считать обычную выборочную дисперсию, а при $n < 100$ следует заменить на исправленную выборочную дисперсию.

Метод моментов (Пирсон)

Зная выборочные моменты можно дать

Пусть имеется выборка $\vec{X} = (X_1 \dots X_n)$ неизвестного распред-я, но при этом знаем, что данные распр-е опред-го типа задаваемого k параметрами $\Theta = (\Theta_1, \dots, \Theta_k)$

Зная параметры можем вычислить теор-ие kе моменты.

Если распр-е

$$m_i = \int_{\mathbb{R}} x^i f(X, \Theta_1, \dots, \Theta_k)\,dx = h_i(\Theta_1, \dots, \Theta_k)$$

Вычислим выборочные моменты и подставим в эти формулы, получаем систему уравнений

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

Находим оценки $\theta^* = \theta_2^* \dots \theta_k^*$ неизвестных параметров

Замечание. При этом обычно получаем состоятельные оценки, но смещённые.

Пример

$\exists\ X \in U(a; b)$ $a < b$

При обработки 100 данных получили оценки 1го и 2го моментов

$\overline{X} = 2{,}25$ $\overline{X^2} = 6{,}75$

Дать оценки неизвестных парам-ов a и b.

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

$$EX = \int_0^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{b^2 - a^2}{(b-a)\,2}$$

$$= \frac{a+b}{2}$$

$$EX^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = \frac{b^3 - a^2}{3(b-a)} =$$

$$= \frac{a^2 + ab + b^2}{3}$$

$$\begin{cases} 2{,}25 = \dfrac{a^* + b^*}{2} \\ 6{,}75 = \dfrac{a^{*2} + a^* b^* + b^{*2}}{3} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} a^* + b^* = 4{,}5 \\ a^{*2} + a^* b^* + b^{*2} = 20{,}2 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} a^* + b^* = 4{,}5 \\ a^* b^* = 0 \end{cases} \Leftrightarrow \begin{cases} b^* = 4{,}5 \\ a^* = 0 \end{cases}$$ т.к. $a < b$