Poslední aktualizace: 29. května 2010

STP022 PRAVDĚPODOBNOST A MATEMATICKÁ STATISTIKA PŘÍKLADY

Pro zdárné absolvování předmětu doporučuji věnovat pozornost zejména příkladům označenými hvězdičkou*. Příklady označené křížkem † jsou o něco náročnější.

1. Přímý výpočet pravděpodobnosti

1. Hodíme 6 kostkami. Určete pravděpodobnost, že

(i) padnou vesměs různá čísla,
(ii) padnou pouze lichá čísla.

2. Házeme třemi kostkami. Který ze součtů je pravděpodobnější 11 nebo 12?

3. Skupina 10 studentů, z nichž 3 jsou z MFF, se náhodně seřadí do fronty. Určete pravděpodobnost, že 3 studenti z MFF budou vedle sebe.
Výsledek: $\frac{1}{15}$

4. Skupina $n \times a$ lidí, z nichž dva jsou naši známí, se rozdělí náhodně do n skupin po a osobách. Jaká je pravděpodobnost, že se naši dva známí dostanou do stejné skupiny?
Výsledek: $\frac{a-1}{an-1}$

5.* Na vánoční večírek přinese každý z n hostů jeden dárek. Přinesené dárky se očíslují a potom si každý vylosuje jeden dárek. Jaká je pravděpodobnost, že si alespoň jeden člověk odnese svůj dárek zase domů? Určete limitu této pravděpodobnosti pro $n \to \infty$.
Návod. Použijte princip inkluze a exkluze.

6. Klobouk obsahuje k černých koulí a jednu červenou. Petr a Pavla se střídají ve vytahování koulí, které nevracejí zpět do klobouku. Vyhrává ten, kdo vytáhne červenou kouli. Peter je gentleman a nabízí Pavle, zda chce začínat nebo ne. Co byste jí doporučili?

7.* Na nástupišti stojí vlak s desíti vagóny, do kterého náhodně nastoupilo 20 lidí. Předpokládejme, že každý cestující si náhodně vybere jeden vagón bez ohledu na to, kam nastupují ostatní cestující. Spočítejte, jaká je pravděpodobnost:
a) že do posledního vagónu nikdo nenastoupil
b) že do posledního vagónu nikdo nenastoupil a zároveň do prvního vagónu nastoupili právě dva lidé.
c) existuje alespoň jeden vagón, do kterého nikdo nenastoupil.

8. Mějme skupinu n osob ($n > 5$) o kterých můžeme předpokládat, že se narodili nezávisle na sobě. Spočítejte pravděpodobnost, že
a) žádný z nich se nenarodil ve středu
b) právě dva se narodili v pondělí a právě tři v sobotu
c) existuje den v týdnu, ve kterém se nikdo z nich nenarodil.

Řešení
Jedná se o Maxwellův-Boltzmannův model. Celkový počet možností – $|\Omega| = 7^n$.
a) Každá osoba má tedy 6 dnů, ve kterých se může narodit, tedy – $|A| = 6^n$ a tudíž

$$\mathrm{P}(A) = \left(\frac{6}{7}\right)^n$$

b) Nejdříve vyberu dva, kteří se narodili v pondělí – $\binom{n}{2}$ způsoby, dále ty, kteří se narodili v sobotu – $\binom{n-2}{3}$ způsobů. Zbývajících $n-5$ lidí má pět dnů, kdy se mohlo narodit. Tedy celkem – $|A| = \binom{n}{2}\binom{n-2}{3}5^{n-5}$. Tudíž,

$$\mathrm{P}\left(A\right) = \frac{\binom{n}{2}\binom{n-2}{3}5^{n-5}}{7^n}$$

b) Označme A_i jev, že se v i-tý den v týdnu nikdo nenarodil. Potom $A = A_1 \cup \ldots \cup A_7$. Dále

$$\mathrm{P}(A_i) = \left(\tfrac{6}{7}\right)^n, i = 1, \ldots, n$$

$$\mathrm{P}(A_i \cap A_j) = \left(\tfrac{5}{7}\right)^n, i \neq j$$

$$\mathrm{P}(A_i \cap A_j \cap A_k) = \left(\tfrac{4}{7}\right)^n, i \neq j \neq k$$

atd ...
Princip inkluze a exkluze nám dává

$$\mathrm{P}(A) = \mathrm{P}\left(\bigcup_{i=1}^{7} A_i\right) = \sum_{i=1}^{7}\mathrm{P}(A_i) - \sum\sum_{1\leq i<j\leq 7}\mathrm{P}(A_i \cap A_j) + \ldots - \sum_{i=1}^{7}\mathrm{P}\left(\bigcap_{j=1, j\neq i}^{n} A_j\right)$$

$$= \binom{7}{1}\left(\frac{6}{7}\right)^n - \binom{7}{2}\left(\frac{5}{7}\right)^n + \ldots - \binom{7}{6}\left(\frac{1}{7}\right)^n = \sum_{i=1}^{6}(-1)^{i+1}\binom{7}{i}\left(\frac{7-i}{7}\right)^n.$$

9. Dobře zamíchaný balíček 32 karet (piket) položíme na stůl obrázky směrem dolů. Otáčíme jednu kartu po druhé.
a) Jaká je pravděpodobnost, že před první károvou kartou otočíme $k = 0, 1, 2, \ldots$ karet jiných barev? Pro které k je tato pravděpodobnost nejvyšší?
b) Jaká je pravděpodobnost, že žádné dvě kárové karty nebudou následovat ihned po sobě?
Pikety obsahují čtyři barvy (srdce, káry, kříže, piky) po osmi kartách.
a) $\frac{\binom{32-k-1}{7}}{\binom{32}{8}}$, *b)* $\frac{\binom{25}{8}}{\binom{32}{8}}$

10. Opakovaně házíme současně dvěma kostkami. Jaká je pravděpodobnost, že součet 5 padne dřív než součet 7?
Výsledek: $\frac{2}{5}$

11.* Hráči A a B střídavě házejí kostkou (začíná hráč A). Hra končí, jestliže
a) hráči A padne 1 (vyhraje hráč A)
b) hráči B padne 2 nebo 3 (vyhraje hráč B) .
Spočítejte pravděpodobnost, že vyhraje hráč A, resp. B.

12. * Určitý člověk vybírá náhodně obraz ze skupiny obsahující 8 originálů a 2 kopie. Konzultuje s expertem, který rozliší originál od kopie s pravděpodobností 5/6. Jestliže expert soudí, že obraz je originál, stanovte pravděpodobnost, že tomu tak skutečně je.

13.* V sáčku se semínky je přimícháno N semínek jiné rostliny, přičemž víme, že $\mathrm{P}[N = i] = 0.1$ pro $i = 1, 2, \ldots, 10$. Pravděpodobnost, že přimíchané semínko vyklíčí je p. Určete
a) rozdělení N, jestliže vyklíčilo 8 z přimíchaných semínek
b) pravděpodobnost, že $N = 10$, jestliže vyklíčilo 8 z přimíchaných semínek a pravděpodobnost vyklíčení je $p = 0.5$.

14. Mějme dva disjunktní jevy A a B. Vyšetřete, jak je to s nezávislostí těchto jevů. Výsledek se pokuste slovně interpretovat.

15. Na následujícím známém vtipu si uvědomte rozdíl mezi pravděpodobností průniku a podmíněnou pravděpodobností:

> Na konferenci se baví dva statistici o tom, jaká byla cesta. Jeden z nich říká, že přijel vlakem, protože letadlem necestuje od té doby, co si přečetl, že pravděpodobnost bomby v letadle je 1/1000, což je pro něj příliš riskantní. Druhý mu na to odpoví: „Tak to dělej jako já. Pravděpodobnost, že by v letadle byly dvě bomby, je 1/1000000, což už je přijatelné. Takže já si vždycky beru jednu bombu svou.“

2. Diskrétní náhodné veličiny

16. Pro $p_k^{(n)} = \binom{n}{k} p_{(n)}^k (1-p_{(n)})^{n-k}$, $k = 0, \dots, n$ ukažte, že

$$\lim_{n\to\infty} p_k^{(n)} = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, \dots,$$

jestliže $p_{(n)} \to 0$ tak, že $np_{(n)} \to \lambda \in (0, \infty)$.

17. V kolejní síti je 923 počítačů, jeden z nich je váš a provozujete na něm FTP server. Každý počítač dá během hodiny na váš server požadavek s pravděpodobností 0.01. Spočtěte pravděpodobnost, že k vám během hodiny přijde požadavek (a) právě z 6, (b) nejvýše z 6 počítačů. Jednak počítejte tyto pravděpodobnosti přesně, tzn. v binomickém rozdělení, a jednak s použitím aproximace Poissonovým rozdělením.

18. Vzhledem k nezvykle vysokému jackpotu bylo 14. 1. 1995 v UK National Lottery vsazeno přibližně 69 800 000 tiketů. V této loterie se tipuje 6 čísel ze 49. Předpokládejme, že každý vyplněný tiket představuje náhodně vybranou šestici z 49 čísel a že šestice vybrané na jednotlivých tiketech jsou nezávislé. Spočítejte pravděpodobnost, že jackpot vyhraje (tj. uhodne všech 6 čísel) více než 10 tiketů. Dá se soudit něco o našich předpokladech (nezávislost a náhodně vybraná šestice), jestliže víme, že jackpot ve skutečnosti vyhrálo 133 tiketů? *Pro numerický výpočet se zde velmi hodí Poissonova aproximace.*

19.* Dva hráči střídavě házejí na koš, první se trefí s pravděpodobností p_1, druhý s p_2. Hra končí, když se někdo trefí. Označme X_1 počet hodů prvního hráče během celé hry, X_2 totéž pro druhého. Určete $\Pr[X_1 = k]$, $\Pr[X_2 = k]$. Spočtěme střední hodnotu počtu hodů prvního a druhého hráče, tj. $\mathsf{E}\, X_1$ a $\mathsf{E}\, X_2$.

3. Spojité náhodné veličiny

20.* Nechť náhodná veličina X má rozdělení s hustotou

$$f(x) = \begin{cases} cx^2, & x \in (0,1) \\ 0, & \text{jinak} \end{cases}$$

Spočítejte konstantu c, distribuční funkci, střední hodnotu, rozptyl a $\mathrm{P}\,[X \in (0.1, 0.5)]$.

21.† Veličina X má Cauchyho rozdělení s hustotou

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Povšimněte si, že veličina nemá střední hodnotu. Ukažte, že $Y = 1/X$ má opět Cauchyho rozdělení.

22. Najděte hustotu obsahu a obvodu čtverce, jehož hrana má náhodnou délku s rovnoměrným rozdělením na intervalu $(0, 5)$.

4. Náhodné vektory

23.* Nechť náhodný vektor $(X,Y)^{\mathsf{T}}$ má rovnoměrné rozdělení v rovnoběžníku s vrcholy $(0,0)$, $(1,0)$, $(1,1)$, $(2,1)$.
a) Určete sdruženou hustotu náhodných veličin X, Y. b) Najděte marginální hustotu náhodných veličin X, Y. Jsou tyto veličiny nezávislé?
c) Najděte marginální distribuční funkci náhodných veličin X, Y.
d) Spočítejte, $\mathsf{E}\,X$, $\mathsf{E}\,Y$ a $\operatorname{cov}(X,Y)$
e) Spočítejte, $\mathrm{P}(X > 2\,Y)$
Nezapomeňte nejdříve dopočítat konstantu c.

24.* Náhodný vektor $(X,Y)^{\mathsf{T}}$ má hustotu

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & (x,y) \in [0,1]^2, \\ 0 & \text{jinak.} \end{cases}$$

Najděte marginální hustoty X a Y, střední hodnoty, kovarianční a korelační matici, sdruženou distribuční funkci, marginální distribuční funkce. Rozhodněte, zda X a Y jsou nezávislé.

25.* Pro nezávislé veličiny $X \sim \mathrm{Exp}(\lambda_1)$ a $Y \sim \mathrm{Exp}(\lambda_2)$

(i) určete rozdělení $Z = \min(X,Y)$,
(ii) spočtěte $\Pr[X < Y]$.

26. Nezávisle hodíme dvěma symetrickými mincemi. Pro každou minci zaznamenáme výsledek 1, když padne panna, 0, když padne orel. Označme S součet výsledků na obou mincích, R jejich rozdíl. Jaké je rozdělení vektoru $(S,R)^{\mathsf{T}}$? (Zapište tabulku pravděpodobností.) Jsou S a R nezávislé? Určete korelaci.

27. Určete rozdělení $S = X + Y$, jestliže $X \sim \mathrm{Geom}(p)$, a $Y \sim \mathrm{Geom}(p)$, a X, Y jsou nezávislé.

28. Najděte rozdělení $X+Y$, jestliže X a Y jsou nezávislé stejně rozdělené náhodné veličiny s hustotou $f(x) = x\,e^{-x}\,\mathbb{I}\{x > 0\}$.

5. Borel–Cantelliho lemma, zákony velkých čísel

29.* Nechť $\{X_n, n \geq 1\}$ je posloupnost nezávislých stejně rozdělených veličin s konečným rozptylem. Uvažujme posloupnost $\{Y_n, n \geq 1\}$ s $Y_n = n^{1/3} X_n$. Platí pro ni SZVČ?

30.† Veličiny $X_1, X_2, \ldots$ nechť jsou nezávislé s rozdělením daným pravděpodobnostmi

$$\Pr[X_n = 1] = \Pr[X_n = -1] = \frac{1}{2} - \frac{1}{2^{n+1}}, \qquad \Pr[X_n = 2^n] = \Pr[X_n = -2^n] = \frac{1}{2^{n+1}}.$$

Dokažte, že pro posloupnost $\{X_n\}$ platí silný zákon velkých čísel.
Návod: Definujte si $Y_n = X_n\,\mathbb{I}\{|X_n| = 1\}$. Dokažte silný zákon velkých čísel pro posloupnost $\{Y_n\}$. Označte si jev $A_n = [Y_n \neq X_n]$ a dokažte, že tento jev nastane nekonečně-krát s pravděpodobností 0.

31. Nechť $X_1, X_2, \ldots$ je posloupnost stejně rozdělených náhodných veličin takových, že $\operatorname{var}\{X_1\} < \infty$ a pro korelační koeficient libovolné dvojice platí $\operatorname{corr}\{X_i, X_j\} = \frac{1}{1+(i-j)^2}$. Dokažte, že posloupnost $\{X_i\}_{i=1}^{\infty}$ splňuje slabý zákon velkých čísel.

6. Centrální limitní věta

32. Nechť $X_1, X_2, \dots$ jsou nezávislé náhodné veličiny takové, že

$$\mathrm{P}(X_n = n) = \mathrm{P}(X_n = -n) = \frac{1}{2\sqrt{n}}, \qquad \text{a} \qquad \mathrm{P}(X_n = 0) = 1 - \frac{1}{\sqrt{n}}.$$

Dokažte, že posloupnost $\{X_n\}$ splňuje centrální limitní větu.

33.* Životnost součástky má exponenciální rozdělení se střední hodnotou 10 hodin. Pomocí centrální limitní věty řešte následující.

(a) Máme 100 součástek. Jakmile se jedna porouchá, nahradíme ji další. Jaká je pravděpodobnost, že celková životnost bude mezi 900 a 1050 hodinami?

(b) Kolik máme koupit součástek, aby nám celkově vydržely aspoň 600 hodin s pravděpodobností aspoň 95 %?

34.* Na server má přístup 100 uživatelů. Z dřívějších zkušeností víme, že uživatel má na serveru průměrně $\mu = 120\,\mathrm{MB}$ dat, směrodatná odchylka množství dat je $\sigma = 40\,\mathrm{MB}$. Jak velký diskový prostor potřebujeme, aby s pravděpodobností 99 % nedošlo k jeho zaplnění? (Užijte CLV.)

35. Odhad π pomocí generování náhodných čísel. Nechť $(X_1, Y_1), \dots, (X_n, Y_n)$ jsou nezávislé vektory s nezávislými složkami s rovnoměrným rozdělením na $[0, 1]$. Označme $U_i = \mathbb{I}\{X_i^2 + Y_i^2 \leq 1\}$, tedy U_i je rovno 1 právě tehdy, když (X_i, Y_i) leží ve čtvrtkruhu se středem 0 a poloměrem 1, jinak je U_i rovno 0. Podle silného zákona velkých čísel $\bar{U}_n = \frac{1}{n}\sum_{i=1}^n U_i \xrightarrow[n\to\infty]{\text{s.j.}} \mathsf{E}\, U_1 = \Pr[U_1 = 1] = \pi/4$. Tudíž číslo π odhadneme hodnotou $\hat{\pi}_n = 4\bar{U}_n$. Určete přibližně (použitím CLV) $\Pr[3.1 \leq \hat{\pi}_{1000} \leq 3.2]$. Určete n tak, aby $|\hat{\pi}_n - \pi| \leq 0.01$ s pravděpodobností alespoň 0.9. Použijte jednak Čebyševovu nerovnost a jednak CLV.

7. Bodové odhady parametrů

36. Pro $X_1, \dots, X_n$ výběr z Poiss(λ) ukažte, že $(\bar{X}_n)^2$ není nestranným odhadem λ^2.

37.* Najděte maximálně věrohodný odhad $\hat{p}$ parametru p pro výběr $X_1, \dots, X_n$ z Geom(p). Ukažte, že tento odhad je konzistetní.

38.* Uvažujte výběr $X_1, \dots, X_n$ z Weibullova rozdělení s parametry $a > 0$ a $p > 0$, přičemž hodnota p je známá. Rozdělení je dáno hustotou

$$f(x; a) = a x^{p-1} \exp\left\{-\tfrac{a}{p} x^p\right\} 1_{(0,\infty)}(x).$$

Najděte maximálně věrohodný odhad $\hat{a}$ neznámého parametru a a vyšetřete jeho konzistenci.

39.* Buďte $X_1, \dots, X_n$ nezávislé náhodné veličiny s logaritmicko-normálním rozdělením s hustotou

$$f(x; \mu) = \begin{cases} \frac{1}{x\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2}\right\}, & x > 0, \\ 0 & \text{jinak.} \end{cases}$$

Najděte maximálně věrohodný odhad parametru μ a vyšetřete jeho nestrannost a konsistenci.

Řešení: ($\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \log(X_i)$, odhad je nestranný a konzistení)

40.* Mějme výběr o rozsahu n z normálního rozdělení $\mathrm{N}(\mu_0, \sigma^2)$, kde $\mu_0 \in \mathbb{R}$ je známá hodnota a $\sigma^2 > 0$ neznámý parametr. Nalezněte metodou maximální věrohodnosti odhad parametru σ^2. Vyšetřete nestrannost a konsistenci.

41.[*] Mějme výběr o rozsahu n z rozdělení s hustotou

$$f(x) = \begin{cases} c\, e^{-x}, & x > \theta, \\ 0 & \text{jinak}, \end{cases}$$

kde $\theta > 0$ je neznámý parametr. Spočítejtě konstantu c a najděte maximálně věrohodný odhad parametru θ. Vyšetřete nestrannost a konzistenci tohoto odhadu.

42.[†] Nechť náhodné veličiny $X_1, X_2, \dots$ jsou stejně rozdělené nezávislé se střední hodnotou μ a nenulovým konečným rozptylem σ^2. Dokažte, že náhodná veličina T_n

$$T_n = \frac{1}{2(n-1)} \sum_{i=2}^{n} (X_i - X_{i-1})^2$$

je nestranným a silně konzistentním odhadem σ^2.

8. Testování hypotes

43.[*] Máme podezření, že nás v hospodě okrádají. Proto koupíme 8 piv a změříme jejich objem: $0.51, 0.462, 0.491, 0.466, 0.461, 0.503, 0.495, 0.488$ (v litrech). Prokazují naše data, že hostinský je nepoctivý? Zformulujte hypotesu a alternativu, doplňte předpoklady a testujte na hladině $\alpha = 5\,\%$.

44.[*] Požadovaná přesnost výroby stogramových čokoládových tabulek je charakterizovaná rozptylem 4 (g^2). Vybereme náhodně 20 tabulek a zjistíme, že výběrový rozptyl jejich hmotnosti je 6 (g^2). Opravňují nás tato data na hladině $0,1$ tvrdit, že výrobní proces není v pořádku? A na hladině $0,05$?

45.[*] Ve dvou městech byla měřena tvrdost vody. V místě A bylo analysováno 40 vzorků, průměrná tvrdost vyšla 4.0 s výběrovým rozptylem 0.25. Ve městě B bylo zkoumáno 50 vzorků, průměr naměřených hodnot zde byl 3.8, výběrový rozptyl 0.24. Liší se tvrdost v těchto dvou městech? Zformulujte hypotesu a alternativu, doplňte předpoklady a testujte na hladině 0.05.

46.[*] Firma provedla test znalostí angličtiny u svých sedmi zaměstnanců. Poté jim zaplatila jazykový kurs a po něm opět otestovala jejich znalosti. Počty bodů získané jednotlivými zaměstnanci před a po kursu zachycuje tabulka:

Zaměstnanec	1	2	3	4	5	6	7
Před kursem	175	133	143	133	132	105	101
Po kursu	172	133	147	135	133	113	109

Došlo k významnému zlepšení? Zformulujte hypotesu a alternativu, doplňte předpoklady a testujte na hladině 0.05 a 0.1.

47. U vybraných dobrovolníků byla sledována hladina bifenylů v krvi před očistnou dietou (1.sloupec) a po očistné dietě druhý sloupec. Úkolem je na hladině 0.01 otestovat nulovou hypotézu, že průměrna hladina bifenylů v krvi se neliší před dietou a po dietě. Zkonstruujte příslušný 99%-ní interval spolehlivosti pro průměrný rozdíl. Uveďte všechny předpoklady, které používáte.

Data:

Č.	Před dietou	Po dietě
1	0.8	1.1
2	1.1	1.8
3	1.7	1.2
4	2.5	1.4
5	2.0	0.3
6	3.7	1.6
7	3.1	1.4
8	2.2	1.8
9	2.9	0.8
10	2.0	1.5
11	0.9	0.9
12	3.4	2.3
13	2.2	0.5
14	3.5	1.3
15	1.5	1.0
16	2.8	2.2
17	2.2	1.6
18	2.4	1.7
19	1.3	1.4
20	1.2	0.9
21	3.8	2.6
22	2.6	0.9
23	2.1	2.0
24	2.6	0.1
25	0.6	1.1

48.* Před francouzským referendem o euroústavě (29. 5. 2006) nechal list Le Figaro 20. a 21. května provést průzkum veřejného mínění. Z 950 dotázaných se 511 vyslovilo proti euroústavě. Prokazují výsledky, že většina je proti? Zformulujte hypotesu a alternativu, doplňte předpoklady a testujte na asymptotické hladině 0.05 (s využitím CLV).