Empirické odhady

6. cvičení

1. Nechť $X_1, \ldots, X_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$ (tj. s hustotou $f(x;\lambda) = \lambda \mathrm{e}\{-\lambda x\}\mathbb{I}\{x > 0\}$), kde $\lambda > 0$ je neznámý parametr. 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
$$\widetilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_i \leq y\} \quad \text{a} \quad \widehat{\theta}_n = 1 - \mathrm{e}^{-\widehat{\lambda}_n y}, \quad \text{kde} \quad \widehat{\lambda}_n = \frac{1}{\overline{X}_n}.$$

   (a) Jaké jsou vlastnosti daných dvou odhadů?
   (b) Který z odhadů se Vám zdá vhodnější a proč?
   (c) Nalezněte intervalový odhad pro parametr $1 - \mathrm{e}^{-2\lambda}$.
   (d) Nalezněte intervalový odhad pro parametr $\frac{1}{1-\mathrm{e}^{-2\lambda}}$.

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 $X_1, \ldots, X_n$ 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 $\mu_3 = \mathsf{E}\,(X_i - \mathsf{E}X_i)^3$.
   (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 $X_1, \ldots, X_n$ 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) Pro $\alpha = 0.25$, $\beta = 0.30$ a $n = 100$ spočtěte (přibližně) pravděpodobnost $\mathsf{P}\big(\widehat{u}_n(\alpha) > u_X(\beta)\big)$.
   (b) Jaká bude pravděpodobnost z (a) pro $n = 1\,000$?

OPAKOVÁNÍ Z PŘEDNÁŠKY

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

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

nazýváme empirická distribuční funkce náhodného výběru $X_1, \ldots, X_n$. Pro tuto funkci platí (pro pevné $x \in \mathbb{R}$)

* $n\widehat{F}_n(x) \sim \mathsf{Bi}(n, F_X(x))$;
* $\mathsf{E}\,\widehat{F}_n(x) = F_X(x)$;
* $\mathsf{var}\left(\widehat{F}_n(x)\right) = F_X(x)(1 - F_X(x))/n$;
* $\sqrt{n}\big(\widehat{F}_n(x) - F_X(x)\big) \xrightarrow[n\to\infty]{\mathsf{d}} \mathsf{N}\big(0, F_X(x)(1 - F_X(x))\big)$.

KVANTILY Uvažujme rozdělení dané distribuční funkcí F_X. Funkci

$$F_X^{-1}(\alpha) = \inf\{x\colon F_X(x) \geq \alpha\} \qquad \text{pro } \alpha \in (0,1)$$

nazýváme kvantilová funkce rozdělení F_X. Hodnotu

$$u_X(\alpha) = F_X^{-1}(\alpha) \qquad \text{pro } \alpha \in (0,1)$$

nazýváme α-kvantilem rozdělení F_X.

Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s distribuční funkcí $F_X(u)$, a nechť $\alpha \in (0,1)$. Potom

$$\widehat{u}_n(\alpha) = \widehat{F}_n^{-1}(\alpha) \qquad \text{pro } \alpha \in (0,1)$$

nazýváme empirickým (výběrovým) α-kvantilem rozdělení F_X.

Empirický (výběrový) α-kvantil $\widehat{u}_n(\alpha)$ můžeme pomocí pořádkových statistik spočítat následovně. Označme

$$k_\alpha = \begin{cases} \alpha n & \text{pokud } \alpha n \text{ je celé číslo,} \\ \lfloor \alpha n \rfloor + 1 & \text{jinak.} \end{cases}$$

Empirický (výběrový) α-kvantil $\widehat{u}_n(\alpha)$ pak definujeme jako k_α-tou pořádkovou statistiku náhodného výběru $X_1, \ldots, X_n$, tedy $\widehat{u}_n(\alpha) = X_{(k_\alpha)}$.