NMFM301 – Statistika pro finanční matematiky

Pořádkové statistiky. Nestrannost a konsistence odhadů

Teoretické cvičenie #2 | Zimní semestr 2024/2025

Pro náhodný výběr $X_1, \dots, X_n$ z nějakého rozdelení F definujeme uspořádaný výběr $X_{(1)}, \dots, X_{(n)}$, kde $X_{(1)}$ je tzv. první pořadová statistika náhodného výběru (t.j. minimum $X_1, \dots, X_n$) a $X_{(n)}$ je n-tá pořadová statistika náhodného výběru (t.j. maximum $X_1, \dots, X_n$). Obecně platí, že náhodná veličina $X_{(k)}$ je teda k-tou nejmenší hodnotou v realizovaném náhodném výběru $X_1, \dots, X_n$.

Špeciálně v případě, že $n = 2k+1$ pro nějaké $k \in \mathbb{N}$, pak definujeme výběrový medián náhodného výberu $X_1, \dots, X_n$ jako $X_m = X_{(k+1)}$ (t.j. prostřední hodnota v uspořádaném výberu $X_{(1)}, \dots, X_{(n)}$). Ak je rozsah náhodného výběru sudý (t.j. $n = 2k$ pro nějaké $k \in \mathbb{N}$), pak je výběrový medián definován jako průměr dvou prostředních pozorovaní, t.j. $X_m = (X_{(k)} + X_{(k+1)})/2$.

A Příklady na cvičení

Nechť $X_1, \dots, X_n$ je náhodný výběr ze spojitého rozdělení s distribuční funkcí F a hustotou f.

A1. Nechť $n = 2k+1$.

- Najděte hustotu prostředního pozorování, t.j. výběrového mediánu $X_{(k+1)}$.
- Nechť X_i má rovnoměrné rozdělení na intervalu $(0,1)$. Spočítejte $\mathsf{E}\, X_{(k+1)}$ a $\mathsf{var}\, X_{(k+1)}$.
- Nechť $X_i \sim \mathsf{R}(0, \theta)$. Je $X_{(k+1)}$ nestranným a/nebo konsistentním odhadem mediánu rozdělení $\mathsf{R}(0, \theta)$? [Použijte tvrzení P.7.5]

A2. Nechť X_i má exponenciální rozdělení s parametrem 1.

- Definujte
$$Z_1 = nX_{(1)}, \quad Z_k = (n-k+1)(X_{(k)} - X_{(k-1)}), \quad k = 2, \dots, n.$$
Ukažte, že $Z_1, \dots, Z_n$ jsou nezávislé stejně rozdělené náhodné veličiny s rozdělením $\mathsf{Exp}(1)$.
- Vyjádřete $X_{(r)}$ pomocí lineární kombinace veličin $Z_1, \dots, Z_n$ a pomocí tohoto vztahu spočítejte $\mathsf{E}\, X_{(r)}$ a $\mathsf{var}\, X_{(r)}$ (pro libovolné $r = 1, \dots, n$).
- Nechť $X_i \sim \mathsf{Exp}(\lambda)$ a $n = 2k+1$. Je $X_{(k+1)}$ nestranným a/nebo konsistentním odhadem mediánu rozdělení $\mathsf{Exp}(\lambda)$?

A3. Nechť X_i má rozdělení $\mathsf{R}(\theta_1, \theta_2)$. Najděte nestranné odhady parametrů θ_1 a θ_2 založené na maximu $X_{(n)}$ a minimu $X_{(1)}$.

A4. Pro náhodný výběr $X_1, \dots, X_n$ z Poissonova rozdělení s parametrem $\lambda > 0$ ukažte, že $\widehat{p}_0 = (1 - \frac{1}{n})^{n\bar{X}_n}$ je nestranným a konzistentním odhadem $p_0 = \mathbb{P}[X_1 = 0]$.

B Doplňující příklady (nahrazování, samostatné procvičování)

Z následujúcich príkladov je potrebné samostatne spočítať **aspoň dva** príklady (podľa vlastného výberu) a riešenie zaslať emailom na adresu cvičiaceho (`[hlavka,maciak]@karlin.mff.cuni.cz`), prípadne doručiť osobne na začiatku tretieho cvičenia.

B1. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení $\mathsf{R}(0, \theta)$. Zjistěte, zdali $X_{(n)}$ je nestranným a/nebo konsistentním odhadem parametru θ.

B2. Uvažujme náhodný výběr $X_1, \dots, X_n$ s hustotou

$$f(x) = 3\theta^{-3} x^2 1_{(0,\theta)}(x), \quad x \in \mathbb{R},\ \theta > 0.$$

(a) Ověřte, že $\widehat{\theta}_n = \frac{4}{3}\overline{X}_n$ je nestranný odhad parametru θ.

(b) Ověřte, že $\widetilde{\theta}_n = \frac{3n+1}{3n} X_{(n)}$ je nestranný odhad parametru θ.

(c) *[Pro nahrazování nepovinný]* Najděte rozptyl $\widehat{\theta}_n$ a $\widetilde{\theta}_n$ a porovnejte rychlost konvergence rozptylů k 0 při $n \to \infty$.

B3. Nechť X_i má rozdělení $\mathsf{Alt}(p)$. Najděte nestranný odhad parametru $\theta = p(1-p)$ založený na $\overline{X}_n$.

B4. Nechť X_i má rozdělení $\mathsf{Exp}(\lambda)$. Ukažte, že

$$\widehat{\theta}_n = 1 - \left(1 - \frac{u}{n\overline{X}_n}\right)^{n-1}$$

je nestranným odhadem parametru $\theta = 1 - \mathrm{e}^{-\lambda u} = F_X(u)$.

B5. Uvažujte nezávislé náhodné veličiny $X_1, X_2, \dots$ s rozdělením $\mathsf{Alt}(p)$.

(a) Ukažte (sporem), že při pevném rozsahu výběru n neexistuje nestranný odhad parametru $\theta = 1/p$ založený na $X_1, \dots, X_n$.

(b) Nechť Z značí počet počet nul předcházejících první jedničce v posloupnosti $X_1, X_2, \dots$. Víme, že Z má rozdělení $\mathsf{Geo}(p)$. Ukažte, že $Z + 1$ je nestranným odhadem parametru $\theta = 1/p$.

B6. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení $\mathsf{Exp}(1)$ a $U_1, \dots, U_n$ je náhodný výběr z rozdělení $\mathsf{R}(0,1)$. Ukažte, že $-\log U_{(n-r+1)}$ má stejné rozdělení jako $X_{(r)}$. Pomocí příkladu **A2** ukažte, že

$$Q_r = \left(\frac{U_{(r)}}{U_{(r+1)}}\right)^r$$

jsou nezávislé stejně rozdělené náhodné veličiny s rozdělením $\mathsf{R}(0,1)$.

B7. Máme-li dva nezávislé náhodné výběry $X_1, \dots, X_n \sim \mathsf{Exp}(\lambda)$ a $Y_1, \dots, Y_m \sim \mathsf{Exp}(\nu)$, odvozte konfidenční interval pro parametr $\varrho = \lambda/\nu$. [Návod: Uvědomte si vztahy mezi exponenciálním, Gamma a χ^2-rozdělením.]

B8. Nechť náhodný vektor $\mathbf{X} = [X_1, X_2, X_3, X_4]^\top$ má multinomické rozdělení s parametry $n \in \mathbb{N}$ a $\mathbf{p} = [p_1, p_2, p_3, p_4]^\top \in (0,1)^4$.

(a) Odvoďte asymptotické rozdělení vektoru $\mathbf{X}/n$ vzhledem k rostoucímu n na základě CLV.

(b) Ukažte asymptotickou normalitu odhadu $\widehat{\theta} = \frac{\frac{X_1}{n}\frac{X_4}{n}}{\frac{X_2}{n}\frac{X_3}{n}}$ parametru $\theta = \frac{p_1 p_4}{p_2 p_3}$ pomocí Delta metody.

(c) Použijte Delta metodu na odhad parametru $\vartheta = \log \theta$.

(d) Na základě asymptotické normality odhadu parametru ϑ skonstruujte přibližný $100(1-\alpha)\%$ interval spolehlivosti pro parametr θ.

(e) Yuleovo $Q = \frac{\theta - 1}{\theta + 1} = \frac{p_1 p_4 - p_2 p_3}{p_1 p_4 + p_2 p_3}$ odhadujeme pomocí $\widehat{Q} = \frac{\widehat{\theta} - 1}{\widehat{\theta} + 1}$. Ukažte, že $1 - \widehat{Q}^2 = \frac{4\widehat{\theta}}{(\widehat{\theta}+1)^2}$.

(f) Skonstruujte přibližný $100(1-\alpha)\%$ interval spolehlivosti pro Yuleovo Q.