
NMFM301 – Statistika pro finanční matematiky

Opakování teoretické části

5. Cvičení – Středa, 27.10.2021 | Online

Piate cvičenie je určené výhradne k opakovaniu teórie z prvých štyroch cvičení a k diskutovaniu správneho riešenia niektorých úloh zo série príkladov označených v zadaniach písmenom "B". Na nasledujúce cvičenie, t.j. **v stredu, 03.11.2021**, je plánovaná zápočtová písomna práca. Tématické zameranie písomnej práce bude výhradne na teóriu diskutovanú na prvých štyroch cvičeniach (t.j., náhodný výber a usporiadaný náhodný výber, pořadové statistiky a ich rozdelenie, bodové a intervalové odhady neznámych parametrov a vyšetrovanie vlastnosti bodových odhadov—konzistencia a asymptotická normalita).

Nižšie je uvedených niekoľko príkladov s podrobným riešením. Tieto príklady boli uvedené v zadaniach v častiach označených písmenom "B" a zároveň bol u týchto príkladov zo strany študentov prejavovaný explicitný záujem prediskutovať, resp. ukázať správne riešenie.

Na záver je uvedených (bez riešenia) niekoľko vzorových príkladov z písomných zápočtových prác z minulých rokov. Tieto príklady slúžia k samostatnému precvičovaniu a tiež k ilustrácii konkrétnych problémov, ktoré lze očakávať na zápočtovej práci.

Príklady zo série úloh B s požadovaným riešením

A1. [Séria 1, Príklad B3]

Pre nezávislé náhodné veličiny X, Y s exponenciálnym rozdelením $Exp(\lambda)$ uvažujte náhodnú veličinu U definovanú nižšie a nájdite jej hustotu.

$$U = \frac{\min(X, Y)}{\max(X, Y)}$$

Riešenie:

Náhodné veličiny X, Y tvoria náhodný výber (keďže sú nezávislé a rovnako rozdelené). Môžeme si tento náhodný výber označiť aj ako X_1, X_2 , kde $X_1 = X$ a $X_2 = Y$. Náhodné veličiny $V = \min(X, Y)$ a $W = \max(X, Y)$ tým pádom ale tvoria usporiadaný náhodný výber, t.j. $(X_{(1)}, X_{(2)}) = (V, W) = (\min(X, Y), \max(X, Y))$.

Veta 1.7 (Statistický větník na stránce kolegu Doc. Pešty) říká, že pro náhodný výber X_1, \dots, X_n z rozdelenia s hustotou $f(x)$, je združená hustota náhodného vektoru, ktorý je zároveň usporiadaným náhodným výberom, t.j. náhodného vektoru $(X_{(1)}, \dots, X_{(n)})^\top$, daná predpisom

$$f(x_1, \dots, x_n) = n!f(x_1) \dots f(x_n) \quad \text{pre } x_1 < \dots < x_n$$

a $f(x_1, \dots, x_n) = 0$ jinak.

Pre náš konkrétny prípad to znamená, že náhodný vektor (V, W) má hustotu

$$f(v, w) = 2!f(v)f(w) \quad \text{pre } 0 < v < w < \infty$$

a $f(v, w) = 0$ inak. Prítom platí, že $f(v) = \lambda \cdot e^{-\lambda v}$, pre $v > 0$ a analogicky aj pre $f(w)$. Hustota náhodného vektoru (V, W) je teda

$$f_{(V,W)}(v, w) = 2\lambda^2 e^{-\lambda(v+w)}, \quad \text{pre } 0 < v < w < \infty.$$

Tým pádom je zároveň zrejme, že náhodné veličiny V a W nie sú vzájomne nezávislé. Následne už len stačí definovať vhodnú transformáciu náhodného vektoru (V, W) tak, aby sme dostali požadovanú náhodnú veličinu $U = V/W$. Použijeme nasledujúcu transformáciu:

$$t : \begin{pmatrix} V \\ W \end{pmatrix} \rightarrow \begin{pmatrix} V/W \\ W \end{pmatrix} = \begin{pmatrix} U \\ W \end{pmatrix}.$$

Príslušné inverzné zobrazenie je

$$t^{-1} : \begin{pmatrix} U \\ W \end{pmatrix} \rightarrow \begin{pmatrix} U \cdot W \\ W \end{pmatrix} = \begin{pmatrix} V \\ W \end{pmatrix}.$$

Jakobián inverzného zobrazenia je

$$|J_{t^{-1}}| = \begin{vmatrix} w & u \\ 0 & 1 \end{vmatrix} = w,$$

tým pádom združená hustota náhodného vektoru (U, W) je

$$f_{(U,W)}(u, w) = f_{(V,W)}(t_1^{-1}(u, w), t_2^{-1}(u, w)) \cdot |J_{t^{-1}}| = 2\lambda^2 e^{-\lambda u w} e^{-\lambda w w},$$

pre $u \in (0, 1)$ a $w > 0$. Zároveň sme využili značnie, kde t_1^{-1} predstavuje prvú zložku inverzného zobrazenia t^{-1} a analogicky aj t_2^{-1} označuje druhú zložku inverzného zobrazenia t^{-1} .

Požadovaná hustota náhodnej veličiny U sa získa preintegrovaním združenej hustoty $f_{(U,W)}(u, w)$ vzhľadom k nadbytočnej premennej, t.j. premennej w . Platí teda:

$$f_U(u) = \int_{\mathbb{R}} f_{(U,W)}(u, w) dw = \int_0^{\infty} 2\lambda^2 e^{-\lambda(u+1)w} w dw = \frac{2\lambda^2}{\lambda(u+1)} \int_0^{\infty} w \cdot \lambda(u+1) e^{-\lambda(u+1)w} dw,$$

pričom integrál nie je nič iné, ako stredná hodnota náhodnej veličiny s exponenciálnym rozdelením s parametrom $\lambda \cdot (u+1)$ (kde $u \in (0, 1)$). Hustota náhodnej veličiny U je preto

$$f_U(u) = \frac{2\lambda^2}{\lambda(u+1)} \cdot \frac{1}{\lambda(u+1)} = \frac{2}{(u+1)^2}, \quad \text{pre } u \in (0, 1)$$

a $f_U(u) = 0$ inak.

A2. [Séria 2, Príklad B2]

Pomocou náhodného výberu X_1, \dots, X_n z rozdelenia daného hustotou $f(x, \theta) = \theta x^{\theta-1}$ pre $x \in (0, 1)$ a $\theta > 0$ je potrebné zostrojiť presný interval spoľahlivosti a pomocou CLV aj približný interval spoľahlivosti (pre nejaké vhodné $\alpha \in (0, 1)$).

Riešenie:

Je dobré si uvedomiť, že sa jedná o Beta rozdelenie $Beta(\theta, 1)$ so strednou hodnotou $EX_1 = \frac{\theta}{\theta+1}$ a je prirodzené, že odhad parametru sa snažíme zostrojiť pomocou výberového priemeru \bar{X} (keďže je zrejma návaznosť neznámeho parametru na strednú hodnotu). Potrebujeme ale poznať presné rozdelenie nejakej vhodnej pivotnej štatistiky - napr. náhodného súčtu $\sum_{i=1}^n X_i$, aby sme boli schopní zostrojiť presný interval spoľahlivosti. Využijeme k tomu transformáciu, ktorú ponúka zadanie, t.j., $t : Y_i = -\log X_i$.

Hustota náhodnej veličiny Y_i je daná (z vety o transformácii) predpisom

$$f_Y(y) = f_X(t^{-1}(y)) \cdot |((t^{-1})'y)| = \theta(e^{-y})^{\theta-1} \cdot |-e^{-y}| = \theta e^{-\theta y},$$

pre $y > 0$ a $f_Y(y) = 0$ inak. Náhodné veličiny $-\log X_1, \dots, -\log X_n$ majú teda exponenciálne rozdelenie s parametrom $\theta > 0$ (tak, že stredná hodnota je $E[-\log X_i] = 1/\theta$). Zároveň už poznáme vzťah medzi exponenciálnym rozdelením a Gamma rozdelením (prípadne χ^2 rozdelením). Exponenciálne rozdelenie s parametrom jedna polovina (t.j., stredná hodnota je 2) je to samé rozdelenie, ako $Gamma(1, \frac{1}{2})$. Zároveň súčet n nezávislých náhodných veličín s exponenciálnym rozdelením s parametrom $\lambda > 0$ má Erlangovo rozdelenie s parametrami n (velikosť součtu) a $\lambda > 0$ – čo je vlastne $Gamma(n, \lambda)$ rozdelenie. A nakoniec vieme, že $Gamma(\frac{n}{2}, \frac{1}{2})$ je vlastne χ^2 rozdelenie s n stupňami voľnosti. Toto všetko potrebujeme využiť ku konštrukcii presného intervalu spoľahlivosti pre neznámy parameter $\theta > 0$.

Náhodný súčet $-2\theta \sum_{i=1}^n \log X_i$ má Erlangovo rozdelenie $Erlang(n, \frac{1}{2})$ (pretože stredná hodnota $E[-2\theta \log X_i] = 2\theta E[-\log X_i] = 2$, čo znamená, že náhodná veličina $-2\theta \log X_i$ má exponenciálne rozdelenie s parametrom $\lambda = \frac{1}{2}$).

Celkovo teda dostaneme, že

$$-2\theta \sum_{i=1}^n \log X_i \sim \Gamma(n, \frac{1}{2}) \equiv \chi_{2n}^2.$$

Pre presný interval spoľahlivosti nám stačí použiť kvantily χ^2 rozdelenia s $2n$ stupňami voľnosti a dostaneme výraz

$$P \left[\chi_{2n}^2(\alpha/2) \leq -2\theta \sum_{i=1}^n \log X_i \leq \chi_{2n}^2(1 - \alpha/2) \right] = 1 - \alpha.$$

Ekvivalentnými úpravami získame z predchádzajúceho výrazu potrebný (presný) interval spoľahlivosti pre neznámy parameter $\theta > 0$.

Pre približný interval spoľahlivosti použijeme Centrálnu limitnú vetu (CLV), vďaka ktorej vieme, že

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (-\log X_i) - \frac{1}{\theta} \right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{\theta^2}\right).$$

To zároveň znamená, že

$$\sqrt{n} \frac{\left(\frac{1}{n} \sum_{i=1}^n (-\log X_i) - \frac{1}{\theta} \right)}{\frac{1}{\theta}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Použitím kvantilov štandardného normálneho rozdelenia získame výraz

$$P \left[u_{\alpha/2} \leq \sqrt{n} \left(\frac{\theta}{n} \sum_{i=1}^n (-\log X_i) - 1 \right) \leq u_{1-\alpha/2} \right]$$

a pomocou ekvivalentných úprav získame konečný výraz pre približný interval spoľahlivosti pre neznámy parameter $\theta > 0$.

A3. [Séria 3, Príklad B6]

Pre náhodný výber X_1, \dots, X_n z rovnomerného rozdelenia na intervale $(0, \theta_X)$, pre $\theta_X > 0$, je potrebné zostrojiť presný interval spoľahlivosti pre výberový medián, definovaný ako $\hat{m}_X = X_{(k+1)}$, pričom platí, že $n = 2k + 1$, pre $n \in \mathbb{N}$.

Riešenie:

V prvom rade je potrebné si uvedomiť, že teoretický medián v rovnomernom rozdelení na intervale $(0, \theta_X)$ je $m_X = \theta_X/2$. Zároveň platí, že ak náhodné veličiny X_1, \dots, X_n majú rovnomerné rozdelenie na intervale $(0, \theta_X)$, tak potom náhodné veličiny $X_1/\theta_X, \dots, X_n/\theta_X$ majú rovnomerné rozdelenie na intervale $(0, 1)$ a teoretický medián je $1/2$. Ak je $X_{(k+1)}$ teda $(k+1)$ -vou pořadovou

statistikou v usporiadanom náhodnom výbere $X_{(1)}, \dots, X_{(n)}$, tak potom $X_{(k+1)}/\theta_X$ je $(k+1)$ -vou pořadovou statistikou v usporiadanom náhodnom výbere $X_{(1)}/\theta_X, \dots, X_{(n)}/\theta_X$.

Obecne platí, že k -ta pořadová statistika v náhodnom výbere X_1, \dots, X_n zo spojitého rozdelenia s distribučnou funkciou $F(x)$ a príslušnou hustotou $f(x)$ má hustotu, ktorú lze vyjádriť ve tvaru

$$f_{(k)}(x) = n \binom{n-1}{k-1} f(x) [F(x)]^{k-1} [1-F(x)]^{n-k}, \quad \text{pre } x \in \mathbb{R}$$

a preto pre $(k+1)$ -vú pořadovú statistiku (kde $n = 2k+1$) dostaneme

$$f_{(k+1)}(x) = n \binom{2k}{k} f(x) [F(x)]^k [1-F(x)]^k, \quad \text{pre } x \in \mathbb{R}.$$

Keďže náhodné veličiny $X_1/\theta_X, \dots, X_n/\theta_X$ majú rovnomerné rozdelenie na intervale $(0, 1)$ s distribučnou funkciou $F(x) = x$ pre $x \in (0, 1)$ (a $F(x) = 0$ pre $x < 0$ a $F(x) = 1$, pre $x > 1$) a hustotou $f(x) = \mathbb{I}_{\{x \in (0,1)\}}$, tak $(k+1)$ -vá pořadová statistika $X_{(k+1)}/\theta_X$ má hustotu

$$f_{(k+1)}(x) = n \binom{2k}{k} x^k (1-x)^k \quad \text{for } x \in (0, 1) \text{ a } f_{(k+1)}(x) = 0 \text{ jinak.}$$

Náhodná veličina $X_{(k+1)}/\theta_X$ má teda Beta rozdelenie s parametrami $\alpha = k+1$ a $\beta = k+1$. Táto náhodná veličina je zároveň naším odhadom pre teoretický medián, teda hodnotu $1/2$ (napr. \tilde{m}_X).

Jedna z možností, ako zostrojiť presný interval spoľahlivosti pre medián, teda parameter $m_X = \theta_X/2$ (ale táto možnosť nebola správna v zmysle požadovaného riešenia) by bolo využiť príslušné kvantily beta rozdelenia. Beta rozdelenie ale nepatrí k štandardným rozdeleniam v zmysle bežne tabulovaných kvantilov. Preto (v zmysle správneho riešenia) je nutné hľadať možnosť, ako využiť kvantily buď normálneho rozdelenia, studentovho t rozdelenia, χ^2 rozdelenia, alebo Fisherovho F rozdelenia (ktoré sú bežne tabulované v štatistických tabuľkách).

V tomto prípade použijeme to posledné—Fisherovo F rozdelenie. Obecně totiž platí, že pre náhodnú veličinu s beta rozdelením $Z \sim \text{Beta}(\alpha/2, \beta/2)$ má transformovaná náhodná veličina $\frac{\beta Z}{\alpha(1-Z)}$ Fisherovo F rozdelenie s α a β stupňami voľnosti. V našom konkrétnom prípade teda dostávame, že ak

$$\frac{X_{(k+1)}}{\theta_X} \sim \text{Beta}(k+1, k+1) \implies \frac{2(k+1)X_{(k+1)}/\theta_X}{2(k+1)(1-X_{(k+1)}/\theta_X)} = \frac{X_{(k+1)}}{\theta_X - X_{(k+1)}} \sim F_{2(k+1), 2(k+1)}.$$

Pre príslušné kvantily (ktoré pre stručnosť označíme ako $f_{\alpha/2}$ a $f_{1-\alpha/2}$) z Fisherovho F rozdelenia s $2k+2$ a $2k+2$ stupňami voľnosti, môžeme písať

$$P \left[f_{\alpha/2} \leq \frac{X_{(k+1)}}{\theta_X - X_{(k+1)}} \leq f_{1-\alpha/2} \right] = 1 - \alpha$$

a použitím ekvivalentných úprav tiež

$$P \left[\frac{X_{(k+1)}}{f_{1-\alpha/2}} \leq \theta_X - X_{(k+1)} \leq \frac{X_{(k+1)}}{f_{\alpha/2}} \right] = P \left[X_{(k+1)} + \frac{X_{(k+1)}}{f_{1-\alpha/2}} \leq \theta_X \leq X_{(k+1)} + \frac{X_{(k+1)}}{f_{\alpha/2}} \right] = 1 - \alpha.$$

Ted' si už jenom stačí uvědomit, že máme konfidenčný interval máme pre parameter $\theta_X > 0$, zatiaľ čo v zadaní sme požadovali konfidenčný interval pre parameter $m_X = \theta_X/2$. Stačí teda hornú a dolnú hranicu intervalu podeliť hodnotou $1/2$ a získame interval spoľahlivosti pro $\theta_X/2$, čo je vlastne parameter teoretického mediánu v rovnomernom rozdelení na intervale $(0, \theta_X)$.

A4. [Séria 3, Príklad B7]

Pre náhodný výber X_1, \dots, X_n z Poissonovho rozdelenia s parametrom $\lambda > 0$ sestrojte pomoci CLV približný interval spoľahlivosti pro $\lambda > 0$. Pomoci CLV sestrojte také približný interval spoľahlivosti pro parameter $\sqrt{\lambda}$ a tento interval využijte pro odvození približného intervalu spoľahlivosti pro $\lambda > 0$.

Riešenie:

Centrální limitní věta říká, že pro náhodný výběr $X_1, \dots, X_n \sim Poiss(\lambda)$ (pre strednú hodnotu a rozptyl platí, že $EX = \lambda$ a tiež $VarX = \lambda$) máme

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{\mathcal{D}} N(0, \lambda),$$

pre $n \rightarrow \infty$. Analogicky tiež

$$\sqrt{n} \frac{(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, 1),$$

opäť pre $n \rightarrow \infty$. S využitím kvantilov normovaného normálneho rozdelenia dostávame

$$P \left[u_{\alpha/2} \leq \sqrt{n} \frac{(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \leq u_{1-\alpha/2} \right] \approx 1 - \alpha.$$

Problém ale nastane pri pokuse aplikovať ekvivalentné úpravy na výraz v zátvorke, aby sme osamostatnili uprostred neznámy parameter $\lambda > 0$ a explicitne tak definovali dolnú a hornú medzu intervalu spoľahlivosti. Riešenie ponúka tzv. Cramér-Slutského veta, ktorá obecně říká, že ak

$$\left[X_n \xrightarrow{\mathcal{D}} X \quad \text{a zároveň} \quad Y_n \xrightarrow{P} c \right] \implies X_n Y_n \xrightarrow{\mathcal{D}} cX.$$

Teda ak postupnosť náhodných veličín konverguje v distribúcii k nejakému limitnému rozdeleniu a postupnosť vynásobíme inou náhodnou veličinou, ktorá v pravdepodobnosti konverguje ku nenulovej konštantne, tak potom súčin konverguje v distribúcii k povodnému limitnému rozdeleniu, ktoré je danou konštantou príslušne preškálované.

Pre potreby tohto príkladu využijeme Cramér-Slutského vetu následovne:

$$\underbrace{\sqrt{n} \frac{(\bar{X}_n - \lambda)}{\sqrt{\lambda}}}_{\rightarrow^{\mathcal{D}} N(0,1)} \cdot \underbrace{\frac{\sqrt{\lambda}}{\bar{X}_n}}_{\rightarrow^P 1} = \sqrt{n} \frac{(\bar{X}_n - \lambda)}{\sqrt{\bar{X}_n}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Konvergenciu v distribúcii pre prvý výraz máme zaručenú centrálnou limitnou vetou. Druhý výraz konferguje k pravdepodobnosti k jednotke, čo plynie z vety o spojitaj transformácii, pretože \bar{X}_n (výberový priemer) je konzistentným odhadom parametru $\lambda > 0$ (teoretickej strednej hodnoty). Inými slovami, neznámy ohad směřodatnej chyby v jmenovateli zlomku jsme nahradili konzistentným odhadem. Konfidenčný interval získame ekvivalentnými úpravami z následujúceho výrazu:

$$P \left[u_{\alpha/2} \leq \sqrt{n} \frac{(\bar{X}_n - \lambda)}{\sqrt{\bar{X}_n}} \leq u_{1-\alpha/2} \right] \approx 1 - \alpha.$$

Pamätať na to, že sa jedná pouze o asymptoticky, resp. približný interval spoľahlivosti.

Pre druhú časť riešenia využijeme transformáciu $g(x) = \sqrt{x}$. Dostaneme teda

$$\sqrt{n}(g(\bar{X}_n) - g(\lambda)) \xrightarrow{\mathcal{D}} N(0, [g'(\lambda)]^2 \cdot \lambda).$$

Jedná sa o tzv. "rozptyl stabilizujúcu transformáciu", pretože $g'(x) = \frac{1}{2\sqrt{x}}$ a $[g'(\lambda)]^2 = \frac{1}{4\lambda}$ a je zrejmé, že neznámy parameter sa v asymptotickom rozptyle vykrátí. Asymptotický rozptyl teda nebude závisieť na neznámom parametri $\lambda > 0$ a dostaneme

$$P \left[u_{\alpha/2} \geq 2\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\lambda}) \geq u_{1-\alpha/2} \right] \approx 1 - \alpha.$$

Aplikáciou ekvivalentných úprav získame požadovaný interval spoľahlivosti parameter $\lambda > 0$.

Vzorové příklady k zápočtové práci (bez řešení)

A1. [Vzorový příklad k zápočtové práci]

Uvažujte náhodný výběr X_1, \dots, X_n z exponenciálního rozdělení s parametrem $\lambda = 1$ a na něm jiný nezávislý náhodný výběr Y_1, \dots, Y_n z rovnoměrného rozdělení na intervalu $(0, 1)$.

- Nájděte združené rozdělení náhodného vektoru $(X_{(1)}, X_{(n)})^\top$, kde $X_{(1)}$ je první a $X_{(n)}$ je n -tá pořádková statistika v uspořádaném náhodném výběru $X_{(1)} \leq \dots \leq X_{(n)}$.
- Uvažujte transformaci $Z_i = -\log Y_i$ a nájděte rozdělení náhodné veličiny Z_i .
- Ukážte, že rozdělení k -té pořádkové statistiky $X_{(k)}$ je stejné, jako rozdělení $(n - k + 1)$ -ní pořádkové statistiky $Z_{(n-k+1)} := -\log Y_{(n-k+1)}$.
- Nájděte združené rozdělení náhodného vektoru $(X_{(k)}, Z_{(n-k+1)})^\top$.

A2. [Vzorový příklad k zápočtové práci] Uvažujte náhodný výběr Y_1, \dots, Y_n z rovnoměrného rozdělení na intervalu $I \equiv (1, \theta) \subset \mathbb{R}$, kde $\theta > 1$ je neznámý parameter.

- Najděte odhad $\hat{\theta}_{MM}$ parametru θ pomocí momentové metody. Vyšetřete jeho nestrannost a konzistenci.
- Najděte asymptotické rozdělení odhadu $\hat{\theta}_{MM}$ a pomocí transformace $g(\theta) = \log(\theta - 1)$ sestrojte přibližný interval spolehlivosti pro neznámý parameter $\theta > 1$.
- Najděte odhad $\hat{\theta}_{ML}$ neznámého parametru θ pomocí metody maximální věrohodnosti a určete jeho rozdělení.
- Modifikujte odhad $\hat{\theta}_{ML}$ z předchozího kroku tak, aby byl nestranný.
- Uvažujte transformaci $t : X_i = -\log\left(\frac{Y_i - 1}{\theta - 1}\right)$ a odvoďte rozdělení náhodné veličiny X_i pro $i \in \{1, \dots, n\}$. Spočítejte střední hodnotu a rozptyl X_i .
- Pomocí centrální limitní věty spočítejte přibližně pravděpodobnost náhodného javu, který je definovaný jako $[W_n \geq n]$, kde $W_n = \sum_{i=1}^n X_i$.
- Spočítejte přesně pravděpodobnost náhodného javu $[W_n \geq n]$, kde $W_n = \sum_{i=1}^n X_i$.
- Nechť $n = 2k + 1$, pro nějaké $k \in \mathbb{N}$. Spočítejte distribuční funkci náhodného vektoru $(X_{(1)}, X_{(n)})^\top$ a nájděte rozdělení pro výběrový median $X_{(k+1)}$.