Výpočetní vzorec pro rozptyl

Platí

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

Odvození:

$$D[X] = E\left[(X - E[X])^2\right] = E\left[X^2 - 2E[X] \cdot X + E[X]^2\right] =$$

$$= E\left[X^2\right] - 2E[X] \cdot E[X] + E[X]^2 = E\left[X^2\right] - E[X]^2$$

kde jsme využili skutečnosti, že $E[X]$ je konstanta, kterou lze vytknout před operátor střední hodnoty a pro niž platí $E[E[X]] = E[X]$.

Pro výběrové charakteristiky platí obdobně

$$\overline{(x - \bar{x})^2} = \overline{x^2} - \bar{x}^2$$

kde $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Statistika

Statistika je funkce výběru taková, že po dosazení výběru její hodnota ukazuje na hodnotu odhadovaného parametru.

Příklad

Uvažujeme náhodnou veličinu X (tedy pokus, na kterém měříme hodnoty x_i - realizace náhodné veličiny X). Tato náhodná veličina má rozdělení $f(x)$ o kterém předpokládáme, že má exponenciální rozdělení $f(x) = a\exp\{-ax\}$, $x \geq 0$ ale hodnotu a neznáme. Provedeme náhodný výběr a získáme datovou množinu.

Konkrétně: sledovali jsme zákazníky přicházející do obchodu a měřili jsme časové mezery mezi nimi. Změřili jsme tyto údaje (ve vteřinách)

15, 21, 35, 7, 12, 38, 5, 43, 48, 32

Chceme odhadnout hodnotu parametru a.

1. Hledáme statistiku

   (a) metoda momentů

   První výběrový moment je průměr $\bar{x}$

   První souborový moment je střední hodnota

$$E[X] = \int_0^\infty x \cdot a\exp\{-ax\}\, dx = |\text{per partes}| = \frac{1}{a}$$

   Odtud (porovnáme momenty)

$$\bar{x} = \frac{1}{a} \quad \rightarrow \quad a = \frac{1}{\bar{x}}$$

   Statistika je tedy převrácená hodnota výběrového průměru

$$T(x) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{n}{\sum_{i=1}^{n} x_i}$$

(a) metoda maximální věrohodnosti

Věrohodnostní funkce

$$L(a) = \prod_{i=1}^{n} a \exp\{-ax_i\} = a^n \exp\left\{-a\sum_{i=1}^{n} x_i\right\} = a^n \exp\{-aS\}$$

kde $S = \sum_{i=1}^{n} x_i$ a při odvození jsme využili skutečnosti, že součin exponenciál je exponenciála se součtem v exponentu $\prod \exp\{ax_i\} = \exp\{\sum ax_i\}$ a a se vytkne.

Hledáme maximum L - spočteme derivaci a položíme rovno nule

$$\frac{\partial L}{\partial a} = na^{n-1}\exp\{-aS\} - a^n S \exp\{-aS\} = 0$$

$$n = aS \;\rightarrow\; a = \frac{n}{S}$$

což je stejný výsledek, jako pro metodu momentů.

2. Bodový odhad je statistika s dosazeným výběrem, tedy

$$a = \frac{n}{S} = \frac{10}{256} = 0.039$$

Ještě zkusíme ověřit vlastnosti statistiky

- nestrannost: $E[T] = \theta \;\rightarrow\; E\left[\frac{n}{S}\right] = a$

$$E\left[\frac{n}{S}\right] = \frac{n}{E[S]}$$

  a dále

$$E[S] = E\left[\sum X_i\right] = \sum E[X_i] = n\frac{1}{a}$$

  protože jsme ukázali, že střední hodnota exponenciálního rozdělení je $\frac{1}{a}$. Odtud dále máme

$$E[T] = \frac{n}{E[S]} = \frac{n}{n\frac{1}{a}} = a$$

  a to jsme právě chtěli - je to tedy nestranná statistika.

- konzistence: pro $n \to \infty$ je $D[T] = 0$

$$D[T] = \frac{D[X]}{n}$$

  Rozptyl exponenciálního rozdělení je $\frac{1}{a^2}$ - zkuste si spočítat. Bude tam 2x per-partes. Potom je

$$D[T] = \frac{1/a^2}{n}$$

  a to pro $n \to \infty$ jde k nule. Je to tedy statistika konzistentní.

- vydatnost: čím větší bude výběr, tím bude statistika vydatnější (bude mít menší rozptyl - viz předchozí výsledek).