

Cvičení z NSTP022
9. týden cvičení (15. – 19. 4. 2013)

CLV, bodový odhad

1. Pojišťovna má pojištěno 1 000 osob stejného věku. Pravděpodobnost úmrtí v daném roce je u každého pojištěného 0,01. Pojištěnci platí roční pojistné 1 200 EUR a v případě úmrtí je oprávněně osobě vyplaceno 80 000 EUR.
 - (a) Jaká je pravděpodobnost, že pojišťovna utrpí v daném roce ztrátu?
 - (b) Jaký je v daném roce očekávaný zisk (resp. ztráta) pojišťovny?
 - (c) Kolik by musela mít pojišťovna klientů, aby s pravděpodobností alespoň 0,99 vydělala alespoň 10 000 EUR?

2. Nechť $\{X_n\}_{n=1}^{\infty}$ jsou nezávislé náhodné veličiny takové, že

$$P(X_n = n^\lambda) = P(X_n = -n^\lambda) = \frac{1}{2} \quad (\lambda \in \mathbb{R}).$$

- (a) Rozhodněte, pro která $\lambda \in \mathbb{R}$ splňuje posloupnost $\{X_n\}_{n=1}^{\infty}$ silný zákon velkých čísel. Zapište explicitně jeho znění.
 - (b) Rozhodněte, pro která $\lambda \geq 0$ splňuje posloupnost $\{X_n\}_{n=1}^{\infty}$ předpoklady CLV (Ljapunovova podmínka). Zapište explicitně její znění.
3. Nechť X_1, \dots, X_n je náhodný výběr z exponenciálního rozdělení se střední hodnotou $1/\lambda$, kde $\lambda > 0$ je neznámé.
 - (a) Odvoďte odhad T_n parametru λ metodou maximální věrohodnosti.
 - (b) Je T_n konzistentní odhad parametru λ ?
 - (c) Je odhad T_n nestranný (resp. asymptoticky nestranný)? Pokud není, jak jej musíme upravit, abychom dostali nestranný odhad V_n parametru λ ?
Nápověda: Součet n nezávislých veličin s $\text{Exp}(\lambda)$ rozdělením má gama rozdělení s hustotou $f_n(x) = [\lambda^n / (n-1)!] x^{n-1} e^{-\lambda x}$, $x > 0$.
 - (d) Spočítejte rozptyl odhadů V_n a T_n .
 - (e) Který z odhadů V_n a T_n parametru λ je „lepší“ a proč?

4. Nechť X_1, \dots, X_n je náhodný výběr z rovnoměrného rozdělení na intervalu $(0, \theta)$, kde $\theta > 0$ je neznámé.
 - (a) Najděte maximálně věrohodný odhad T_n parametru θ a rozhodněte o nestrannosti a konzistenci tohoto odhadu.
 - (b) Nechť $n \geq 2$. Je odhad $U_n = X_1 + X_2$ nestranný a konzistentní?

5. Nechť X_1, \dots, X_n je náhodný výběr z absolutně spojitého rozdělení s hustotou f , kde

$$f(x) = \frac{1}{\sqrt{\pi b}} \exp\left\{-\frac{x^2}{b}\right\}, \quad x \in \mathbb{R}$$

a $b > 0$ je neznámý parametr.

- (a) Odhadněte parametr b metodou maximální věrohodnosti.
 - (b) Zjistěte, zda je tento odhad nestranný a konzistentní.

Opakování z přednášky

Ljapunovova centrální limitní věta: Necht X_1, \dots, X_n jsou nezávislé náhodné veličiny. Označme

$$\mu_i = \mathbb{E} X_i, \quad \sigma_i^2 = \text{var} X_i, \quad \rho_i^3 = \mathbb{E} |X_i - \mathbb{E} X_i|^3, \quad \text{pro } i = 1, \dots, n.$$

Jestliže $\lim_{n \rightarrow \infty} \frac{(\sum_{i=1}^n \rho_i^3)^{1/3}}{(\sum_{i=1}^n \sigma_i^2)^{1/2}} = 0$, pak pro každé $x \in \mathbb{R}$ platí $\mathbb{P} \left(\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq x \right) \xrightarrow{n \rightarrow \infty} \Phi(x)$. (9.1)

Bodový odhad. Necht X_1, \dots, X_n je náhodný výběr (nezávislé stejně rozdělené veličiny) z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$. Odhadem parametru θ (nebo obecněji parametrické funkce $g(\theta)$) je libovolná borelovská funkce náhodného výběru X_1, \dots, X_n , jejíž funkční předpis nezávisí na θ . **Odhad** $T_n = T_n(X_1, \dots, X_n)$ je tedy **náhodná veličina**.

V praxi pak pracujeme pouze s odhady, které mají „pěkné“ vlastnosti.

„Pěkné“ vlastnosti.

- Řekneme, že odhad $T_n = T_n(X_1, \dots, X_n)$ je **nestranný** odhad parametrické funkce $g(\theta)$, jestliže

$$\mathbb{E} T_n = \mathbb{E}_\theta T_n = g(\theta), \quad \text{pro všechna } \theta \in \Theta. \quad (9.2)$$

Odhad T_n je **asymptoticky nestranný** odhad $g(\theta)$, jestliže $\mathbb{E}_\theta T_n \rightarrow g(\theta)$ při $n \rightarrow \infty$ pro všechna $\theta \in \Theta$.

- Řekneme, že odhad $T_n = T_n(X_1, \dots, X_n)$ je **konzistentní** odhad parametrické funkce $g(\theta)$, jestliže $T_n \rightarrow g(\theta)$ s.j. pro $n \rightarrow \infty$ pro všechna $\theta \in \Theta$, t.j.

$$\mathbb{P}_\theta \left(\lim_{n \rightarrow \infty} T_n = g(\theta) \right) = 1 \quad \text{pro všechna } \theta \in \Theta. \quad (9.3)$$

Stručný popis konstrukce odhadu metodou maximální věrohodnosti. Necht X_1, \dots, X_n je náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$. Předpokládejme, že toto rozdělení má hustotu $f(x, \theta)$ vzhledem k nějaké σ -koněčné míře ν (ν bude čítací míra v případě diskrétního rozdělení a Lebesguova míra v případě spojitého rozdělení). Odhad $T_n = T_n(X_1, \dots, X_n)$ je **odhad metodou maximální věrohodnosti**, jestliže T_n maximalizuje tzv. věrohodnost

$$L_n(\theta) = \prod_{i=1}^n f(X_i, \theta) \quad (9.4)$$

přes všechna $\theta \in \Theta$. Ekvivalentně, T_n maximalizuje logaritmicou věrohodnost

$$l_n(\theta) = \log [L_n(\theta)] = \sum_{i=1}^n \log [f(X_i, \theta)]. \quad (9.5)$$

Máme tedy $T_n = \arg \max_{\theta \in \Theta} L(\theta) = \arg \max_{\theta \in \Theta} l(\theta)$.

Za určitých dalších předpokladů lze T_n najít jako řešení věrohodnostní rovnice $\frac{d}{d\theta} l_n(\theta) = 0$, tj.

$$\frac{d}{d\theta} \sum_{i=1}^n \log [f(X_i, \theta)] = 0. \quad (9.6)$$

Výsledky

- (a) pravděpodobnost ztráty je 0,056, (b) očekávaný zisk je 400 000 EUR, (c) potřebný počet klientů je alespoň 2193.
- (a) Pro $\lambda < 1/2$ platí $\bar{X}_n \rightarrow 0$ s.j.
(b) Pro $\lambda \geq 0$ platí $Z_n = \frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n k^{2\lambda}}} \stackrel{asympt.}{\sim} \mathbf{N}(0, 1)$.
- (a) $T_n = 1/\bar{X}_n = n/(\sum_{i=1}^n X_i)$,
(b) T_n je konzistentní odhad λ ,
(c) T_n není nestranný odhad λ , protože $\mathbf{E} T_n = \frac{n\lambda}{n-1}$;
 $V_n = \frac{(n-1)T_n}{n} = \frac{n-1}{\sum_{i=1}^n X_i}$ je nestranný odhad λ ,
(d) pro $n > 2$ je $\mathbf{var} T_n = \lambda^2 \frac{n^2}{(n-2)(n-1)^2}$, $\mathbf{var} V_n = \frac{(n-1)^2}{n^2} \mathbf{var} T_n = \frac{\lambda^2}{n-2}$,
(e) V_n je lepší, jelikož je nestranný (navíc má i menší rozptyl).
- (a) $T_n = \max_{1 \leq i \leq n} X_i$, $\mathbf{E} T_n = \frac{n\theta}{n+1}$, T_n tedy není nestranným odhadem parametru θ . T_n je však konzistentním odhadem parametru θ .
(b) Odhad U_n je nestranný, ale není konzistentní.
- $\hat{b}_n = \frac{2}{n} \sum_{i=1}^n X_i^2$. Odhad je nestranný i konzistentní.