Katedra pravděpodobnosti a matematické statistiky

Eliška Vlková a Anna Švarcová

Sekvenční intervaly dané délky

Statistický seminář

25. dubna 2024

Úvod do sekvenčních odhadů

- Mějme $\{X_i\}_{i=1}^{\infty}$ i.i.d. s hustotou $f(x,\theta)$, kde $\theta \in (\underline{\theta}, \bar{\theta})$ a θ je neznámý parametr.
- Vycházíme z dvojné posloupnosti $\{\theta_n, \varphi_n\}_{n=1}^{\infty}$, kde
  - $\theta_n = \theta_n(X_1, \ldots, X_n) \ldots$ bodový odhad
  - $\varphi_n = \varphi_n(X_1, \ldots, X_n) \ldots$ statistika pro ukončení výběru
    - $\varphi_n = 0 \ldots$ pokračujeme ve výběru
    - $\varphi_n = 1 \ldots$ ukončíme výběr
- **Sekvenčním bodovým odhadem** pro θ rozumíme

$$\theta_N = \theta_N(X_1, \ldots, X_N),$$

kde

$$N = min\{n; \varphi_n(X_1, \ldots, X_n) = 1\}.$$

Sekvenční intervalový odhad

- Budeme vycházet z trojné posloupnosti $\{L_n, U_n, \varphi_n\}_{n=1}^{\infty}$, kde

$$L_n = L_n(X_1, \dots, X_n) \quad \text{a} \quad U_n = U_n(X_1, \dots, X_n)$$

  jsou statistiky pro intervalové odhady.

- **Sekvenčním intervalovým odhadem** pro θ rozumíme

$$\Big(L_N(X_1, \dots, X_N), U_N(X_1, \dots, X_N)\Big).$$

- Sekvenční intervalový odhad (L_N, U_N) pro θ má hladinu spolehlivosti $1 - \alpha$, pokud

$$P\Big(\theta \in (L_N, U_N); \theta\Big) = 1 - \alpha.$$

Využití

- intervalový odhad dané délky a hladiny
- bodový odhad s rozptylem menším nebo rovným danému číslu
- při částečné znalosti o rozdělení

Sekvenční intervalové odhady dané délky

- Uvažujme předpoklady zmíněné na začátku.
- Cíl: zkonstruovat sekvenční intervalový odhad pro θ dané délky $2d$, kde $d > 0$, s hladinou spolehlivosti $1 - \alpha$.
- Předpokládejme

$$
\begin{aligned}
&(A1) \quad L_n = L_n(X_1, \dots, X_n) < U_n = U_n(X_1, \dots, X_n) \quad \text{sk.j.}, \\
&(A2) \quad \sqrt{n}(U_n - L_n) \longrightarrow 2u_{1-\frac{\alpha}{2}} A \quad \text{sk.j. pro} \quad n \longrightarrow \infty \quad \text{a} \quad A > 0, \\
&(A3) \quad \sqrt{n}(L_n - \theta) - Z_n A + u_{1-\frac{\alpha}{2}} A \longrightarrow 0, \quad \text{sk.j. pro} n \longrightarrow \infty,
\end{aligned}
$$

kde $Z_n \dots$ standardizovaný součet i.i.d. náhodných veličin s konečným druhým momentem.

Sekvenční intervalové odhady

- Pravidlo pro ukončení výběru

$$N(d) = min\{n; U_n - L_n \leq 2d\}, \quad d > 0.$$

- Za hledaný sekvenční interval budeme uvažovat

$$\left(L_{N(d)}, U_{N(d)}\right).$$

- Dále budeme uvažovat předpoklad

  $(A4)$ Náhodné veličiny $\{N(d)d^2; d > 0\}$ jsou stejnoměrně integrovatelné.

Věta. Vlastnosti

Za výše uvedených předpokladů platí:

a) $N(d) < +\infty$ sk.j., $EN(d) < +\infty$ pro. vš. $d > 0$,
$N(d)$ je nerostoucí funkce d , $\lim_{d\to 0} N(d) = +\infty$ sk.j.,

$$\lim_{d\to 0} EN(d) = +\infty,$$

b) $$\lim_{d\to 0} N(d)d^2 = u^2_{1-\frac{\alpha}{2}} A^2, \quad \text{sk.j.},$$

$$\lim_{d\to 0} EN(d)d^2 = u^2_{1-\frac{\alpha}{2}} A^2,$$

c) $$\lim_{d\to 0} P\left(L_{N(d)} \le \theta \le U_{N(d)}\right) = 1 - \alpha.$$

Příklad

- Nechť $\{X_i\}_{i=1}^{\infty}$ je posloupnost i.i.d. náhodných veličin s rozdělením $N(\mu, \sigma^2)$, kde jsou μ, σ^2 neznámé.
- Cíl: zkonstruovat interval spolehlivosti pro μ s délkou $\leq 2d$ a hladinou spolehlivosti $1 - \alpha$.
- Položme

$$L_n = \overline{X_n} - \frac{S_n u_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \quad U_n = \overline{X_n} + \frac{S_n u_{1-\frac{\alpha}{2}}}{\sqrt{n}},$$

kde

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{a} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2 + \frac{1}{n}.$$

Sekvenční interval pro lokační parametr

- Uvažujme posloupnost $X_1, X_2, \dots$ i.i.d. náhodných veličin se symetrickým rozdělením s d.f. $F(x - \theta)$.
- Označme $I_n = I_n(X_1, \dots, X_n)$ interval spolehlivosti pro θ a D_n jeho délku.
- Cíl: zkonstruovat interval spolehlivosti pro θ s hladinou spolehlivosti $1 - \alpha$ (pro $n \to \infty$), takový že

$$D_n \leq 2d, \quad \text{pro } d > 0 \text{ dané.}$$

- Mějme odhad T_n parametru θ, který je konzistentní a translačně ekvivariantní.
- Obvykle má $\sqrt{n}(T_n - \theta)$ asymptotické normální rozdělení (pro $n \to \infty$), značme $N(0, \sigma^2(T, F))$.

Sekvenční interval pro lokační parametr

- Dostáváme

$$\lim_{n\to\infty} P_\theta\left(T_n - \frac{1}{\sqrt{n}}\sigma u_{1-\frac{\alpha}{2}} \le \theta \le T_n + \frac{1}{\sqrt{n}}\sigma u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

- Uvažujme interval

$$I_n = (T_n - d, T_n + d).$$

- Pokud víme $\sigma^2 = \sigma^2(T, F))$, můžeme položit

$$n = n_d = \frac{u^2_{1-\frac{\alpha}{2}}\sigma^2(T, F)}{d^2}.$$

- Tedy dostáváme

$$\lim_{d\to 0} P_\theta(\theta \in I_{n_d}) = 1-\alpha.$$

Sekvenční interval pro lokační parametr

- Avšak F ani $\sigma^2(T, F)$ neznáme
- Využijeme odhad $S_n^2 = S_n^2(X_1, \ldots, X_n)$ a n_d nahradíme

$$N(d) = min\{n \geq n_0 : n \geq \frac{u_{1-\frac{\alpha}{2}}^2 S_n^2}{d^2}\},$$

kde n_0 je počáteční velikost výběru.

- Budeme uvažovat interval

$$I_{N(d)} = (T_{N(d)} - d, T_{N(d)} + d).$$

- Platí navíc vlastnosti

$$\frac{N(d)}{n_d} \xrightarrow{d \to 0} 1 \quad \text{v pravd. nebo sk.j.},$$

$$E\left(\frac{N(d)}{n_d}\right) \xrightarrow{d \to 0} 1.$$

- Máme-li odhad T_n zadefinovaný implicitně jako řešení rovnice.
- Například M-odhad získaný jako řešení

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i - t) = 0, \quad \text{pro } t \in \mathbb{R}.$$

- $V_n(\theta)$ má as. normální rozdělení s rozptylem ρ^2.
- Uvažujme interval $I_n = (\theta_n^-, \theta_n^+)$, kde
  $\theta_n^- = sup\{t : V_n(t) > r_n u_{1-\frac{\alpha}{2}}\}, \quad \theta_n^+ = inf\{t : V_n(t) < r_n u_{1-\frac{\alpha}{2}}\}$ a
  r_n^2 je odhad ρ^2.
- Pravidlo pro ukončení

$$N(d) = min\{n \geq n_0 : D_n = \theta_n^+ - \theta_n^- \leq 2d\}.$$

Chow- Robbins sekvenční konfidenční interval

- Mějme posloupnost $X_1, X_2, \ldots$ i.i.d. náhodných veličin s d.f. $F(x-\theta)$ t.ž.

$$0 < \int_{-\infty}^{\infty} x^2\, dF(x) = \sigma^2 < \infty \quad \text{a} \quad \int_{-\infty}^{\infty} x\, dF(x) = 0.$$

- Mějme $\overline{X_n}$ a S_n. Poté máme konfidenční interval

$$I_{N(d)} = (\overline{X_{N(d)}} - d, \overline{X_{N(d)}} + d),$$

kde N(d) je pravidlo pro ukončení

$$N(d) = min\{n \geq n_0 : n \geq \frac{u^2_{1-\frac{\alpha}{2}} S_n^2}{d^2}\}.$$

- Platí všechny zmíněné vlastnosti.

- Pokud navíc platí

$$\mu_4 = \int_{-\infty}^{\infty} x^4 \, dF(x) < +\infty,$$

pak

$$\lim_{d \to 0} P_\theta(\sqrt{N(d)} - \sqrt{n_d} \leq x) = \Phi\left(\frac{x}{\rho}\right),$$

kde

$$\rho^2 = \frac{1}{4}\left(\frac{\mu_4}{\sigma^2} - 1\right).$$

Sekvenční konfidenční interval založen na M-odhadu

- M-odhadem M_n značíme řešení rovnice

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i - t) = 0,$$

kde ψ je neklesající funkce a platí $\psi(-x) = -\psi(x)$.

- Zároveň předpokládáme, že $\psi = \psi_1 + \psi_2$, kde ψ_1 je absolutně spojitá neklesající funkce a ψ_2 je skoková neklesající funkce.

- Předpokládejme, že

$$0 < \gamma = \int_{\mathbb{R}} \psi_1'(x)\, dF(x) + \int_{\mathbb{R}} f(x)\, d\psi_2(x),$$

kde $f(x) = \frac{dF}{x}$.

Sekvenční konfidenční interval založen na M-odhadu

- Navíc předpokládejme, že
$$\sigma_0^2 = \int \psi^2(x)\,dF(x) < +\infty.$$
- $V_n(\theta)$ má as. normální rozdělení $N(0, \sigma_0^2)$
- Označme
$$\theta_n^- = sup\{t : V_n(t) > S_n u_{1-\frac{\alpha}{2}}\} \quad \text{a} \quad \theta_n^+ = inf\{t : V_n(t) < S_n u_{1-\frac{\alpha}{2}}\},$$
kde
$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} \psi^2(X_i - M_n), \quad n = 1, 2, \ldots.$$
- Konfidenční interval má tvar $I_{N(d)} = (\theta^+_{N(d)}, \theta^-_{N(d)})$, kde N(d) je pravidlo pro ukončení
$$N(d) = min\{n \geq n_0 : D_n = \theta_n^- - \theta_n^+ \leq 2d\}.$$

Sekvenční konfidenční interval založen na M-odhadu

- Poté $n_d = \frac{d^{-2} u_{1-\frac{\alpha}{2}}^2 \sigma_0^2}{\gamma^2}$.
- Pravidlo ukončení splňuje všechny vlastnosti, kromě

$$E\left(\frac{N(d)}{n_d}\right) \xrightarrow{d \to 0} 1.$$

- Pokud navíc $\psi = \psi_1$ a $\int_{\mathbb{R}} \psi^4(x)\, dF(x) < +\infty$, pak

$$\mathcal{L}(\sqrt{N(d)} - \sqrt{n_d}) \xrightarrow{d \to 0} N(0, \sigma^{*2}).$$

Sekvenční konfidenční interval založen na M-odhadu

- Pokud ovšem skoková funkce ψ_2 nezmizí, pak se as. rozdělení změní následovně:

$$\mathcal{L}\left(\sqrt{d}(\sqrt{N(d)} - \sqrt{n_d})\right) \xrightarrow{d \to 0} N(0, \frac{2\rho^2}{\gamma^2}).$$

- Skoková funkce má tvar

$$\psi_2(x) = \begin{cases} \alpha_0 & pro \quad \text{-}\infty < x < a_1, \\ \alpha_j & pro \quad a_j < x < a_{j+1}, \quad j = 1, \ldots, p-1, \\ \alpha_p & pro \quad a_p < x < +\infty. \end{cases}$$

.

- Potom $\rho^2 = \sum_{j=1}^{p} (\alpha_j - \alpha_{j-1})^2 f(a_j)$.

Děkujeme za pozornost!