Zadání 13. cvičení, podzim 2019
Popisná statistika, nerovnosti, limitní věty

Exponenciální rozdělení $\mathrm{Exp}(\lambda)$ je rozdělení s hustotou

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

Markovova nerovnost. Pro libovolnou nezápornou náhodnou veličinu X se střední hodnotou $\mathrm{E}(X)$ platí

$$P(X \geq \varepsilon) \leq \frac{\mathrm{E}(X)}{\varepsilon}$$

pro všechna $\varepsilon > 0$.

Čebyševova nerovnost. Pro libovolnou náhodnou veličinu X se střední hodnotou $\mathrm{E}(X)$ a roztylem $\mathrm{D}(X)$ platí

$$P(X - \mathrm{E}(X)) \geq \varepsilon) \leq \frac{\mathrm{D}(X)}{\varepsilon^2}$$

nebo ekvivalentně

$$P(X - \mathrm{E}(X)) < \varepsilon) \geq 1 - \frac{\mathrm{D}(X)}{\varepsilon^2}$$

pro všechna $\varepsilon > 0$.

Centrální limitní věta. Uvažujme posloupnost nezávislých náhodných veličin X_i, které mají společnou střední hodnotu $EX_i = \mu$, společný rozptyl $\operatorname{var} X_i = \sigma^2 > 0$ a stejně omezený třetí absolutní moment $E|X_i| < C$. Pro rozdělení náhodných veličin

$$S_n = \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n}\sigma}$$

platí limitní vztah

$$\lim P(S_n \leq x) = \Phi(x),$$

kde Φ je distribuční funkce normovaného normálního rozdělení $\mathrm{N}(0, 1)$.

Moivreova-Laplaceova věta je speciálním případem centrální limitní věty: Nechť X_n jsou náhodné veličiny s binomickým rozdělením $\mathrm{Bi}(n, p)$. Pak pro normované náhodné veličiny

$$S_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

platí vztah

$$\lim_{n\to\infty} P(S_n \leq x) = \Phi(x),$$

kde Φ je distribuční funkce normovaného normálního rozdělení $\mathrm{N}(0, 1)$.

Příklad 1. *Byly naměřeny následující hodnoty nějakého znaku*

$$10; 7; 7; 8; 8; 9; 10; 9; 4; 9; 10; 9; 11; 9; 7; 8; 3; 9; 8; 7.$$

Určete aritmetický průměr, medián, kvartily, rozptyl.

Příklad 2. *Mějme nezápornou náhodnou veličinu X se střední hodnotou μ.*

(1) *Bez dalších informací o rozdělení X odhadněte $P(X > 3\mu)$.*
(2) *Víte-li, že $X \sim \mathrm{Ex}(\frac{1}{\mu})$, vypočtěte $P(X > 3\mu)$.*

Příklad 3. *Určete pravděpodobnost, že při 1200 hodech kostkou padne šestka alespoň 150 krát a nejvýše 250 krát*

a) *pomocí Čebyševovy nerovnosti,*
b) *pomocí de Moivreovy-Laplaceovy věty.*

Příklad 4. *Průměrná rychlost větru je na určitém místě 20 km/hod.*

a) *Bez ohledu na rozdělení rychlosti větru jako náhodné veličiny určete pravděpodobnost, že při jednom pozorování rychlost větru nepřesáhne 60 km/h.*
b) *Určete interval, v němž se bude rychlost větru nacházet s pravděpodobností alespoň 0,9, víte-li navíc, že směrodatná odchylka* $\sigma = 1$ *km/hod.*

Příklad 5. *Na FI je* 10% *studentů s prospěchem do* $1,2$. *Jak velkou skupinu je třeba vybrat, aby s pravděpodobností* $0,95$ *v ní bylo 8-12*$\%$ *studentů s prospěchem do* $1,2$*? Úlohu řešte zvlášť pomocí Čebyševovy a zvlášť pomocí Moivreovy-Laplaceovy věty.*

Příklad 6. *Dokažte, že pro kvantily normovaného normálního rozdělení platí vztah*

$$-u_{\frac{\alpha}{2}} = u_{1-\frac{\alpha}{2}}.$$

Příklad 7. *Pravděpodobnost, že se zasazený strom ujme, je 0,8. Jaká je pravděpodobnost, že z 500 zasazených stromů se jich ujme aspoň* 380*?*

Výsledek. 0,987.

Příklad 8. *Pravděpodobnost, že semeno vyklíčí, je* 0,9*. Kolik semen je třeba zasadit, aby s pravděpodobností aspoň* 0,995 *vyklíčilo cca 90% semen (což přesněji formulujeme se zpřesňujícím požadavkem, aby odchylka podílu vyklíčených semen od* 0,9 *nepřevýšila* 0,034*).*

Výsledek. $n \approx 614$.

Příklad 9. *Životnost (v hodinách) určité elektrické součástky má exponenciální rozdělení s parametrem* $\lambda = \frac{1}{10}$*. Pomocí centrální limitní věty odhadněte pravděpodobnost, že celková životnost 100 takových součastek bude mezi 900 a 1050 hodinami.*

Výsledek. $\mu = 10, \sigma^2 = 100, P(900 \le \sum X_i \le 1050) = P\left(\frac{900-n\mu}{\sigma\sqrt{n}} \le \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \le \frac{1050-n\mu}{\sigma\sqrt{n}}\right) = \Phi(0{,}5) - \Phi(-1) \approx 0{,}533$.

Náhodný výběr z normálního rozdělení

Náhodným výběrem rozsahu n rozumíme n-tici nezávislých náhodných veličin $X_1, \ldots, X_n$, které mají totéž rozdělení. S náhodným výběrem se obvykle setkáváme při opakovaném provádění téhož pokusu.

Statistika je náhodná veličina vzniklá transformací náhodného výběru.

- Výběrový průměr $M = \frac{1}{n}\sum_{i=1}^n X_i$, a jsou-li navíc $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, pak $M \sim N(\mu, \sigma^2/n)$.
- Výběrový rozptyl $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - M)^2 = \frac{1}{n-1}\big(\sum_{i=1}^n X_i^2 - nM\big), S = \sqrt{S^2}$.

Intervalovým odhadem parametru θ rozumíme interval (T_L, T_U), kde $T_L(X_1, \ldots, X_n)$ a $T_U(X_1, \ldots, X_n)$ jsou statistiky výběru $(X_1, \ldots, X_n)$. Platí-li

$$P(T_L \le \theta \le T_U) = 1 - \alpha,$$

říkáme, že (T_L, T_U) je $100 \cdot (1-\alpha)\%$ interval spolehlivosti pro parametr θ.

Horním odhadem parametru θ na hladině významnosti $1 - \alpha$ je statistika U, pro niž

$$P(\theta < U) \ge 1 - \alpha.$$

Dolním odhadem θ na hladině významnosti $1-\alpha$ je pak statistika L, pro niž
$$P(L < \theta) \geq 1-\alpha.$$

Případ, kdy je $X_1, \dots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$:

- M a S^2 jsou nezávislé náhodné veličiny.
- $M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M-\mu)/(\sigma/\sqrt{n}) \sim N(0,1)$.
- $K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.
- $\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.
- $T = (M-\mu)/(S/\sqrt{n}) \sim t(n-1)$.

Intervaly spolehlivosti:

μ (známe σ^2)	$(M - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, M + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2})$
μ (neznáme σ^2)	$(M - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1), M + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1))$
σ^2 (neznáme μ)	$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right)$
σ^2 (známe μ)	$\left(\frac{\sum(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}, \frac{\sum(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}\right)$
$\mu_1 - \mu_2$ (známe σ_1^2, σ_2^2)	$M_1 - M_2 \pm \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \cdot u_{1-\alpha/2}$
$\mu_1 - \mu_2$ (neznáme $\sigma_1^2 = \sigma_2^2$)	$M_1 - M_2 \pm S_* \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot t_{1-\alpha/2}(m+n-2)$
podíl rozptylů σ_1^2/σ_2^2	$\left(\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(m-1,n-1)}, \frac{S_1^2/S_2^2}{F_{\alpha/2}(m-1,n-1)}\right)$

Příklad 10. *Při 600 hodech kostkou padla jednička pouze 45 krát. Rozhodněte, jestli je možné tvrdit, že jde o ideální kostku na hladině $\alpha = 0{,}01$. Vše zdůvodněte a svůj závěr explicitně formulujte.*

Příklad 11. *Do bedny ukládáme výrobky se střední hodnotou 3 kg a směrodatnou odchylkou $0{,}8\,kg$. Jaký maximální počet výrobků můžeme do bedny uložit, aby celková hmotnost s pravděpodobností $0{,}9738$ nepřekročila jednu tunu?*

Výsledek. $n \approx 324$.

Příklad 12. *Odběratel provádí kontrolu jakosti námi dodaných výrobků namátkovou kontrolou testovaného rozměru u 21 náhodně vybraných výrobků. Dodávka bude přijata, pokud nebude výběrová směrodatná odchylka překračovat hodnotu 0,2 mm. Víme přitom, že naše stroje produkují výrobky, u nichž má sledovaný rozměr normální rozdělení tvaru $N(10\,\mathrm{mm}; 0,0737\,\mathrm{mm}^2)$. S využitím statistických tabulek určete pravděpodobnost, s níž bude dodávka přijata. Jak se změní odpověď, pokud odběratel kvůli nákladům na testy začne testovat pouze 4 výrobky?*
(V případě chybějících údajů v tabulce hodnoty, které máte k dispozici, lineárně interpolujte).

Výsledek. Při 21 testech je pravděpodobnost $\chi^2_{20}(10,9) = 0,05$, při 4 testech $\chi^2_3(1,63) = 0,24$.

Příklad 13. *Ze základního souboru, z rozdělení $N(\mu, \sigma^2)$, kde $\sigma^2 = 0{,}06$ jsme pořídili náhodný výběr s realizacemi $1{,}3; 1{,}8; 1{,}4; 1{,}2; 0{,}9; 1{,}5; 1{,}7$. Určete oboustranný 95% interval spolehlivosti pro neznámou střední hodnotu.*

Výsledek. $1{,}22 \leq \mu \leq 1{,}58$.

Příklad 14. *Náhodný veličina X má normální rozdělení $N(\mu, \sigma^2)$, kde μ, σ^2 nejsou známy. V následující tabulce jsou uvedeny četnosti jednotlivých realizací této náhodné veličiny.*

x_i	8	11	12	14	15	16	17	18	20	21
četnost	1	2	3	4	7	5	4	3	2	1

Vypočtěte:

a) *výběrový průměr,*
b) *výběrový rozptyl a výběrovou směrodatnou odchylku,*
c) *99% interval spolehlivosti pro střední hodnotu μ.*

Výsledek. $13{,}954 \leq \mu \leq 16{,}671$

Příklad 15. *Nechť $X_1, \ldots, X_n$ je náhodný vvběr z rozdělení $N(\mu, 0{,}04)$. Určete nejmenší počet měření, která je třeba provést, aby šířka 95% intervalu spolehlivosti pro μ nepřesáhla* 0,16.

Výsledek. 25

Příklad 16. *Byla provedena čtyři nazávislá měření obsahu manganu u dvou vzorků oceli a byly získány výsledky:*

1. vzorek	*0,31%*	*0,30%*	*0,29%*	*0,32%*
2. vzorek	*0,59%*	*0,57%*	*0,58%*	*0,57%*

Stanovte 95% interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$ za předpokladu, že jde o realizace náhodného výběru z normálního rozdělení s neznámými, ale shodnými rozptyly.

1. Testování hypotéz:

Testování hypotéz umožňuje na základě náhodného výběru s danou pravděpodobností ověřovat domněnky o rozdělení, z něhož pochází daný náhodný výběr.
Hypotézou budeme rozumět nějaké tvrzení o parametrech tohoto rozdělení.
H_0 ... nulová hypotéza (např. $\theta = c$, kde c je domněnka o hodnotě parametru θ)
H_1 ... (oboustranná) alternativní hypotéza (obvykle negace nulové)
Testováním H_0 oproti alternativní hypotéze rozumíme postup založený na náhodném výběru, s jehož pomocí platnost H_0 *zamítneme* nebo *nezamítneme* (= připouštíme).

Chyba 1. druhu ... H_0 platí a my ji zamítneme (závažnější)
Chyba 2. druhu ... H_0 neplatí a my ji nezamítneme

Pravděpodobnost chyby 1. druhu se nazývá *hladina významnosti* (α, obvykle $\alpha = 0,05$), pravděpodobnost chyby 2. druhu se značí β a číslo $1 - \beta$ se nazývá *síla testu*. Hypotézy budeme testovat pomocí příslušnosti do intervalu spolehlivosti – na základě realizace náhodného výběru sestrojíme $100(1 - \alpha)\%$ interval spolehlivosti pro neznámý parametr θ a zjistíme, zda c patří do tohoto intervalu. Pokud ano, hypotézu H_0 nezamítáme (v opačném případě zamítáme) na hladině významnosti α.

Příklad 17. *Z velkého souboru resistorů téhož typu bylo náhodně vybráno 16 kusů s výběrovým průměrem hodnot odporu* 9,3 kΩ. *Na hladině významnosti* 0,05 *testujte hypotézu, že výběr pochází z normálního rozdělení se střední hodnotou $\mu = 10\,\mathrm{k}\Omega$, za předpokladu, že:*

a) $\sigma^2 = 4\,\mathrm{k}\Omega$,
b) σ^2 *není známo a* $S^2 = 6{,}25\,\mathrm{k}\Omega$.

Výsledek. a) nezamítáme, b) nezamítáme.

Příklad 18. *Na dvou soustruzích se vyrábějí tytéž součástky, u nichž se měříme vnitřní průměr (předpokládá se normální rozdělení se stejnými rozptyly u obou soustruhů). Byl pořízen náhodný výběr rozsahu 16 z produkce prvního soustruhu a rozsahu 25 z produkce druhého soustruhu. Příslušné vyběrové průměry jsou* 37,5 *mm, resp.* 36,8 *mm a výběrový rozptyly* 1,21 *mm*^{2}, *resp.* 1,44 *mm*^{2}. *Testujte hypotézu o rovnosti střední hodnoty kontrolovaných rozměrů v produkci obou strojů oproti oboustranné alternativě při* $\alpha = 0{,}1$.

Příklad 19. *Na šachový turnaj má být vybrán jeden zástupce ze dvou oddílových šachistů, a to ten, jehož výkon je stabilnější (má menší rozptyl). Procentuální úspěšnost z posledních turnajů je:*

A	*49,6*	*59,4*	*59,5*	*76,8*	*69,4*	*70,9*	*68,1*	*66,3*
B	*38,5*	*51,2*	*79,5*	*72,3*	*86,5*			

Na hladině významnosti 0,05 *testujte, zda je možno rozhodnout o tom, který z hráčů se má turnaje zůčastnit.*