EMPIRICKÉ ODHADY
9.11.2017

1. Nechť $X_1, \dots, X_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$, $\lambda > 0$. Naším cílem je odhadnout hodnotu distribuční funkce v nějakém daném bodě $y > 0$, tj. $\theta_X = 1 - \mathrm{e}^{-\lambda y}$. Uvažujte následující dva odhady

$$\widetilde{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq y\} \qquad \widehat{\theta}_n = 1 - \mathrm{e}^{-\widehat{\lambda}_n y}, \qquad \text{kde} \quad \widehat{\lambda}_n = \frac{1}{\overline{X}_n}.$$

   (a) Jaké jsou vlastnosti daných dvou odhadů?
   (b) Který z odhadů se Vám zda vhodnější a proč?

2. V následující tabulce jsou zachyceny porodní hmotnosti chlapců (narozených v daném roce v daném regionu).

Hmotnost [kg]	$(1.5, 2.0]$	$(2.0, 2.5]$	$(2.5, 3.0]$	$(3.0, 3.5]$	$(3.5, 4.0]$	$(4.0, 4.5]$	$(4.5, 5.0]$	$(5.0, 5.5]$	$\sum$
Počet	4	101	769	1 904	1 651	369	37	3	4 838

   (a) Bodově i intervalově odhadněte o kolik je větší pravděpodobnost, že se narodí chlapec s hmotností do 4 kg (včetně) než pravděpodobnost že se narodí chlapec s hmotností do 3 kg (včetně).
   (b) Bodově i intervalově odhadněte, kolikrát je větší pravděpodobnost, že se narodí chlapec s hmotností do 4 kg (včetně) než pravděpodobnost že se narodí chlapec s hmotností do 3 kg (včetně).

3. Nechť $X_1, \dots, X_n$ je náhodný výběr.

   (a) Najděte asymptotické rozdělení třetího centrálního momentu $\widehat{\mu}_3 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^3$.
   (b) Na základě znalosti z (a) sestavte intervalový odhad pro μ_3.
   (c) Najděte asymptotické rozdělení třetího centrálního momentu $\widehat{\mu}_3$ za předpokladu, že X_i má normální rozdělení $\mathsf{N}(\mu, \sigma^2)$.

4. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení se spojitou distribuční funkcí F_X. Označme $\widetilde{m}$ medián rozdělení F_X. Najděte největší možné přirozené číslo k_1 a nejmenší možné přirozené číslo k_2 takové, že
$$\mathsf{P}\big(X_{(k_1)} > \widetilde{m}\big) \leq \tfrac{\alpha}{2}, \qquad \mathsf{P}\big(X_{(k_2)} < \widetilde{m}\big) \leq \tfrac{\alpha}{2}.$$
   Ukažte, že $\big(X_{(k_1)}, X_{(k_2)}\big)$ je intervalový odhad pro $\widetilde{m}$ se spolehlivostí alespoň $1 - \alpha$.

5. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení se spojitou distribuční funkcí F_X. Nechť $u_X(\beta)$ je β-kvantilem rozdělení F_X a $\widehat{u}_n(\alpha)$ je empirický (výběrový) α-kvantil.

   (a) Vyjádřete (přibližně) pravděpodobnost $\mathsf{P}\big(\widehat{u}_n(\alpha) > u_X(\beta)\big)$.
   (b) Vyčíslete pravděpodobnost z (a) pro $\alpha = 0.25$, $\beta = 0.30$ a $n = 100$ a pro $n = 1\,000$.

Opakování z přednášky

- Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s distribuční funkcí F_X. Funkci

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \le x\} \qquad \text{pro } x \in \mathbb{R}$$

nazýváme empirickou distribuční funkcí náhodného výběru $X_1, \ldots, X_n$.
- Zjevně $n\widehat{F}_n(x) \sim \mathsf{Bi}(n, F_X(x))$ a pro všechna $x \in \mathbb{R}$ platí

$$\sqrt{n}\left(\widehat{F}_n(x) - F_X(x)\right) \xrightarrow{D} \mathsf{N}(0, F_X(x)(1 - F_X(x))).$$

- Nechť $\alpha \in (0,1)$ je dáno. Kvantilová funkce rozdělení F_X je definována jako

$$F_X^{-1}(\alpha) = \inf\{x : F_X(x) \ge \alpha\}$$

a α-kvantilem rozdělení F_X rozumíme číslo $u_X(\alpha) = F_X^{-1}(\alpha)$.
- Výběrový (empirický) kvantil pro $\alpha \in (0,1)$ definujeme jako jako $\hat{u}_n(\alpha) = \widehat{F}_n^{-1}(\alpha)$, kde $\widehat{F}_n$ je empirická distribuční funkce.
Speciálně platí, že $\hat{u}_n(\alpha) = X_{(k_\alpha)}$, kde $k_\alpha = \alpha n$, pokud αn je celé číslo, a $k_\alpha = \lfloor n\alpha \rfloor + 1$ pokud αn není celé číslo.
- Pro $\alpha = 1/2$ dostáváme výběrový medián

$$\widehat{m}_n = \begin{cases} X_{(n+1)/2} & n \text{ liché}, \\ X_{n/2} & n \text{ sudé}. \end{cases}$$