Лекция 2

Ilya Yaroshevskiy

13 сентября

Содержание

$$\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_n)$$

Определение. **Статистикой** называется измеримая функция:

$$\Theta^* = \Theta^*(\mathbb{X}_1, \dots, \mathbb{X}_n)$$

Примечание. Пусть требуется найти оценку параметра Θ случайную величины $\mathbb{X}$ по данной выборке. Оценку будет считать с помощью некоторой статистики:

$$\Theta^* = \Theta^*(\mathbb{X}_1, \dots, \mathbb{X}_n)$$

1 Свойства статистических оценок

1. Состоятельность

   Определение. Статистика Θ^* называется **состоятельной оценкой** параметра Θ, если

   $$\Theta^* \xrightarrow{p} \Theta \text{ при } n \to \infty$$

2. Несмещенность

   Определение. Статистика Θ^* называется несмещенной оценкой параметра Θ, если

   $$E\Theta^* = \Theta$$

   Примечание. С равной вероятностью можем ошибиться как в меньшую так и в большую сторону. Нет систематической ошибки.

   Определение. Статистика Θ^* называется **асимптотически несмещенной**, если

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

3. Эффективность

   Определение. Оценка Θ_1^* **не хуже** Θ_2^*, если

   $$E(\Theta_1^* - \Theta) \le E(\Theta_2^* - \Theta)$$

   или оценки несмещенные

   $$D\Theta_1^* \le D\Theta_2^*$$

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

   Теорема 1.1. Не существует эффективной оценки в классе всех возможных оценок

   Теорема 1.2. В классе не смещенных оценок существует эффективная оценка

2 Точечные оценки моментов

Определение. Выборочным средним $\overline{\mathbb{X}_c}$ называется величина равная среднему арифметическому данных

$$\overline{\mathbb{X}_c} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{X}_i$$

Определение. Выборочная дисперсией D_c называется величина

$$D_c = \frac{1}{n}\sum_{i=1}^{n}(\mathbb{X}_1 - \mathbb{X}_c)^2$$

Определение. Исправленной выборочной дисперсией S^2 называется величина

$$S^2 = \frac{n}{n-1}D_c$$

Определение. Выборочным средним квадратическим отклонением σ_c называется величина:

$$\sigma_c = \sqrt{D_c}$$

Определение. Исправленным выборочным средним квадратическим отклонением называется величина:

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

Определение. Выборочным k-ым моментом называется величина

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

Определение. Модой M_0^* называется величина с наибольшей частотой

Определение. Медианой Me^* вариационного ряда называется значение варианты в середине ряда. Если число четное, то берем среднее от средних.

Теорема 2.1. $\overline{\mathbb{X}_c}$ является несмещенной состоятельной оценкой для математического ожидания

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

Доказательство. Доделать □

Теорема 2.2. $\overline{\mathbb{X}^k}$ является несмещенной состоятельной оценкой для теоретического k-го момента

1. $E\overline{\mathbb{X}^k} = E\mathbb{X}^k = m_k$ — несмещенность
2. $\overline{\mathbb{X}^k} \xrightarrow{p} E\mathbb{X}^K = m_k$ — состоятельность

Теорема 2.3. D_c и S^2 являются состоятельными оценками для дисперсии

1. D_c — смещенная оценка
2. S^2 — несмещенная оценка

Доказательство. Доделать □

Примечание. Пусть имеется выборка $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ неизвестного распределения, знаем тип распределения. Пусть этот тип определяется k неизвестными параметрами. Теоретическое распределение задает теоретический k-тый момент. *Например, если распределение непрерывное, то оно задается плотность*:

$$f(\mathbb{X}, \Theta_1, \ldots, \Theta_k) \text{ и } m_k = \int_{-\infty}^{\infty}\mathbb{X}^k f(\mathbb{X}, \Theta_1, \ldots, \Theta_k)d\mathbb{X} = f_k(\Theta_1, \ldots, \Theta_k)$$

Метод моментов состоит в следующем: вычисляем выборочные моменты и подставляем их в уравнение вместо теоретичских. В результате получаем систему уравнений:

$$\begin{cases} \overline{\mathbb{X}} = f_1(\Theta_1, \dots, \Theta_k) \\ \overline{\mathbb{X}}^2 = f_2(\Theta_1, \dots, \Theta_k) \\ \vdots \\ \overline{\mathbb{X}}^k = f_k(\Theta_1, \dots, \Theta_k) \end{cases}$$

Решив эту систему получим оценки параметров $\Theta_1, \dots, \Theta_k$. При этом как правило получаем оценки состоятельные но смещенные

Пример. Пусть случайная величина $\mathbb{X} \in U(a, b)$, $a < b$. Обработав стат. данные получили оценки:

$$\overline{\mathbb{X}} = 2.25 \quad \overline{X}^2 = 6.75$$

Решение.

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

$$E\mathbb{X} = \int_a^b \mathbb{X} f(\mathbb{X}) d\mathbb{X} = \dots = \frac{a+b}{2}$$

$$E\mathbb{X}^2 = \int_a^b \mathbb{X}^2 \frac{1}{b-a} d\mathbb{X} = \dots = \frac{a^2 + ab + b^2}{3}$$

$$\begin{cases} 2.25 = \frac{a+b}{2} \\ 6.75 = \frac{a^2+ab+b^2}{3} \end{cases} \quad \Leftrightarrow \dots \Leftrightarrow \begin{cases} b = 4.5 \\ a = 0 \end{cases}$$