NMFM301 – Statistika pro finanční matematiky

Pořádkové statistiky. Konstrukce intervalových odhadů

Podrobné riešenie príkladov z 3. cvičenia

A Příklady na cvičení

A1. *[Procvičovací]*

Nechť $X_1, \dots, X_n$ je náhodný výběr z rovnoměrného rozdělení na intervalu $(0, \theta_X)$, pro $\theta_X > 0$. Sestrojte přesný interval spolehlivosti pro θ_X založený na pořadové statistike $X_{(n)}$.

Řešení:

Distribučná funkcia náhodnej veličiny $X_{(n)}$ (n-tej poradovej štatistiky) je $[F(x)]^n$ a príslušná hustota je $nf(x)[F(x)]^{n-1}$, pričom $F(x)$ a $f(x)$ je distribučná funkcia, resp. hustota náhodnej velčiny X_i.

Navyše pre náhodné veličiny $X_1, \dots, X_n$ z rovnomerného rozdelenia na intervale $(0, \theta_X)$ platí, že transformované náhodné veličiny $\frac{X_1}{\theta_X}, \dots, \frac{X_n}{\theta_X}$ majú rovnomerné rozdelenie na intervale $(0, 1)$. Náhodná veličina $\frac{X_{(n)}}{\theta_X}$ má preto rozdelenie s distribučnou funkciou

$$F_{(n)}(x) = x^n \qquad \text{pre } x \in (0, 1),$$

a $F_{(n)}(x) = 0$ pre $x < 0$ a $F_{(n)}(x) = 1$ pre $x > 1$. Príslušná hustota náhodnej veličiny $\frac{X_{(n)}}{\theta_X}$ je

$$f_{(n)}(x) = nx^{n-1}\mathbb{I}_{\{x \in (0,1)\}},$$

čo je vlastne hustota Beta rozdelenia s parametrami $\alpha = n$ a $\beta = 1$ ($Beta(n, 1)$). Z definície kvantilu zároveň vieme, že platí

$$P\Big[c_{\alpha/2} < \frac{X_{(n)}}{\theta_X} < c_{1-\alpha/2}\Big] = 1 - \alpha, \tag{1}$$

kde $c_{\alpha/2}$ a $c_{1-\alpha/2}$ sú príslušné kvantily daného beta rozdelenia. Ich presné hodnoty získame pomocou distribučnej funkcie náhodnej veličiny $\frac{X_{(n)}}{\theta_X}$ následovne:

$$\frac{\alpha}{2} = P\Big[\frac{X_{(n)}}{\theta_X} \leq c_{\alpha/2}\Big] = F_{(n)}(c_{\alpha/2}) = c_{\alpha/2}^n;$$
$$1 - \frac{\alpha}{2} = P\Big[\frac{X_{(n)}}{\theta_X} \leq c_{1-\alpha/2}\Big] = F_{(n)}(c_{1-\alpha/2}) = c_{1-\alpha/2}^n,$$

kde v každom riadku prvá rovnosť plynie z definície kvantilu, druhá rovnosť z definície distribučnej funkcie, a posledná, tretia rovnosť z konkrétneho tvaru distribučnej funkcie $F_{(n)}(x) = x^n$, pre $x \in (0, 1)$. Pre kvantily teda dostávame rovnosti: $c_{\alpha/2} = \sqrt[n]{\frac{\alpha}{2}}$ a $c_{1-\alpha/2} = \sqrt[n]{1 - \frac{\alpha}{2}}$ a presný interval spoľahlivosti pre parameter $\theta_X > 0$ dostaneme z rovnosti (1) (pomocou vhodných ekvivalentných úprav) v tvare

$$P\Big[\frac{X_{(n)}}{\sqrt[n]{1 - \alpha/2}} < \theta_X < \frac{X_{(n)}}{\sqrt[n]{\alpha/2}}\Big] = 1 - \alpha.$$

A2. *[Procvičovací]*

Nechť $X_1, \dots, X_n$ je náhodný výběr z exponenciálního rozdělení s parametrem $\lambda_X > 0$. Sestrojte přesný interval spolehlivosti pro $\mathsf{E}\, X_i = 1/\lambda_X$ založený na náhodné veličině $\sum_{i=1}^n X_i$.

Řešení:
Opäť využijeme možnosť transformovať pôvodné náhodné veličiny $X_1, \dots, X_n$. Zároveň využijeme fakt, že exponenciálne rozdelenie s parametrom $\lambda_X = \frac{1}{2}$ je vlastne Gamma rozdelenie s parametrami $\frac{1}{2}$ a 1 (t.j. $Exp(\frac{1}{2}) \equiv \Gamma(\frac{1}{2}, 1)$) a súčet nezávislých náhodných veličín s takýmto rozdelením má potom Gamma rozdelenie, s parametrami $\frac{1}{2}$ a n. Postupne teda dostaneme nasledujúce riadky:

$$
\begin{aligned}
X_1, \dots, X_n &\sim Exp(\lambda_X) \\
\lambda_X X_1, \dots, \lambda_X X_n &\sim Exp(1) \qquad \text{(lebo } E[\lambda_X X_i] = \lambda_X E X_i = 1\text{)} \\
2\lambda_X X_1, \dots, 2\lambda_X X_n &\sim Exp\Big(\frac{1}{2}\Big) \equiv \Gamma\Big(\frac{1}{2}, 1\Big) \\
\sum_{i=1}^n 2\lambda_X X_i = 2\lambda_X \sum_{i=1}^n X_i &\sim \Gamma\Big(\frac{1}{2}, n\Big) \equiv \chi^2_{2n}
\end{aligned}
$$

Posledna rovnosť ($\equiv$) plynie z faktu, že $\Gamma(1/2, n/2) \equiv \chi^2_n$ (t.j., χ^2 rozdelenie s n stupňami voľnosti). S využitím definície kvantilov môžeme napísať nasledujúcu rovnosť:

$$
P\left[\chi^2_{2n}(\alpha/2) < 2\lambda_X \sum_{i=1}^n X_i < \chi^2_{2n}(1-\alpha/2)\right] = 1 - \alpha,
$$

kde $\chi^2_{2n}(\alpha/2)$ a $\chi^2_{2n}(1-\alpha/2)$ sú príslušné kvantily χ^2 rozdelenia s $2n$ stupňami voľnosti. Ekvivalentnými úpravami dostaneme vzťah (a zároveň presný interval spoľahlivosti pre strednú hodnotu $\mu = EX_i = \frac{1}{\lambda_x} > 0$)

$$
P\left[\frac{2\sum_{i=1}^n X_i}{\chi^2_{2n}(1-\alpha/2)} < \frac{1}{\lambda_X} < \frac{2\sum_{i=1}^n X_i}{\chi^2_{2n}(\alpha/2)}\right] = 1 - \alpha.
$$

Hodnotu $\alpha \in (0,1)$ volíme dostatočne malú, aby mal interval spoľahlivosti rozumné pokrytie. Najčastejšie voľby pre $\alpha \in (0,1)$ sú napr. $\alpha = 0.05$, alebo $\alpha \in \{0.1, 0.01, 0.005\}$.

A3. *[Procvičovací]*

Nechť $X_1, \dots, X_n$ je náhodný výběr z exponenciálního rozdělení s neznámým parametrem $\lambda_X > 0$.

(a) Pomocí centrální limitní věty sestrojte přibližný interval spolehlivosti pro λ_X.

Řešení:
Interval spoľahlivosti pre strednú hodnotu $EX_i = \frac{1}{\lambda_X}$ (odvodený vyššie), je možné priamo využiť aj pre odvodenie intervalu spoľahlivosti pre neznámy parameter $\lambda_X > 0$. Jedná sa ale o presný interval spoľahlivosti, pretože sme poznali presné rozdelenie vhodnej štatistiky, v tomto prípade náhodnej veličiny $2\lambda_X \sum_{i=1}^n X_i \sim \chi^2_{2n}$. Namiesto presného rozdelenia ale teraz využijeme asymtotické vlastnosti – konkrétne centrálnu limitnú vetu (CLV), ktorá nám pre súčet nezávislých a rovnako rozdelených náhodných veličín s konečným rozptylom dáva

$$
\sqrt{n}(\overline{X} - EX_i) \xrightarrow{\mathcal{D}} N(0, \mathsf{var}\, X_i),
$$

kde $\overline{X}$ je klasický výberový priemer. Jedná sa o konvergenciu v distribúcii. Špeciálne pre náhodné veličiny s exponenciálnym rozdelením s parametrom $\lambda_X > 0$ (tak, že $EX_i = \frac{1}{\lambda_X}$) dostaneme z CLV

$$
\sqrt{n}\Big(\overline{X} - \frac{1}{\lambda_X}\Big) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\lambda_X^2}\right), \tag{2}
$$

resp. ekvivalentný zápis v tvare

$$\sqrt{n}\lambda_X\left(\overline{X} - \frac{1}{\lambda_X}\right) \xrightarrow{\mathcal{D}} N\left(0,1\right).$$

S využitím kvantilov normálneho (štandardizovaného) rozdelenia môžeme napísať rovnosť

$$P[-u_{1-\alpha/2} < \sqrt{n}(\lambda_X\overline{X} - 1) < u_{1-\alpha/2}] \to 1-\alpha,$$

pre $n \to \infty$. Hodnota $u_{1-\alpha/2}$ je $(1-\alpha/2)$ kvantil normálneho $N(0,1)$ rozdelenia. Ekvivalentnými úpravami dostaneme asymtotický interval spoľahlivosti

$$P\Big[\frac{1}{\overline{X}} - \frac{u_{1-\alpha/2}}{\sqrt{n}\overline{X}} < \lambda_X < \frac{1}{\overline{X}} + \frac{u_{1-\alpha/2}}{\sqrt{n}\overline{X}}\Big] \to 1-\alpha,$$

opäť pre $n \to \infty$. Čím je rozsah náhodného výberu väčší, tým je pokrytie neznámeho parametru (v pravdepodobnosti) bližšie k požadovanej hodnote $1-\alpha$.

(b) Pomocí centrální limitní věty sestrojte přibližný interval spolehlivosti pro $\log\lambda_X$ a z něho odvoďte přibližný interval spolehlivosti pro λ_X.

Řešení:
Opäť nás zaujíma asymptotický interval spoľahlivosti pre neznámy parameter λ_X zostrojený na základe náhodného výberu $X_1,\ldots,X_n$ z exponenciálneho rozdelenia $Exp(\lambda_X)$ (tak, že $EX_i = \frac{1}{\lambda_X}$). Využijeme k tomu transformáciu $g(x) = \log(x)$. Keďže stále platí centrálna limitná veta v (2) , dostaneme s použitím transformácie aj

$$\sqrt{n}\left(\log(\overline{X} - \log\left(\frac{1}{\lambda_X}\right)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\lambda_X^2}\cdot\left[g'\left(\frac{1}{\lambda_X}\right)\right]^2\right).$$

To vlastne plynie z Taylorovho rozvoju funkcie $\log(x)$ v okolí bodu $x = EX_i = \frac{1}{\lambda_X}$, pretože

$$g(\overline{X}) \approx g(EX_i) + g'(EX_i)(\overline{X} - EX_i),$$

čo po dosadení a jednoduchých úpravach dáva výraz

$$\sqrt{n}\left(\log(\overline{X}) - \log\left(\frac{1}{\lambda_X}\right)\right) \approx \sqrt{n}g'(EX_i)(\overline{X} - EX_i),$$

pričom z CLV už vieme, že náhodná veličina na pravej strane od znamienka '≈' má asymptoticky normálne rozdelenie $N\left(0, \frac{1}{\lambda_X^2}\cdot[g'\left(EX_i\right)]^2\right)$. A preto aj ľavá strana má rovnaké asymptotické rozdelenie. Keďže $g'(x) = 1/x$ a $g'(1/\lambda_X) = \lambda_X$, tak dosadením a využitím kvantilov štandardizovaného normálneho rozdelenia dostaneme výraz

$$P\left[-u_{1-\alpha/2} < \sqrt{n}\left(\log(\overline{X}) - \log\left(\frac{1}{\lambda_X}\right)\right) < u_{1-\alpha/2}\right] \to 1-\alpha,$$

a jednoduchými ekvivalentnými úpravami dostaneme

$$P\left[-\log(\overline{X}) - \frac{u_{1-\alpha/2}}{\sqrt{n}} < \log(\lambda_X) < -\log(\overline{X}) + \frac{u_{1-\alpha/2}}{\sqrt{n}}\right] \to 1-\alpha,$$

čo je asymptotický interval spoľahlivosti pre neznámy parameter $\log(\lambda_X)$. Aplikovaním exponenciály na všetky tri členy v zátvorke dostaneme aj asymptotický interval spoľahlivosti pre parameter $\lambda_X > 0$ v tvare

$$P\left[\frac{1}{\overline{X}}e^{-\frac{u_{1-\alpha/2}}{\sqrt{n}}} < \lambda_X < \frac{1}{\overline{X}}e^{\frac{u_{1-\alpha/2}}{\sqrt{n}}}\right] \to 1-\alpha,$$

pre $n \to \infty$.

A4. *[Procvičovací]*

Máme-li dva nezávislé náhodné výběry $X_1, \dots, X_n \sim \mathsf{Exp}(\lambda_X)$ a $Y_1, \dots, Y_m \sim \mathsf{Exp}(\lambda_Y)$. Odvoďte přesný interval spolehlivosti pro parametr $\varrho = \lambda_X/\lambda_Y$.

<u>Řešení:</u>
Analogicky, ako v príklade A2 možeme využiť vhodné transformácie oboch náhodných výberov:

$$\begin{aligned} 2\lambda_X X_1, \dots, 2\lambda_X X_n &\sim Exp(1/2) \equiv \Gamma(1/2, 1) \\ 2\lambda_Y Y_1, \dots, 2\lambda_Y Y_m &\sim Exp(1/2) \equiv \Gamma(1/2, 1). \end{aligned}$$

Zároveň platí, že

$$2\lambda_X \sum_{i=1}^{n} X_i \sim \chi^2_{2n} \qquad \text{a} \qquad 2\lambda_Y \sum_{i=1}^{m} Y_i \sim \chi^2_{2m},$$

pričom obe náhodné veličiny sú vzájomne nezávislé. S využitím definície F rozdelenia a kvantilov tohto rozdelenia môžeme písať, že

$$\frac{\frac{2\lambda_x \sum_{i=1}^{n} X_i}{n}}{\frac{2\lambda_Y \sum_{i=1}^{m} Y_i}{m}} \sim F_{n,m}$$

a tiež

$$P\left[f_{n,m}(\alpha/2) < \frac{\lambda_X}{\lambda_Y} \cdot \frac{m\sum_{i=1}^{n} X_i}{n\sum_{i=1}^{m} Y_i} < f_{n,m}(1-\alpha/2)\right] = 1-\alpha,$$

kde $f_{n,m}(\alpha/2)$ a $f_{n,m}(1-\alpha/2)$ sú príslušné kvantily F rozdelenia s n a m stupňami voľnosti. Ekvivalentnými úpravami sa získa presný interval spoľahlivosti pre neznámi parameter (pomer) $\varrho = \lambda_X/\lambda_Y$.