MATEMATICKÁ STATISTIKA 1, CVIČENÍ (NMSA331)

Příklady (nejen) pro přípravu na písemnou zápočtovou práci

Poslední úprava dokumentu: 17. listopadu 2016

Poslední úprava dokumentu: 17. listopadu 2016

1 Mnohorozměrné normální rozděleni

Příklad 1. Mnohorozměrné normální rozdělení

Nechť $\boldsymbol{X} = (X_1, \dots, X_n)^\mathsf{T}$ je náhodný výběr z mnohorozměrného normálního rozdělení $\mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(i) Najděte rozdělení náhodné veličiny $Y = \sum_{i=1}^n i\, X_i$.

(ii) Najděte rozdělení náhodné veličiny $Y = (X_i - \mu_i)^2/\sigma_i^2$, kde σ_i^2 je i-tý diagonální prvek matice $\boldsymbol{\Sigma}$.

(iii) Za jakýcho dotačných předpokladů má náhodná veličina $Z = \sum_{i=1}^n \frac{(X_i-\mu_i)^2}{\sigma_i^2}$ rozdělení χ^2 o n stupních volnosti?

2 Limitní věty

Příklad 2. Konvergence normálního rozdělení

Nechť pro každé $n \in \mathbb{N}$ je X_n náhodná veličina s normálním rozdělením $\mathsf{N}(\mu, \sigma_n^2)$, kde $\sigma_n > 0$. Dále předpokládejte, že $\sigma_n \xrightarrow[n\to\infty]{} \sigma$.

(i) Dokažte, že pokud $\sigma > 0$, potom $X_n \xrightarrow[n\to\infty]{d} \mathsf{N}(\mu, \sigma^2)$.

(ii) Dokažte, že pokud $\sigma = 0$, potom $X_n \xrightarrow[n\to\infty]{d} \delta_\mu$, kde δ_μ je Diracova míra v bodě μ.

Příklad 3. Konvergence Studentova rozdělení

Nechť pro každé $n \in \mathbb{N}$ je X_n náhodná veličina se (Studentovým) t-rozdělení o n stupních volnosti. Dokažte, že $X_n \xrightarrow[n\to\infty]{d} N(0,1)$.

Příklad 4. Konvergence Fisherova F-rozdělení

Nechť pro každé $n \in \mathbb{N}$ je X_n náhodná veličina s (Fisherovým) F-rozdělením o m a n stupních volnosti. Dokažte, že pro $n \to \infty$ posloupnost náhodných veličin $Y_n = m\, X_n$ konverguje v distribuci k náhodné veličině Y se χ^2-rozdělením o m stupních volnosti.

Příklad 5. Poissonovo rozdělení

Nechť $X_1, \ldots, X_n$ je náhodný výběr z Poissonova rozdělení, tj.

$$\mathsf{P}(X = k) = \frac{e^{-\lambda}\,\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

(i) Podrobně dokažte, že

$$\lim_{n\to\infty} \mathsf{P}\left(u_{\alpha/2} \leq \frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\overline{X}_n}} \leq u_{1-\alpha/2}\right) = 1 - \alpha,$$

kde $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ a u_β je β-kvantil normovaného normálního rozdělení.

Příklad 6. Odhad parametru p v geometrickém rozdělení

Nechť $X_1, \ldots, X_n$ je náhodný výběr z geometrického rozdělení, tj.

$$\mathsf{P}(X_1 = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \ldots$$

(i) Dokažte konzistenci a odvoďte asymptotické rozdělení následujícího odhadu parametru p daného předpisem

$$\widehat{p}_n = \frac{1}{1 + \frac{1}{n}\sum_{i=1}^n X_i}.$$

Příklad 7. Maximálně věrohodný a momentový odhad v Paretově rozdělení

Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení s hustotou

$$f(x) = \begin{cases} \frac{p}{x^{p+1}}, & x \geq 1, \\ 0, & \text{jinak.} \end{cases}$$

(i) Dokažte konzistenci a odvoďte asymptotické rozdělení maximálně věrohodného odhadu parametru p daného předpisem

$$\widehat{p}_n = \frac{n}{\sum_{i=1}^n \log(X_i)}.$$

(ii) Dokažte konzistenci a odvoďte asymptotické rozdělení momentového odhadu parametru p daného předpisem

$$\widetilde{p}_n = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i - n}.$$

Příklad 8. Odhad rozptylu v alternativním rozdělení

Nechť $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení, tj.

$$\mathsf{P}(X_1 = j) = p^j(1-p)^{1-j}, \quad j \in \{0, 1\}.$$

(i) Odvoďte asymptotické rozdělení odhadu rozptylu $\overline{X}_n(1 - \overline{X}_n)$.

Příklad 9. Transformace průměru

Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení s nulovou střední hodnotou a konečným nenulovým rozptylem σ^2.

(i) Odvoďte asymptotické rozdělení náhodné veličiny $\overline{X}_n \exp\left\{\overline{X}_n^3\right\}$.

Příklad 10. Podíl průměrů

Nechť $X_1, \dots, X_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda_X)$ s hustotou

$$f(x) = \begin{cases} \lambda_X \, \mathrm{e}^{-\lambda_X x}, & x \geq 0, \\ 0, & \text{jinak.} \end{cases}$$

a $Y_1, \dots, Y_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda_Y)$. Předpokládejte, že oba náhodné výběry jsou na sobě nezávislé.

(i) Najděte asymptotické rozdělení odhadu poměru středních hodnot daného předpisem $\frac{\overline{Y}_n}{\overline{X}_n}$, kde $\overline{X}_n$ a $\overline{Y}_n$ jsou příslušné výběrové průměry.

Příklad 11. Momentové odhady v gamma rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z gama rozdělení $\Gamma(a, p)$, tj.

$$f(x) = \begin{cases} \dfrac{a^p}{\Gamma(p)} \, x^{p-1} \, \mathrm{e}^{-a\,x}, & x \geq 0, \\ 0, & \text{jinak.} \end{cases}$$

(i) Najděte sdružené asymptotické rozdělení odhadu

$$\left(\widehat{a}_n, \widehat{p}_n\right)^{\mathsf{T}} = \left(\frac{\overline{X}_n}{S_n^2}, \frac{\left(\overline{X}_n\right)^2}{S_n^2}\right)^{\mathsf{T}},$$

kde $\overline{X}_n$, a S_n^2 jsou výběrový průměr a výběrový rozptyl.

Návod. Využijte věty z přednášky o sdružené asymptotické normalitě $\left(\overline{X}_n, S_n^2\right)^{\mathsf{T}}$ a dále pak toho, že pro gama rozdělení platí $\mathsf{E}X_1 = \frac{p}{a}$, $\mathsf{var}(X_1) = \frac{p}{a}$, $\gamma_3 = \frac{2}{\sqrt{p}}$ a $\gamma_4 = \frac{6}{p} + 3$.

3 Bodové odhady

Příklad 12. Nesmyslný nestranný odhad (podmíněné Poissonovo rozdělení)

Nechť X je diskrétní náhodná veličina s rozdělením

$$\mathsf{P}(X_1 = k) = \frac{1}{1-e^{-\lambda}} \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 1, 2, 3, \ldots$$

(i) Dokažte, že $T(X) = 1 + (-1)^X$ je nestranný odhad parametru $\theta_X = 1 - e^{-\lambda}$.
(ii) Dokažte, že $T(X) = 1 + (-1)^X$ je jediný nestranný odhad parametru $\theta_X = 1 - e^{-\lambda}$.

Příklad 13. Nesmyslný nestranný odhad (geometrické rozdělení)

Nechť X je diskrétní náhodná veličina s geometrickým rozdělením, tj.

$$\mathsf{P}(X_1 = k) = (1-p)^k\, p \quad k = 0, 1, 2, \ldots$$

(i) Ověřte, že $T(X) = (-1)^X$ je nestranný odhad parametrické funkce $\frac{p}{2-p}$. Zamyslete se nad užitečností takového odhadu.

Příklad 14. Neexistuje nestranný odhad (binomické rozdělení)

Nechť $X_1, \ldots, X_n$ je náhodný výběr z binomického rozdělení $\mathsf{Bi}(m, p)$, tj.

$$\mathsf{P}(X_1 = k) = \binom{m}{k} p^k (1-p)^{m-k} \quad k = 0, 1, \ldots, m.$$

(i) Ukažte, že $\widehat{p}_n = \frac{1}{n\,m} \sum_{i=1}^n X_i$ je nestranný odhad parametru p.
(ii) Ukažte, že neexistuje nestranný odhad parametru $\theta_X = \frac{1}{\sqrt{p(1-p)}}$.

Příklad 15. Odhad rozptylu v alternativním rozdělení

Nechť $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení, tj.

$$\mathsf{P}(X_1 = j) = p^j (1-p)^{1-j}, \quad j \in \{0, 1\}.$$

(i) Rozhodněte, zda je odhad $X_1(1-X_2)$ nestranný a konzistentní odhad parametru $\theta_X = p(1-p)$.
(ii) Rozhodněte, zda je odhad $\overline{X}_n(1-\overline{X}_n)$ nestranný a konzistentní odhad parametru $\theta_X = p(1-p)$.

Příklad 16. Odhad parametru λ v Poissonově rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z Poissonova rozdělení, tj.

$$\mathsf{P}(X_1 = k) = \frac{e^{-\lambda}\,\lambda^k}{k!}, \qquad k = 0, 1, \dots$$

(i) Dokažte, že $\widehat{\lambda}_n = \frac{1}{n}\sum_{i=1}^n X_i$ a $\widetilde{\lambda}_n = \frac{1}{n-1}\sum_{i=1}^n \big(X_i - \overline{X}_n\big)^2$ jsou nestranné a konzistentní odhady parametru λ.

(ii) Který z odhadů $\widehat{\lambda}_n$, $\widetilde{\lambda}_n$ byste spíše doporučili a proč?

Návod. *Ve (ii) porovnejte asymptotické rozptyly. Můžete využít toho, že špičatost (γ_4) Poissonova rozdělení je $\frac{1}{\lambda} + 3$.*

Příklad 17. Odhad pravděpodobnosti $\mathsf{P}(X_1 = 0)$ v Poissonově rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z Poissonova rozdělení, tj.

$$\mathsf{P}(X_1 = k) = \frac{e^{-\lambda}\,\lambda^k}{k!}, \qquad k = 0, 1, \dots$$

Uvažujte parametr $\theta_X = \mathsf{P}(X_1 = 0) = \mathrm{e}^{-\lambda}$.

(i) Rozhodněte, zda je odhad $\widehat{\theta}_n = \mathrm{e}^{-\overline{X}_n}$ nestranný a konzistentní odhad parametru θ_X. Pokud není nestranný, spočítejte jeho vychýlení.

(ii) Rozhodněte, zda je odhad $\widetilde{\theta}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i = 0\}$ nestranný a konzistentní odhad parametru θ_X. Pokud není nestranný, spočítejte jeho vychýlení.

(iii) Porovnejte odhady $\widehat{\theta}_n$ a $\widetilde{\theta}_n$ na základě jejich středních čtvercových chyb.

(iv) Porovnejte odhady $\widehat{\theta}_n$ a $\widetilde{\theta}_n$ na základě jejich asymptotických rozptylů.

Příklad 18. Normální rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z normálního rozdělení s hustotou

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\tfrac{(x-\mu)^2}{2\,\sigma^2} \Big\}, \qquad x \in \mathbb{R}.$$

(i) Rozhodněte, zda $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ je konzistentní a nestranný odhad parametru σ^2. Pokud není nestranný, spočítejte jeho vychýlení.

(ii) Rozhodněte, zda $S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2}$ je konzistentní a nestranný odhad parametru $\theta_X = \sigma$. Pokud není nestranný, spočítejte jeho vychýlení.

(iii) Spočtěte střední čtvercovou chybu odhadu S_n.

Příklad 19. Odhad parametru exponenciálního rozdělení při různé parametrizaci

Mějme náhodný výběr $X_1, \dots, X_n$ z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$, tj. X_i má hustotu

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \in (0, \infty), \\ 0, & \text{jinak}, \end{cases}$$

kde $\lambda > 0$ je neznámé.

(i) Rozhodněte, zda $\widehat{\lambda}_n = \frac{n}{\sum_{i=1}^n X_i}$ je nestranný a konzistentní odhad parametru λ.

(ii) Rozhodněte, zda $\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$ je nestranný a konzistentní odhad parametru $\theta_X = \frac{1}{\lambda}$.

Příklad 20. Odhad θ v rovnoměrném rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z rovnoměrného rozdělení $R(0, \theta)$ s hustotou

$$f(x) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{jinak}, \end{cases}$$

kde $\theta > 0$.

(i) Rozhodněte, zda $\widehat{\theta}_n = \max_{1 \le i \le n} X_i$ je nestranný a konzistentní odhad parametru θ. Pokud není, spočtěte jeho vychýlení.

(ii) Spočtěte střední čtvercovou chybu odhadu $\widehat{\theta}_n$.

(iii) Rozhodněte, zda $\widetilde{\theta}_n = 2\,\overline{X}_n$ je nestranný a konzistentní odhad parametru θ. Pokud není, spočtěte jeho vychýlení.

(iv) Spočtěte střední čtvercovou chybu odhadu $\widetilde{\theta}_n$.

(v) Kterému z odhadů $\widehat{\theta}_n$, $\widetilde{\theta}_n$ byste dali přednost?

Příklad 21. Odhad posunutí exponenciálního rozdělení

Mějme náhodný výběr $X_1, \dots, X_n$ z rozdělení

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda(x-\delta)}, & x \in (\delta, \infty), \\ 0, & \text{jinak}, \end{cases}$$

kde $\delta \in \mathbb{R}$ a $\lambda > 0$.

(i) Rozhodněte, zda $\widehat{\delta}_n = \min_{1 \le i \le n} X_i$ je nestranný a konzistentní odhad parametru δ. Pokud není, spočtěte jeho vychýlení.

(ii) Spočtěte střední čtvercovou chybu odhadu $\widehat{\delta}_n$.

(iii) Rozhodněte, zda odhad $\widetilde{\delta}_n = \max_{1 \le i \le n} X_i$ je konzistentní odhad parametru δ.

(iv) Rozhodněte, zda odhad $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ je nestranný a konzistentní odhad parametru δ.

Příklad 22. Odhad parametru p v geometrickém rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z geometrického rozdělení, tj.

$$\mathsf{P}(X_1 = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \dots$$

(i) Rozhodněte, zda $\check{p}_n = \mathbb{I}\{X_1 = 0\}$ je nestranný a konzistentní odhad parametru p.

(ii) Rozhodněte, zda $\widetilde{p}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i = 0\}$ je nestranný a konzistentní odhad parametru p.

(iii) Rozhodněte, zda $\widehat{p}_n = \frac{n+1}{n+\sum_{i=1}^n X_i}$ je konzistentní odhad parametru p.

Příklad 23. Odhad parametrů a, b v rovnoměrném rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z rovnoměrného rozdělení $R(a, b)$ s hustotou

$$f(x) = \begin{cases} \dfrac{1}{b}, & a - \frac{b}{2} < x < a + \frac{b}{2}, \\ 0, & \text{jinak}, \end{cases}$$

kde $a < b$. Označme $X_{(1)} = \min_{1 \le i \le n} X_i$ a $X_{(n)} = \max_{1 \le i \le n} X_i$.

(i) Rozhodněte, zda $\widehat{a}_n = \frac{X_{(1)} + X_{(n)}}{2}$ je nestranný a konzistentní odhad parametru a. Pokud není, spočtěte jeho vychýlení.

(ii) Spočtěte střední čtvercovou chybu odhadu $\widehat{a}_n$.

(iii) Rozhodněte, zda $\widehat{b}_n = X_{(n)} - X_{(1)}$ je nestranný a konzistentní odhad parametru b. Pokud není, spočtěte jeho vychýlení.

(iv) Spočtěte střední čtvercovou chybu odhadu $\widehat{b}_n$.

Návod. *Uvědomte si, že náhodný výběr $X_1, \dots, X_n$ můžeme „vyrobit“ pomocí lineární transformace jako $X_i = a + b\big(Y_i - \frac{1}{2}\big), i = 1, \dots, n$, kde $Y_1, \dots, Y_n$ je náhodný výběr z rovnoměrného rozdělení $R(0,1)$. Sdružená hustota náhodného vektoru $(Y_{(1)}, Y_{(n)})$ pak je*

$$f_{(Y_{(1)}, Y_{(n)})}(y_1, y_2) = n(n-1)(y_2 - y_1)^{n-2}\, \mathbb{I}\big\{0 < y_1 < y_2 < 1\big\}.$$

Příklad 24. Korelační koeficient

Nechť $(X_1, Y_1)^{\mathsf{T}}, \dots, (X_n, Y_n)^{\mathsf{T}}$ je náhodný výběr z dvourozměrného rozdělení s regulární varianční maticí. Dokažte, že odhad $\widehat{\rho}_n = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)\,(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sum_{i=1}^n (Y_i - \overline{Y}_n)^2}}$ je konzistentním odhadem korelačního koeficientu $\rho = \frac{\mathsf{cov}(X_1, Y_1)}{\sqrt{\mathsf{var}(X_1)\,\mathsf{var}(Y_1)}}$.

4 Intervalové odhady (intervaly spolehlivosti)

Předpokládejme, že chceme zkonstruovat intervalový odhad pro parametr θ, přičemž máme odhad $\widehat{\theta}_n$ tohoto parametru, pro který platí

$$\sqrt{n}\big(\widehat{\theta}_n - \theta_X\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, \sigma^2(\theta_X)\big), \tag{1}$$

kde $\sigma^2(\cdot)$ je funkce spojitá ve skutečné hodnotě parametru θ_X.

Interval spolehlivosti „Waldova" typu

Tento interval spolehlivosti je založen na tom, že díky (1) a Cramérově-Sluckého větě dostáváme $\frac{\sqrt{n}\,(\widehat{\theta}_n - \theta_X)}{\sigma(\widehat{\theta}_n)} \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0,1\big)$. Tudíž

$$\lim_{n\to\infty} \mathsf{P}\Big(- u_{1-\alpha/2} \le \tfrac{\sqrt{n}\,(\widehat{\theta}_n - \theta_X)}{\sigma(\widehat{\theta}_n)} \le u_{1-\alpha/2}\Big) = 1 - \alpha,$$

a tedy

$$\Big(\widehat{\theta}_n - \tfrac{u_{1-\alpha/2}\,\sigma(\widehat{\theta}_n)}{\sqrt{n}}, \widehat{\theta}_n + \tfrac{u_{1-\alpha/2}\,\sigma(\widehat{\theta}_n)}{\sqrt{n}}\Big) \tag{2}$$

je intervalový odhad parametru θ_X o asymptotické spolehlivosti $1 - \alpha$.

Interval spolehlivosti „Wilsonova" typu

Tento interval vychází z toho, že díky (1) máme

$$\lim_{n\to\infty} \mathsf{P}\Big(- u_{1-\alpha/2} \le \tfrac{\sqrt{n}\,(\widehat{\theta}_n - \theta_X)}{\sigma(\theta_X)} \le u_{1-\alpha/2}\Big) = 1 - \alpha.$$

Odtud dostáváme, že pro množinu

$$B_n = \Big\{\theta : \Big|\frac{\sqrt{n}\,(\widehat{\theta}_n - \theta)}{\sigma(\theta)}\Big| \le u_{1-\alpha/2}\Big\} \tag{3}$$

platí, že $\mathsf{P}\big(B_n \ni \theta_X\big) = 1 - \alpha$. Zpravidla je B_n interval, tudíž se (s trochou nepřesnosti) o ní mluví jako o intervalu spolehlivosti.

Ukazuje se, že pro konečné rozsahy výběrů je skutečné pokrytí intervalu (3) zpravidla blíže předepsané hladině $1-\alpha$ než pro interval spolehlivosti (2). Na druhou stranu interval spolehlivosti (3) je obecně zadán pouze implicitně a a jeho sestavení může být numericky výrazně náročnější.

Interval spolehlivosti založený na transformace stabilizující asymptotický rozptyl

Uvažujme transformaci g takovou, že funkce $[g'(\theta)]^2\,\sigma^2(\theta)$ již nezávisí na θ. Bez újmy na obecnosti nechť $[g'(\theta)]^2\,\sigma^2(\theta) = 1$. Potom pomocí (1) a Δ-metody dostáváme, že $\sqrt{n}\,\big(g(\widehat{\theta}_n) - g(\theta)\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0,1\big)$. Tudíž $\Big(g(\widehat{\theta}_n) - \frac{u_{1-\alpha/2}}{\sqrt{n}}, g(\widehat{\theta}_n) + \frac{u_{1-\alpha/2}}{\sqrt{n}}\Big)$ je intervalový odhad pro $g(\theta)$ a

$$\Big(g^{-1}\Big(g\big(\widehat{\theta}_n\big) - \tfrac{u_{1-\alpha/2}}{\sqrt{n}}\Big), g^{-1}\Big(g\big(\widehat{\theta}_n\big) + \tfrac{u_{1-\alpha/2}}{\sqrt{n}}\Big)\Big) \tag{4}$$

je intervalový odhad parametru θ o asymptotické spolehlivosti $1 - \alpha$.

Ukazuje se, že pro konečné rozsahy výběrů je skutečné pokrytí intervalu (4) zpravidla blíže předepsané hladině $1 - \alpha$ než pro interval spolehlivosti (2), i když ne tak blízko jako pro interval (3). Výhodou intervalového odhadu (4) oproti (3) však je, že pokud umíme invertovat funkci g, tak dostáváme explicitní předpis. Navíc transformace stabilizující asymptotický rozptyl je zpravidla vhodnější pro jednostranné intervalové odhady.

Příklad 25. Poissonovo rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z Poissonova rozdělení, tj.

$$\mathsf{P}(X_1 = k) = \frac{e^{-\lambda}\,\lambda^k}{k!}, \qquad k = 0, 1, \dots$$

(i) Najděte asymptotické rozdělení $\overline{X}_n$ a pomocí tohoto rozdělení sestavte oboustranný (a levostranný) intervalový odhad o asymptotické spolehlivosti $1 - \alpha$.

(ii) Najděte transformaci, která stabilizuje asymptotický rozptyl $\overline{X}_n$ a pomocí této transformace sestavte oboustranný (a pravostranný) intervalový odhad o asymptotické spolehlivosti $1 - \alpha$.

Příklad 26. Alternativní rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z alternativního rozdělení, tj.

$$\mathsf{P}(X_1 = j) = p^j (1-p)^{1-j}, \quad j \in \{0, 1\}.$$

(i) Najděte asymptotické rozdělení $\overline{X}_n$ a pomocí tohoto rozdělení sestavte oboustranný (a pravostranný) interval spolehlivosti.

(ii) Najděte asymptotické rozdělení $\arcsin\left(\sqrt{\overline{X}_n}\right)$ a pomocí tohoto rozdělení sestavte oboustranný (a levostranný) interval spolehlivosti.

Příklad 27. Odhad parametru exponenciálního rozdělení

Mějme náhodný výběr $X_1, \dots, X_n$ z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$, tj. X_i má hustotu

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \in (0, \infty), \\ 0, & \text{jinak}, \end{cases}$$

kde $\lambda > 0$.

(i) Najděte asymptotické rozdělení výběrového průměru $\overline{X}_n$ a pomocí tohoto rozdělení sestavte intervalový odhad pro parametr λ.

(ii) Najděte asymptotické rozdělení odhadu parametru λ daného předpisem $\widehat{\lambda}_n = \frac{1}{\overline{X}_n}$ a pomocí tohoto rozdělení sestavte intervalový odhad pro parametr λ.

(iii) Najděte transformaci, která stabilizuje asymptotický rozptyl náhodné veličiny $\frac{1}{\overline{X}_n}$ a pomocí této transformace sestavte intervalový odhad pro parametr λ.

Příklad 28. Odhad parametru p v geometrickém rozdělení

Nechť $X_1, \ldots, X_n$ je náhodný výběr z geometrického rozdělení, tj.

$$\mathsf{P}(X_1 = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \ldots$$

(i) Najděte asymptotické rozdělení odhadu $\widehat{p}_n = \frac{1}{1+\overline{X}_n}$ a na základě tohoto rozdělení odvoďte interval spolehlivosti („Waldova“ typu).

(ii) Najděte asymptotické rozdělení odhadu $\widetilde{p}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{X_i = 0\}$ a na základě tohoto rozdělení odvoďte interval spolehlivosti („Waldova“ typu).

(iii) Který z intervalů spolehlivosti odvozených v (i) a (ii) byste si vybrali a proč?

Příklad 29. Podíl průměrů

Nechť $X_1, \ldots, X_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda_X)$ s hustotou

$$f(x) = \begin{cases} \lambda_X\, \mathrm{e}^{-\lambda_X\, x}, & x \geq 0, \\ 0, & \text{jinak}, \end{cases}$$

a $Y_1, \ldots, Y_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda_Y)$. Předpokládejte, že oba náhodné výběry jsou na sobě nezávislé.

(i) S využitím asymptotického rozdělení $\frac{\overline{Y}_n}{\overline{X}_n}$ sestavte oboustranný intervalový odhad pro $\frac{\lambda_X}{\lambda_Y}$.

(ii) Najděte asymptotické rozdělení náhodné veličiny $\log\big(\frac{\overline{Y}_n}{\overline{X}_n}\big)$ a využijte této znalosti ke konstrukci oboustranného intervalového odhadu pro $\frac{\lambda_X}{\lambda_Y}$.

Příklad 30. Korelační koeficient

Nechť $(X_1, Y_1)^{\mathsf{T}}, \ldots, (X_n, Y_n)^{\mathsf{T}}$ je náhodný výběr z dvourozměrného normálního rozdělení s regulární varianční maticí a korelačním koeficientem ρ. Označme $\widehat{\rho}_n = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)\,(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\,\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2}}$ odhad korelačního koeficientu. Potom se dá pomocí Δ-metody dokázat, že

$$\sqrt{n}\,\big(\widehat{\rho}_n - \rho\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, (1-\rho^2)^2\big).$$

(i) Pro parametr ρ sestavte intervalový odhad „Waldova“ typu.

(ii) Pro parametr ρ sestavte intervalový odhad „Wilsonova“ typu.

(iii) Najděte transformaci, která stabilizuje asymptotický rozptyl odhadu korelačního koeficientu $\widehat{\rho}_n$ a pomocí této transformace sestavte intervalový odhad pro parametr ρ.

Příklad 31. Odhad θ v rovnoměrném rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z rovnoměrného rozdělení $R(0, \theta)$ s hustotou

$$f(x) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{jinak}, \end{cases}$$

kde $\theta > 0$.

(i) Na základě výběrového maxima $X_{(n)} = \max_{1 \le i \le n} X_i$ najděte intervalový odhad pro parametr θ.

(ii) Sestavte oboustranný (a levostranný) intervalový odhad na základě asymptotického rozdělení výběrového průměru $\overline{X}_n$ a porovnejte s intervalovým odhadem z (i).

Návod. *V (i) využijte toho, že $\frac{X_{(n)}}{\theta}$ je pivotální statistika (tj. statistika jejíž rozdělení nezávisí na neznámých parametrech).*

Příklad 32. Odhad posunutí exponenciálního rozdělení

Mějme náhodný výběr $X_1, \dots, X_n$ z rozdělení

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda(x-\delta)}, & x \in (\delta, \infty), \\ 0, & \text{jinak}, \end{cases}$$

kde $\delta \in \mathbb{R}$ je neznámý parametr a $\lambda > 0$ je známá konstanta.

(i) Ověřte, že $Z_n = \min_{1 \le i \le n} X_i - \delta$ je pivotální statistika.

(ii) Pomocí Z_n najděte oboustranný intervalový odhad pro parametr δ.

5 Empirické odhady

Příklad 33. Odhady distribuční funkce exponenciálního rozdělení

Nechť $X_1, \dots, X_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$ s hustotou

$$f(x) = \begin{cases} \lambda\,\mathrm{e}^{-\lambda x}, & x \geq 0, \\ 0, & \text{jinak.} \end{cases}$$

Naším cílem je odhadnout hodnotu distribuční funkce v nějakém daném bodě y, tj. $\theta_X = 1 - \mathrm{e}^{-\lambda y}$. Uvažujte následující dva odhady

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

Který z odhadů se Vám zda vhodnější a proč?

Příklad 34. Porodní hmotnost chlapců

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

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

Příklad 35. Asymptotické rozdělení třetího centrálního momentu

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

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

Návod. *Všimněte si, že v (i) můžete bez újmy na obecnosti předpokládat, že* $\mathsf{E}X_1 = 0$.

Příklad 36. Asymptotické rozdělení empirického odhadu šikmosti

Nechť $X_1, \dots, X_n$ je náhodný výběr z normálního rozdělení $\mathsf{N}(\mu, \sigma^2)$.

(i) Najděte asymptotické rozdělení empirického odhadu šikmosti $\widehat{\gamma}_3 = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^3}{\left(\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right)^{3/2}}$.

(ii) Zkuste promyslet, jak by se informace z (i) dala využít k testování normality (tj. testování, že náhodný výběr pochází z normálního rozdělení).

Návod. *Všimněte si, že v (i) můžete bez újmy na obecnosti předpokládat, že* $\mu = 0$ *a* $\sigma^2 = 1$.

Příklad 37. Empirický kvantil

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.

(i) Pro $\alpha = 0.25$, $\beta = 0.30$ a $n = 100$ spočtěte (přibližně) pravděpodobnost $\mathsf{P}\big(\widehat{u}_n(\alpha) > u_X(\beta)\big)$.

(ii) Jaká bude pravděpodobnost z (i) pro $n = 1\,000$?

Příklad 38. Intervalový odhad pro medián

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 $1 - \alpha$.

6 Testování hypotéz

Příklad 39. P-hodnota oboustranného testu

Nechť náhodná veličina U má rovnoměrné rozdělení na intervalu $(0, 1)$. Dokažte, že potom také náhodná veličina $V = 2\min(U, 1 - U)$ má rovnoměrné rozdělení na $(0, 1)$.

7 Výsledky některých příkladů

Příklad 6

(i) $\sqrt{n}\big(\widehat{p}_n - p\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, p^2(1-p)\big)$.

Příklad 7

(i) $\sqrt{n}\big(\widehat{p}_n - p\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, p^2\big)$.

(ii) $\sqrt{n}\big(\widetilde{p}_n - p\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, \frac{p(p-1)^3}{(p-2)}\big)$.

Příklad 8

(i) $\sqrt{n}\big(\overline{X}_n(1-\overline{X}_n) - p(1-p)\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, (1-2p)^2 p(1-p)\big)$ pro $p \neq \frac{1}{2}$. Všimněte si, že pro $p = \frac{1}{2}$ dostáváme $\sqrt{n}\big(\overline{X}_n(1-\overline{X}_n) - p(1-p)\big) \xrightarrow[n\to\infty]{\mathsf{P}} 0$.

Příklad 9

(i) $\sqrt{n}\,\overline{X}_n \exp\big\{\overline{X}_n^3\big\} \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, \sigma^2\big)$

Příklad 10

(i) $\sqrt{n}\big(\frac{\overline{Y}_n}{\overline{X}_n} - \frac{\lambda_X}{\lambda_Y}\big) \xrightarrow[n\to\infty]{d} \mathsf{N}\big(0, \frac{2\,\lambda_X^2}{\lambda_Y^2}\big)$.

Příklad 11

(i) $\sqrt{n}\left[\begin{pmatrix}\widehat{a}_n\\ \widehat{p}_n\end{pmatrix} - \begin{pmatrix}a\\ p\end{pmatrix}\right] \xrightarrow[n\to\infty]{d} \mathsf{N}_2\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \mathbb{V}\right)$, kde $\mathbb{V} = \mathbb{D}\,\Sigma\,\mathbb{D}^{\mathsf{T}}$, přičemž

$$\mathbb{D} = \begin{pmatrix}\frac{a^2}{p}, & -\frac{a^2}{p}\\ 2a, & -a^2\end{pmatrix} \qquad \text{a} \qquad \Sigma = \begin{pmatrix}\frac{p}{a^2}, & -\frac{2p}{a^3}\\ \frac{2p}{a^3}, & \frac{p^2}{a^4}\big(\frac{6}{p}+2\big)\end{pmatrix}.$$

Příklad 18

(i)

(ii) $\mathsf{bias}(S_n) = \sigma\Big(1 - \frac{\sqrt{2\,n}\,\Gamma\big(\frac{n}{2}\big)}{\Gamma\big(\frac{n-1}{2}\big)}\Big)$

Příklad 23

(i) Odhad $\widehat{a}_n$ je nestranný a konzistentní odhad parametru a.

(ii) $\mathsf{MSE}(\widehat{a}_n) = \mathsf{var}(\widehat{a}_n) = \frac{b^2}{2(n+2)(n+1)}$.

(iii) Odhad $\widehat{b}_n$ není nestranný, ale je konzistentní odhad parametru b.

(iv) $\mathsf{MSE}(\widehat{b}_n) = \mathsf{var}(\widehat{b}_n) + \big[\mathsf{E}(\widehat{b}_n - b)\big]^2 = \frac{b^2(n-1)}{2(n+2)(n+1)^2} + \frac{b^2}{(n+1)^2} = \frac{3\,b^2}{2(n+2)(n+1)}$.

Příklad 25

(i) Oboustranný interval „Waldova“ typu $\Big(\overline{X}_n - \frac{u_{1-\alpha/2}\sqrt{\overline{X}_n}}{\sqrt{n}}, \overline{X}_n + \frac{u_{1-\alpha/2}\sqrt{\overline{X}_n}}{\sqrt{n}}\Big)$. Levostranný interval by byl $\Big(\overline{X}_n - \frac{u_{1-\alpha}\sqrt{\overline{X}_n}}{\sqrt{n}}, \infty\Big)$

Oboustranný interval „Wilsonova“ typu

$$\left(\overline{X}_n + \frac{u^2_{1-\alpha/2}}{2\,n} - \sqrt{\overline{X}_n \frac{u^2_{1-\alpha/2}}{n} + \frac{u^4_{1-\alpha/2}}{4\,n^2}}, \overline{X}_n + \frac{u^2_{1-\alpha/2}}{2\,n} + \sqrt{\overline{X}_n \frac{u^2_{1-\alpha/2}}{n} + \frac{u^4_{1-\alpha/2}}{4\,n^2}}\right).$$

Levostranný interval by byl $\Big(\overline{X}_n + \frac{u^2_{1-\alpha}}{2\,n} - \sqrt{\overline{X}_n \frac{u^2_{1-\alpha}}{n} + \frac{u^4_{1-\alpha}}{4\,n^2}}, \infty\Big)$.

(ii) $\Big(\overline{X}_n - \frac{u_{1-\alpha/2}\sqrt{\overline{X}_n}}{\sqrt{n}} + \frac{u^2_{1-\alpha/2}}{4\,n}, \overline{X}_n + \frac{u_{1-\alpha/2}\sqrt{\overline{X}_n}}{\sqrt{n}} + \frac{u^2_{1-\alpha/2}}{4\,n}\Big)$, přičemž levou stranu intervalu spolehlivosti nahradíme nulou, pokud $\overline{X}_n < \frac{u^2_{1-\alpha/2}}{4\,n}$. Pravostranný interval by byl $\Big(0, \overline{X}_n + \frac{u_{1-\alpha}\sqrt{\overline{X}_n}}{\sqrt{n}} + \frac{u^2_{1-\alpha}}{4\,n}\Big)$,

Příklad 27

(i) Interval „Waldova“ typu $\left(\frac{1}{\overline{X}_n\left(1+\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)}, \frac{1}{\overline{X}_n\left(1-\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)}\right)$,

Interval „Wilsonova“ typu $\left(\frac{1-\frac{u_{1-\alpha/2}}{\sqrt{n}}}{\overline{X}_n}, \frac{1+\frac{u_{1-\alpha/2}}{\sqrt{n}}}{\overline{X}_n}\right)$,

(ii) Interval „Waldova“ typu $\left(\frac{1-\frac{u_{1-\alpha/2}}{\sqrt{n}}}{\overline{X}_n}, \frac{1+\frac{u_{1-\alpha/2}}{\sqrt{n}}}{\overline{X}_n}\right)$,

Interval „Wilsonova“ typu $\left(\frac{1}{\overline{X}_n\left(1+\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)}, \frac{1}{\overline{X}_n\left(1-\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)}\right)$,

(iii) $\left(\frac{1}{\overline{X}_n \exp\left(\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)}, \frac{1}{\overline{X}_n \exp\left(-\frac{u_{1-\alpha/2}}{\sqrt{n}}\right)}\right)$

Příklad 33

Podíl asymptotických rozptylů je

$$\frac{\mathsf{avar}(\widetilde{\theta}_n)}{\mathsf{avar}(\widehat{\theta}_n)} = \frac{\mathrm{e}^{\lambda\,y} - 1}{(\lambda\,y)^2} > 1.$$

Příklad 34

(i) Bodový odhad je přibližně 0.735. Intervalový odhad se spolehlivostí 0.95 je přibližně (0.722, 0.747).

(ii) Bodový odhad je přibližně 5.07. Intervalový odhad se spolehlivostí 0.95 je přibližně (4.71, 5.42).

Příklad 35

(i) $\sqrt{n}\left(\widehat{\mu}_3 - \mu_3\right) \xrightarrow[n\to\infty]{d} \mathsf{N}(0, v^2)$, kde $v^2 = \mu_6 - \mu_3^2 - 6\,\mu_2\,\mu_4 + 9\,\mu_2^3$

(iii) $\sqrt{n}\,\widehat{\mu}_3 \xrightarrow[n\to\infty]{d} \mathsf{N}(0, 6\,\sigma^6)$

Příklad 36

(i) $\sqrt{n}\,\widehat{\gamma}_3 \xrightarrow[n\to\infty]{d} \mathsf{N}(0, 6)$