Příklady k procvičení z Matematické statistiky 1

Poslední úprava 9. listopadu 2017- oprava zadání příkladu 22.

1 Konvergence posloupnosti náhodných veličin

1. Nechť $X_1, X_2 \dots$ jsou nezávislé veličiny s rovnoměrným rozdělením na $[0,1]$. Definujme $Y_n = \min\{X_1, \dots, X_n\}$. Ukažte, že pro $n \to \infty$ platí (a) $Y_n \xrightarrow{D} 0$, (b) $Y_n \xrightarrow{P} 0$.

2. Nechť X je nějaká náhodná veličina a $X_n = X(1 + 1/n)$. Ukažte, že $X_n \xrightarrow{P} X$.

3. Nechť $X \sim \mathsf{N}(0,1)$ a $X_n = (-1)^n X$. Ukažte, že $X_n \xrightarrow{D} X$ ale neplatí $X_n \xrightarrow{P} X$.

4. Nechť X_n jsou nezávislé náhodné veličiny a nechť X_n má alternativní rozdělení $\mathsf{Alt}(1/n)$, tj. $\mathsf{P}(X_n = 1) = 1/n$ a $\mathsf{P}(X_n = 0) = 1 - 1/n$. Ukažte, že $X_n \xrightarrow{P} 0$, ale neplatí $X_n \xrightarrow{s.j.} 0$.

5. Nechť X_n má exponenciální rozdělení se střední hodnotou $1/n$. Ukažte, že $X_n \xrightarrow{P} 0$.

6. Nechť $\{X_n\}$ je posloupnost náhodných veličin takových, že X_n má geometrické rozdělení s parametrem λ/n. Definujme $Y_n = 1/n \cdot X_n$. Vyšetřete konvergenci v distribuci veličin $\{Y_n\}$ pro $n \to \infty$.

7. Uvažujte posloupnost náhodných veličin $\{X_n\}$, kde X_n má distribuční funkci

$$F_n(x) = \begin{cases} 1 - \left(1 - \frac{1}{n}\right)^{nx}, & x \geq 0, \\ 0 & x < 0. \end{cases}$$

   Vyšetřete konvergenci v distribuci veličin $\{X_n\}$.

8. Uvažujte posloupnost náhodných veličin $X_n \sim \mathsf{Bi}(n, \lambda/n)$. Vyšetřete konvergenci v distribuci posloupnosti $\{X_n\}$ pro $n \to \infty$.

9. Nechť X_n má distribuční funkci

$$F_n(x) = \frac{e^{n(x-1)}}{1 + e^{n(x-1)}}, \quad x \geq 1,$$

   a $F_n(x) = 0$ pro $x < 1$.

   (a) Vyšetřete konvergenci v distribuci $\{X_n\}$.
   (b) Vyšetřete konvergenci v pravděpodobnosti $\{X_n\}$.

10. Nech $X_1, X_2, \dots$ tvoria náhodný výber z rozdelenia $\mathsf{N}(0,1)$. Vyšetrite konvergenciu postupnosti $\{X_n\}$ v zmysle

    (a) skoro iste,
    (b) v pravdepodobnosti,
    (c) v distribúcii.

11. Nech $(X_1, X_2)^{\mathsf{T}}$ je náhodný vektor s rozdelením $\mathsf{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$.
    (a) Nájdite rozdelenie náhodného vektoru $(X_1, -X_1)^{\mathsf{T}}$.
    (b) Ukážte, že obe marginálne rozdelenia vektorov $(X_1, X_2)^{\mathsf{T}}$ a $(X_1, -X_1)^{\mathsf{T}}$ sú identické.
    (c) Z výsledkov v častiach (a) a (b) usúďte, že z konvergencie $Y_n \xrightarrow{D} Y$ a $Z_n \xrightarrow{D} Z$ neplynie $(Y_n, Z_n)^{\mathsf{T}} \xrightarrow{D} (Y, Z)^{\mathsf{T}}$.
    (d) Môžeme povedať niečo viac ak v časti (c) navyše predpokladáme nezávislosť náhodných veličín Y_n a Z_n pre každé n?

12. 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 $\sigma_n \to \sigma$.
    (a) Dokažte, že pokud $\sigma > 0$, potom $X_n \xrightarrow{D} \mathsf{N}(\mu, \sigma^2)$.
    (b) Dokažte, že pokud $\sigma = 0$, potom $X_n \xrightarrow{D} \delta_\mu$, kde δ_μ je Diracova míra v bodě μ.

2 Vlastnosti a asymptotické rozdělení odhadů

13. Nechť $X_1, \ldots, X_n$ je náhodný výběr z Poissonova rozdělení $\mathsf{Po}(\lambda)$, $\lambda > 0$. Podrobně dokažte, že
$$\lim_{n\to\infty} \mathsf{P}\left(u_{\alpha/2} \le \frac{\sqrt{n}\big(\overline{X}_n - \lambda\big)}{\sqrt{\overline{X}_n}} \le u_{1-\alpha/2}\right) = 1 - \alpha,$$
kde $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ a u_β je β-kvantil normovaného normálního rozdělení.

14. Nechť $X_1, \ldots, X_n$ je náhodný výběr z geometrického rozdělení, tj.
$$\mathsf{P}(X_1 = k) = p\,(1-p)^k, \qquad k = 0, 1, 2, \ldots$$
    (a) Najděte odhad T_n parametru p metodou maximální věrohodnosti. Prověřte jeho konzistenci a odvoďte asymptotické rozdělení. Umíte rozhodnout i o nestrannosti?
    (b) Rozhodněte, zda $U_n = \mathbb{I}\{X_n = 0\}$ je nestranný a konzistentní odhad parametru p.
    (c) Rozhodněte, zda $V_n = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i = 0\}$ je nestranný a konzistentní odhad parametru p.
    (d) Jaké je přesné rozdělení nV_n? Jaké je asymptotické rozdělení V_n?
    (e) Rozhodněte, zda $W_n = \dfrac{n+1}{n + \sum_{i=1}^n X_i}$ je konzistentní odhad parametru p.

15. Nechť $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení s parametrem $p \in (0, 1)$.
    (a) Uvažujte odhad rozptylu $\mathsf{Var}\, X_1 = p(1-p)$ ve tvaru $U_n = \overline{X}_n(1 - \overline{X}_n)$. Vyšetřete jeho nestrannost, konzistenci a odvoďte asymptotické rozdělení.
    (b) Rozhodněte, zda je odhad $V_n = X_1(1 - X_n)$ nestranný a konzistentní odhad parametru $p(1-p)$.
    (c) Uvažujte výběrový rozptyl S_n^2. Rozepište si, jak se S_n^2 liší od U_n pro tento případ alternativního rozdělení.

16. Nechť X je diskrétní náhodná veličina s useknutým Poissonovým rozdělením, tj.

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

   (a) Dokažte, že $T(X) = 1 + (-1)^X$ je nestranný odhad parametru $\theta_X = 1 - e^{-\lambda}$.
   (b) Dokažte, že $T(X) = 1 + (-1)^X$ je jediný nestranný odhad parametru $\theta_X = 1 - e^{-\lambda}$.
   (c) Zamyslete se nad užitečností tohoto odhadu.

17. Nechť $X_1, \dots, X_n$ je náhodný výběr z binomického rozdělení $\mathsf{Bi}(m, p)$, $p \in (0, 1)$.
   (a) Ukažte, že $\widehat{p}_n = \frac{1}{n\,m} \sum_{i=1}^n X_i$ je nestranný odhad parametru p.
   (b) Ukažte, že neexistuje nestranný odhad parametru $\theta_X = \frac{1}{\sqrt{p(1-p)}}$.

18. Nechť $X_1, \dots, X_n$ je náhodný výběr z Poissonova rozdělení $\mathsf{Po}(\lambda)$, $\lambda > 0$. Uvažujte parametr $\theta_X = \mathsf{P}(X_1 = 0) = \mathrm{e}^{-\lambda}$.
   (a) 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.
   (b) 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.
   (c) Porovnejte odhady $\widehat{\theta}_n$ a $\widetilde{\theta}_n$ na základě jejich středních čtvercových chyb.
   (d) Porovnejte odhady $\widehat{\theta}_n$ a $\widetilde{\theta}_n$ na základě jejich asymptotických rozptylů.

19. Mějme náhodný výběr $X_1, \dots, X_n$ z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$, $\lambda > 0$.
   (a) Rozhodněte, zda $\widehat{\lambda}_n = \frac{n}{\sum_{i=1}^n X_i}$ je nestranný a konzistentní odhad parametru λ.
   (b) Rozhodněte, zda $\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$ je nestranný a konzistentní odhad parametru $\theta_X = \frac{1}{\lambda}$.

20. Mějme náhodný výběr $X_1, \dots, X_n$ z rovnoměrného rozdělení na $[0, \theta]$, kde $\theta > 0$. Uvažujte odhad θ ve tvaru $T_n = \max_{1 \le i \le n} X_i$. Najděte asymptotické rozdělení $n(T_n - \theta)$.
   Návod: Konvergenci v distribuci vyšetřujte na základě její definice.

21. 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 $\lambda > 0$.
   (a) 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.
   (b) Spočtěte střední čtvercovou chybu odhadu $\widehat{\delta}_n$.
   (c) Rozhodněte, zda odhad $\widetilde{\delta}_n = \max_{1 \le i \le n} X_i$ je konzistentní odhad parametru δ.
   (d) Rozhodněte, zda odhad $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ je nestranný a konzistentní odhad parametru δ.

22. Nechť $X_1, \dots, X_n$ je náhodný výběr z rovnoměrného rozdělení 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 $X_{(n)} = \max_{1 \le i \le n} X_i$.

(a) Rozhodněte, zda $\widehat{a}_n = \frac{X_{(1)} + X_{(n)}}{2}$ je nestranný a konzistentní odhad parametru a. Pokud není, spočtěte jeho vychýlení.
(b) Spočtěte střední čtvercovou chybu odhadu $\widehat{a}_n$.
(c) Rozhodněte, zda $\widehat{b}_n = X_{(n)} - X_{(1)}$ je nestranný a konzistentní odhad parametru b. Pokud není, spočtěte jeho vychýlení.
(d) Spočtěte střední čtvercovou chybu odhadu $\widehat{b}_n$.

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

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

23. Nechť $(X_1, Y_1)^\mathsf{T}, \ldots, (X_n, Y_n)^\mathsf{T}$ 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 $\rho = \frac{\mathsf{Cov}\,(X_1, Y_1)}{\sqrt{\mathsf{Var}\,(X_1)\,\mathsf{Var}\,(Y_1)}}$.

3 INTERVALOVÉ ODHADY

24. Nechť $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení s parametrem $p \in (0,1)$.

(a) Najděte asymptotické rozdělení $\overline{X}_n$ a pomocí tohoto rozdělení sestavte oboustranný (a pravostranný) interval spolehlivosti pro p.
(b) Najděte asymptotické rozdělení $\arcsin\big(\sqrt{\overline{X}_n}\big)$ a pomocí tohoto rozdělení sestavte oboustranný (a levostranný) interval spolehlivosti pro p.

25. Mějme náhodný výběr $X_1, \ldots, X_n$ z exponenciálního rozdělení s hustotou

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

kde $\lambda > 0$.

(a) Najděte asymptotické rozdělení výběrového průměru $\overline{X}_n$ a pomocí něj zkonstruujte intervalový odhad pro parametr λ.
(b) 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.
(c) Najděte transformaci, která stabilizuje asymptotický rozptyl náhodné veličiny $\frac{1}{\overline{X}_n}$ a pomocí této transformace sestavte intervalový odhad pro parametr λ.

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

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

(a) Najděte asymptotické rozdělení odhadu $\widehat{p}_n = \dfrac{1}{1+\overline{X}_n}$ a na základě tohoto rozdělení odvoďte klasický asymptotický interval spolehlivosti pro p.
(b) 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 klasický asymptotický interval spolehlivosti pro p.
(c) Porovnejte intervaly z (a) a (b). Který vám přijde vhodnější?

27. Nechť $X_1, \ldots, X_n$ je náhodný výběr z exponenciálního rozdělení s parametrem $\lambda_X > 0$ a $Y_1, \ldots, Y_n$ je náhodný výběr z exponenciálního rozdělení s parametrem $\lambda_Y > 0$. Předpokládejte, že oba náhodné výběry jsou na sobě nezávislé.
(a) S využitím asymptotického rozdělení $\dfrac{\overline{Y}_n}{\overline{X}_n}$ sestavte oboustranný intervalový odhad pro $\dfrac{\lambda_X}{\lambda_Y}$.
(b) Najděte asymptotické rozdělení náhodné veličiny $\log\big(\dfrac{\overline{Y}_n}{\overline{X}_n}\big)$ a využijte této znalosti ke konstrukci oboustranného intervalového odhadu pro $\dfrac{\lambda_X}{\lambda_Y}$.

28. 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

$$\sqrt{n}\,\big(\widehat{\rho}_n - \rho\big) \xrightarrow{D} \mathsf{N}\big(0, (1-\rho^2)^2\big).$$

(a) Pro parametr ρ sestavte klasický asymptotický intervalový odhad.
(b) Pro parametr ρ nalezněte spolehlivostní množinu B_n.
(c) Najděte transformaci, která stabilizuje asymptotický rozptyl odhadu korelačního koeficientu $\widehat{\rho}_n$ a pomocí této transformace sestavte intervalový odhad pro parametr ρ.

29. Mějme náhodný výběr $X_1, \ldots, X_n$ z rozdělení s hustotou

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

kde $\delta \in \mathbb{R}$ je neznámý parametr a $\lambda > 0$ je známá konstanta.
(a) Ověřte, že $Z_n = \min_{1\le i\le n} X_i - \delta$ je pivotální statistika.
(b) Pomocí Z_n najděte oboustranný intervalový odhad pro parametr δ.

4 Empirické odhady

30. Nechť $X_1, \ldots, X_n$ je náhodný výběr z exponenciálního rozdělení $\mathsf{Exp}(\lambda)$, $\lambda > 0$. 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?

31. 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).
(c) 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.
(d) 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.

32. Nechť $X_1, \dots, X_n$ 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 μ_3.
(c) 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 (a) můžete bez újmy na obecnosti předpokládat, že* $\mathsf{E}X_1 = 0$.

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

(a) Najděte asymptotické rozdělení empirického odhadu šikmosti $\widehat{\gamma}_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right)^{3/2}}$.
(b) Zkuste promyslet, jak by se informace z (a) dala využít k testování normality (tj. testování, že náhodný výběr pochází z normálního rozdělení).

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

34. 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.

(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 $n = 1\,000$?

35. 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

$$\mathsf{P}\big(X_{(k_1)} > \widetilde{m}\big) \le \tfrac{\alpha}{2}, \qquad \mathsf{P}\big(X_{(k_2)} < \widetilde{m}\big) \le \tfrac{\alpha}{2}.$$

Ukažte, že $\big(X_{(k_1)}, X_{(k_2)}\big)$ je intervalový odhad pro $\widetilde{m}$ se spolehlivostí alespoň $1 - \alpha$.