Opakování a rozšíření pojmů z matematické statistiky

Parametrický model: Náhodný výběr $X_1, \dots, X_n$ z rozdělení s distribuční funkcí $F(x, \theta)$. Tu známe až na hodnotu neznámého parametru $\theta \in \mathbb{R}^p$.

Úlohy matematické statistiky:

- bodový odhad parametru θ
- intervalový odhad -||-
- testy hypotéz o parametru θ

Poznámka

Někdy nás místo parametru θ zajímá nějaká jeho funkce, tzv. parametrická funkce $\gamma: \mathbb{R}^p \to \mathbb{R}$.

Př.: sledujeme počet nehod v Brně za 1 den $= X_i$, necht $X_i \sim Po(\lambda)$

$\lambda =$ průměrný počet nehod za 1 den

$\gamma(\lambda) = e^{-\lambda} = P(X_i = 0)$ .. pravděpodobnost, že daný den nestane žádná nehoda

Bodové odhady

Řekneme, že $T: \mathbb{R}^n \to \mathbb{R}^p$ je bodový odhad parametru θ, jestliže $T = T(X_1, \dots, X_n)$ je měřitelná funkce náh. výběru $X_1, \dots, X_n$. Tedy T je náhodný vektor.

Vlastnosti bodových odhadů:

- T je nestranný odhad parametru θ, jestliže $ET = \theta$, $\forall \theta$.
- T je asymptoticky nestranný -||-, jestliže $\lim\limits_{n \to \infty} ET = \theta$, $\forall \theta$.
- T je konzistentní -||- , jestliže $T \xrightarrow[n\to\infty]{P} \theta$, $\forall \theta$, tj. $\forall \varepsilon > 0$ $P(\|T - \theta\| > \varepsilon) \xrightarrow{n \to \infty} 0$, $\forall \theta$.

Dále uvažujme jednorozměrný parametr $\theta \in \mathbb{R}$.

Příklad

$X_1 \sim Bi(N,p)$ je náh. výběr o rozsahu 1 z binomického rozdělení s parametry N (známé) a $p \in (0,1)$ neznámé.

Chceme odhad parametrické funkce $\frac{1}{p}$.

$$P(X_1 = x) = \binom{N}{x} p^x (1-p)^{N-x}, \quad x = 0,1,\dots,N.$$

Nechť $T = T(X_1)$ je nestranný odhad par. fce $\frac{1}{p}$.

$$ET = \underbrace{\sum_{x=0}^{N} T(x) \binom{N}{x} p^x (1-p)^{N-x}}_{\text{polynom stupně nejvýše } N \text{ (proměnné } p\text{)}} = \frac{1}{p}, \quad \forall p \in (0,1)$$

$\Rightarrow$ taková funkce T neexistuje, tedy neexistuje nestranný odhad param. funkce $\frac{1}{p}$.

Který odhad je (nej)lepší?

Nechť T_1 a T_2 jsou nestranné odhady parametru θ. Řekneme, že T_1 je vydatnější (lepší) než T_2, jestliže $DT_1 \le DT_2$, $\forall \theta$.

Řekneme, že T_1 je nejlepší nestranný odhad parametru θ, jestliže $DT_1 \le DT^*$, $\forall \theta$ a pro všechny T^* nestranné odhady parametru θ.

A co pro odhady, které nejsou nestranné?

Pak se jako měřítko přesnosti bere střední čtvercová chyba $MSE(T_1) = E(T_1 - \theta)^2 = DT_1 + (ET_1 - \theta)^2$.

T_1, T_2 jsou odhady parametru θ. T_1 je vydatnější než T_2, jestliže $MSE(T_1) \le MSE(T_2)$, $\forall \theta$.

Stejnoměrně nejlepší odhad parametru θ neexistuje!

Příklad

Nechť $X_1,\dots,X_n$ je náh. výběr z $N(\mu,\sigma^2)$, kde $\mu\in\mathbb{R}$ a $\sigma^2>0$ jsou neznámé parametry. Zajímá nás odhad parametru σ^2 (μ je rušivý parametr).

$$T_1 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad \dots \text{ výběrový rozptyl} \qquad \text{JE NESTRANNÝ}$$

$$T_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad \dots \text{ maximálně věrohodný odhad} \qquad \text{NENÍ NESTRANNÝ}$$

Který z nich je „lepší"?

$$MSE(T_1) = DT_1 = D\Big(\frac{1}{n-1}\overbrace{\sum_{i=1}^{n}(X_i-\bar{X})^2}^{\text{"}Y\text{"}}\Big) = \frac{1}{(n-1)^2}\cdot DY = \frac{\sigma^4\, 2(n-1)}{(n-1)^2} = \frac{2\sigma^4}{n-1}.$$

víme, že $\frac{Y}{\sigma^2}\sim\chi^2_{n-1} \Rightarrow$

$$EY = \sigma^2\cdot E\chi^2_{n-1} = \sigma^2(n-1)$$
$$DY = \sigma^4 D\chi^2_{n-1} = \sigma^4 2(n-1)$$

$$MSE(T_2) = DT_2 + (ET_2-\sigma^2)^2 = \frac{1}{n^2}DY + \Big(\frac{n-1}{n}\sigma^2-\sigma^2\Big)^2 = \frac{\sigma^4 2(n-1)}{n^2} + \frac{\sigma^4}{n^2} = \frac{\sigma^4(2n-1)}{n^2}.$$

$$ET_2 = E\Big(\frac{n-1}{n}T_1\Big) = \frac{n-1}{n}\cdot\sigma^2.$$

ukážeme, že $MSE(T_1) > MSE(T_2)$

$$\frac{2\cancel{\sigma^4}}{n-1} > \frac{\cancel{\sigma^4}(2n-1)}{n^2}$$

$$\cancel{2n^2} > \cancel{2n^2} - n - 2n + 1$$

$$n > \frac{1}{3}$$

Tedy střední čtvercová chyba odhadu T_2 je menší než odhadu T_1.

Metody pro odhady parametrů

I) Metoda momentů

uvažujme opět $\Theta \in \mathbb{R}^p$ p-rozměrný neznámý parametr

nechť existují obecné momenty: $\mu_k' = \mu_k'(\Theta) = EX_i^k$ pro $k = 1, 2, \dots, p$

jejich přirozenými odhady jsou výběrové momenty $M_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ pro $k = 1, 2, \dots, p$

$\tilde{\Theta}$ je odhad parametru Θ metodou momentů, jestliže řeší soustavu p rovnic o p neznámých: $\mu_k'(\tilde{\Theta}) = M_k'$, $k = 1, \dots, p$.

Příklad

Nechť $X_1, \dots, X_n$ je náhodný výběr ze ZIP rozdělení s pravd. funkcí:

$$\left.\begin{aligned} P(X=0) &= \pi + (1-\pi)e^{-\lambda} \\ P(X=x) &= (1-\pi)e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 1, 2, \dots \end{aligned}\right\}$$

kde $\pi \in [0, 1]$ a $\lambda > 0$ jsou neznámé parametry. Odhadněte je metodou momentů.

$$EX_i = (1-\pi)\cdot\lambda \overset{!}{=} M_1'$$

$$EX_i^2 = (1-\pi)(\lambda + \lambda^2) \overset{!}{=} M_2'$$

$$M_1'(1+\lambda) = M_2'$$

$$\tilde{\lambda} = \frac{M_2'}{M_1'} - 1, \quad \tilde{\pi} = 1 - \frac{(M_1')^2}{M_2' - M_1'}$$