Точечная оценка

Пусть имеется выборка объёма n: $\vec{x}(x_1, x_2, \ldots, x_n)$

Статистикой называется (измеримая - Борелевская) функция

$$\Theta^* = \Theta^*(X_1, \ldots, X_n)$$

Пусть требуется найти приближённую оценку неизвестного параметра Θ по выборке $(x_1, \ldots, x_n)$

Оценка считается при помощи некоторой статистики

Свойства статистических оценок

- состоятельность. При увеличении объёма данных повышается точность.

Статистика
$\Theta^* = \Theta^*(X_1, \ldots, X_n)$
неизвестного параметра Θ называется **состоятельной**, если $\Theta^* \to \Theta \;\; n \to \infty$

- Статистика
  $\Theta^* = \Theta^*(X_1, \ldots, X_n)$
  неизвестного параметра Θ называется **несмещённой,**если

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

  ... асимптотически несмещённой, если вместо = стоит $\to, n \to \infty$

Оценка $\Theta_1^* = \Theta_1^*(X_1, \ldots, X_n)$ не хуже оценки $\Theta_2^* = \Theta_2^*(X_1, \ldots, X_n)$, если

$$E(\Theta_1^* - \Theta*)^2 \leq E(\Theta_2^* - \Theta*)^2$$

Если это несмещённый оценки, то можно эквивалентно записать

$$DQ_1^* \leq DQ_2^*$$

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

В классе всех возможных оценок не существует эффективной оценки.

Теорема: В классе несмещённых оценок *существует* эффективная оценка причём *единственная*.

Точные оценки моментов

> **Среднее выборочное** — число
>
> $$\overline{x} = \frac{x_1 + \cdots + x_n}{n}$$
>
> **Выборочная дисперсия** — число

> $$D_B = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Исправленное выборочное среднее — число

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

Если выборка задана в виде частотного вариационного ряда

- $$\begin{array}{c|c|c|c} x_i & x_1 & \dots & x_k \\ \hline n_i & n_1 & \dots & n_k \end{array}$$

  , то удобнее использовать несколько другие формулы $X = \frac{1}{n}\sum x_i n_i$ и $D_B = \frac{1}{n}\sum (X_i - X)^2 n_i$

Выборочным средним квадратическим отклонением называется величина $\sigma^* = \sqrt{D_B}$, а исправленным $S = \sqrt{S^2}$

Выборочный k-ый момент — величина

$$X^k = \frac{1}{n}\sum X_i^k$$

Мода Mo^* вариационного ряда — варианта с наибольшей частотой. $Mo^* = x_i$, где $n_i = max(n_1, \dots, n_k)$

Медиана Me^* вариационного ряда — значение варианты в середине ряда:

- $n = 2k - 1$ - нечётный, то медина это x_k
- $n = 2k$ - чётный, то $\frac{x_k + x_{k+1}}{2}$

Note: соответствующие функции в Excel: СРЗНАЧ, ДИСП.Г, ДИСП.В, СТАНДОТКЛОН.{Г,В}, МЕДИАНА, МОДА.ОДН

Теорема 1:
Выборочное среднее является несмещённой состоятельной оценкой для математического ожидания:

- Несмещённость $E\overline{X} = EX = a$
- Состоятельность $\overline{X} \to_{n\to\infty}^{p} EX = a$
  Доказательство:
- $E\overline{X} = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum EX_i = \frac{1}{n}\cdot nEX = EX$
- $\overline{X} = \frac{\overline{X_1} + \dots + \overline{X_n}}{n} \to_{n\to\infty}^{p} EX$ по закону больших чисел Хинчина.

Теорема 2:
Выборочный k-ый момент $\overline{X^k}$ является несмещённой состоятельной оценкой теоретического k-ого момента M_k

- Несмещённость $E\overline{X^k} = m_k$

- Состоятельность $\overline{X^k} \to^p_{n\to\infty} m_k$
  Доказательство: Это следствие пред. теоремы если в качестве случайно величины взять X^k

Теорема 3:
Выборочные дисперсии являются состоятельными оценками для дисперсии.
При этом D_B это смещённая вниз оценка, а S^2 несмещённая оценка.
Доказательство:

- Смещённость. Заметим, что $D_B = \frac{1}{n}\sum(X_i - \overline{X}) = \overline{X^2} - \overline{X}^2$.
  $D\overline{X} = E\overline{X}^2 - (E\overline{X})^2$
  Будем смотреть

$$\begin{aligned}ED_B = E(\overline{X^2} - \overline{X}^2) = E\overline{X^2} - E\overline{X}^2 = \\ = E\overline{X^2} - ((E\overline{X})^2 + D\overline{X}) = EX^2 - (EX)^2 - D\overline{X} = \\ = DX - D\overline{X} = DX - D\left(\frac{1}{n}\sum X_i\right) = \\ = DX - \frac{1}{n^2}\sum DX_i = DX - \frac{1}{n^2}\cdot nDX = \\ = Dx - \frac{1}{n}DX = \frac{n-1}{n}DX\end{aligned}$$

  $ES^2 = E\left(\frac{n}{n-1}D_B\right) = \frac{n}{n-1}\cdot\frac{n-1}{n}DX = DX$
- Состоятельность:
  $D_B = \overline{X^2} - \overline{X}^2 \to^p_{n\to\infty} EX^2 - (EX)^2 = DX$
  $S^2 = \frac{n}{n-1}D_B \to DX$

Отсюда видим, что D_B асимптотически несмещённая оценка, т.к. $\frac{n}{n-1} \to 1, n\to\infty$, поэтому при больших объёмах выборки (начиная со 100) обычно считают просто выборочную дисперсию. Если объём выборки мал (30-50), то обязательно заменять её на исправленную дисперсию.

Метод моментов (Пирсона)

Пусть имеет выборка неизвестного распределения. При этом из теории знаем тип распределения и что оно определяется набором параметром $\vec{\Theta} = (\Theta_1, \dots, \Theta_n)$. Цель — дать оценку данных параметров.

Если знаем параметры распределения и его тип, то теоретические k-ые моменты можем вычислить по известным формулам. Например: Если распределение абсолютно непрерывное с плотностью $f(x, \Theta_1, \dots, \Theta_n)$, то

$$m_i = \int_{-\infty}^{\infty} X^i f(X, \Theta_1, \dots, \Theta_k)\,dx = h_i(\Theta_1, \dots, \Theta_k)$$

Суть метода: в этих уравнениях теоретические моменты заменяем их оценками, после чего находим неизвестные k параметров, решая систему из k уравнений.

$$\begin{cases}\overline{X} = h_1(\Theta_1, \dots, \Theta_k) \\ \overline{X^2} = h_2(\Theta_1, \dots, \Theta_k) \\ \dots \\ \overline{X^k} = h_k(\Theta_1, \dots, \Theta_k)\end{cases}$$

Такие оценки как правило состоятельные, но смещённые.

Пример: Пусть $X \in U(a;b), a < b$ – равномерное распределение. При обработке стат. данных получили оценки первого и второго моментов.
$\overline{X} = 2.25; \overline{X^2} = 6.75$. Дать оценки параметрам a, b.
Плотность

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

$EX = \int_a^b X \frac{1}{b-a}\, dx = \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2}$
$EX^2 = \int_a^b X^2 \frac{1}{b-a}\, dx = \frac{a^2+ab+b^2}{3}$

$$\begin{cases}\overline{X} = 2.25 = \frac{a^*+b^*}{2} \\ \overline{X}^2 = 6.75 = \frac{a^{*2}+a^*b^*+b^{*2}}{3}\end{cases} \iff \begin{cases}a^* + b^* = 4.5 \\ a^*b^* = 0\end{cases}$$

Литература:

- Как не ошибаться
- Чернова
- Логутин - Наглядная математическая статистикаю

> ## 2
>
> Известно, что данное распределение является показательным. При обработке данных получили среднее выборочное $\overline{X} = 0.86$. Дать оценку неизвестному параметру α. Является ли эта оценка а) состоятельной б) несмещённой. Если она смещена, то куда?
>
> $$f_\alpha(x) = \begin{cases}0 & , x < 0 \\ \alpha e^{-\alpha x} & , x \ge 0\end{cases}$$
>
> $EX = \frac{1}{\alpha} \implies \alpha = \frac{1}{EX} \implies \alpha^* = \frac{1}{\overline{X}} = \frac{1}{0.86} \approx 1.162$
> Состоятельность:
> $\alpha^* = \frac{1}{\overline{X}} \to^p_{n\to\infty} \frac{1}{EX} = \alpha$
> Несмещённость:
> $E\alpha^* = E\frac{1}{\overline{X}} \le \frac{1}{E\overline{X}} = \frac{1}{EX} = \alpha$ – оценка является смещённой вниз
> (по неравенству Йенсена)
>
> > **Математическая статистика с 2022-09-12**
> >
> > Посмотреть экстремумыфункций нескольких переменных, дифференциалы второго порядка, ..

> Конкурсная задача:
> Оценка математического ожидания ищется в виде
> $f\left(\vec{x}\right) = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n$
> а) При каких λ эта оценка состоятельная. б) При каких λ она будет эффективной?

$E\overline{X} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$ – частный случай линейной оценки — выборочное среднее.
$\hat{X} = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n = \sum \lambda_i x_i$
Несмещённость:

$$E(\hat{X}) = E\left(\sum_{i=1}^{n} \lambda_i x_i\right) = \sum_{i=1}^{n} E(\lambda_i x_i) = \sum_{i=1}^{n} \lambda_i E x_i =$$
$$= \sum_{i=1}^{n} \lambda_i EX = EX \cdot \sum_{i=1}^{n} \lambda_i$$
$$E(\hat{X}) = EX \implies EX \cdot \sum_{i=1}^{n} \lambda_i = EX \implies \sum_{i=1}^{n} \lambda_i = 1$$

Рассмотрим оценку, где $\lambda_i = \frac{1}{n}$. В таком случае оценка принимает вид выборочного среднего. Покажем, что такая оценка эффективная среди данного класса оценок.

$$E()$$