České vysoké učení technické v Praze
Fakulta elektrotechnická

Sbírka příkladů z pravděpodobnosti a matematické statistiky

Kateřina Helisová
Miroslav Korbelář
Mirko Navara
Matěj Novotný

Praha 2016

Předmluva

Tento text je určen pro studenty technických vysokých škol, zejména elektrotechnických fakult, k procvičení teorie pravděpodobnosti a metod matematické statistiky. Neobsahuje výklad pojmů a metod, v tomto navazuje na skriptum [6], včetně použitého značení.

Skriptum bylo vytvořeno zčásti v prostředí *Scientific Workplace* (s výpočetní podporou *MuPAD*) a vysázeno v systému LaTeX s využitím šablony nakladatelství Springer. Jelikož se jedná o mezinárodní prostředí, prosíme čtenáře o shovívavost při zjištěných nedostatcích české sazby (nicméně o nahlášení závad). Statistické tabulky (převzaté z [6]) byly vygenerovány tabulkovým procesorem *MS Excel*. Část výpočtů byla provedena programem *Maple*®. Tvorba sbírky byla podpořena rozvojovým projektem RPAPS č. 1051603C004.

Většinu úloh vytvořili vyučující z katedry matematiky FEL ČVUT v Praze; kromě vyjmenovaných autorů zejména Tomáš Kroupa, Libor Nentvich, Milan Petrík a Ladislav Průcha. Tímto děkujeme jim i mnoha dalším učitelům i studentům (zejména Svetlaně Ulitině), že svými nápady a postřehy přispěli k obohacení sbírky. V neposlední řadě patří náš dík rodinným příslušníkům, kteří s námi měli trpělivost, když jsme věnovali čas této sbírce.

Praha, 31. prosince 2016 *autoři*

Obsah

Přehled značení (dle [6])

Odchylky od českého standardu

V následujících příkladech je použit anglický standard místo českého; podobný trend však lze pozorovat i v jiných současných českých učebnicích, např. [5, 8].

značka	příklad použití	význam
$\subseteq$	$A \subseteq B$	podmnožina (místo $\subset$, pokud připouštíme rovnost)
.	$\langle 0.1, 0.9\rangle$	desetinná tečka místo čárky

Obecné matematické symboly

$\mathbb{N}$	$\mathbb{N} = \{1, 2, \ldots\}$	množina všech přirozených čísel
$\mathbb{Z}$	$x \in \mathbb{Z}$	množina všech celých čísel
$\mathbb{R}$	$x \in \mathbb{R}$	množina všech reálných čísel
$\mathbb{C}$	$x \in \mathbb{C}$	množina všech komplexních čísel
i	$\mathrm{e}^{\mathrm{i}\,\omega\,t} = \cos\omega\,t + \mathrm{i}\,\sin\omega\,t$	imaginární jednotka
$\doteq$	$\pi \doteq 3.14$	přibližná rovnost
$\lvert . \rvert$	$\lvert M \rvert$	počet prvků (=kardinalita) množiny
$\cap, \cup, \overline{}$	$A \cap B = \overline{\overline{A} \cup \overline{B}}$	množinové operace (průnik, sjednocení, doplněk), ale i výrokové operace s jevy (konjunkce, disjunkce, negace)
,	$P\,[X > 0, X \leq 1]$	konjunkce výroků (pouze v kontextu pravděpodobnosti hodnot náhodných veličin)
$\setminus$	$A \setminus B = A \cap \overline{B}$	rozdíl množin
exp	$\exp \varOmega = \{A \mid A \subseteq \varOmega\}$	potenční množina (=množina všech podmnožin) množiny
$.\restriction.$	$X \restriction_{\varOmega_j}$	zúžení funkce na podmnožinu jejího definičního oboru
$*$	$(f_X * f_Y)\,(u)$	konvoluce

Základní pojmy teorie pravděpodobnosti

$\varOmega$	$\varOmega$	množina všech elementárních jevů
$\mathbf{1}, \varOmega$	$P(\mathbf{1}) = P(\varOmega) = 1$	jev jistý
$\mathbf{0}, \emptyset$	$P(\mathbf{0}) = P(\emptyset) = 0$	jev nemožný
$P(.)$	$P(A)$	pravděpodobnost (pravděpodobnostní míra) jevu
$P[.]$	$P[X \leq 0]$	pravděpodobnost jevu (pouze v kontextu pravděpodobnosti hodnot náhodných veličin)
$P(.\vert.)$	$P(A\vert B)$	podmíněná pravděpodobnost jevu A za podmínky B
$\mathcal{P}\,(.)$	$\mathcal{P}\,(\mathcal{A})$	množina všech pravděpodobností na σ-algebře

Popis náhodných veličin

$P_.$	P_X	pravděpodobnostní míra popisující rozdělení náhodné veličiny
$F_.$	F_X	distribuční funkce náhodné veličiny nebo rozdělení (zprava spojitá)
$\varPhi$	$\varPhi = F_{\mathrm{N}(0,1)}$	distribuční funkce normovaného normálního rozdělení
Mix	$\mathrm{Mix}_w(X, Y)$	směs dvou náhodných veličin

Mix	$\mathrm{Mix}_{(w_1,\dots,w_n)}(X_1,\dots,X_n)$	směs více náhodných veličin
$p_.$	p_X	pravděpodobnostní funkce diskrétní náhodné veličiny
$f_.$	f_X	hustota spojité náhodné veličiny nebo rozdělení
φ	$\varphi = f_{\mathrm{N}(0,1)}$	hustota normovaného normálního rozdělení
$q_.$	q_X	kvantilová funkce náhodné veličiny nebo rozdělení
$q_.(.)$	$q_X(\alpha)$	kvantil náhodné veličiny nebo rozdělení
$q_.\,(1/2)$	$q_X\,(1/2)$	medián náhodné veličiny nebo rozdělení
Φ^{-1}	$\Phi^{-1} = q_{\mathrm{N}(0,1)}$	kvantilová funkce normovaného normálního rozdělení
$O_.$	$O_X = \{u \in \mathbb{R} \mid P[X = u] \neq 0\}$	obor hodnot, které náhodná veličina nabývá s nenulovou pravděpodobností

Charakteristiky náhodných veličin

E.	$\mathrm{E}X$	střední hodnota náhodné veličiny
D.	$\mathrm{D}X$	rozptyl (=disperze) náhodné veličiny
$\sigma_.$	σ_X	směrodatná odchylka
norm .	$\mathrm{norm}\,X$	normovaná náhodná veličina
$\Psi_.$	Ψ_X	charakteristická funkce náhodné veličiny
$\mathrm{cov}(.,.)$	$\mathrm{cov}(X,Y) = \mathrm{E}\,(X\,Y) - \mathrm{E}X \cdot \mathrm{E}Y$	kovariance náhodných veličin
$\varrho(.,.)$	$\varrho(X,Y) = \mathrm{cov}(\mathrm{norm}\,X, \mathrm{norm}\,Y)$	korelace náhodných veličin
$\Sigma_.$	$\Sigma_{\boldsymbol{X}}$	kovarianční matice náhodného vektoru
$\varrho_.$	$\varrho_{\boldsymbol{X}}$	korelační matice náhodného vektoru

Typy rozdělení

Bi	$\mathrm{Bi}(m, p)$	binomické
Alt	$\mathrm{Alt}(p) = \mathrm{Bi}(2, p)$	alternativní
Po	$\mathrm{Po}(w)$	Poissonovo
Geom	$\mathrm{Geom}(q)$	geometrické
R	$\mathrm{R}(a, b)$	spojité rovnoměrné
N	$\mathrm{N}(\mu, \sigma^2)$	normální (=Gaussovo)
$\mathrm{N}(0, 1)$	$\mathrm{N}(0, 1)$	normované normální
LN	$\mathrm{LN}(\mu, \sigma^2)$	logaritmicko-normální
Ex	$\mathrm{Ex}(w)$	exponenciální
N	$\mathrm{N}(\boldsymbol{\mu}, \varSigma)$	vícerozměrné normální (=Gaussovo)
χ^2	$\chi^2(\eta)$	χ^2 (=„chí kvadrát“)
t	$\mathrm{t}(\eta)$	t (=Studentovo)
F	$\mathrm{F}(\xi, \eta)$	F (=Fisherovo-Snedecorovo)

Lineární prostor náhodných veličin

$\mathcal{L}$	$X \in \mathcal{L}$	množina všech náhodných veličin na pravděpodobnostním prostoru
$\mathcal{L}_2$	$\mathcal{L}_2 \subseteq \mathcal{L}$	množina všech náhodných veličin na pravděpodobnostním prostoru, které mají rozptyl
$\bullet$	$X \bullet Y = \mathrm{E}\,(X\,Y)$	skalární součin náhodných veličin
$\|\|.\|\|$	$\|\|X\|\| = \sqrt{X \bullet X}$	norma náhodné veličiny
d	$d(X, Y) = \|\|X - Y\|\|$	metrika na prostoru náhodných veličin $\mathcal{L}_2$
$\mathcal{R}$	$\mathcal{R}$	prostor všech konstantních náhodných veličin
$\mathcal{N}$	$\mathcal{N}$	prostor všech náhodných veličin s nulovou střední hodnotou

Výběrové statistiky

$\overline{}$	$\overline{\boldsymbol{X}}, \overline{\boldsymbol{X}}_n$	výběrový průměr
$\overline{}$	$\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}_n$	realizace výběrového průměru
$S^2_{.}, S^2$	$S^2_{\boldsymbol{X}}, S^2$	výběrový rozptyl
$s^2_{.}, s^2$	$s^2_{\boldsymbol{x}}, s^2$	realizace výběrového rozptylu
$\widehat{\varSigma}^2_{.}$	$\widehat{\varSigma}^2_{\boldsymbol{X}}$	výběrový 2. centrální moment (rozptyl empirické rozdělení)
$\widehat{\sigma}^2_{.}$	$\widehat{\sigma}^2_{\boldsymbol{x}}$	realizace výběrového 2. centrálního momentu (alternativního odhadu rozptylu)
$S_{.}, S$	$S_{\boldsymbol{X}}, S$	výběrová směrodatná odchylka
$s_{.}, s$	$s_{\boldsymbol{x}}, s$	realizace výběrové směrodatné odchylky
M	M_{X^k}	výběrový k-tý obecný moment
m	m_{x^k}	realizace výběrového k-tého obecného momentu
$R_{.,.}$	$R_{\boldsymbol{X},\boldsymbol{Y}}$	výběrový koeficient korelace
$r_{.,.}$	$r_{\boldsymbol{x},\boldsymbol{y}}$	realizace výběrového koeficientu korelace
$n_{.}$	n_u	četnost výsledku
$r_{.}$	$r_u = \frac{n_u}{n}$	relativní četnost výsledku
$\mathrm{Emp}(.)$	$\mathrm{Emp}(\boldsymbol{x})$	empirické rozdělení dané realizací náhodného výběru
$q_{\mathrm{Emp}(.)}\,(1/2)$	$q_{\mathrm{Emp}(\boldsymbol{x})}\,(1/2)$	výběrový medián (=medián empirického rozdělení)

Pojmy z teorie odhadu

Π	$\Pi \subseteq \mathbb{R}^k$	parametrický prostor
ϑ	$\vartheta \in \mathbb{R}$	skutečný parametr
$\widehat{\Theta}$	$\widehat{\Theta} \in \mathcal{L}$	odhad parametru
$\widehat{\vartheta}$	$\widehat{\vartheta} \in \mathbb{R}$	realizace odhadu parametru
$\boldsymbol{\vartheta}, (\vartheta_1, \ldots, \vartheta_k)$	$\boldsymbol{\vartheta} \in \mathbb{R}^k$	vektor skutečných parametrů
$\widehat{\boldsymbol{\Theta}}, (\widehat{\Theta}_1, \ldots, \widehat{\Theta}_k)$	$\widehat{\boldsymbol{\Theta}} \in \mathcal{L}^k$	odhad vektoru parametrů
$\widehat{\boldsymbol{\vartheta}}, (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_k)$	$\widehat{\boldsymbol{\vartheta}} \in \mathbb{R}^k$	realizace odhadu vektoru parametrů
$p_X(.;\boldsymbol{\vartheta})$	$p_X(u;a,b)$	pravděpodobnostní funkce závislá na vektoru parametrů
$f_X(.;\boldsymbol{\vartheta})$	$f_X(u;a,b)$	hustota závislá na vektoru parametrů
$F_X(.;\boldsymbol{\vartheta})$	$F_X(u;a,b)$	distribuční funkce závislá na vektoru parametrů
L	$L(\boldsymbol{\vartheta})$	věrohodnost diskrétního rozdělení
ℓ	$\ell(\boldsymbol{\vartheta})$	logaritmus věrohodnosti diskrétního rozdělení
Λ	$\Lambda(\boldsymbol{\vartheta})$	věrohodnost spojitého rozdělení
λ	$\lambda(\boldsymbol{\vartheta})$	logaritmus věrohodnosti spojitého rozdělení

Pojmy z testování hypotéz

H_0	$H_0\colon \mathrm{E}X \leq c$	nulová hypotéza
H_1	$H_1\colon \mathrm{E}X > c$	alternativní hypotéza
κ	$\kappa = \Phi^{-1}\,(1-\alpha)$	kritická hodnota testu
$\alpha(\kappa), \alpha$	$\alpha(\kappa), \alpha$	pravděpodobnost chyby 1. druhu
α	$\alpha = 5\%$	hladina významnosti
$\beta(\kappa), \beta$	$\beta(\kappa), \beta$	pravděpodobnost chyby 2. druhu
$1-\beta$	$1-\beta$	síla testu

Část I
Teorie pravděpodobnosti

1 Motivační příklady

Cvičení 1.1 (Monty Hall Problem). *Hráč má uhádnout, za kterými z trojích dveří se skrývá výhra. Řekne svůj tip, poté mu moderátor (který ví, kde výhra je) otevře jiné dveře, za kterými výhra není. Poté dá hráči možnost změnit svůj tip. Má to hráč udělat?*

Řešení. Pokud hráč trvá na svém prvním odhadu, je pravděpodobnost výhry 1/3.
Pokud změní tip, volí ze 2 možností, ale jeho šance se zvýší na 2/3:
S pravděpodobností 1/3 byl první odhad správný a druhý chybný.
S pravděpodobností 2/3 byl první odhad chybný a druhý správný.

Cvičení 1.2 (4 PINy). [1]*Banka poslala jednomu klientovi ke 4 jeho kontům přístupová hesla (PIN), ale neuvedla, které heslo patří ke kterému účtu. Ke každému účtu lze vyzkoušet 3 kódy, po 3 chybách se zablokuje. Navrhněte postup, který dovolí zpřístupnit (v průměru) co nejvíce kont.*

Řešení. První heslo zkoušíme postupně k jednotlivým kontům, dokud neuspějeme. Pak postupujeme stejně s druhým heslem. V nejnepříznivějším případě (pokud jsme správné konto našli vždy až na poslední pokus) nyní máme pravděpodobnost 1/2, že zablokujeme jedno konto, všechna ostatní se podaří otevřít. (Toto není jediný postup s tímto výsledkem.)

2 Kombinatorické pojmy a vzorce

Cvičení 2.1 (druhy náhodných výběrů). *Kolika způsoby lze z populace velikosti n vybrat*

a) 12 finalistek soutěže krásy,

b) 4-členné družstvo na závod Dolomitenmann,

c) 1000 výherců spotřebitelské soutěže?

U těch probíraných druhů náhodných výběrů, které zde nejsou zastoupeny, najděte vlastní příklad.

Řešení. a) 12 finalistek soutěže krásy: neuspořádaný výběr bez vracení, $\binom{n}{12}$.

b) 4-členné družstvo na závod Dolomitenmann: uspořádaný výběr bez vracení, $\frac{n!}{(n-4)!}$.

c) 1000 výherců spotřebitelské soutěže: neuspořádaný výběr s vracením, $\binom{n+1000-1}{1000}$.

Není zde zastoupen uspořádaný výběr s vracením, např. výherci prvních tří cen v literární soutěži, n^3, a

permutace s opakováním, např. počet možných způsobů rozmístění osmi bílých šachových figur (bez pěšců) na 1. řadě šachovnice, $\frac{8!}{2!\cdot 2!\cdot 2!\cdot 1\cdot 1!} = 5040$.

Cvičení 2.2. *Na party se sešlo 14 studentů, z toho 8 vysokoškoláků a 6 středoškoláků. Jaká je pravděpodobnost, že v náhodně vybrané čtveřici budou*

a) všichni čtyři vysokoškoláci,

b) právě jeden vysokoškolák?

Řešení. Počet všech možných čtveřic utvořených ze 14 studentů je $\binom{14}{4}$.

a) Počet všech čtveřic vytvořených z vysokoškoláků je $\binom{8}{4}$, tudíž hledaná pravděpodobnost je

$$P = \frac{\binom{8}{4}}{\binom{14}{4}}.$$

b) Počet všech trojic vytvořených ze středoškoláků je $\binom{6}{3}$, počet všech „jednic“ vytvořených z vysokoškoláků je $\binom{8}{1} = 8$. K vytvoření příznivých čtveřic je třeba všechny trojice zkombinovat se všemi „jednicemi“, tudíž hledaná pravděpodobnost je

$$P = \frac{8 \cdot \binom{6}{3}}{\binom{14}{4}}.$$

[1] Autorem úlohy je V. Smutný, a ač to zní neuvěřitelně, je to skutečný příběh.

Cvičení 2.3. *K úspěšnému absolvování zkoušky je potřeba nadpoloviční počet bodů z písemky. Každý příklad je hodnocen* $0, 1$*, nebo* 2 *body a student odhadl, že všechna bodová hodnocení jsou stejně pravděpodobná a nezávislá na výsledcích v ostatních příkladech. Kdy má větší šanci na úspěch, pokud bude mít zkouška* 2 *příklady, nebo* 3*?*

Výsledky. Pro 3 příklady, neboť $\frac{10}{27} > \frac{1}{3}$.

Cvičení 2.4 (hypergeometrické rozdělení)**.** *Mezi* M *výrobky je* K *vadných. Jaká je pravděpodobnost, že mezi* m *náhodně vybranými výrobky je právě* k *vadných?*

Řešení. Všechny možné výběry m z M výrobků představují $\binom{M}{m}$ elementárních jevů.
Z K vadných vybereme k výrobků $\binom{K}{k}$ způsoby,
z $M - K$ dobrých vybereme $m - k$ výrobků $\binom{M-K}{m-k}$ způsoby,
celkový počet možností je $\binom{K}{k}\binom{M-K}{m-k}$.
Výsledná pravděpodobnost je

$$\frac{\binom{K}{k}\binom{M-K}{m-k}}{\binom{M}{m}}, \qquad k \in \{0, 1, 2, \ldots, m\}.$$

Odvodili jsme tzv. *hypergeometrické rozdělení.*

Cvičení 2.5 (pravděpodobnosti zařazení do průzkumu)**.** *Alice a Bob žijí ve státě, který má* $n = 10^7$ *obyvatel. Do statistického průzkumu bude vybráno* $k = 10\,000$ *respondentů. Pro všechny čtyři typy výběrů vypočtěte počet všech možností výběru a pravděpodobnost, že do výběru bude vybrána*

a) Alice aspoň jednou,

b) Alice i Bob,

c) Alice více než jednou.

Řešení. *Uspořádaný výběr bez vracení:*
Celkový počet možností $\frac{n!}{(n-k)!} \doteq 6.730 \cdot 10^{69\,997}$.

a) Alice (stejně jako Bob) není vybrána v $\frac{(n-1)!}{(n-1-k)!} \doteq 6.723 \cdot 10^{69\,997}$ případech, tj. s pravděpodobností $\frac{(n-1)!}{(n-1-k)!} \cdot \frac{(n-k)!}{n!} = \frac{n-k}{n} = 0.999$,
je vybrána s pravděpodobností $1 - \frac{(n-1)!}{(n-1-k)!} \cdot \frac{(n-k)!}{n!} = 1 - \frac{n-k}{n} = \frac{k}{n} = 0.001$.

b) Alice ani Bob nejsou vybráni v $\frac{(n-2)!}{(n-2-k)!}$ případech, tj. s pravděpodobností $\frac{(n-2)!}{(n-2-k)!} \cdot \frac{(n-k)!}{n!} = \frac{(n-k)\,(n-k-1)}{n\,(n-1)} \doteq 0.998\,001$.
Od jednotky odečteme pravděpodobnost, že není vybrána Alice, a také, že není vybrán Bob, tj. $1 - 2\,\frac{n-k}{n}$.
To jsme ale dvakrát odečetli výběry bez Alice i Boba, a musíme je jednou přičíst. Pravděpodobnost, že bude vybrána Alice i Bob, je $1 - 2\,\frac{n-k}{n} + \frac{(n-k)\,(n-k-1)}{n\,(n-1)} = \frac{k\,(k-1)}{n\,(n-1)} \doteq 9.999 \cdot 10^{-7}$.

Alternativní řešení: Alice bude vybrána s pravděpodobností $\frac{k}{n}$, ze zbývajících obyvatel do zbytku výběru Bob s pravděpodobností $\frac{k-1}{n-1}$.

Neuspořádaný výběr bez vracení:
Celkový počet možností $\binom{n}{k} = \frac{n!}{(n-k)!\,k!} \doteq 2.365 \cdot 10^{34\,338}$.

a) Alice je vybrána v $\binom{n-1}{k-1} = \frac{(n-1)!}{(n-k)!\,(k-1)!} \doteq 2.365 \cdot 10^{34\,335}$ případech, tj. s pravděpodobností $\frac{(n-1)!}{(n-k)!\,(k-1)!} \cdot \frac{(n-k)!\,k!}{n!} = \frac{k}{n} = 0.001$.

b) Alice i Bob jsou vybráni v $\binom{n-2}{k-2}$ případech, tj. opět s pravděpodobností $\frac{(n-2)!}{(n-k)!\,(k-2)!} \cdot \frac{(n-k)!\,k!}{n!} = \frac{k\,(k-1)}{n\,(n-1)} \doteq 9.999 \cdot 10^{-7}$.

Uspořádaný výběr s vracením:
Celkový počet možností $n^k = 10^{70\,000}$.

a) Alice není vybrána v $(n-1)^k \doteq 9.99 \cdot 10^{69\,999}$ případech, tj. s pravděpodobností $\left(\frac{n-1}{n}\right)^k \doteq 0.999$,
je vybrána s pravděpodobností $1 - \left(\frac{n-1}{n}\right)^k \doteq 0.001$.

b) Alice ani Bob nejsou vybráni v $(n-2)^k \doteq 9.98 \cdot 10^{69\,999}$ případech, tj. s pravděpodobností $\left(\frac{n-2}{n}\right)^k \doteq 0.998\,001$.
Obdobně jako u výběru bez vracení, pravděpodobnost, že bude vybrána Alice i Bob, je $1-2\left(\frac{n-1}{n}\right)^k + \left(\frac{n-2}{n}\right)^k = \frac{k\,(k-1)}{n\,(n-1)} \doteq 9.989 \cdot 10^{-7}$.

c) Pokud je Alice vybrána právě jednou, může se to stát při k příležitostech; zbývajících $k-1$ respondentů je vybráno z $n-1$ obyvatel, možností je $k\,(n-1)^{k-1}$. Pravděpodobnost, že Alice bude vybrána více než jednou, je $1-\left(\frac{n-1}{n}\right)^k - \frac{k\,(n-1)^{k-1}}{n^k} \doteq 4.996 \cdot 10^{-7}$.
(U výběrů bez vracení byla nulová.)

Neuspořádaný výběr s vracením:
Celkový počet možností $\binom{n+k-1}{k} \doteq 5.203 \cdot 10^{34\,342}$, ale *nejsou stejně pravděpodobné.* Počty možností bychom mohli vypočítat, ale pravděpodobnosti z nich nelze snadno určit. Pravděpodobnosti jsou stejné jako pro uspořádaný výběr s vracením.

Cvičení 2.6. *Na šachovnici* 8×8 *náhodně rozmístíte* 8 *pěšců. Jaká je pravděpodobnost,*

a) že všech 8 *pěšců bude v jedné řadě nebo na jednom sloupci nebo na diagonále,*

b) že žádní dva pěšci nebudou na stejném sloupci ani na stejné řadě?

Výsledky. a) $\frac{18}{\binom{64}{8}}$; b) $\frac{8!}{\binom{64}{8}}$

Cvičení 2.7. *Určete pravděpodobnost toho, že při n hodech korunou v lichém počtu hodů padne lev. (Návod: Lze použít mj. matematickou indukci.)*

Řešení. První řešení: Označme jev A_n, že „při n hodech korunou v lichém počtu hodů padne lev“ a jev L, že „v posledním hodu padne lev.“ Je-li $n = 1$, pak zřejmě $P[A_1] = 1/2$.
Předpokládejme platnost tvrzení pro n. Potom pro $n+1$ platí

$$P[A_{n+1}] = P[L] \cdot P[A_{n+1}|L] + P[\bar{L}] \cdot P[A_{n+1}|\bar{L}]\,.$$

Z indukčního předpokladu víme, že

$$P[A_{n+1}|\bar{L}] = P[A_n] = 1/2\,, \qquad P[A_{n+1}|L] = P[\overline{A_n}] = 1 - P[A_n] = 1/2\,.$$

Dosadíme do předchozího vzorce a dostaneme

$$P[A_{n+1}] = 1/2\,.$$

Druhé řešení: Je-li n liché, pak ke každé posloupnosti, při které padne lichý počet lvů, je jednoznačně přiřazena posloupnost se sudým počtem padlých lvů $(0 \mapsto 1,\ 1 \mapsto 0)$. Proto je tato pravděpodobnost $1/2$. Je-li n sudé, použijeme větu o úplné pravděpodobnosti. Nechť X je jev, že „padl lichý počet lvů,“ a L je jev, že „v prvním hodu padl lev.“ Pak zřejmě $P[X|L] = P[X|\bar{L}] = 1/2$ (v prvním hodu něco padlo a zbývá lichý počet hodů).

$$P[X] = P[X|L] \cdot P[L] + P[X|\bar{L}] \cdot P[\bar{L}] = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}\,.$$

Cvičení 2.8. *Posloupnost délky* 20 *obsahuje* 17 *nul a* 3 *jedničky. Určete pravděpodobnost, že žádné dvě jedničky nejsou vedle sebe. Dále určete podmíněnou pravděpodobnost, že žádné dvě jedničky nejsou vedle sebe, jestliže víme, že jedna jednička je na páté pozici.*

Řešení. Uvažujme posloupnost délky 17 složenou ze samých nul. Doplníme mezery mezi libovolnými dvěma nulami a na začátku a na konci posloupnosti. Například pro posloupnost 5 nul a 3 jedniček

`_0_0_0_0_0_.`

Umístíme-li jedničky na místa mezer (bez opakování), žádné dvě nebudou vedle sebe. Počet mezer je $1+16+1$. Počet všech možností, jak rozmístit 3 jedničky do 18 mezer, je

$$\binom{18}{3}.$$

Počet všech možností je

$$\binom{20}{3}.$$

Výsledná pravděpodobnost je

$$\frac{\binom{18}{3}}{\binom{20}{3}} = \frac{17 \cdot 16}{20 \cdot 19} \doteq 0.7158\,.$$

Počet všech možností, kdy jednička je na páté pozici, je dán počtem možností volby dvou zbývajících jedniček na 19 pozicích, $\binom{19}{2}$. Počet všech možností, kdy jednička je na páté pozici a žádné dvě jedničky nejsou vedle sebe, je součet následujících položek:

1 (počet možností, kdy dvě nesousedící jedničky jsou před pozicí 5, tj. na pozicích 1 a 3),

$\binom{13}{2}$ (počet možností, kdy dvě nesousedící jedničky jsou za pozicí 5, tj. na pozicích 7 až 20; výpočet obdobný jako v první části),

$3 \cdot 14$ (počet možností, kdy jedna jednička je na pozici 1 až 3, druhá na pozici 7 až 20).

Výsledná podmíněná pravděpodobnost je

$$\frac{121}{171} \doteq 0.7076\,.$$

Cvičení 2.9. *Jaká je pravděpodobnost, že ve skupině n lidí mají alespoň dva narozeniny ve stejný den (neuvažujte přestupné roky a předpokládejte, že se během celého roku děti rodí rovnoměrně)?*

Řešení. Nechť A je jev, že „ve skupině n lidí mají alespoň dva narozeniny ve stejný den.“ Pak jev A^c znamená, že „ve skupině n lidí má každý člověk narozeniny v jiný den.“

$$P(A^c) = \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n},$$

a tudíž

$$P(A) = 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

Cvičení 2.10. *Během dne se v porodnici narodilo 10 dětí. Pravděpodobnost narození chlapce je $p = 0,514$. Jaká je pravděpodobnost, že se během tohoto dne narodil v porodnici:*

a) alespoň jeden chlapec,

b) maximálně dva chlapci?

Řešení. Nechť $A_i, i = 1, \ldots, 10$, značí jev „i-té narozené dítě je chlapec.“ Jev A pak znamená „alespoň jedno narozené dítě je chlapec“ a jev A^c „žádné narozené dítě není chlapec.“ Jelikož $A_1, \ldots, A_{10}$ jsou nezávislé, pak

$$\begin{aligned} P(A) = P(\cup_{i=1}^{10} A_i) = 1 - P(A^c) = 1 - \prod_{i=1}^{10}(1 - P(A_i)) = \\ = 1 - (0,486)^{10}. \end{aligned}$$

Jev $P(B)$ pak znamená, že se v ten den narodily buď samé holčičky nebo že jedno narozené dítě byl chlapec a zbytek holčičky (na pořadí nezáleží) nebo že dvě z narozených děti byli chlapci a zbytek holčičky (na pořadí rovněž nezáleží). Pak

$$P(B) = (0,486)^{10} + \binom{10}{1} 0,514 \cdot (0,486)^9 + \binom{10}{2} (0,514)^2 \cdot (0,486)^8$$

Cvičení 2.11. *V urně jsou dvě koule, bílá a černá. Provádí se výběr po jedné kouli do doby, než se vytáhne černá koule. Kdykoliv se vytáhne bílá koule, vrátí se do urny a přidají se ještě dvě bílé koule. Určete pravděpodobnost toho, že se při prvních 50 tazích černá koule nevytáhne.*

Řešení. Použijeme následující vztah, který si obecně hodí, pokud provádíme sérii pokusů:

Mějme posloupnost jevů $A_1 \supseteq A_2 \cdots \supseteq A_n$ kde $P(A_n) \neq 0$. Protože máme rovnost $A_i = A_i \cap A_{i-1}$, tak dostáváme vztah

$$P(A_i) = \frac{P(A_i \cap A_{i-1})}{P(A_{i-1})} \cdot P(A_{i-1}) = P(A_i | A_{i-1}) \cdot P(A_{i-1})\ .$$

Postupně tak iterací dostaneme

$$P(A_n) = P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_{n-1})\ .$$

Při praktickém použití pak jev A_i znamená „výsledky série prvních i pokusů." Podmíněná pravděpodobnost $P(A_i|A_{i-1})$ pak je pravděpodobnost výsledku i-tého pokusu za předpokladu, že už nastaly dané výsledky prvních $i-1$ pokusů.

V našem případě budeme mít jevy

$A_i =$ „v prvních i tazích se vytáhne bílá koule,"

podmíněné pravděpodobnosti pak budou

$$P(A_i|A_{i-1}) = \frac{2i-1}{2i} \ ,$$

protože po prvních $i-1$ pokusech je v urně $2i-1$ bílých koulí a 1 černá koule. Zřejmě je $P(A_1) = \frac{1}{2}$. Pro $n = 50$ tedy máme

$$P(A_n) = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdots \frac{2n-3}{2n-2} \cdot \frac{2n-1}{2n} = \frac{1 \cdot 2 \cdots (2n-1) \cdot 2n}{\left(2 \cdot 4 \cdots (2n-2) \cdot 2n\right)^2} = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}} = \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} \ .$$

Použitím Stirlingova vzorce $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ dostaneme přibližnou hodnotu

$$P(A_n) = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}} \doteq \frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}}{2\pi n\left(\frac{n}{e}\right)^{2n} \cdot 2^{2n}} = \frac{1}{\sqrt{\pi n}} \ ,$$

tedy pro $n = 50$ to je

$$P(A_{50}) = \frac{100!}{(50!)^2 \cdot 2^{100}} \doteq \frac{1}{\sqrt{50\pi}} \doteq 0.0798 \ .$$

Současně je vidět, že pro $n \to \infty$ se $P(A_n)$ blíží k nule. Podmíněné pravděpodobnosti $P(A_i|A_{i-1}) = \frac{2i-1}{2i}$ vytažení bílé koule v dalším tahu se přitom postupně blíží k jedné pro $i \to \infty$.

Cvičení 2.12. *Tomáš bude hrát s Ivanem a Janem tenis. Bude hrát tři sety. Může si vybrat pořadí protihráčů, v jakém se sety budou hrát: Ivan-Jan-Ivan nebo Jan-Ivan-Jan. Jestliže vyhraje dva sety po sobě, získá prémii. Jaké pořadí má Tomáš zvolit, aby měl větší šanci získat prémii, jestliže Ivan je lepší hráč než Jan?*

Řešení. Nechť a je pravděpodobnost výhry s Ivanem a b je pravděpodobnost výhry s Janem. Přitom je $a < b$. K získání prémie se musí odehrát buď série „výhra-výhra-cokoliv" nebo „prohra-výhra-výhra." Pravděpodobnost získání prémie v pořadí Ivan-Jan-Ivan je tak $ab + (1-a)ba = ab(2-a)$ a pravděpodobnost získání prémie v pořadí Jan-Ivan-Jan je $ba + (1-b)ab = ab(2-b)$. Vyšší pravděpodobnost získání prémie je tak pro pořadí Ivan-Jan-Ivan, což může působit proti-intuitivně.

3 Vlastnosti pravděpodobnosti

Cvičení 3.1. *Náhodně vybrané číslo (s rovnoměrným rozdělením) od* 1 *do* 100 *je dělitelné dvěma. Určete pravděpodobnost toho, že je toto číslo dělitelné třemi nebo pěti.*

Řešení. Označme A náhodný jev, že „vybrané číslo je dělitelné dvěma (sudé)," pak $P(A) = 0.5$. Dále označme B náhodný jev, že „vybrané číslo je dělitelné třemi" a C, že „vybrané číslo je dělitelné pěti." Počítáme tedy pravděpodobnost náhodného jevu $(B \cup C)|A$. Ze vzorce pro podmíněnou pravděpodobnost dostaneme

$$P((B \cup C)|A) = \frac{P((B \cup C) \cap A)}{P(A)} = \frac{P((A \cap B) \cup (A \cap C))}{P(A)} \ .$$

Protože

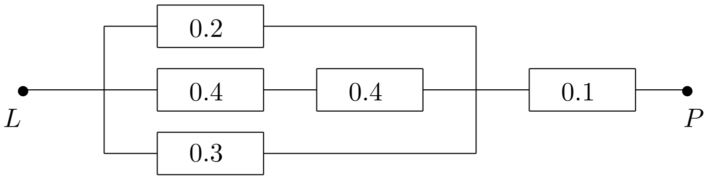
$$P((A \cap B) \cup (A \cap C)) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) = 0.16 + 0.1 - 0.03 = 0.23,$$

je

$$P((B \cup C)|A) = \frac{0.23}{0.5} = 0.46 \, .$$

Náhodně vybrané sudé číslo bude s pravděpodobností 46 % dělitelné třemi nebo pěti.

Cvičení 3.2. *Signál prochází zařízením zleva (L) doprava (P). Jednotlivé bloky mají poruchu s vyznačenou pravděpodobností a výskyty poruch jsou na sobě nezávislé. Určete pravděpodobnost toho, že vyslaná zpráva bude zařízením přenesena.*



Řešení. Budeme počítat pravděpodobnost toho, že signál zařízením neprojde. Pro paralelní zapojení dostaneme sjednocení a pro sériové zapojení průniky. Označme větve paralelního zapojení po řadě A, B, C od shora dolů. Pro prostřední větev dostaneme pravděpodobnost rozpojené cesty jako

$$P(B) = P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = 0.4 + 0.4 - 0.4 \cdot 0.4 = 0.64\,,$$

jestliže využijeme nezávislosti funkčnosti jednotlivých bloků.

Paralelní zapojení bloků bude rozpojené s pravděpodobností

$$P(D_1) = P(A \cap B \cap C) = 0.2 \cdot 0.64 \cdot 0.3 = 0.0384\,,$$

kde jsme jako D_1 označili paralelní spojení trojice bloků. Výsledné zařízení je sériovým zapojením dvou bloků D_1 a D_2, kde pravděpodobnosti rozpojení jsou rovny $P(D_1) = 0.0384$ a $P(D_2) = 0.1$. Signál neprojde s pravděpodobností

$$P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2) = 0.0384 + 0.1 - 0.0384 \cdot 0.1 = 0.13456\,.$$

Přenos zprávy se tedy uskuteční s pravděpodobností

$$P = 1 - P(D_1 \cup D_2) = 1 - 0.13456 = 0.86544\,.$$

Cvičení 3.3. *Náhodné jevy A a B jsou nezávislé a $P(A \cup B) = 0.545$, $P(A \cap B) = 0.105$. Určete pravděpodobnosti $P(A)$, $P(B)$ a $P(A \cap \overline{B})$.*

Řešení. Jestliže využijeme nezávislosti náhodných jevů A a B, pak dostaneme

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B), \qquad P(A \cap B) = P(A) \cdot P(B).$$

Označme si $P(A) = x$ a $P(B) = y$. Pro hledané pravděpodobnosti tak dostaneme soustavu rovnic

$$0.545 = x + y - 0.105, \qquad x \cdot y = 0.105 \Rightarrow y = \frac{0.105}{x}\,.$$

Po dosazení do první rovnice dostaneme kvadratickou rovnici pro neznámou x ve tvaru

$$x^2 - 0.65x + 0.105 = 0 \Rightarrow x_1 = 0.35, \quad x_2 = 0.3\,.$$

Ze symetrie vztahů plyne, že je

$$P(A) = 0.35, \quad P(B) = 0.3 \qquad \text{nebo} \qquad P(A) = 0.3, \quad P(B) = 0.35\,.$$

Dále je $P(A \cap \overline{B}) = P(A) - P(A \cap B) = 0.35 - 0.105 = 0.245$, nebo pro druhou volbu řešení $P(A \cap \overline{B}) = 0.3 - 0.135 = 0.195$.

4 Geometrická pravděpodobnost

Cvičení 4.1. *Mějme dvě zcela náhodná čísla x a y mezi 0 a 1. Jaká je pravděpodobnost, že jejich součet je menší než 1 a zároveň rozdíl $x - y$ větší než 0.5?*

Řešení. Nakreslíme-li si čtverec o hraně délky 1 značící (náhodné) souřadnice x a y, pak odpovídající plocha je část čtverce ohraničená přímkami $y = -x + 1$ směrem k bodu $[0; 0]$ a $y = x - 0,5$ směrem k bodu $[1; 0]$. Její plocha je rovna $1/16$, tudíž hledaná pravděpodobnost je

$$P = \frac{\frac{1}{16}}{1} = \frac{1}{16}.$$

Cvičení 4.2 (Buffonova úloha). *Na linkovaný papír hodíme jehlu, jejíž délka je rovna vzdálenosti mezi linkami. Jaká je pravděpodobnost, že jehla protne nějakou linku?*

Řešení. BÚNO: Délka jehly (a vzdálenost linek) je jednotková.
Označme $x \in \langle 0, \pi/2\rangle$ úhel mezi linkou a jehlou a $y \in \langle 0, 1/2\rangle$ vzdálenost středu jehly od nejbližší linky (za jednotku bereme vzdálenost mezi linkami). Předpokládáme, že tyto náhodné veličiny jsou nezávislé a mají rovnoměrná rozdělení na příslušných intervalech. Za množinu elementárních jevů vezmeme dvojrozměrný interval (obdélník) $\Omega = \langle 0, 1/2\rangle \times \langle 0, \pi/2\rangle$, na kterém máme rovnoměrné rozdělení. Jev A – protnutí linky – nastane, pokud $y < \frac{1}{2}\sin x$,

$$A = \left\{(y, x) \in \langle 0, 1/2\rangle \times \langle 0, \pi/2\rangle \mid y \le \tfrac{1}{2}\sin x\right\} .$$

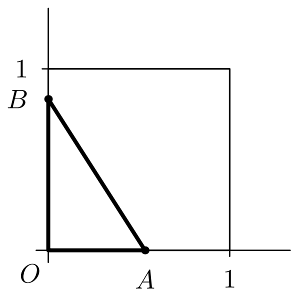
Hledaná pravděpodobnost je poměr obsahů množin A a Ω, přičemž A je plocha pod křivkou, jejíž integrací dostaneme obsah, a Ω je obdélník:

$$P(A) = \frac{1}{\frac{1}{2}\cdot\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{1}{2}\sin x \,\mathrm{d}x = \frac{2}{\pi} \doteq 0.636\,619\,772 .$$

Cvičení 4.3. *Na rovnoměrnou nekonečnou čtvercovou mřížku, kde vzdálenost průsečíků je a, hodíme minci o průměru b, $b < a$. Jaká je pravděpodobnost, že mince zakryje část některé z linek této mřížky?*

Řešení. Jev zakrytí je určen polohou středu mince. K zakrytí *nedojde*, pokud se střed mince bude nacházet uvnitř některého čtverce o straně a ve vzdálenosti alespoň b od každé jeho strany. Pravděpodobnost zakrytí tak je $1 - \frac{(a-b)^2}{a^2} = \frac{b}{a}\left(2 - \frac{b}{a}\right)$.

Cvičení 4.4. *Na sousedních stranách jednotkového čtverce jsou náhodně zvoleny dva body A a B (viz obrázek). Určete pravděpodobnost jevu, že plocha trojúhelníka $\triangle OAB$ je větší než $3/8$.*



Řešení. Počítáme pravděpodobnost $P[a\,b/2 > 3/8]$. Ta je rovna obsahu té části plochy čtverce $\langle 0, 1\rangle \times \langle 0, 1\rangle$, která leží nad hyperbolou s rovnicí $a = \frac{3}{4b}$, tj. integrálu

$$\int_{\{(a,b)|ab>3/4\}} 1 \,\mathrm{d}a\,\mathrm{d}b = \int_{3/4}^{1}\int_{3/4b}^{1} 1 \,\mathrm{d}a\,\mathrm{d}b = \int_{3/4}^{1} \left(1 - \frac{3}{4b}\right) \mathrm{d}b = \frac{1}{4} + \frac{3}{4}\ln\frac{3}{4} \doteq 0.03424 .$$

Cvičení 4.5. *Dvě osoby přijdou na smluvené místo úplně náhodně nezávisle na sobě mezi 12. a 13. hodinou, počkají 20 minut a pak odejdou. Jaká je pravděpodobnost, že se potkají?*

Řešení. Nakreslíme-li si čtverec o hraně délky 60 značící čas (uvedený v minutách) po dvanácté hodině, kdy přijde první (osa x), resp. druhá (osa y) osoba, pak odpovídající plocha je část čtverce ohraničená přímkami $y = x + 20$ a $y = x - 20$. Její plocha je rovna 2000, tudíž hledaná pravděpodobnost je

$$P = \frac{2000}{3600} = \frac{5}{9}.$$

5 Nezávislé jevy

Cvičení 5.1. *Jevy A, B, C jsou po dvou nezávislé. Platí $P(A) = P(B) = \frac{1}{2}$, dále $P(A \cup B \cup C) = 1$ a $P(A \cup C) = \frac{3}{4}$. Určete*

a) $P(A \cup B)$

b) $P(C)$

c) $P(A \cap B \cap C)$

Řešení. Máme
$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{1}{2}+\frac{1}{2}-\frac{1}{4}=\frac{3}{4},$$
neboť jevy A a B jsou nezávislé. Dále
$$P(C)=P(A\cup C)+P(A\cap C)-P(A)=\frac{3}{4}+P(A)\cdot P(C)-\frac{1}{2}=\frac{1}{4}+\frac{1}{2}\cdot P(C),$$
z čehož plyne $P(C)=\frac{1}{2}$. Celkově
$$\begin{aligned}P(A\cap B\cap C)&=P(A\cup B\cup C)-P(A)-P(B)-P(C)+P(A\cap B)+P(A\cap C)+P(B\cap C)\\&=1-3\cdot\frac{1}{2}+3\cdot\frac{1}{4}=\frac{1}{4}.\end{aligned}$$

Cvičení 5.2. *Nezávislé jevy A, B, C mají po řadě pravděpodobnosti* 0.2, 0.3, 0.4. *Určete pravděpodobnost jevu $X=(A\cup B)\cap C$.*

Řešení. Pro nezávislé jevy
$$P(A\cup B)=P(A)+P(B)-P(A)\cdot P(B)=0.2+0.3-0.2\cdot 0.3=0.44\ ,$$
$$P(X)=P((A\cup B)\cap C)=P(A\cup B)\cdot P(C)=0.44\cdot 0.4=0.176\,.$$

Cvičení 5.3 (vylepšení náhodného generátoru)**.** *Alice a Bob chtějí spravedlivě vybrat jednoho z nich. Mohou si hodit mincí, ale ta je zdeformovaná, takže jsou pochyby, zda padají oba výsledky se stejnou pravděpodobností. Dohodnou se, že hodí mincí dvakrát. Alice vyhrává, pokud padnou stejné výsledky, Bob při různých výsledcích. Kdo z nich má větší naději na výhru?*

Řešení. Líc padá s pravděpodobností $1/2+\varepsilon$,
rub s pravděpodobností $1/2-\varepsilon$, kde $\varepsilon\in(-1/2,1/2)$.
$2\times$ líc s pravděpodobností $(1/2+\varepsilon)^2$,
$2\times$ rub s pravděpodobností $(1/2-\varepsilon)^2$.
Alice vyhrává s pravděpodobností
$$\left(\tfrac{1}{2}+\varepsilon\right)^2+\left(\tfrac{1}{2}-\varepsilon\right)^2=\tfrac{1}{2}+2\,\varepsilon^2>\tfrac{1}{2}\,,$$
Bob s pravděpodobností
$$2\,\left(\tfrac{1}{2}+\varepsilon\right)\,\left(\tfrac{1}{2}-\varepsilon\right)=\tfrac{1}{2}-2\,\varepsilon^2<\tfrac{1}{2}\,.$$
Pravděpodobnost se od 1/2 liší od 1/2 o $h(\varepsilon)=2\,\varepsilon^2$ namísto ε.
Např. pro $\varepsilon=0.01$ Alice vyhrává s pravděpodobností $1/2+2\,\varepsilon^2=0.500\,2$. Udělali jsme *ze špatného náhodného generátoru lepší.*

Cvičení 5.4 (vylepšení náhodného generátoru 2)**.** *Vylepšete předchozí příklad.*

Řešení. Potřebujeme více než dva hody.
Např. při sudém počtu líců vyhrává Alice, při lichém Bob.
Pro 3 hody vyhrává Alice s pravděpodobností
$$\left(\tfrac{1}{2}-\varepsilon\right)^3+3\,\left(\tfrac{1}{2}-\varepsilon\right)\,\left(\tfrac{1}{2}+\varepsilon\right)^2=\tfrac{1}{2}-4\,\varepsilon^3\,.$$
Pro $\varepsilon=0.01$ je to 0.499 996. Pro 4 hody
$$\left(\tfrac{1}{2}-\varepsilon\right)^4+6\,\left(\tfrac{1}{2}-\varepsilon\right)^2\,\left(\tfrac{1}{2}+\varepsilon\right)^2+\left(\tfrac{1}{2}+\varepsilon\right)^4=\tfrac{1}{2}+8\,\varepsilon^4=0.500\,000\,08\,.$$
Pro velmi špatnou minci, u níž líc padá s pravděpodobností 0.9, tj. $\varepsilon=0.4$,
provedeme např. $2^5=32$ pokusů.
Pravděpodobnost, že počet líců je sudý, se liší od 1/2 o
$$h\,(h\,(h\,(h\,(h\,(\varepsilon)))))\doteq 3.96\cdot 10^{-4}\,.$$
Takto lze vytvořit *z velmi špatného náhodného generátoru libovolně dobrý* (byť pomalejší).

Cvičení 5.5. *Nechť p je prvočíslo a $\Omega=\{1,\dots,p\}$ je množina všech elementárních jevů. Pravděpodobnost každého elementárního jevu je $1/p$. Nalezněte všechny dvojice nezávislých jevů. Nestačí dvojice jevů vyjmenovat, je nutné podrobně zdůvodnit, proč to jsou pouze vámi popsané dvojice.*

Řešení. Všechny dvojice nezávislých jevů jsou

Ω a A, kde $A \subseteq \Omega$ a

$\emptyset$ a A, kde $A \subseteq \Omega$.

Aby jevy A a B byly nezávislé, musí platit

$$\frac{z}{p} = P[A \cap B] = P[A] \cdot P[B] = \frac{x}{p} \cdot \frac{y}{p},$$

z čehož plyne, že p dělí x nebo y. Tedy $x \in \{0, p\}$ nebo $y \in \{0, p\}$.

Cvičení 5.6. *Hodíme dvěma kostkami. Označme jevy*

$A =$ *„na první kostce padlo liché číslo,“*

$B =$ *„na druhé kostce padla nejvýše trojka,“*

$C =$ *„součet hodů je lichý.“*

Určete, zda jsou jevy A, B, dále B, C a A, C nezávislé. Co lze říci o nezávislosti jevů A, B, C?

Řešení. Především $P(A) = P(B) = P(C) = \frac{1}{2}$ (u C si lze například rozepsat všech 36 možností, co může padnout, z čehož dává polovina sudý součet a polovina lichý).
Jevy A, B jsou nezávislé, neboť každý je vázaný na jinou kostku a kostky hází nezávisle.
Jev $A \cap C$ znamená, že na první kostce padlo liché číslo a na druhé sudé číslo. Tyto dva jevy jsou nezávislé a pravděpodobnost každého z nich je $\frac{1}{2}$, proto je $P(A \cap C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A) \cdot P(C)$. Tedy jevy A, C jsou nezávislé.
Při jevu $B \cap C$ mohou nastat možnosti

a) Na druhé kostce padla jednička a na první sudé číslo. Tyto dva jevy jsou nezávislé a mají po řadě pravděpodobnosti $\frac{1}{6}$ a $\frac{1}{2}$.

b) Na druhé kostce padla dvojka a na první liché číslo. Tyto dva jevy jsou nezávislé a mají po řadě pravděpodobnosti $\frac{1}{6}$ a $\frac{1}{2}$.

c) Na druhé kostce padla trojka a na první sudé číslo. Tyto dva jevy jsou nezávislé a mají po řadě pravděpodobnosti $\frac{1}{6}$ a $\frac{1}{2}$.

Celkem je tedy

$$P(B \cap C) = \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{4} = P(B) \cdot P(C),$$

a proto jsou jevy B, C nezávislé.
Jev $A \cap B \cap C$ znamená, že na první kostce padlo liché číslo a na druhé dvojka. Tyto jevy jsou nezávislé a mají po řadě pravděpodobnosti $\frac{1}{2}$ a $\frac{1}{6}$. je tedy $P(A \cap B \cap C) = \frac{1}{12} \neq \frac{1}{8} = P(A) \cdot P(B) \cdot P(C)$. Jevy A, B, C tedy nejsou nezávislé, jsou pouze po dvou nezávislé.

Cvičení 5.7. *Automat vyrábí součástky ve tvaru obdélníka. Tolerance v šířce je překročena v 8 %, tolerance v délce je překročena v 7 % a v obou rozměrech ve 3 % vyrobených součástek.*

a) Rozhodněte, zda jsou porušení rozměru v délce a šířce nezávislá či závislá?

b) Vypočtěte pravděpodobnost toho, že náhodně vybraný výrobek má oba rozměry v toleranci.

Řešení. Označme A porušení tolerance v šířce a B porušení tolerance v délce.
a) Potom je

$$P(A) = 0.08, \qquad P(B) = 0.07, \qquad P(A \cap B) = 0.03\,.$$

Náhodné jevy A a B budou nezávislé, právě když platí

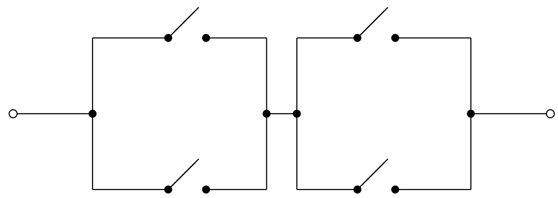
$$P(A \cap B) = P(A) \cdot P(B)\ .$$

To ale zjevně není pro náš případ pravda, protože $0.03 \neq 0.08 \cdot 0.07$. Porušení rozměrů jsou tedy závislé náhodné veličiny.
b) Dobrý výrobek je jev $\bar{A} \cap \bar{B}$, jehož pravděpodobnost je

$$\begin{aligned} P(\bar{A} \cap \bar{B}) &= P(\overline{A \cup B}) = 1 - P(A \cup B) = \\ &= 1 - P(A) - P(B) + P(A \cap B) = 1 - 0.08 - 0.07 + 0.03 = 0.88\,. \end{aligned}$$

Dobrý výrobek získáme s pravděpodobností 88 %.

Cvičení 5.8. *Hladina je kontrolována 4 spínači dle obrázku. Při nízké hladině mají být všechny sepnuty, při vysoké vypnuty. Každý z nich (nezávisle) je s pravděpodobností 10 % v opačném stavu, než by měl být. Jaká je pravděpodobnost poruchy celého zapojení v sepnutém, resp. vypnutém stavu? Porovnejte s použitím jednoho spínače.*



Řešení. Označme p pravděpodobnost, že spínač je sepnutý. Paralelní spojení dvou nezávislých spínačů je spojené s pravděpodobností $q = 1-(1-p)^2$, sériové spojení dvou takových obvodů s pravděpodobností $r = q^2$.

Nízká hladina: $p = 0.9$, $q = 0.99$, $r = 0.9801$, pravděpodobnost poruchy je $1 - r = 0.0199$.

Vysoká hladina: $p = 0.1$, $q = 0.19$, $r = 0.0361$, což je i pravděpodobnost poruchy.

V obou stavech se pravděpodobnost poruchy několikanásobně snížila oproti jednomu systému, kde byl jen jeden spínač.

Cvičení 5.9. *Stykač má být zapnut 8 hodin denně. Má-li být vypnutý, je s pravděpodobností 10 % zapnutý, má-li být zapnutý, je s pravděpodobností 5 % vypnutý.*

a) S jakou pravděpodobností nepracuje správně?

b) Jaká bude tato pravděpodobnost, pokud použijeme dva nezávislé stykače a spojíme je sériově?

c) Jaká bude tato pravděpodobnost, pokud použijeme dva nezávislé stykače a spojíme je paralelně?

Výsledky. a) $1/12 \doteq 0.083$, b) $0.1175/3 \doteq 0.039$, c) 0.1275.

6 Podmíněná pravděpodobnost a Bayesova věta

Cvičení 6.1. *Náhodně vybereme kladné celé číslo N, rozdělení pravděpodobnosti je $P[N = i] = 2^{-i}$. Poté hodíme N kostkami. Nechť S je součet hodnot, které při hodu padly.*

Určete podmíněné pravděpodobnosti $P[N = 2 \mid S = 4]$ a $P[S = 4 \mid N$ je sudé$]$.

Řešení.

$$P[N = 2 \mid S = 4] = \frac{P[N = 2, S = 4]}{P[S = 4]} = \frac{\frac{1}{4} \cdot \frac{1}{12}}{\frac{1}{2} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{12} + \frac{1}{8} \cdot \frac{3}{216} + \frac{1}{16} \cdot \frac{1}{1296}},$$

$$P[S = 4 \mid N \text{ je sudé}] = \frac{\frac{1}{4} \cdot \frac{1}{12} + \frac{1}{16} \cdot \frac{1}{1296}}{\frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \cdots} = \frac{4^2 \cdot 3^3 + 1}{4^4 \cdot 3^3}.$$

Cvičení 6.2. [2]

a) Rodina má dvě děti, starší je dcera. Jaká je pravděpodobnost, že mají dvě dcery?

b) Rodina má dvě děti, (aspoň) jedno z nich je dcera. Jaká je pravděpodobnost, že mají dvě dcery?

Řešení. a) Jde o pravděpodobnost, že mladší z dětí je dcera, což nastává s pravděpodobností q blízkou 1/2, přesněji asi 0.52. (Předpokládáme, že pohlaví dětí jsou nezávislá, což je přibližně správně.)

b) **Pozor, nejde o stejnou úlohu jako A!** Pokud pro jednoduchost předpokládáme $q = 1/2$, pak předpoklad J, že „rodina má aspoň 1 dceru," je splněn s pravděpodobností $P(J) = 1-(1-q)^2 = 3/4$, ale to, že „rodina má 2 dcery," je podjev $D \subseteq J$ s pravděpodobností $P(D) = q^2 = 1/4 = P(D \cap J)$. Podmíněná pravděpodobnost je

$$P(D|J) = \frac{P(D \cap J)}{P(J)} = \frac{P(D)}{P(J)} = \frac{1/4}{3/4} = \frac{1}{3}\,.$$

Obecněji pro pravděpodobnost narození dívky q

$$P(D|J) = \frac{q^2}{1-(1-q)^2}\,,$$

pro $q = 0.52$

$$P(D|J) = \frac{q^2}{1-(1-q)^2} \doteq 0.35\,.$$

[2]Dle David Grudl: Mozek se vzpouzí uvěřit. http://www.latrine.cz/ 5.6.2008. Jako autoři jsou uvedeni Pixy a Arthur.

Cvičení 6.3. *U 10% řidičů, kteří způsobili dopravní nehodu, bylo prokázáno požití alkoholu. Rozsáhlý průzkum ukázal, že riziko nehody se požitím alkoholu zvyšuje 7×. Odhadněte, kolik procent řidičů požilo alkohol.*

Řešení. Označme jevy

$A =$ „požil alkohol,“

$H =$ „způsobil nehodu.“

Pak máme $P(A|H) = 0.1$ a $P(H|A) = 7 \cdot P\left(H|\bar{A}\right)$. Tudíž

$$0.1 = P(A|H) = \frac{P(H|A) \cdot P(A)}{P(H|A) \cdot P(A) + P\left(H|\bar{A}\right) \cdot P\left(\bar{A}\right)} = \frac{7 \cdot P(A)}{7 \cdot P(A) + (1 - P(A))} \ .$$

Výsledek je

$$P(A) = \frac{1}{64} \ .$$

Cvičení 6.4. *Požití alkoholu bylo prokázáno u 1 % všech řidičů a u 10 % řidičů, kteří způsobili dopravní nehodu. Kolikrát se požitím alkoholu zvyšuje riziko nehody?*

Řešení. Opět si označme jevy

$A =$ „požil alkohol,“

$H =$ „způsobil nehodu.“

Pak máme $P(A) = 0.01$ a $P(A|H) = 0.1$. Tudíž

$$0.1 = P(A|H) = \frac{P(H|A) \cdot P(A)}{P(H|A) \cdot P(A) + P\left(H|\bar{A}\right) \cdot P\left(\bar{A}\right)} =$$

$$= \frac{P(H|A) \cdot 0.01}{P(H|A) \cdot 0.01 + P\left(H|\bar{A}\right) \cdot 0.99} = \frac{1}{1 + \frac{P\left(H|\bar{A}\right)}{P(H|A)} \cdot 99} \ .$$

A výsledek je

$$\frac{P(H|A)}{P\left(H|\bar{A}\right)} = 11 \ .$$

Cvičení 6.5. *V populaci je infikována 1/4 jedinců, ale jen u 2/3 infikovaných se nákaza projevuje (a u žádných neinfikovaných). Jaká je pravděpodobnost, že jedinec bez příznaků není infikovaný?*

Řešení. Označme jevy

$I =$ „člověk je infikovaný,“

$Pr =$ „člověk má příznaky.“

Pak máme $P(I) = 1/4$, $P(Pr|I) = 2/3$ a $P(Pr|\bar{I}) = 0$. Tudíž

$$P(\bar{I}|\overline{Pr}) = \frac{P(\overline{Pr}|\bar{I}) \cdot P(\bar{I})}{P(\overline{Pr})} = \frac{P(\overline{Pr}|\bar{I}) \cdot P(\bar{I})}{P(\overline{Pr}|\bar{I}) \cdot P(\bar{I}) + P(\overline{Pr}|I) \cdot P(I)} =$$

$$= \frac{1 \cdot 3/4}{1 \cdot 3/4 + 1/3 \cdot 1/4} = 0.9 \ .$$

Cvičení 6.6 (pozitivní test na nemoc). *Test nemoci je u 1 % zdravých falešně pozitivní a u 10 % nemocných falešně negativní. Nemocných je v populaci 0.001. Jaká je pravděpodobnost, že pacient s pozitivním testem je nemocný?*

Řešení. Označme jevy

$T =$ „pacient má pozitivní test,“

$N =$ „pacient je nemocný.“

Pak máme $P(N) = 0.001$, $P(T|\overline{N}) = 0.01$ a $P(\bar{T}|N) = 0.1$. Tudíž

$$P(N|T) = \frac{P(T|N) \cdot P(N)}{P(T)} = \frac{P(T|N) \cdot P(N)}{P(T|N) \cdot P(N) + P(T|\overline{N}) \cdot P(\overline{N})} =$$

$$= \frac{0.9 \cdot 0.001}{0.9 \cdot 0.001 + 0.01 \cdot 0.999} = \frac{1}{12.1} \doteq 0.08264 \ .$$

Cvičení 6.7 (výskyt nemoci v populaci). *Modifikace předchozího příkladu: Nevíme, kolik nemocných je v populaci, ale víme, že pravděpodobnost pozitivního testu je* 0.02. *(Test nemoci je u* 1 % *zdravých falešně pozitivní a u* 10 % *nemocných falešně negativní.) Odhadněte podíl nemocných je v populaci.*

Řešení. Opět označme jevy

$T =$ „pacient má pozitivní test,“

$N =$ „pacient je nemocný.“

Víme, že $P(T) = 0.02$, $P(T|\overline{N}) = 0.01$ a $P(\overline{T}|N) = 0.1$. Tudíž

$$0.02 = P(T) = P(T|\overline{N}) \cdot P(\overline{N}) + P(T|N) \cdot P(N) =$$

$$= 0.01 \cdot (1 - P(N)) + 0.9 \cdot P(N) = 0.89 \cdot P(N) + 0.01$$

a tedy

$$P(N) = \frac{0.01}{0.89} \doteq 0.011236 \ .$$

Cvičení 6.8. *Pravděpodobnost onemocnění cukrovkou je* 5 % *u těch, jejichž rodiče tuto nemoc neměli,* 10 % *tam, kde ji měl jeden z rodičů, a* 30 %, *pokud měli cukrovku oba rodiče. Předpokládáme, že onemocnění otce a matky jsou nezávislé jevy.*

a) Jaký je rovnovážný podíl c nemocných cukrovkou v populaci (stejný u generace rodičů i dětí)?

b) Jestliže pacient onemocněl cukrovkou, jaká je pravděpodobnost, že tuto nemoc měl aspoň jeden z jeho rodičů, pokud předpokládáme, že v populaci je rovnovážný výskyt?

Řešení. Označme jevy

$D =$ „potomek má cukrovku,“

$R_0 =$ „žádný z rodičů potomka nemá cukrovku,“

$R_1 =$ „právě jeden z rodičů potomka má cukrovku,“

$R_2 =$ „oba rodiče potomka mají cukrovku.“

a) Pro rovnovážný podíl pak máme vztah

$$c = P(D) = P(D|R_0) \cdot P(R_0) + P(D|R_1) \cdot P(R_1) + P(D|R_2) \cdot P(R_2) =$$

$$= 0.05 \cdot (1 - c)^2 + 0.1 \cdot 2c(1 - c) + 0.3 \cdot c^2 \ .$$

Tedy

$$c = 5.608 \cdot 10^{-2} \ .$$

b) Máme

$$P(R_1 \cup R_2|D) = 1 - P(R_0|D) = 1 - \frac{P(D|R_0) \cdot P(R_0)}{P(D)} = 1 - \frac{0.05 \cdot (1 - c)^2}{c} \doteq 1 - 0.7944 = 0.2056 \ .$$

Cvičení 6.9. *Jazykový korektor změní* 99 % *chybných slov na správná a* 0.01 % *správných na chybná. Změnil* 2 % *slov. Odhadněte množství chybných slov v jeho výstupu.*

Řešení. Předtím pravděpodobnost chybného slova p. Opraveno $0.99 \cdot p + 10^{-4} \cdot (1 - p) = 0.02$, $p = 2.0103 \cdot 10^{-2}$. Po opravě chybně $0.01 \cdot p + 10^{-4} \cdot (1 - p) = 2.9902 \cdot 10^{-4}$.

Cvičení 6.10. *Z* 60 *žijících členů klubu vysloužilých námořních kapitánů jich* 5 *zažilo ztroskotání (jednou). Podle statistiky při ztroskotání lodi v této oblasti třetina kapitánů zahyne. Odhadněte pravděpodobnost, že kapitán zažije ztroskotání (aspoň jednou za život – možnost opakovaného ztroskotání téhož kapitána i předčasného úmrtí z jiné příčiny zanedbáváme).*

Řešení. Označme jevy

$A =$ „kapitán se dožil důchodu,“

$B =$ „kapitán zažil ztroskotání.“

Pak máme $P(A|B) = \frac{2}{3}$, $P(A|\overline{B}) = 1$, $P(B|A) = \frac{5}{60} = \frac{1}{12}$ (odhad)
Bayesova věta:

$$\begin{aligned} P(B|A) &= \frac{P(B) \cdot P(A|B)}{P(B) \cdot P(A|B) + P(\overline{B}) \cdot P(A|\overline{B})} \\ \frac{1}{12} &= \frac{\frac{2}{3}P(B)}{\frac{2}{3}P(B) + (1 - P(B))} \\ P(B) &= \frac{3}{25} = 0.12 \end{aligned}$$

Alternativní řešení: Na 5 přeživších námořníků připadá v průměru $5 \cdot \frac{3}{2} = 7.5$ účastníků ztroskotání, z toho 2.5 nepřežilo, celkový počet je $60+2.5 = 62.5$ a pravděpodobnost, že se jedná o účastníka ztroskotání, je $\frac{7.5}{62.5} = \frac{3}{25}$ (tyto četnosti nám jen názorněji nahrazují pravděpodobnosti, proto není nutné, aby byly celočíselné, pokud vycházíme z toho, že statistika úmrtnosti při ztroskotáních je založena i na dalších případech kromě zde uvažovaných; z těch by nemohla vyjít $\frac{1}{3}$).

Cvičení 6.11. *Do obchodu dodávají čipy tři výrobci, po řadě* 50 %, 30 % *a* 20 % *zásoby obchodu. Pravděpodobnosti výroby dobrého čipu od jednotlivých výrobců jsou po řadě* $p_1 = 0.98$, $p_2 = 0.95$ *a* $p_3 = 0.99$.

a) Určete pravděpodobnosti, že zakoupený náhodně vybraný čip je vadný.

b) Určete pravděpodobnost, že čip je od druhého výrobce, za předpokladu, že je dobrý.

Řešení. Označme H_i, $1 \leq i \leq 3$, jev, že zakoupený čip byl od i-tého výrobce a A náhodný jev, že zakoupený čip byl dobrý.

a) Počítáme pravděpodobnost jevu $\overline{A}$, který se realizuje ve třech navzájem se vylučujících možnostech. Podle vzorce pro úplnou pravděpodobnost je

$$P(\overline{A}) = \sum_{i=1}^{3} P(\overline{A}|H_i) \cdot P(H_i) = 0.02 \cdot 0.5 + 0.05 \cdot 0.3 + 0.01 \cdot 0.2 = 0.027\,.$$

Pravděpodobnost, že náhodně vybraný čip je vadný, je rovna 0.027.

b) Požadovanou pravděpodobnost vypočteme pomocí Bayesova vzorce, kde využijeme skutečnosti, že $P(A) = 1 - P(\overline{A}) = 0.973$. Potom je

$$P(H_2|A) = \frac{P(A|H_2) \cdot P(H_2)}{P(A)} = \frac{0.95 \cdot 0.3}{0.973} = 0.293\,.$$

Cvičení 6.12. *Máme* 4 *krabice stejného vzhledu. V první jsou* 3 *bílé a* 2 *černé koule, ve druhé jsou* 2 *bílé a* 2 *černé koule, ve třetí je* 1 *bílá a* 4 *černé koule, ve čtvrté* 5 *bílých a* 1 *černá koule. Pokud nám někdo řekl, že náhodně vybral jednu z krabic a vytáhl* 1 *kouli, která byla bílá, s jakou pravděpodobností můžeme usuzovat, že v téže krabici se nachází alespoň* 3 *černé koule?*

Výsledky. Jediná krabice obsahující alespoň 3 černé koule je třetí krabice. Podmíněná pravděpodobnost pak bude $\frac{3}{32} = 0.09375$.

Cvičení 6.13. *Cestovatel přijede do města, kde je* 30% *lhářů,* 15% *náladových a* 55% *normálních lidí. Lháři lžou s pravděpodobností* 0.9. *Normální lidi mluví s pravděpodobností* 0.75 *pravdu. Náladový lidé v polovině případů lžou a v polovině říkají pravdu. Cestovatel potkal jednoho z obyvatel města a zeptal se ho, jestli je normální. Jaká je pravděpodobnost, že mu cizinec odpoví, že je normální?*

Řešení. Označme si jevy:

$A =$ „cestovatel se zeptal lháře,“

$B =$ „cestovatel se zeptal náladového člověka,“

$C =$ „cestovatel se zeptal normálního člověka,“

$W =$ „dotyčný odpoví, že je normální“

(doplňkový jev: $\overline{W} =$ „dotyčný odpoví, že není normální“).

Víme, že A, B, C je úplný disjunktní systém jevů a

$$P(A) = 0.3 \quad P(B) = 0.15 \quad P(C) = 0.55$$

$$P(W|A) = 0.9 \quad P(W|B) = 0.5 \quad P(W|C) = 0.75$$

a zajímá nás $P(W)$. Z věty o úplné pravděpodobnosti máme:

$$P(W) = P(W|A) \cdot P(A) + P(W|B) \cdot P(B) + P(W|C) \cdot P(C) =$$

$$= 0.9 \cdot 0.3 + 0.5 \cdot 0.15 + 0.75 \cdot 0.55 = 0.7575 \ .$$

Cizinec tedy odpoví, že je normální, s pravděpodobností $75,75\%$.
Je dobré si ještě všimnout, jak souvisí jev $W =$ „dotyčný odpoví, že je normální“ s jevy

$T =$ „dotyčný odpovídá pravdivě,“

$F =$ „dotyčný odpovídá nepravdivě.“

Vztah je zřejmě tento:

$$W = (A \cap F) \cup (B \cap F) \cup (C \cup T) \ .$$

Cvičení 6.14. *Ve sbírce* 50 *obrazů je* 5 *padělků. Jestliže je obraz falešný, znalec to pozná s pravděpodobností* 80%. *Je-li obraz originál, znalec ho mylně posoudí s pravděpodobností* 5%. *Určete*

a) pravděpodobnost, že obraz je originál, jestliže byl znalcem označen za originál,

b) pravděpodobnost, že obraz je padělaný, jestliže byl znalcem označen za padělek.

Řešení. Označme jevy:

$O =$ „obraz je originál,“

(doplňkový jev: $\bar{O} =$ „obraz je padělek“)

$Z =$ „znalec obraz označí za originál,“

(doplňkový jev: $\bar{Z} =$ „znalec obraz označí za padělek“).

Víme, že

$$P(\bar{O}) = \frac{5}{50} = 0.1 \quad P(\bar{Z}|\bar{O}) = 0.8 \quad P(\bar{Z}|O) = 0.05 \ .$$

a) Zajímá nás $P(O|Z)$. Z Bayesových vět máme:

$$P(O|Z) = \frac{P(Z|O) \cdot P(O)}{P(Z)} = \frac{\big(1 - P(\bar{Z}|O)\big) \cdot \big(1 - P(\bar{O})\big)}{1 - P(\bar{Z})}$$

$$P(\bar{Z}) = P(\bar{Z}|O) \cdot P(O) + P(\bar{Z}|\bar{O}) \cdot P(\bar{O}) = 0.05 \cdot 0.9 + 0.8 \cdot 0.1 = 0.125$$

a tedy

$$P(O|Z) = \frac{0.95 \cdot 0.9}{0.875} = 0.9771 \ .$$

b) Zajímá nás $P(\bar{O}|\bar{Z})$. Z Bayesovy věty a předchozího máme:

$$P(\bar{O}|\bar{Z}) = \frac{P(\bar{Z}|\bar{O}) \cdot P(\bar{O})}{P(\bar{Z})} = \frac{0.8 \cdot 0.1}{0.125} = 0.64 \ .$$

Cvičení 6.15. *Když vypravěč vypráví dětem nějaký svůj zážitek, je šance* 40%, *že si vymýšlí. Děti to poznají v* 70% *případů. Na druhou stranu, i když si zrovna nevymýšlí, tak mu zase ve* 20% *případů nevěří.*

a) Jaká je pravděpodobnost, že následující zážitek budou děti vypravěči věřit?

b) Další vypravěčův příběh mu děti uvěřili. Jaká je pravděpodobnost, že vypravěč si nevymýšlel?

Výsledky. a) 0.6 ;   b) 0.8.

Cvičení 6.16. *Student přišel na zkoušku na poslední chvíli, ale nemůže si vzpomenout, který ze tří možných (stejně pravděpodobných) předmětů (analýza, pravděpodobnost, algebra) se vlastně zkouší. Ví, že neumí* 40% *otázek z analýzy,* 35% *z pravděpodobnosti a* 20% *z algebry.*

a) Jaká je pravděpodobnost, že zkoušku udělá?

b) Jestliže zkoušku udělal, jaká je pravděpodobnost toho, že to bylo z pravděpodobnosti?

Výsledky. a) $\frac{41}{60} \doteq 0.6833$; b) $\frac{13}{41} \doteq 0.3171$.

Cvičení 6.17. *Máme testovací metodu pro výrobek, která ale nefunguje vždy. Pokud je výrobek defektní, test ukáže, že je defektní s pravděpodobností* 0.9*. Pokud je výrobek funkční, test řekne, že je defektní s pravděpodobností* 0.2*. Daný výrobek je funkční s pravděpodobností* 0.7*. Nyní nezávisle otestujeme* 3 *výrobky s výsledkem testu: funkční. Jaká je pravděpodobnost, že všechny výrobky jsou skutečně funkční?*

Výsledky. $\left(\frac{56}{59}\right)^3 \doteq (0.9492)^3 \doteq 0.8552$.

Cvičení 6.18. *Kocourkovští obyvatelé mluví v sudých týdnech pravdu s pravděpodobností* $2/3$ *a v týdnech lichých pouze s pravděpodobností* $1/5$*. Vy jste se vydali do Kocourkova navštívit tamní vyhlášenou restauraci a na cestu se zeptáte někoho místního.*

a) Jaká je pravděpodobnost, že je zrovna lichý týden, jestliže jste se podle popisu skutečně do restaurace dostali?

b) Jaká je pravděpodobnost, že je sudý týden, když jste podle popisu do restaurace nedorazili?

Výsledky. Předpokládáme, že sudé a liché týdny mají stejnou pravděpodobnost (je jich 52 v roce).
a) $\frac{3}{13} \doteq 0.2308$; b) $\frac{5}{17} \doteq 0.2941$.

Cvičení 6.19. *Na technické vysoké škole se čtyřmi fakultami A,B,C,D studuje* 13.85 % *dívek, přičemž na fakultě A je třikrát snazší potkat dívku než na fakultě B. Na fakultě C studuje* 15 % *dívek (tedy* 85 % *kluků), na fakultě D jich je* 10 % *(tedy* 90 % *kluků). Poměr počtu studentů na jednotlivých fakultách uvádí následující tabulka. Každý student studuje právě na jedné fakultě.*

fakulta	*A*	*B*	*C*	*D*
studentů	10 %	35 %	24 %	31 %

a) Jakou máte šanci, že když potkáte studenta z fakulty A, že to bude dívka?

b) Pokud jste právě potkali dívku z technické vysoké školy, na jaké fakultě nejpravděpodobněji/nejméně pravděpodobně studuje? Jaká je šance, že studuje fakultu D?

c) Pokud za závěsem stojí student technické vysoké školy, s jakou pravděpodobností je to kluk studující na fakultě C?

Řešení. a) Označme po řadě A, B, C, D jevy, že vybraný student studuje na fakultě A,B,C,D a F, že student je ženského pohlaví. Protože jsou jevy A, B, C, D možné (tj. žádný z nich není nemožný), po dvou disjunktní, a $P(A \cup B \cup C \cup D) = 1$, platí dle věty o úplné pravděpodobnosti

$$P(F) = P(F|A) \cdot P(A) + P(F|B) \cdot P(B) + P(F|C) \cdot P(C) + P(F|D) \cdot P(D).$$

Víme, že $P(F|A) = 3 \cdot P(F|B)$, a dále $P(F|C) = 0.15$, $P(F|D) = 0.1$. Dosazením dostáváme

$$\begin{aligned} 0.1385 &= 3 \cdot P(F|B) \cdot 0.1 + P(F|B) \cdot 0.35 + 0.15 \cdot 0.24 + 0.1 \cdot 0.31 \\ 0.0715 &= 0.65 \cdot P(F|B) \\ P(F|B) &= 0.11. \end{aligned}$$

Tedy pravděpodobnost, že náhodně vybraný student z fakulty B je dívka, činí 11 %. Z toho plyne, že pravděpodobnost, že náhodně vybraný student z fakulty A je dívka, činí $3 \cdot 11$ % = 33 %.

b) Chceme uspořádat pravděpodobnosti $P(A|F)$, $P(B|F)$, $P(C|F)$, $P(D|F)$ dle velikosti a znát přesnou hodnotu $P(D|F)$. Dle definice dostáváme

$$P(D|F) = \frac{P(D \cap F)}{P(F)} = \frac{P(F|D) \cdot P(D)}{P(F)} = \frac{0.1 \cdot 0.31}{0.1385} \doteq 0.224.$$

Tedy pravděpodobnost, že dívka bude studovat na fakultě D, je přibližně rovna 22 %. Chceme-li znát pouze uspořádání pravděpodobností, že dívka studuje na té které fakultě, a ne jejich přesné hodnoty, lze porovnávané výrazy jistě pronásobit členem $P(F)$: to na uspořádání nic nezmění. Tedy namísto výrazů $P(A|F)$, $P(B|F), ...,$ se budeme zabývat výrazy $P(A \cap F)$, $P(B \cap F), ...$ Máme tedy

$$P(A \cap F) = P(F|A) \cdot P(A) = 0.33 \cdot 0.1 = 0.033$$

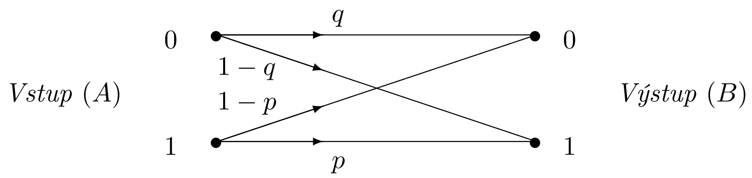
a obdobným způsobem zjistíme, že $P(B \cap F) = 0.0385$, $P(C \cap F) = 0.036$ a $P(D \cap F) = 0.031$. Porovnáním zjistíme, že s největší pravděpodobností dívka studuje na fakultě B, naopak nejméně pravděpodobné je, že studuje na fakultě D.

c) Chceme zjistit hodnotu výrazu $P(C \cap \bar{F})$. Čili

$$P(\bar{F} \cap C) = P(\bar{F}|C) \cdot P(C) = (1 - P(F|C)) \cdot P(C) = 0.85 \cdot 0.24 = 0.204,$$

a tedy pravděpodobnost, že za závěsem stojí kluk z fakulty C, je asi 20 %.

Cvičení 6.20. *Binární komunikační kanál přenáší symboly* 0 *a* 1 *ze vstupu (jevy* A_0 *a* A_1*) na výstup (jevy* B_0 *a* B_1*) podle uvedeného schématu, kdy je vlivem šumu s určitou pravděpodobností zaměněn symbol* 1 *za symbol* 0 *nebo naopak.*



Jestliže je $q = P(B_0|A_0) = 0.9$, $p = P(B_1|A_1) = 0.8$ *a* $P(A_0) = 0.7$, $P(A_1) = 0.3$, *určete:*

a) pravděpodobnosti $P(B_0)$ *a* $P(B_1)$ *výskytu symbolů na výstupu;*

b) pravděpodobnost toho, že vyslaný symbol bude přenesen správně.

Řešení. Pro $i = 0, 1$ máme jevy:

A_i = „byl vyslán znak i,“

B_i = „byl přijat znak i.“

a) Pravděpodobnosti výskytu symbolů na výstupu určíme pomocí vzorce pro úplnou pravděpodobnost. Je tedy

$$\begin{bmatrix} P(B_0) & P(B_1) \end{bmatrix} = \begin{bmatrix} P(A_0) & P(A_1) \end{bmatrix} \cdot \begin{bmatrix} P(B_0|A_0) & P(B_1|A_0) \\ P(B_0|A_1) & P(B_1|A_1) \end{bmatrix} =$$
$$= \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.69 & 0.31 \end{bmatrix}$$

b) Pravděpodobnost P správného přenosu je rovna součtu pravděpodobností správného přenosu symbolů 0 a 1, tedy

$$P = P(A_0 \cap B_0) + P(A_1 \cap B_1) = P(B_0|A_0) \cdot P(A_0) + P(B_1|A_1) \cdot P(A_1) = 0.9 \cdot 0.7 + 0.8 \cdot 0.3 = 0.87\,.$$

Cvičení 6.21. *Na vstupu binárního informačního kanálu mohou být znaky* 0 *a* 1 *(jevy* A_0 *a* A_1*). Na výstupu (jevy* B_0 *a* B_1*) jsou znaky přečteny s nezávislou pravděpodobností chyby* $p = 0.1$. *Určete podmíněné pravděpodobnosti vstupu při známém výstupu, je-li apriorní pravděpodobnost jedničky*

a) $P(A_1) = 0.4$,

b) $P(A_1) = 0.1$,

c) $P(A_1) = 0.05$

Řešení. a)

$$\begin{bmatrix} P(B_0) & P(B_1) \end{bmatrix} = \begin{bmatrix} P(A_0) & P(A_1) \end{bmatrix} \cdot \begin{bmatrix} P(B_0|A_0) & P(B_1|A_0) \\ P(B_0|A_1) & P(B_1|A_1) \end{bmatrix} =$$

$$= \begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.58 & 0.42 \end{bmatrix},$$

$$P(A_0|B_0) = \frac{P(B_0|A_0) \cdot P(A_0)}{P(B_0)} = \frac{0.9 \cdot 0.6}{0.58} \doteq 0.93103,$$

$$P(A_1|B_0) = \frac{P(B_0|A_1) \cdot P(A_1)}{P(B_0)} = \frac{0.1 \cdot 0.4}{0.58} \doteq 6.8966 \cdot 10^{-2},$$

$$P(A_0|B_1) = \frac{P(B_1|A_0) \cdot P(A_0)}{P(B_1)} = \frac{0.1 \cdot 0.6}{0.42} \doteq 0.14286,$$

$$P(A_1|B_1) = \frac{P(B_1|A_1) \cdot P(A_1)}{P(B_1)} = \frac{0.9 \cdot 0.4}{0.42} \doteq 0.85714.$$

b)

$$\begin{bmatrix} P(B_0) & P(B_1) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.82 & 0.18 \end{bmatrix},$$

$$P(A_0|B_0) = \frac{P(B_0|A_0) \cdot P(A_0)}{P(B_0)} = \frac{0.9 \cdot 0.9}{0.82} = 0.9878,$$

$$P(A_1|B_0) = \frac{P(B_0|A_1) \cdot P(A_1)}{P(B_0)} = \frac{0.1 \cdot 0.1}{0.82} = 1.2195 \cdot 10^{-2},$$

$$P(A_0|B_1) = \frac{P(B_1|A_0) \cdot P(A_0)}{P(B_1)} = \frac{0.1 \cdot 0.9}{0.18} = 0.5,$$

$$P(A_1|B_1) = \frac{P(B_1|A_1) \cdot P(A_1)}{P(B_1)} = \frac{0.9 \cdot 0.1}{0.18} = 0.5.$$

c)

$$\begin{bmatrix} P(B_0) & P(B_1) \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.86 & 0.14 \end{bmatrix},$$

$$P(A_0|B_0) = \frac{P(B_0|A_0) \cdot P(A_0)}{P(B_0)} = \frac{0.9 \cdot 0.95}{0.86} \doteq 0.99419,$$

$$P(A_1|B_0) = \frac{P(B_0|A_1) \cdot P(A_1)}{P(B_0)} = \frac{0.1 \cdot 0.05}{0.86} \doteq 5.8140 \cdot 10^{-3},$$

$$P(A_0|B_1) = \frac{P(B_1|A_0) \cdot P(A_0)}{P(B_1)} = \frac{0.1 \cdot 0.95}{0.14} \doteq 0.67857,$$

$$P(A_1|B_1) = \frac{P(B_1|A_1) \cdot P(A_1)}{P(B_1)} = \frac{0.9 \cdot 0.05}{0.14} \doteq 0.32143$$

Závěr: Je-li výstup 1, pak v případě (b) je stejně pravděpodobné, že vstup je 0 nebo 1; v případě (c) je dokonce pravděpodobnější, že vstup je 0 (takže bayesovské rozhodování vede k závěru, že na vstupu jsou samé nuly).

Cvičení 6.22. *Binárním informačním kanálem jsou zasílány symboly* 0 *a* 1 *(jevy* A_0 *a* A_1*). Pravděpodobnost správné detekce symbolu* 1 *je* $p = 0.95$ *a symbolu* 0 *je* $q = 0.8$ *(jevy* B_0 *a* B_1*).*

Jaké budou pravděpodobnosti hodnot na výstupu po průchodu dvěma články, jsou-li na vstupu počáteční pravděpodobnosti symbolů 1 *a* 0 *po řadě* 0.4 *a* 0.6*?*

Řešení. Máme dány tyto pravděpodobnosti $P(A_1) = 0.4$, $P(A_0) = 0.6$ a $P(B_1|A_1) = 0.95$, $P(B_0|A_1) = 0.05$, $P(B_0|A_0) = 0.8$, $P(B_1|A_0) = 0.2$.

Pravděpodobnosti na výstupu po průchodu jedním článkem jsou

$$(P(B_1), P(B_0)) = (P(A_1), P(A_0)) \begin{pmatrix} P(B_1|A_1), & P(B_0|A_1) \\ P(B_1|A_0), & P(B_0|A_0) \end{pmatrix}.$$

Označme si náhodné jevy

$$C_i = \text{„po průchodu dvěma články na výstupu je zachycen symbol } i\text{,“}$$

pro $i = 0, 1$. Po průchodu dvěma články se matice transformace násobí, tedy hledané pravděpodobnosti jsou dány vzorcem

$$(P(C_1), P(C_0)) = (P(A_1),\, P(A_0)) \begin{pmatrix} P(B_1|A_1), & P(B_0|A_1) \\ P(B_1|A_0), & P(B_0|A_0) \end{pmatrix}^2 .$$

Po dosazení zadaných hodnot dostaneme

$$(P(C_1), P(C_0)) = (0.4,\, 0.6) \begin{pmatrix} 0.95, & 0.05 \\ 0.2, & 0.8 \end{pmatrix}^2 = (0.4,\, 0.6) \begin{pmatrix} 0.9125, & 0.0875 \\ 0.35, & 0.65 \end{pmatrix} = (0.575,\, 0.425)\ .$$

Cvičení 6.23. *Binárním informačním kanálem na obrázku jsou zasílány symboly* 0 *a* 1 *(jevy* A_0 *a* A_1*) a zachytávány na výstupu (jevy* B_0 *a* B_1*). Pravděpodobnost správné detekce symbolu* 1 *je* $p = 0.8$ *a symbolu* 0 *je* $q = 0.9$*. Známe pravděpodobnosti zachycena výstupu* $(P(B_0),\, P(B_1)) = (0.7,\, 0.3)$*.*

a) Jaké jsou pravděpodobnosti $(P(A_0),\, P(A_1))$ *dvojice* $(0,\, 1)$ *na vstupu vstupních hodnot?*

b) Jaká je pravděpodobnost toho, že detekovaná 0 *na výstupu byla odeslána jako* 0 *ze vstupu?*

c) Jaká je pravděpodobnost správné detekce, tedy situace, že detekovaný symbol na výstupu byl jako takový odeslán ze vstupu?

d) Jaká je pravděpodobnost chybné detekce, tedy situace, že detekovaný symbol na výstupu byl jako opačný odeslán ze vstupu?

Řešení. Máme zadány uvedené podmíněné pravděpodobnosti

$$P(B_0|A_0) = 0.9, \qquad P(B_1|A_0) = 0.1, \qquad P(B_0|A_1) = 0.2, \qquad P(B_1|A_1) = 0.8.$$

a) Ze vzorce pro úplnou pravděpodobnost dostaneme pro hledané pravděpodobnosti na vstupu soustavu rovnic

$$\begin{aligned} P(B_0) &= P(A_0)\cdot P(B_0|A_0) + P(A_1)\cdot P(B_0|A_1)\,, \qquad & 0.7 &= 0.9\cdot P(A_0) + 0.2\cdot P(A_1)\,, \\ P(B_1) &= P(A_0)\cdot P(B_1|A_0) + P(A_1)\cdot P(B_1|A_1)\,, \qquad & 0.3 &= 0.1\cdot P(A_0) + 0.8\cdot P(A_1)\,. \end{aligned}$$

Soustava má řešení

$$\begin{aligned} P(A_0) &= \frac{0.7\cdot 0.8 - 0.3\cdot 0.2}{0.9\cdot 0.8 - 0.1\cdot 0.2} = \frac{0.5}{0.7} \doteq 0.7143\,, \\ P(A_1) &= \frac{0.9\cdot 0.3 - 0.7\cdot 0.1}{0.7} = \frac{0.2}{0.7} \doteq 0.2857\,. \end{aligned}$$

Pravděpodobnosti symbolů na vstupu jsou $(P(A_0),\, P(A_1)) \doteq (0.7143,\, 0.2857)$.

b) Hledáme pravděpodobnost jevu, kdy detekovaná 0 na výstupu byla jako 0 vyslána ze vstupu, tedy podmíněnou pravděpodobnost $P(A_0|B_0)$:

$$P(A_0|B_0) = \frac{P(B_0|A_0)\cdot P(A_0)}{P(B_0)} \doteq \frac{0.9\cdot 0.7143}{0.7} \doteq 0.9184\,.$$

c) Pravděpodobnost P_s správné detekce bude

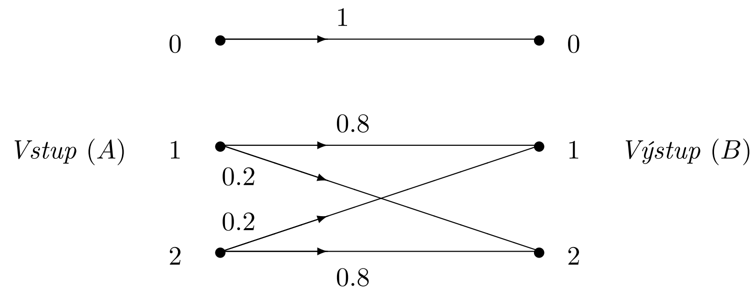
$$P_s = P(A_0\cap B_0) + P(A_1\cap B_1) = P(B_0|A_0)\cdot P(A_0) + P(B_1|A_1)\cdot P(A_1) \doteq$$

$$\doteq 0.9\cdot 0.7143 + 0.8\cdot 0.2857 = 0.87143\,.$$

d) Jedná se o doplňkovou pravděpodobnost k části c), $P_n = 1 - P_s = \frac{9}{70} \doteq 0.12857$.
Výpočet jako v části c) by vedl k následujícím výsledkům:

$$\begin{aligned} P_n &= P(A_1\cap B_0) + P(A_0\cap B_1) \doteq \\ &\doteq 0.2\cdot 0.2857 + 0.1\cdot 0.7143 = 0.05714 + 0.07143 = 0.12857\,. \end{aligned}$$

Cvičení 6.24. *Možné znaky na vstupu jsou* $0, 1, 2$ *(jevy* A_0*,* A_1 *a* A_2*). Zařízení je přečte a uloží (jevy* B_0*,* B_1 *a* B_2*). Pravděpodobnost, že znak* 1 *nebo* 2 *přečte a uloží správně, je* $80\ \%$*, přičemž pokud zařízení udělá chybu, pak vždy zamění* 1 *za* 2 *a naopak. Chyby se vyskytují nezávisle. Nulu přečte a uloží správně vždy.*



Pravděpodobnost, že na vstupu se objeví 0*, je rovna* 40 %*, pravděpodobnost každého ze znaků* 1 *a* 2 *je* 30 %*.*

a) Jaká je pravděpodobnost, že se znak na vstupu uloží jako 2*?*

b) Uložily se znaky $(1,1)$*. Jaká je pravděpodobnost, že byly oba na vstupu?*

Řešení. Pro $i = 0,1,2$ si označme jevy

$A_i =$ „znak na vstupu je i,“

$B_i =$ „znak se uloží jako i.“

a) Počítáme dle věty o úplné pravděpodobnosti

$$P(B_2) = P(B_2|A_1)\cdot P(A_1) + P(B_2|A_2)\cdot P(A_2) + P(B_2|A_0)\cdot P(A_0) = 0.2\cdot 0.3 + 0.8\cdot 0.3 + 0\cdot 0.4 = 0.3.$$

Neboť role 1 a 2 je symetrická, máme také $P(B_1) = 0.3$.

b) Zjistíme nejprve pomocí Bayesovy věty, že

$$P(A_1|B_1) = \frac{P(B_1|A_1)\cdot P(A_1)}{P(B_1)} = \frac{0.8\cdot 0.3}{0.3} = 0.8.$$

Protože chyby se vyskytují nezávisle, platí

$$P(A_{(1,1)}|B_{(1,1)}) = (P(A_1|B_1))^2 = 0.8^2 = 0.64.$$

Cvičení 6.25. *Zpráva je kódována znaky „tečka“, „čárka“, „mezera“, jejich pravděpodobnosti vyslání jsou po řadě* 0.4, 0.4, 0.2*. Vyslaná tečka je s pravděpodobností* 0.2 *přijata jako čárka, s pravděpodobností* 0.2 *jako mezera. Vyslaná čárka je s pravděpodobností* 0.2 *přijata jako tečka, s pravděpodobností* 0.1 *jako mezera. Mezera je vždy přijata správně. Jaké jsou pravděpodobnosti jednotlivých vyslaných znaků, jestliže byla přijata čárka?*

Řešení. Označme jevy vyslání znaku $A_{tečka}$, $A_{čárka}$, A_{mezera}, přijetí znaku $B_{tečka}$, $B_{čárka}$, B_{mezera}.

$$\begin{aligned}
&\begin{bmatrix} P(B_{tečka}) & P(B_{čárka}) & P(B_{mezera}) \end{bmatrix} = \\
&= \begin{bmatrix} P(A_{tečka}) & P(A_{čárka}) & P(A_{mezera}) \end{bmatrix} \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0 & 1 \end{bmatrix} = \\
&= \begin{bmatrix} 0.4 & 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.32 & 0.36 & 0.32 \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
P\left(A_{tečka}|B_{čárka}\right) &= \frac{P\left(B_{čárka}|A_{tečka}\right)\cdot P\left(A_{tečka}\right)}{P\left(B_{čárka}\right)} = \frac{0.2\cdot 0.4}{0.36} = \frac{2}{9} \doteq 0.222\,, \\
P\left(A_{čárka}|B_{čárka}\right) &= \frac{P\left(B_{čárka}|A_{čárka}\right)\cdot P\left(A_{čárka}\right)}{P\left(B_{čárka}\right)} = \frac{0.7\cdot 0.4}{0.36} = \frac{7}{9} \doteq 0.778\,, \\
P\left(A_{mezera}|B_{čárka}\right) &= \frac{P\left(B_{čárka}|A_{mezera}\right)\cdot P\left(A_{mezera}\right)}{P\left(B_{čárka}\right)} = \frac{0\cdot 0.2}{0.36} = 0\,.
\end{aligned}$$

7 Směs náhodných veličin

Cvičení 7.1. *Máme 2 hrací kostky, na jedné padají pouze lichá čísla* $1, 3, 5$*, na druhé pouze sudá,* $2, 4, 6$*, všechna se stejnou pravděpodobností* $1/3$*. Najděte rozdělení a střední hodnotu výsledků následujících pokusů:*

a) hodíme oběma kostkami a vezmeme aritmetický průměr obou čísel,

b) náhodně (s pravděpodobností $1/2$*) vybereme jednu kostku a tou hodíme.*

Řešení. a) Označme si náhodné veličiny:

X = „výsledek na 1. kostce,“

Y = „výsledek na 2. kostce.“

Zajímá nás rozdělení náhodné veličiny $Z = \frac{X+Y}{2}$. Rozlišíme 9 stejně pravděpodobných možností, vedoucích k následujícím výsledkům:

$y \backslash x$	1	3	5
2	1.5	2.5	3.5
4	2.5	3.5	4.5
6	3.5	4.5	5.5

Možné výsledky veličiny $Z = \frac{X+Y}{2}$ a jejich pravděpodobnosti:

i	1.5	2.5	3.5	4.5	5.5
$p_Z(i)$	1/9	2/9	3/9	2/9	1/9

Střední hodnota je

$$\mathrm{E}Z = \frac{1}{9} \cdot 1.5 + \frac{2}{9} \cdot 2.5 + \frac{3}{9} \cdot 3.5 + \frac{2}{9} \cdot 4.5 + \frac{1}{9} \cdot 5.5 = 3.5\,.$$

b) Tentokrát s pravděpodobností $1/2$ určuje výsledek první kostka, se stejnou pravděpodobností druhá. Opět máme veličiny

X = „výsledek na 1. kostce,“

Y = „výsledek na 2. kostce.“

Nyní ale půjde o směs náhodných veličin $Z = \mathrm{Mix}_{1/2}(X, Y)$. Dostáváme 6 stejně pravděpodobných výsledků, rozdělení je stejné jako u normální hrací kostky.

i	1	2	3	4	5	6
$p_Z(i)$	1/6	1/6	1/6	1/6	1/6	1/6

Střední hodnota je stejná,

$$\mathrm{E}Z = \frac{1}{6}\,(1 + 2 + 3 + 4 + 5 + 6) = 3.5\,.$$

Cvičení 7.2. *V urně je* 15 *hracích kostek, z toho* 10 *správných, na nichž padají všechna čísla se stejnou pravděpodobností, a* 5 *vadných, na nichž padá šestka s pravděpodobností* $1/2$*, ostatní čísla s pravděpodobností* $1/10$*. Náhodně vybereme jednu kostku a hodíme; jaká je pravděpodobnost možných výsledků?*

Řešení. Označme si náhodné veličiny:

X = „výsledek na správné kostce,“

Y = „výsledek na vadné kostce.“

Zajímá nás výsledek celého pokusu, tedy veličina $Z = \mathrm{Mix}_{2/3}(X, Y)$, což je směs náhodných veličin X, Y s koeficientem $c = 10/15 = 2/3$.

$$P[Z = t] = \frac{2}{3} \cdot P[X = t] + \frac{1}{3} \cdot P[Y = t]$$
$$P[Z = 1] = \frac{2}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{10} = \frac{13}{90}$$
$$P[Z = 6] = \frac{2}{3} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{18}$$

t	1	2	3	4	5	6
$P[X=t]$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
$P[Y=t]$	$1/10$	$1/10$	$1/10$	$1/10$	$1/10$	$1/2$
$P[Z=t]$	$13/90$	$13/90$	$13/90$	$13/90$	$13/90$	$5/18$

Cvičení 7.3. *Náhodná veličina X má diskrétní rovnoměrné rozdělení na množině $\{1,2,3\}$. Náhodná veličina Y má spojité rovnoměrné rozdělení na intervalu $\langle 0,3\rangle$. Popište a znázorněte rozdělení směsi* $\mathrm{Mix}_{1/2}(X,Y)$.

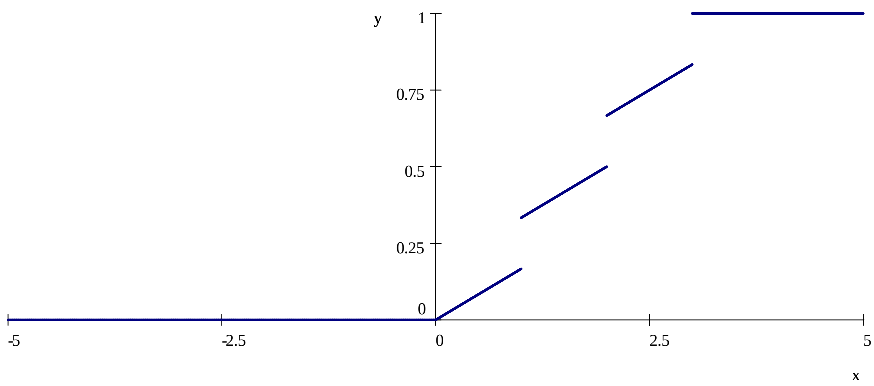
Řešení. Máme $F_{\mathrm{Mix}_{1/2}(X,Y)} = \frac{1}{2}\cdot F_X + \frac{1}{2}\cdot F_Y$. Distribuční funkce jednotlivých veličin pak jsou

$$F_X(t) = \begin{cases} 0, & t\in(-\infty,1),\\ \frac{1}{3}, & t\in\langle 1,2),\\ \frac{2}{3}, & t\in\langle 2,3),\\ 1, & t\in\langle 3,\infty), \end{cases}$$

$$F_Y(t) = \begin{cases} 0, & t\in(-\infty,0),\\ \frac{t}{3}, & t\in\langle 0,3),\\ 1, & t\in\langle 3,\infty), \end{cases}$$

a

$$F_{\mathrm{Mix}_{1/2}(X,Y)}(t) = \begin{cases} 0, & t\in(-\infty,0),\\ \frac{t}{6}, & t\in\langle 0,1),\\ \frac{1}{6}+\frac{t}{6}, & t\in\langle 1,2),\\ \frac{1}{3}+\frac{t}{6}, & t\in\langle 2,3),\\ 1, & t\in\langle 3,\infty). \end{cases}$$



8 Druhy náhodných veličin

8.1 Diskrétní náhodné veličiny

Cvičení 8.1. *Uvažujme hod mincí, tj.* $\Omega = \{\omega_1, \omega_2, \omega_3\}$*, kde*
ω_1 *je jev, že padl rub (*$P(\omega_1) = 0.49$*),*
ω_2 *je jev, že padl líc (*$P(\omega_2) = 0.49$*),*
ω_3 *je jev, že nastala výjimečná situace (hrana, ukradení mince :-) apod.)*
($P(\omega_3) = 0.02$*).*
Sestrojte dvě různé náhodné veličiny a nakreslete jejich distribuční funkce.

Řešení. Veličiny mohou být např.

- $X : X(\omega_1) = 1, X(\omega_2) = -1, X(\omega_3) = 0$. Distribuční funkce je pak skokovitá se skoky v bodech -1,0 a 1 o velikostech 0.49, 0.02 a 0.49, tj.

$$F_X(x) = \begin{cases} 0\,, & x < -1\,, \\ 0.49\,, & -1 \le x < 0\,, \\ 0.51\,, & 0 \le x < 1\,, \\ 1\,, & x \ge 1\,. \end{cases}$$

- $Y : Y(\omega_1) = Y(\omega_2) = 1, Y(\omega_3) = 7$. Distribuční funkce je pak skokovitá se skoky v bodech 1 a 7 o velikostech 0.98 a 0.02 (neboť obraz = 1 mají dva elementární jevy, jejichž souhrná pravděpodobnost je $0.49 + 0.49 = 0.98$), tj.

$$F_Y(x) = \begin{cases} 0\,, & x < 1\,, \\ 0.98\,, & 1 \le x < 7\,, \\ 1\,, & x \ge 7\,. \end{cases}$$

Cvičení 8.2. *Pravděpodobnost, že atlet v oddíle skočí do dálky přes* $5\,\mathrm{m}$*, je* 0.7*. V oddíle je* 6 *atletů.*

a) Jaké rozdělení má náhodná veličina X *popisující, zda atlet skočil (*$X = 1$*) nebo neskočil (*$X = 0$*) přes* $5\,\mathrm{m}$*? Určete i střední hodnotu* $\mathrm{E}X$ *a rozptyl* $\mathrm{D}X$*.*

b) Jaké rozdělení má náhodná veličina Y *popisující počet atletů v oddíle, kteří skočili přes* $5\,\mathrm{m}$*? Určete i střední hodnotu* $\mathrm{E}Y$ *a rozptyl* $\mathrm{D}Y$*.*

c) Jaká je pravděpodobnost, že přes $5\,\mathrm{m}$ *skočí v oddíle alespoň* 4 *atleti?*

Řešení. Jelikož se jedná o sérii nezávislých pokusů, v nichž nastává pouze úspěch nebo neúspěch, řešení je následující:

a) $X \sim \mathrm{Alt}(0.7)$, tj. $P(X = 1) = 0.7$ a $P(X = 0) = 0.3$.
$\mathrm{E}X = 0.7$, $\mathrm{D}X = 0.7 \cdot 0.3 = 0.21$.

b) $X \sim \mathrm{Bi}(6, 0.7)$, tj. $P(Y = k) = \binom{6}{k} 0.7^k 0.3^{6-k}$ pro $k = 0, 1, ..., 6$.
$\mathrm{E}Y = 6 \cdot 0.7 = 4.2$, $\mathrm{D}Y = 6 \cdot 0.7 \cdot 0.3 = 1.26$.

c) $P(Y \ge 4) = \sum_{k=4}^{6} \binom{6}{k} 0.7^k 0.3^{6-k} = \binom{6}{4} 0.7^4 0.3^2 + \binom{6}{5} 0.7^5 0.3^1 + \binom{6}{6} 0.7^6 0.3^0$.

Cvičení 8.3. *V testu je* 15 *otázek s možnými odpověďmi a)-e), právě jedna je správná. Jaká je pravděpodobnost, že při zcela náhodném tipování trefí student správně aspoň tři otázky? Popište náhodnou veličinu popisující počet správných odpovědí.*

Řešení. Označme X náhodnou veličinu popisující počet správných odpovědí. Ta má binomické rozdělení $\mathrm{Bi}(15; 0.2)$, tj.

$$P(X = k) = \binom{15}{k} 0,2^k (0,8)^{15-k}, \text{ pro } k = 0, 1, \ldots, 15.$$

$$\mathrm{E}X = 15 \cdot 0,2 = 3,$$

$$\mathrm{D}X = 15 \cdot 0,2 \cdot 0,8 = 2,4.$$

Pravděpodobnost, že při zcela náhodném tipování trefí student správně aspoň tři otázky, je tedy

$$P(X \ge 3) = \sum_{k=3}^{15} \binom{15}{k} 0,2^k (0,8)^{15-k}.$$

Cvičení 8.4. *Telefonní ústředna zapojí během hodiny průměrně* 15 *hovorů. Jaká je pravděpodobnost, že během* 4 *minut zapojí ústředna*

a) právě jeden hovor,

b) alespoň dva hovory a nejvýše pět hovorů?

Řešení. Označme X náhodnou veličinu popisující počet příchozích hovorů během 4 minut. Ta má Poissonovo rozdělení Po(1), neboť střední hodnota počtu příchozích hovorů během 4 minut odpovídající parametru λ je 1, tj.

$$P(X=k)=\frac{1^k}{k!}e^{-1},\quad k=0,1,\dots,$$

tudíž

$$P(X=1)=\frac{1^1}{1}e^{-1}=e^{-1}=0.3678,$$

a

$$P(2\le X\le 5)=\frac{1^2}{2!}e^{-1}+\frac{1^3}{3!}e^{-1}+\frac{1^4}{4!}e^{-1}+\frac{1^5}{5!}e^{-1}=0.2636.$$

Cvičení 8.5. *Na látce (pevné šířky* 1 m*) je průměrně jeden kaz na* 10 m *délky. Předpokládáme, že počet kazů se řídí Poissonovým rozdělením. Jaká je pravděpodobnost, že na* 50 m *délky látky bude*

a) přesně 10 *kazů,*

b) maximálně 3 *kazy,*

c) přesně 5 *kazů, z toho* 4 *na prvních* 20 m*?*

Řešení. Označme X náhodnou veličinu popisující počet kazů na 50 m délky látky. Pak

$$P(X=k)=\frac{5^k}{k!}e^{-5},k=0,1,...$$

neboť $\mathrm{E}X=\lambda=5$. Tudíž

a) $P(X=10)=\frac{5^{10}}{10!}e^{-5}$.

b) $P(X\le 3)=e^{-5}\left(\frac{5^0}{0!}+\frac{5^1}{1!}+\frac{5^2}{2!}+\frac{5^3}{3!}\right)$.

c) Označme X_1 náhodnou veličinu popisující počet kazů na prvních 20 m délky látky a X_2 náhodnou veličinu popisující počet kazů na zbylých 30 m délky látky. Analogicky k předešlému

$$P(X_1=k)=\frac{2^k}{k!}e^{-2}\quad \text{a}\quad P(X_2=k)=\frac{3^k}{k!}e^{-3}.$$

Hledaná pravděpodobnost je tedy

$$P(X_1=4,X_2=1)=P(X_1=4)P(X_2=1)=\frac{2^4}{4!}e^{-2}\frac{3^1}{1!}e^{-3}=2e^{-5}.$$

Cvičení 8.6. *Pravděpodobnost narození chlapce je* 0.51*. Jaká je pravděpodobnost, že v dané porodnici bylo nejpozději (v časovém pořadí) čtvrté narozené dítě holčička?*

Řešení. Označme X náhodnou veličinu popisující počet narozených chlapců před první narozenou holčičkou. Pak

$$X\sim \mathrm{Geom}(0.49),\ \text{tj. } P(X=k)=0.51^k\cdot 0.49,\ k=0,1,...$$

Hledaná pravděpodobnost je tedy

$$P(X<4)=\sum_{k=0}^{3}0.51^k\cdot 0.49=0.49\sum_{k=0}^{3}0.51^k=0.49\cdot\frac{1-0.51^4}{1-0.51}=1-0.51^4.$$

Jiný způsob výpočtu:
Označme Y náhodnou veličinu popisující počet narozených holčiček v prvních čtyřech narozených dětech. Pak

$$Y\sim \mathrm{Bi}(4,0.49),\ \text{tj. } P(Y=k)=\binom{4}{k}0.49^k\cdot 0.51^{4-k},\ k=0,1,2,3,4.$$

Hledaná pravděpodobnost je tedy

$$P(Y\ge 1)=1-P(Y=0)=1-\binom{4}{0}0.49^0\cdot 0.51^4=1-0.51^4.$$

Cvičení 8.7. *Pan Novák má svazek pěti klíčů, z nichž právě jeden je od jeho bytu. Po příchodu z hospody ke dveřím bytu je ve stavu, že správný klíč nehledá, ale zkouší strčit do zámku klíč náhodně vybraný, přičemž vždy, když netrefí ten pravý, mu svazek klíčů spadne na zem. Po desátém neúspěšném pokusu se pak vzbudí jeho manželka a…(důsledky raději nedomýšlet). Jaká je pravděpodobnost, že se manželka pana Nováka nevzbudí?*

Řešení. Označme X náhodnou veličinu popisující počet neúspěšných pokusů. Pak $X \sim \mathrm{Geom}(0.2)$, tj.

$$\begin{aligned} P(\text{spící manželka}) &= P(X \leq 9) = \sum_{k=0}^{9} 0.8^k 0.2 = 0.2\frac{1-0.8^{10}}{1-0.8} \\ &= 1 - 0.8^{10} = 0.893. \end{aligned}$$

Nebo jinak:
Označme Y náhodnou veličinu popisující počet úspěšných pokusů v prvních 10 pokusech. Pak $Y \sim \mathrm{Bi}(10, 0.2)$, tj.

$$\begin{aligned} P(\text{spící manželka}) &= P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{10}{0} 0.2^0 0.8^{10} \\ &= 1 - 0.8^{10} = 0.893. \end{aligned}$$

Cvičení 8.8. *Semena mají klíčivost $p \in (0, 1)$. Jaký je optimální počet n semen v jamce, aby byla co nejvyšší pravděpodobnost, že vyklíčí právě jedno? Řešte obecně a pro $p = 1/3$.*

Řešení. Jedná se o binomické rozdělení $\mathrm{Bi}(n, p)$, které pro $n \in \mathbb{N}$ $(n \geq 1)$ nabývá hodnotu 1 s pravděpodobností

$$g(p) = \tbinom{n}{1}\, p^1\, (1-p)^{n-1} = n\, p\, (1-p)^{n-1}\,.$$

Funkce g je definována i pro všechna kladná reálná n a je *unimodální* (do maxima rostoucí, pak klesající), takže maximum nastává v jediném bodě s nulovou derivací

$$g'(n) = p\,(1-p)^{n-1} + n\,p\,(1-p)^{n-1}\,\ln(1-p) = p\,(1-p)^{n-1}\big(1 + n\,\ln(1-p)\big)\,.$$

Nulová může být pouze poslední závorka, a to pro

$$n = \frac{-1}{\ln(1-p)}\,.$$

Maximum v oboru přirozených čísel nastává pro jedno ze dvou celých čísel, která jsou nejblíže této hodnotě. Pro $p = 1/3$ je g' nulová v

$$n = \frac{-1}{\ln\frac{2}{3}} \doteq 2.466$$

a hodnoty pro blízká přirozená čísla jsou v tabulce (stačí 2 a 3, ale uvádíme víc):

n	1	2	3	4
$n\,p\,(1-p)^{n-1}$	$\frac{1}{3}$	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{32}{81}$

V tomto případě nastává maximum pro $n \in \{2, 3\}$.

Pro $p \to 0$ je $\ln(1-p) \approx -p$ a $n \approx 1/p$ v souladu s očekáváním. Pro $p \to 1$ vychází $n \to 0$, což vypadá překvapivě. Znamená to ale jen, že funkce g je na množině přirozených čísel klesající a maxima nabývá v 1.

8.2 Spojité náhodné veličiny

Cvičení 8.9. *Pošta chodí pravidelně mezi* 10:00 *až* 12:00 *rovnoměrně. Jaká je pravděpodobnost, že obdržíme poštu mezi* 11:30 *až* 12:30*?*

Řešení. Označme X náhodnou veličinu popisující čas příchodu pošty. Ta má rovnoměrné rozdělení $\mathrm{R}(10, 12)$. Hustota je tedy

$$f(x) = \begin{cases} \frac{1}{2}\,, & x \in (10, 12)\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

a hledaná pravděpodobnost je

$$\int_{11{,}5}^{12{,}5} f(x)\mathrm{d}x = \int_{11{,}5}^{12} \frac{1}{2}\mathrm{d}x = \frac{1}{4}.$$

Cvičení 8.10. *Při numerickém výpočtu se reálná čísla zaokrouhlují na jedno desetinné místo. Jaká je pravděpodobnost, že vzdálenost skutečného čísla od zaokrouhleného bude větší než* 0.04*?*

Řešení. Označme X náhodnou veličinu popisující rozdíl „skutečná - zaokrouhlená hodnota“. Pak

$$X \sim \mathrm{R}(-0.05, 0.05),$$

tj. X má hustotu $f(x) = 10$ pro $x \in (-0.05, 0.05)$ a $f(x) = 0$ jinde.
Hledaná pravděpodobnost je tedy

$$\begin{aligned} P(|X| > 0.04) &= 1 - P(|X| \leq 0.04) = 1 - P(-0.04 \leq X \leq 0.04) \\ &= 1 - \int_{-0.04}^{0.04} 10 \mathrm{d}x = 0.2. \end{aligned}$$

Cvičení 8.11. *Do pojišťovny přijdou průměrně* 2 *hlášení škody denně. Jaká je pravděpodobnost, že do pojišťovny přijde nejbližší hlášení škody nejdříve třetí den?*

Řešení. Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení.

a) Náhodná veličina X „doba čekání na příchod dalšího hlášení“ má exponenciální rozdělení. Parametr $\lambda = 2$, protože $\mathrm{E}X = 1/\lambda = 1/2$ (ze zadání víme, že v průměru přijdou 2 hlášení škody denně, tj. střední doba čekání je půl dne). Hledaná pravděpodobnost je pak

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - (1 - e^{-2 \cdot 2}) = e^{-4}$$

nebo také počítáno jako

$$P(X > 2) = \int_2^\infty f(x) dx = \int_2^\infty 2e^{-2x} dx = e^{-4}.$$

b) Náhodná veličina Y „počet hlášení během dvou dní“ má Poissonovo rozdělení. Parametr $\lambda = 4$, protože $\mathrm{E}Y = \lambda$ a ze zadání víme, že během dvou dní přijdou v průměru 4 hlášení. Hledaná pravděpodobnost je pak

$$P(Y = 0) = e^{-4} \cdot \frac{4^0}{0!} = e^{-4}.$$

Cvičení 8.12. *Na zákaznickou linku přichází průměrně* 12 *hovorů za hodinu. Doba čekání na hovor má exponenciální rozdělení.*

a) Jaká je pravděpodobnost, že nejbližší hovor přijde nejdříve za 10 *minut?*

b) Určete čas t takový, že nejbližší hovor přijde nejdříve za t minut s pravděpodobností 0.7.

Řešení. K úloze je opět možno přistupovat více způsoby.

a) Řešíme analogicky jako cvičení 8.11.

1. Označme X náhodnou veličinu popisující dobu čekání (v minutách) na nejbližší hovor. Pak

$$X \sim \mathrm{Ex}(1/5),$$

tj. X má hustotu

$$f(x) = \begin{cases} \frac{1}{5} e^{-x/5}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

a distribuční funkci

$$F(x) = \begin{cases} 1 - e^{-x/5}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Hodnotu $1/5$ jsme získali z faktu, že pro $X \sim \mathrm{Ex}(\lambda)$ je $\mathrm{E}X = 1/\lambda$ a přitom střední doba čekání na hovor je 5 minut.
Hledaná pravděpodobnost je tedy

$$P(X > 10) = 1 - P(X \leq 10) = 1 - F(10) = 1 - (1 - e^{-10/5}) = e^{-2},$$

kterou lze spočítat také pomocí integrálu z hustoty jako

$$P(X > 10) = \int_{10}^\infty \frac{1}{5} e^{-x/5} dx = e^{-2},$$

2. Označme Y náhodnou veličinu popisující počet hovorů během 10 minut. Pak

$$Y \sim \mathrm{Po}(2), \text{ tj. } P(Y = k) = \frac{2^k}{k!}e^{-2}, k = 0, 1, ...$$

Hodnotu 2 jsme získali z faktu, že pro $Y \sim \mathrm{Po}(\lambda)$ je $\mathrm{E}X = \lambda$ a přitom střední počet hovorů za 10 minut jsou 2.
Hledaná pravděpodobnost je tedy

$$P(Y = 0) = \frac{2^0}{0!}e^{-2} = e^{-2}.$$

b) Uvažujme opět X náhodnou veličinu popisující dobu čekání (v minutách) na nejbližší hovor. Pak

$$\begin{aligned} P(X > t) &= 0.7 \\ 1 - P(X \le t) &= 0.7 \\ 1 - F(t) &= 0.7 \\ 1 - (1 - e^{-t/5}) &= 0.7 \\ e^{-t/5} &= 0.7 \\ t &= -5 \ln 0.7 \\ t &= 1.78. \end{aligned}$$

I zde se nabízí řešení pomocí Poissonova rozdělení, a to při označení Y náhodné veličiny popisující počet hovorů během t minut. Pak $Y \sim \mathrm{Po}(t/5)$ a řešíme rovnici

$$P(Y = 0) = \frac{(t/5)^0}{0!}e^{-t/5} = e^{-t/5} = 0.7.$$

Cvičení 8.13. *Spojitě rozdělená náhodná veličina X má hustotu*

$$f_X(u) = e^{-c|u|}, \ u \in \mathbb{R}.$$

Určete konstantu c a pravděpodobnosti $P(X \le 0)$, $P(-1 \le X \le 1)$, $P(X > \ln 3)$ a $P(X = 1)$.

Řešení. Neboť f_X je sudá, máme

$$1 = \int_{-\infty}^{\infty} e^{-c|u|} \, \mathrm{d}u = 2\int_0^{\infty} e^{-cu} \, \mathrm{d}u = 2\left[\frac{e^{-cu}}{-c}\right]_0^{\infty} = \frac{2}{c},$$

za (rozumného) předpokladu, že $c > 0$. Porovnáním pravé a levé strany vyjde $c = 2$. Dále máme

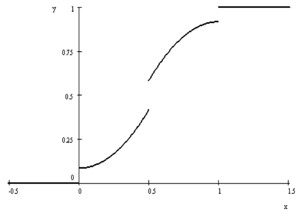
$$\begin{aligned} &P(X \le 0) = \frac{1}{2}, \text{ (hustota je sudá)}, \\ &P(-1 \le X \le 1) = \int_{-1}^{1} e^{-2|u|} \, \mathrm{d}u = 2\int_0^1 e^{-2u} \, \mathrm{d}u = \left[\frac{e^{-2u}}{-2}\right]_0^1 = 1 - e^{-2}, \\ &P(X > \ln 3) = \int_{\ln 3}^{\infty} e^{-2u} \, \mathrm{d}u = \left[\frac{e^{-2u}}{-2}\right]_{\ln 3}^{\infty} = \frac{1}{18}, \\ &P(X = 1) = 0, \text{ (X je spojitě rozdělená)}. \end{aligned}$$

8.3 Náhodné veličiny se smíšeným rozdělením

Cvičení 8.14. *Náhodná veličina X má distribuční funkci*

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{12} + \frac{4t^2}{3}\,, & 0 \le t < 1/2\,, \\ \frac{11}{12} - \frac{4(1-t)^2}{3}\,, & 1/2 \le t < 1\,, \\ 1\,, & t \ge 1\,. \end{cases}$$

Vyjádřete ji jako směs náhodných veličin U, V, z nichž U je diskrétní a V spojitá; popište a znázorněte jejich rozdělení.



Obrázek 1:

Řešení. Nespojitosti distribuční funkce jsou v bodech $0, 1/2, 1$,

$$F_X(0) - F_X(0-) = F_X(1) - F_X(1-) = \frac{1}{12},$$
$$F_X(1/2) - F_X(1/2-) = \frac{1}{6},$$

$$c\,F_U(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{12}, & 0 \le t < \frac{1}{2}, \\ \frac{1}{4}, & \frac{1}{2} \le t < 1, \\ \frac{1}{3}, & t \ge 1, \end{cases}$$

$$c = \lim_{t\to\infty} F_U(t) = \frac{1}{3},$$

$$F_U(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{4}, & 0 \le t < \frac{1}{2}, \\ \frac{3}{4}, & \frac{1}{2} \le t < 1, \\ 1, & t \ge 1, \end{cases}$$

$$p_U(t) = \begin{cases} \frac{1}{4}, & t \in \{0,1\}, \\ \frac{1}{2}, & t = \frac{1}{2}, \\ 0, & \text{jinde}. \end{cases}$$

$$(1-c)F_V(t) = F_X(t) - cF_U(t) = \begin{cases} 0, & t < 0, \\ \frac{4\,t^2}{3}, & 0 \le t < 1/2, \\ \frac{2}{3} - \frac{4\,(1-t)^2}{3}, & 1/2 \le t < 1, \\ \frac{2}{3}, & t \ge 1, \end{cases}$$

$$F_V(t) = \begin{cases} 0, & t < 0, \\ 2t^2, & 0 \le t < 1/2, \\ 1 - 2\,(1-t)^2, & 1/2 \le t < 1, \\ 1, & t \ge 1, \end{cases}$$

$$f_V(t) = \begin{cases} 4\,t, & 0 \le t < 1/2, \\ 4\,(1-t), & 1/2 \le t < 1, \\ 0, & \text{jinak}. \end{cases}$$

Cvičení 8.15. *Náhodná veličina X má alternativní rozdělení (s hodnotami $0, 1$), $P[X = 1] = 2/3$. Náhodná veličina Y má spojité rovnoměrné rozdělení na intervalu $\langle 0, 2\rangle$. Popište rozdělení jejich směsi $Z = \mathrm{Mix}_{2/3}(X, Y)$.*

Řešení.

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{3}\,, & 0 \le t < 1\,, \\ 1\,, & t \ge 1\,, \end{cases}$$

$$F_Y(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{t}{2}\,, & 0 \le t < 2\,, \\ 1\,, & t \ge 2\,, \end{cases}$$

$$F_Z(t) = \frac{2}{3}\,F_X(t) + \frac{1}{3}\,F_Y(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{2}{9} + \frac{t}{6}\,, & 0 \le t < 1\,, \\ \frac{2}{3} + \frac{t}{6}\,, & 1 \le t < 2\,, \\ 1\,, & \text{pro } t \ge 2\,. \end{cases}$$

Cvičení 8.16. *Náhodná veličina X má distribuční funkci*

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{5}{6} - \frac{2}{3}\,\exp(-2\,t)\,, & 0 \le t < 2\,, \\ 1 - \frac{2}{3}\,\exp(-2\,t)\,, & t \ge 2\,. \end{cases}$$

Vyjádřete její rozdělení jako směs diskrétního a spojitého rozdělení.

Řešení. $X = \mathrm{Mix}_c(U, V)$, U diskrétní, V spojitá.
Nespojitosti distribuční funkce jsou v bodech 0, 2, obě stejné velikosti

$$F_X(0) - F_X(0-) = F_X(2) - F_X(2-) = \frac{1}{6}\,,$$

$$c\,F_U(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{6}\,, & 0 \le t < 2\,, \\ \frac{1}{3}\,, & t \ge 2\,, \end{cases}$$

$$c = \lim_{t\to\infty} F_U(t) = \frac{1}{3}\,,$$

$$F_U(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1}{2}\,, & 0 \le t < 2\,, \\ 1\,, & t \ge 2\,, \end{cases}$$

$$p_U(t) = \begin{cases} \frac{1}{2}\,, & t \in \{0, 2\}\,, \\ 0\,, & \text{jinde}\,. \end{cases}$$

$$(1-c)\,F_V(t) = F_X(t) - c\,F_U(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{2}{3} - \frac{2}{3}\,\exp(-2\,t)\,, & 0 \le t < 2\,, \\ \frac{2}{3} - \frac{2}{3}\,\exp(-2\,t)\,, & t \ge 2\,, \end{cases}$$

$$F_V(t) = \begin{cases} 0\,, & t < 0\,, \\ 1 - \exp(-2\,t)\,, & t \ge 0\,, \end{cases}$$

$$f_V(t) = \begin{cases} 0\,, & t < 0\,, \\ 2\,\exp(-2\,t)\,, & t \ge 0\,. \end{cases}$$

9 Operace s náhodnými veličinami

Cvičení 9.1. *Výška dětí v* 1. *třídě je náhodná veličina $X \sim \mathrm{N}(130, 36)$. Jaká je pravděpodobnost, že náhodně vybrané dítě bude* *a) větší než* 136 cm,

b) menší než 118 cm,

c) mít výšku mezi 127 *a* 133 cm*?*

Řešení. Označme Z náhodnou veličinu s normovaným normálním rozdělením, tj. $Z \sim \mathrm{N}(0,1)$, a Φ její distribuční funkci, přičemž hodnoty $\Phi(x)$ pro různá kladná X lze vyčíst ze statistických tabulek. Pak

a)

$$\begin{aligned} P(X>136) &= P\left(\frac{X-130}{\sqrt{36}} > \frac{136-130}{\sqrt{36}}\right) = P(Z>1) = 1-P(Z\le 1) \\ &= 1-\Phi(1) = 1-0.8413 = 0.1587. \end{aligned}$$

b)

$$\begin{aligned} P(X<118) &= P\left(\frac{X-130}{\sqrt{36}} < \frac{118-130}{\sqrt{36}}\right) = P(Z<-2) = \Phi(-2) \\ &= 1-\Phi(2) = 1-0.9772 = 0.0228. \end{aligned}$$

c)

$$\begin{aligned} P(127<X<133) &= P\left(\frac{127-130}{\sqrt{36}} < \frac{X-130}{\sqrt{36}} < \frac{133-130}{\sqrt{36}}\right) \\ = P(-0.5<Z<0.5) &= P(-0.5<Z<0.5) = P(Z<0.5) - P(Z\le -0.5) \\ &= \Phi(0.5) - \Phi(-0.5) = \Phi(0.5) - (1-\Phi(0.5)) \\ &= 2\Phi(0.5) - 1 = 2\cdot 0.6915 - 1 = 0.383. \end{aligned}$$

Cvičení 9.2. *Systematická chyba měření je* 5 *a směrodatná odchylka je* 36*. Jaká je pravděpodobnost toho, že chyba měření nepřekročí v absolutní hodnotě* 5*? (Předpokládáme normální rozdělení chyby.)*

Řešení.

$$P[-5\le X\le 5] = P\left[\frac{-5-5}{36} \le \frac{X-5}{36} \le \frac{0}{36}\right] \doteq 0.5 - (1-\Phi(0.28)) \doteq 0.5 - 0.39 = 0.11.$$

Cvičení 9.3. *Délka zubního kartáčku se řídí normálním rozdělením o střední hodnotě* 14 cm *a rozptylu* 6 cm^2*. Pouzdro je dlouhé* 13.5 cm*. Jaká je pravděpodobnost, že vezmeme-li libovolný jeden kartáček, že se do pouzdra akorát vejde, tj. nebude o více než* 1 cm *kratší?*

Řešení. Označíme-li délku zubního kartáčku X, potom platí $X \sim \mathrm{N}(14,6)$. Chceme znát $P(12.5 < X < 13.5)$. Náhodná veličina

$$U = \frac{X-14}{\sqrt{6}}$$

bude mít normované normální rozdělení, proto

$$\begin{aligned} P(12.5<X<13.5) &= P\left(\frac{12.5-14}{\sqrt{6}} < \frac{X-14}{\sqrt{6}} < \frac{13.5-14}{\sqrt{6}}\right) \\ &= P\left(\frac{-1.5}{\sqrt{6}} < U < \frac{-0.5}{\sqrt{6}}\right) = \Phi\left(\frac{-0.5}{\sqrt{6}}\right) - \Phi\left(\frac{-1.5}{\sqrt{6}}\right) \\ &= \left(1-\Phi\left(\frac{0.5}{\sqrt{6}}\right)\right) - \left(1-\Phi\left(\frac{1.5}{\sqrt{6}}\right)\right) = \\ &= \Phi\left(\frac{1.5}{\sqrt{6}}\right) - \Phi\left(\frac{0.5}{\sqrt{6}}\right) = 0.729 - 0.579 = 0.15. \end{aligned}$$

Pravděpodobnost, že se kartáček do pouzdra akorát vejde, činí 15 %.

Cvičení 9.4. *Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě* 67 *a* 75 m *a směrodatné odchylky* 6 *a* 3 m*. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál.*

Řešení. Náhodná veličina A má rozdělení $\mathrm{N}(67, 36)$, B má $\mathrm{N}(75, 9)$, $A - B$ má $\mathrm{N}(67 - 75, 36 + 9) = \mathrm{N}(-8, 45)$, kladných hodnot nabývá s pravděpodobností

$$1 - F_{\mathrm{N}(-8,45)}(0) = 1 - \Phi\left(\frac{0 - (-8)}{\sqrt{45}}\right) \doteq 1 - \Phi(1.192\,6) \doteq 1 - 0.883 = 0.117\,.$$

Cvičení 9.5. *Délka hrany krychle je náhodná veličina* $X \sim \mathrm{R}(1, 2)$. *Určete distribuční funkci náhodné veličiny* Y *popisující plochu povrchu této krychle.*

Řešení. Pro distribuční funkci náhodné veličiny X platí

$$F_X(x) = \begin{cases} 0\,, & x < 1\,,\\ x - 1\,, & 1 \le x \le 2\,,\\ 1\,, & x > 2\,. \end{cases}$$

Pro distribuční funkci náhodné veličiny Y tak máme

$$F_Y(y) = P(Y \le y) = P(6X^2 \le y) = \begin{cases} 0\,, & y < 0\,,\\ P\left(-\sqrt{\frac{y}{6}} \le X \le \sqrt{\frac{y}{6}}\right) = F_X\left(\sqrt{\frac{y}{6}}\right)\,, & y \ge 0\,, \end{cases}$$

tj.

$$F_Y(y) = \begin{cases} 0\,, & y < 6\,,\\ \sqrt{\frac{y}{6}} - 1\,, & 6 \le y \le 24\,,\\ 1\,, & x > 24\,. \end{cases}$$

Cvičení 9.6. *Průměrný počet zákazníků během dne v první prodejně je* 20, *ve druhé prodejně* 25 *(předpokládáme, že oba počty se řídí Poissonovým rozdělením). Odvoďte rozdělení počtu zákazníků v obou prodejnách dohromady.*

Řešení. Označme
X počet zákazníků během dne v první prodejně,
Y počet zákazníků během dne ve druhé prodejně,
Z počet zákazníků během dne v obou prodejnách dohromady.
Pak

$$P(X = i) = \frac{20^i}{i!}e^{-20} \quad \text{a} \quad P(Y = j) = \frac{25^j}{j!}e^{-25}$$

Jelikož $Z = X + Y$, dostáváme

$$\begin{aligned} P(Z = k) &= \sum_{i=0}^{k} P(X = i)P(Y = k - i) = \sum_{i=0}^{k} \frac{20^i}{i!}e^{-20}\frac{25^{k-i}}{(k-i)!}e^{-25}\\ &= e^{-45}\sum_{i=0}^{k}\frac{20^i}{i!}\frac{25^{k-i}}{(k-i)!} = e^{-45}\sum_{i=0}^{k}\frac{20^i}{i!}\frac{25^{k-i}}{(k-i)!}\frac{k!}{k!}\\ &= e^{-45}\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}20^i 25^{k-i} = e^{-45}\frac{45^k}{k!}, \end{aligned}$$

tj. $Z \sim \mathrm{Po}(45)$.

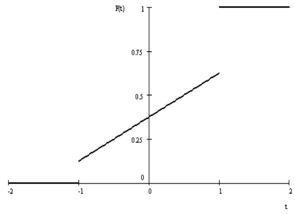
Cvičení 9.7. *Náhodná veličina má spojité rovnoměrné rozdělení na intervalu* $[-3, 5]$. *Zobrazte ji funkcí*

$$h(x) = \begin{cases} -1\,, & x < -2\,,\\ x/2\,, & x \in [-2, 2]\,,\\ 1\,, & x > 2\,, \end{cases}$$

výsledné rozdělení popište a znázorněte.

Řešení. Výstup -1 odpovídá vstupu v intervalu $\langle -3, -2\rangle$, a má tedy pravděpodobnost $1/8$, $P[h(X) = -1] = 1/8$. Výstup 1 odpovídá vstupu v intervalu $\langle 2, 5\rangle$, a má tedy pravděpodobnost $3/8$, $P[h(X) = 1] = 3/8$. Zbývající hodnoty vedou na spojité rovnoměrné rozdělení na $\langle -1, 1\rangle$ (jako složku směsi, která tvoří rozdělení výstupu a má váhu $1/2$), distribuční funkce je

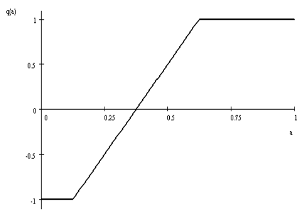
$$F_{h(X)}(t) = \begin{cases} 0\,, & t < -1\,,\\ 3/8 + t/4\,, & t \in [-1, 1)\,,\\ 1\,, & t \ge 1\,. \end{cases}$$



Obrázek 2:

Snazší je řešení přes kvantilovou funkci; původní kvantilová funkce $q_X(a) = 8\,a - 3$ složená s funkcí h dá kvantilovou funkci

$$q_{h(X)}(a) = h(q_X(a)) = \begin{cases} -1\,, & a \leq 1/8\,, \\ 4\,a - 3/2\,, & a \in (1/8, 5/8)\,, \\ 1\,, & a \geq 5/8\,. \end{cases}$$



Obrázek 3:

Cvičení 9.8. *Nezáporná náhodná veličina X má hustotu*

$$f_X(t) = \begin{cases} 0\,, & t < 1\,, \\ c \cdot t^{-3}\,, & t \geq 1\,, \end{cases}$$

kde c je vhodná konstanta. Určete hodnotu c a napište vzorec pro distribuční funkci F_X a distribuční funkci F_Y veličiny $Y = \frac{1}{X+1}$.

Řešení. Pro určení konstanty c potřebujeme, aby

$$1 = \int_{-\infty}^{\infty} f_X(t)\ dt = c \int_1^{\infty} t^{-3}\ dt = c\Big[\frac{t^{-2}}{-2}\Big]_1^{\infty} = \frac{c}{2}$$

tedy $c = 2$. Distribuční funkce je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt$$

a pro $u \geq 1$ tedy máme

$$F_X(u) = \int_1^u 2t^{-3}\ dt = \Big[-t^{-2}\Big]_1^u = 1 - \frac{1}{u^2},$$

celkově pak

$$F_X(u) = \begin{cases} 0\,, & u \leq 1\,, \\ 1 - \frac{1}{u^2}\,, & u \geq 1\,, \end{cases}$$

s grafem

Protože X je nezáporná spojitá veličina, pro distribuční funkci F_Y pro $u > 0$ máme

$$F_Y(u) = P(Y \leq u) = P\Big(\frac{1}{X+1} \leq u\Big) = P\Big(\frac{1}{u} \leq X + 1\Big) = P\Big(\frac{1}{u} - 1 \leq X\Big) =$$

$$= 1 - P\Big(X < \frac{1}{u} - 1\Big) = 1 - F_X\Big(\frac{1-u}{u}\Big)\ .$$

Pro $\frac{1-u}{u} \geq 1$ a $u > 0$ tak dostáváme

$$F_Y(u) = 1 - F_X\Big(\frac{1-u}{u}\Big) = 1 - \Big(1 - \frac{1}{(\frac{1-u}{u})^2}\Big) = \frac{u^2}{(1-u)^2}\ .$$

Pro $\frac{1-u}{u} \le 1$ a $u > 0$ pak máme

$$F_Y(u) = 1 - F_X\Big(\frac{1-u}{u}\Big) = 1\ .$$

A nakonec, X je nezáporná veličina, takže $Y = \frac{1}{X+1} > 0$ a tudíž pro $u \le 0$ je

$$F_Y(u) = P(Y \le 0) = 0\ .$$

Celkově pak dostáváme

$$F_Y(u) = \begin{cases} 0\,, & u \le 0\,, \\ \frac{u^2}{(1-u)^2}\,, & u \in \big(0, \frac{1}{2}\big)\,, \\ 1\,, & u \ge \frac{1}{2}\,, \end{cases}$$

s grafem

Distribuční funkce F_Y je opět (absolutně) spojitá.

Cvičení 9.9. *Náhodná veličina X má exponenciální rozdělení s hustotou*

$$f_X(u) = \begin{cases} 2^{-u}\ln 2\,, & u > 0\,, \\ 0\,, & \textit{jinak}. \end{cases}$$

Určete a znázorněte rozdělení náhodných veličin

a) $2 + X$,

b) $2 - X$,

c) $2\,X$.

Řešení.

$$F_X(x) = \begin{cases} 0\,, & x < 0\,, \\ 1 - 2^{-x}\,, & x \ge 0\,. \end{cases}$$

a)

$$F_{2+X}(x) = F_X(x-2) = \begin{cases} 0\,, & x < 2\,, \\ 1 - 2^{2-x}\,, & x \ge 2\,, \end{cases}$$

$$f_{2+X}(x) = \begin{cases} 0\,, & x < 2\,, \\ 2^{2-x}\ln 2\,, & x \ge 2\,. \end{cases}$$

b)

$$F_{2-X}(x) = 1 - F_X(2-x) = \begin{cases} 2^{x-2}\,, & x < 2\,, \\ 1\,, & x \ge 2\,, \end{cases}$$

$$f_{2-X}(x) = \begin{cases} 2^{x-2}\ln 2\,, & x < 2\,, \\ 0\,, & x \ge 2\,. \end{cases}$$

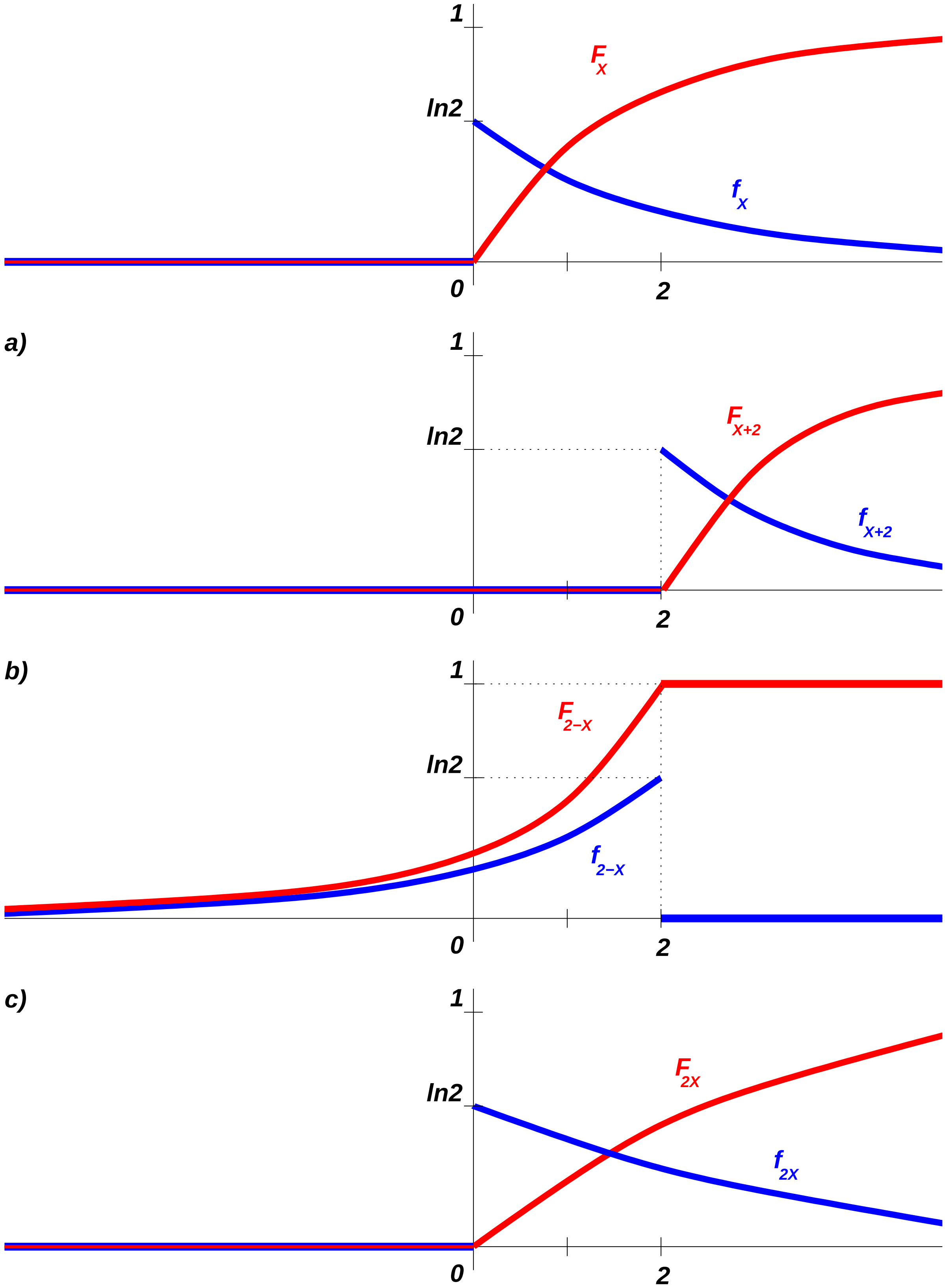
c)

$$F_{2\,X}(x) = F_X(x/2) = \begin{cases} 0\,, & x < 0\,, \\ 1 - 2^{-x/2}\,, & x \ge 0\,, \end{cases}$$

$$f_{2+X}(x) = \begin{cases} 0\,, & x < 0\,, \\ 2^{-x/2-1}\ln 2\,, & x \ge 0\,. \end{cases}$$

Grafy funkcí jsou znázorněné na Obr. 4.

Cvičení 9.10. *Nechť F_X je distribuční funkce náhodné veličiny X, která má nenulovou hustotu $f_X(t) = t/8$ na intervalu $\langle 0, 4\rangle$. Sestrojte distribuční funkci F_X a funkce $1 - F_X$, F_X^2 a $2\,F_X$. Rozhodněte, které z těchto funkcí jsou distribuční funkce spojitých náhodných veličin, a v těchto případech určete jejich hustoty.*



Obrázek 4: Grafy distribučních funkcí a hustot pravděpodobnosti.

Řešení.

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{t^2}{16}\,, & 0 \le t < 4\,, \\ 1\,, & 4 \le t\,. \end{cases}$$

Funkce $1 - F_X$ a $2F_X$ nejsou distribuční funkce. Funkce F_X^2 je distribuční funkce a má hustotu h danou předpisem

$$h(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{t^3}{64}\,, & 0 \le t < 4\,, \\ 0\,, & 4 \le t\,. \end{cases}$$

Cvičení 9.11. *Náhodná veličina X má rovnoměrné rozdělení na intervalu $(-2, 2)$. Zobrazíme ji funkcí h, definovanou následovně:*

$$h(x) = \begin{cases} 0\,, & x < 0\,, \\ x\,, & 0 \le x \le 1\,, \\ 1\,, & x > 1\,. \end{cases}$$

Nalezněte rozdělení náhodné veličiny $h(X)$.

Výsledky. Označme $Y = h(X)$. Potom je

$$F_Y(y) = \begin{cases} 0\,, & y < 0\,, \\ \frac{2+y}{4}\,, & 0 \le y < 1\,, \\ 1\,, & y \ge 1\,. \end{cases}$$

10 Základní charakteristiky náhodných veličin

Cvičení 10.1. *Náhodná veličina X představuje hodnoty, které padají na pravidelné hrací kostce. Pro veličinu $Y = 2X - 1$ určete její střední hodnotu, rozptyl a distribuční funkci.*

Řešení. Veličina X má hodnoty $i = 1, 2, 3, 4, 5, 6$ a pravděpodobnostní funkcí

$$p_X(i) = \begin{cases} \frac{1}{6}\,, & i = 1, 2, 3, 4, 5, 6\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

Protože $E(Y) = E(2X - 1) = 2E(X) - 1$ a $D(Y) = D(2X - 1) = D(2X) = 2^2 D(X)$, stačí zjistit střední hodnotu a rozptyl pro veličinu X:

$$E(X) = \sum_{i\in\mathbb{R}} i \cdot p_X(i) = \frac{1}{6}\big(1 + 2 + 3 + 4 + 5 + 6\big) = \frac{1}{2}(1 + 6) = \frac{7}{2} = 3.5$$

$$E(X^2) = \sum_{i\in\mathbb{R}} i^2 \cdot p_X(i) = \frac{1}{6}\big(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2\big) = \frac{91}{6} \doteq 15.166$$

$$D(X) = E(X^2) - \big(E(X)\big)^2 = \frac{91}{6} - \Big(\frac{7}{2}\Big)^2 = \frac{35}{12} \doteq 2.9167\,.$$

Takže máme

$$E(Y) = 6 \quad \text{a} \quad D(Y) = \frac{35}{3} \doteq 11,67\,.$$

Veličina Y nabývá hodnot $1, 3, 5, 7, 9, 11$ se stejnou pravděpodobností, takže distribuční funkce bude

$$F_Y(t) = \begin{cases} 0\,, & t < 1\,, \\ \frac{1}{6}i\,, & t \in [2i - 1, 2i + 1)\,, i = 1, 2, 3, 4, 5 \\ 1\,, & t \ge 11\,. \end{cases}$$

Cvičení 10.2. *Náhodná veličina W má hustotu*

$$f_W(u) = \begin{cases} \frac{c-1}{3}\,, & u \in [-1, 5)\,, \\ 0\,, & \textit{jinak}\,. \end{cases}$$

Určete konstantu $c \in \mathbb{R}$ a nalezněte střední hodnotu $\mathrm{E}W$.

Řešení. Musí platit

$$1 = \int_{-\infty}^{\infty} f_W(u)\ \mathrm{d}u = \int_{-1}^{5} \frac{c-1}{3}\ \mathrm{d}u = 6 \cdot \frac{c-1}{3},$$

tedy $c = \frac{3}{2}$. Dále

$$\mathrm{E}W = \int_{-\infty}^{\infty} u f_W(u)\ \mathrm{d}u = \int_{-1}^{5} \frac{u}{6}\ \mathrm{d}u = 2.$$

Cvičení 10.3. *Náhodná veličina X nabývá pouze hodnot z tabulky:*

X	-1	0	1	3
p_X	c	$3c$	0.4	0.2

Určete konstantu $c \in [0,1]$. Určete distribuční funkci F_X a střední hodnotu $\mathrm{E}X$.

Řešení. Musí platit

$$1 = \sum_{X=a} p_X(a) = c + 3c + 0.4 + 0.2,$$

z čehož vyřešením rovnice plyne $c = 0.1$. Distribuční funkce je tedy rovna

$$F_X(u) = \begin{cases} 0\,, & u < -1\,, \\ 0.1\,, & u \in [-1,0)\,, \\ 0.4\,, & u \in [0,1)\,, \\ 0.8\,, & u \in [1,3)\,, \\ 1\,, & u \geq 3\,. \end{cases}$$

Střední hodnotu spočteme jednoduše:

$$\mathrm{E}X = \sum_{X=a} a \cdot p_X(a) = -1 \cdot 0.1 + 0 \cdot 0.3 + 1 \cdot 0.4 + 3 \cdot 0.2 = 0.9.$$

Cvičení 10.4. *Náhodná veličina X má rozdělení dané hustotou*

$$f_X(t) = \begin{cases} c \cdot e^t\,, & t < 0\,, \\ 0\,, & jinak\,. \end{cases}$$

Určete konstantu c, distribuční funkci, střední hodnotu a rozptyl náhodné veličiny X.

Řešení. Pro určení konstanty c potřebujeme, aby

$$1 = \int_{-\infty}^{\infty} f_X(t)\ dt = c \int_{-\infty}^{0} e^t\ dt = c\Big[e^t\Big]_{-\infty}^{0} = c$$

tedy $c = 1$. Distribuční funkce je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt$$

a pro $u < 0$ tedy máme

$$F_X(u) = \int_{-\infty}^{u} e^t\ dt = \Big[e^t\Big]_{-\infty}^{u} = e^u$$

celkově pak

$$F_X(u) = \begin{cases} e^u\,, & u < 0\,, \\ 1\,, & u \geq 0\,. \end{cases}$$

Pomocí hustoty spočítáme střední hodnotu

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\ dt = \int_{-\infty}^{0} t \cdot e^t\ dt = \Big[(t-1)e^t\Big]_{-\infty}^{0} = -1$$

a rozptyl

$$D(X) = E(X^2) - \big(E(X)\big)^2\ .$$

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)\ dt = \int_{-\infty}^{0} t^2 \cdot e^t\ dt = \Big[(t^2 - 2t + 2)e^t\Big]_{-\infty}^{0} = 2.$$

Tedy

$$D(X) = E(X^2) - \big(E(X)\big)^2 = 2 - (-1)^2 = 1\ .$$

Cvičení 10.5. *Náhodná veličina X má rozdělení dané hustotou*

$$f_X(t) = \begin{cases} c(t-1)^2\,, & t \in (0,1)\,, \\ 0\,, & \textit{jinak}\,. \end{cases}$$

Určete konstantu c, distribuční funkci, střední hodnotu a rozptyl náhodné veličiny X.

Řešení. Pro určení konstanty c potřebujeme, aby

$$1 = \int_{-\infty}^{\infty} f_X(t)\ dt = c\int_0^1 (t-1)^2\ dt = c\Big[\frac{(t-1)^3}{3}\Big]_0^1 = \frac{c}{3}$$

tedy $c = 3$. Distribuční funkce je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt$$

a pro $u \in (0,1)$ tedy máme

$$F_X(u) = \int_0^u 3(t-1)^2\ dt = \Big[(t-1)^3\Big]_0^u = (u-1)^3 + 1$$

celkově pak

$$F_X(u) = \begin{cases} 0\,, & u \leq 0\,, \\ (u-1)^3 + 1\,, & u \in (0,1)\,, \\ 1\,, & u \geq 1\,. \end{cases}$$

Pomocí hustoty spočítáme střední hodnotu

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\ dt = \int_0^1 t \cdot 3(t-1)^2\ dt = 3\Big[\frac{t^4}{4} - \frac{2t^3}{3} + \frac{t^2}{2}\Big]_0^1 = \frac{1}{4} = 0.25$$

a rozptyl

$$D(X) = E(X^2) - \big(E(X)\big)^2\ .$$

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)\ dt = \int_0^1 t^2 \cdot 3(t-1)^2\ dt = 3\Big[\frac{t^5}{5} - \frac{2t^4}{4} + \frac{t^3}{3}\Big]_0^1 = \frac{1}{10} = 0.1\ .$$

Tedy

$$D(X) = E(X^2) - \big(E(X)\big)^2 = \frac{1}{10} - \Big(\frac{1}{4}\Big)^2 = \frac{3}{80} = 0.0375\ .$$

Cvičení 10.6. *Náhodná veličina X má rozdělení dané hustotou*

$$f_X(t) = \begin{cases} a \cdot \ln t\,, & t \in (1,e)\,, \\ 0\,, & \textit{jinak}\,. \end{cases}$$

Určete konstantu a, distribuční funkci, střední hodnotu a rozptyl náhodné veličiny X.

Řešení. Pro určení konstanty a potřebujeme, aby

$$1 = \int_{-\infty}^{\infty} f_X(t)\ dt = a\int_1^e \ln t\ dt = a\Big[t(\ln t - 1)\Big]_1^e = a$$

tedy $a = 1$. Distribuční funkce je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt$$

a pro $u \in (1,e)$ tedy máme

$$F_X(u) = \int_1^u \ln t\ dt = \Big[t(\ln t - 1)\Big]_1^u = u\ln u - u + 1$$

celkově pak

$$F_X(u) = \begin{cases} 0\,, & u \leq 1\,, \\ u\ln u - u + 1\,, & u \in (1,e)\,, \\ 1\,, & u \geq e\,. \end{cases}$$

Pomocí hustoty spočítáme střední hodnotu

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\ dt = \int_{1}^{e} t \cdot \ln t\ dt = \Big[\frac{t^2}{4}(2\ln t - 1)\Big]_1^e = \frac{e^2+1}{4}$$

a rozptyl

$$D(X) = E(X^2) - \big(E(X)\big)^2 .$$

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)\ dt = \int_{1}^{e} t^2 \cdot \ln t\ dt = \Big[\frac{t^3}{9}(3\ln t - 1)\Big]_1^e = \frac{2e^3+1}{9} .$$

Tedy

$$D(X) = E(X^2) - \big(E(X)\big)^2 = \frac{2e^3+1}{9} - \Big(\frac{e^2+1}{4}\Big)^2 .$$

Cvičení 10.7. *Náhodná veličina X má hustotu*

$$f_X(u) = \begin{cases} 0\,, & u < 1\,, \\ u^{-2}\,, & u \geq 1\,. \end{cases}$$

a) Napište vzorec pro distribuční funkci F_X. Určete $\mathrm{E}X$*, DX.*

b) Napište vzorec pro hustotu veličiny $Y = \frac{1}{X}$.

Řešení. Integrací hustoty po částech dostaneme předpis pro F_X,

$$\int \frac{1}{u^2}\mathrm{d}u = \frac{-1}{u} + c\,.$$

Má platit $F(u) \to 1$ pro $u \to \infty$, čili $c = 1$. Tedy

$$F_X(u) = \begin{cases} 0\,, & u < 1\,, \\ 1 - \frac{1}{u}\,, & u \geq 1\,. \end{cases}$$

Dále

$$\mathrm{E}X = \int_{1}^{\infty} \frac{1}{u}\mathrm{d}u = [\ln u]_{u=1}^{\infty} = \infty\,,$$

proto $\mathrm{D}X$ neexistuje.
Jelikož X nabývá pouze hodnot z intervalu $\langle 1, \infty)$, je $F_Y(u) = 0$ pro $u \in (-\infty, 0\rangle$. Pro $u > 0$ platí

$$\begin{aligned} F_Y(u) &= P[Y \leq u] = P\left[\frac{1}{X} \leq u\right] = P\left[\frac{1}{u} \leq X\right] = 1 - P\left[X < \frac{1}{u}\right] = \\ &= 1 - F_X\left(\frac{1}{u}\right) = \begin{cases} 1 - 0\,, & \frac{1}{u} < 1\,, \\ 1 - (1 - \frac{1}{u})\,, & \frac{1}{u} \geq 1\,, \end{cases} = \begin{cases} u\,, & u < 1\,, \\ 1\,, & u \geq 1\,. \end{cases} \end{aligned}$$

Celkově dostáváme, že Y má rovnoměrné rozdělení na intervalu $(0, 1)$ s hustotou

$$f_Y(u) = \begin{cases} 0\,, & u \notin (0,1)\,, \\ 1\,, & u \in (0,1)\,. \end{cases}$$

Cvičení 10.8. *Náhodná veličina X má rozdělení dané hustotou*

$$f_X(t) = \begin{cases} a \cdot t^2\,, & t \in (0,3)\,, \\ 0\,, & \textit{jinak}. \end{cases}$$

Určete konstantu a, distribuční funkci náhodné veličiny X, střední hodnotu a rozptyl náhodné veličiny $Y = 3X - 1$ a $P(0 \leq X < 2)$.

Řešení. Pro určení konstanty a potřebujeme, aby

$$1 = \int_{-\infty}^{\infty} f_X(t)\ dt = a\int_{0}^{3} t^2\ dt = a\Big[\frac{t^3}{3}\Big]_0^3 = 9a$$

tedy $a = \frac{1}{9}$. Distribuční funkce je rovna

$$F_X(u) = \int_{-\infty}^{u} f_X(t)\ dt$$

a pro $u \in (0,3)$ tedy máme

$$F_X(u) = \int_0^u \frac{t^2}{9}\, dt = \Big[\frac{t^3}{27}\Big]_0^u = \frac{u^3}{27}$$

celkově pak

$$F_X(u) = \begin{cases} 0\,, & u \le 0\,, \\ \frac{u^3}{27}\,, & u \in (0,3)\,, \\ 1\,, & u \ge 3\,. \end{cases}$$

Pro střední hodnotu a rozptyl veličiny $Y = 3X - 1$ platí

$$\mathrm{E}Y = \mathrm{E}(3X-1) = 3 \cdot \mathrm{E}X - 1,$$

$$\mathrm{D}Y = \mathrm{D}(3X-1) = \mathrm{D}(3X) = 3^2 \cdot \mathrm{D}X = 9 \cdot \big(\mathrm{E}(X^2) - (\mathrm{E}X)^2\big)\;.$$

Pomocí hustoty tudíž spočítáme momenty $\mathrm{E}X$ a $E(X^2)$:

$$\mathrm{E}X = \int_{-\infty}^{\infty} t \cdot f_X(t)\; dt = \int_0^3 t \cdot \frac{t^2}{9}\; dt = \Big[\frac{t^4}{36}\Big]_0^3 = \frac{9}{4},$$

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t)\; dt = \int_0^3 t^2 \cdot \frac{t^2}{9}\; dt = \Big[\frac{t^5}{45}\Big]_0^3 = \frac{27}{5}.$$

Tedy $E(Y) = 3 \cdot \frac{9}{4} - 1 = \frac{23}{4} = 5.75$ a $D(Y) = 9 \cdot (\frac{27}{5} - (\frac{9}{4})^2) = \frac{567}{20} = 28.35$. Díky spojitosti veličiny X máme:

$$P(0 \le X < 2) = P(X < 2) - P(X < 0) = F_X(2) - F_X(0) = \frac{2^3}{27} - 0 = \frac{8}{27} \doteq 0.296\,,$$

nebo alternativní řešení:

$$P(0 \le X < 2) = \int_0^2 f_X(t)\; dt = \int_0^2 \frac{t^2}{9}\; dt = \frac{8}{27}\,.$$

Cvičení 10.9. *Nezávislé náhodné veličiny $X_1, X_2, \ldots, X_n$, $n \in \mathbb{N}$ mají (stejné) rovnoměrné rozdělení na intervalu $(-a, 2\,a)$, $a \in (0,\infty)$. Určete střední hodnotu a rozptyl náhodné veličiny*

$$Y = -\frac{5}{n}\sum_{i=1}^{n} X_i\,.$$

Výsledky.

$$\mathrm{E}X_i = \frac{a}{2}, \quad \mathrm{E}Y = \frac{-5a}{2}, \quad \mathrm{D}X_i = \frac{3a^2}{4}, \quad \mathrm{D}Y = \frac{75a^2}{4n}.$$

Cvičení 10.10. *Z kruhu o poloměru $a > 0$ vybereme náhodně bod (předpokládáme rovnoměrné rozdělení). Náhodná veličina X je rovna vzdálenosti bodu od středu. Určete její rozdělení X, včetně střední hodnoty a rozptylu.*

Řešení. Po použití selského rozumu v kombinaci s obrázkem vidíme, že distribuční funkce je rovna

$$F_X(u) = P(X \le u) = \begin{cases} 0\,, & u < 0\,, \\ \frac{\pi u^2}{\pi a^2} = \frac{u^2}{a^2}\,, & u \in [0,a)\,, \\ 1\,, & u \ge a\,. \end{cases}$$

Tedy

$$\mathrm{E}X = \int_0^a u\frac{2u}{a^2}\;\mathrm{d}u = \left[\frac{2u^3}{3a^2}\right]_0^a = \frac{2a}{3}$$

a

$$\mathrm{E}X^2 = \int_0^a u^2\frac{2u}{a^2}\;\mathrm{d}u = \left[\frac{u^4}{2a^2}\right]_0^a = \frac{a^2}{2},$$

tudíž $\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{a^2}{18}$.

Cvičení 10.11. *Náhodně vybereme číslo a z intervalu $[0,5]$ (předpokládáme rovnoměrné rozdělení). Jaký je průměrný obsah rovnostranného trojúhelníka se stranou a?*

Řešení. Hustota veličiny a je rovna

$$f_a(u) = \begin{cases} \frac{1}{5}\,, & u \in [0,5]\,, \\ 0\,, & u \notin [0,5]\,, \end{cases}$$

a obsah rovnostranného trojúhelníka o straně a je z Pythagorovy věty roven $S = \frac{\sqrt{3}a^2}{4}$. Proto je

$$\mathrm{E}S = \mathrm{E}\frac{\sqrt{3}a^2}{4} = \int_{-\infty}^{\infty} \frac{\sqrt{3}u^2}{4} f_a(u)\;\mathrm{d}u = \frac{\sqrt{3}}{20}\int_0^5 u^2\;\mathrm{d}u = \frac{\sqrt{3}}{20}\left[\frac{u^3}{3}\right]_0^5 = \frac{25\sqrt{3}}{12}.$$

Cvičení 10.12. *Náhodně zvolím reálné číslo z intervalu $[2,3]$ (předpokládám rovnoměrné rozdělení) a odečtu od něj jedničku. Výsledné číslo M udává délku strany čtverce. Jaký je průměrný obsah tohoto čtverce?*

Řešení. Délka strany M je náhodná veličina s rovnoměrným rozdělení na intervalu $[1,2]$, má tedy hustotu

$$f_M(u) = \begin{cases} 1\,, & u \in [1,2)\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

Chceme-li znát průměrnou hodnotu obsahu čtverce, musíme počítat

$$\mathrm{E}M^2 = \int_{-\infty}^{\infty} u^2 f_M(u)\;\mathrm{d}u = \int_1^2 u^2\;\mathrm{d}u = \frac{7}{3}.$$

Cvičení 10.13. *Z jednotkového kruhu vybereme náhodně bod (předpokládáme rovnoměrné rozdělení). Veličina U vyjadřuje vzdálenost bodu od kružnice, tj. od okraje kruhu. Nalezněte distribuční funkci F_U a střední hodnotu* $\mathrm{E}U$.

Řešení. Jelikož obsah kruhu o poloměru α je $\pi\alpha^2$, po nakreslení obrázku lze vytušit, že distribuční funkce F_U má předpis

$$F_U(u) = P(U \le u) = \begin{cases} 0\,, & u < 0\,, \\ \frac{1^2\pi - \pi(1-u)^2}{\pi} = 2u - u^2\,, & u \in [0,1)\,, \\ 1\,, & u \ge 1\,. \end{cases}$$

Z toho plyne, že hustota f_U a střední hodnota jsou rovny

$$f_U(u) = \begin{cases} 2-2u\,, & u \in [0,1)\,, \\ 0\,, & \text{jinak}\,, \end{cases} \qquad \mathrm{E}U = \int_{-\infty}^{\infty} u f_U(u)\;\mathrm{d}u = \int_0^1 2u - 2u^2\;\mathrm{d}u = \frac{1}{3}.$$

Cvičení 10.14. *Z obdélníku s rohy $(0,0),(a,0),(0,b),(a,b)$, kde $a,b>0$ vybereme náhodně bod a spojíme jej úsečkami s rohy $(0,0)$, $(a,0)$. Určete průměrný obsah právě vzniklého trojúhelníka.*

Výsledky. Označíme-li obsah S, pak máme $\mathrm{E}S = \frac{ab}{4}$.

Cvičení 10.15. *Náhodně vyberu jedno z čísel $1,2,3,4$ (všechny mají stejnou pravděpodobnost) a označím jej a. Jaká je průměrná hodnota veličiny $S = \frac{a^3+2a}{2}$?*

Řešení. Veličina a má pravděpodobnostní funkci

a	1	2	3	4
p_a	0.25	0.25	0.25	0.25

Je třeba spočítat

$$\mathrm{E}S = \mathrm{E}\left(\frac{a^3+2a}{2}\right) = \sum_{\alpha=1,2,3,4} \frac{\alpha^3+2\alpha}{2}\cdot p_a(\alpha) = \frac{1}{4}\sum_{\alpha=1,2,3,4}\frac{\alpha^3+2\alpha}{2} = 15.$$

Cvičení 10.16. *Náhodná veličina S má distribuční funkci*

$$F_S(u) = \begin{cases} 0\,, & u < -2\,, \\ \frac{1}{2}\,, & u \in [-2,0)\,, \\ \frac{2+u}{4}\,, & u \in [0,2)\,, \\ 1\,, & u \ge 2\,. \end{cases}$$

Určete její střední hodnotu.

Řešení. Náhodná veličina má jak diskrétní část (skok v bodě -2 o velikosti $\frac{1}{2}$), tak spojitou (vše ostatní). Proto musíme do střední hodnoty započítat obě části. Nechť p_d označuje pravděpodobnostní funkci „diskrétní části“ a nechť f_s označuje hustotu „spojité části“ (Ani jedna z funkcí není pravděpodobnostní funkce ani hustota v pravém slova smyslu). Pak derivací F_S skoro všude dostáváme

$$f_s(u) = \begin{cases} \frac{1}{4}\,, & u \in [0,2)\,, \\ 0\,, & \text{jinak}\,, \end{cases}$$

tedy

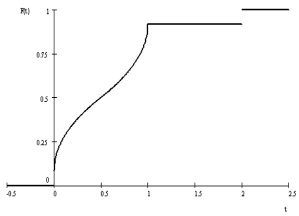
$$\mathrm{E}S = \sum_{\text{v } a \text{ je skok}} a \cdot p_d(a) + \int_{-\infty}^{\infty} u f_s(u)\,\mathrm{d}u = -2 \cdot \frac{1}{2} + \int_0^2 \frac{u}{4}\,\mathrm{d}u = -\frac{1}{2}.$$

Cvičení 10.17. *Náhodná veličina X má distribuční funkci*

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ \frac{1+5\sqrt{2t}}{12}\,, & 0 \le t < 1/2\,, \\ \frac{11-5\sqrt{2(1-t)}}{12}\,, & 1/2 \le t < 1\,, \\ \frac{11}{12}\,, & 1 \le t < 2\,, \\ 1\,, & t \ge 2\,. \end{cases}$$

Najděte její střední hodnotu.

Řešení.



Integrací kvantilové funkce vyjde $0.5 + 1/12 \doteq 0.583\,33$.

Cvičení 10.18. *Náhodná veličina X má distribuční funkci danou předpisem*

$$F_X(x) = \begin{cases} 0\,, & x < -4\,, \\ \frac{x+5}{5}\,, & -4 \le x < -3\,, \\ \frac{2}{5}\,, & -3 \le x < 0\,, \\ \frac{2x+4}{5}\,, & 0 \le x < \frac{1}{2}\,, \\ 1\,, & x \ge \frac{1}{2}\,. \end{cases}$$

a) Určete hodnoty $P(X < -10)$, $P(X = -4)$, $P(X = 0)$, $P\left(X \in \left\{-3, \frac{1}{2}\right\}\right)$, $P(-4 \le X \le \frac{1}{2})$.

b) Když rozdělíme X na směs náhodných veličin $Mix_w(U,V)$, kde U je diskrétní, V absolutně spojitá a $w \in [0,1]$, nalezněte hustotu veličiny V.

Řešení. a) Pravděpodobnosti jsou následující

$$P(X < -10) = 0,\; P(X = -4) = \frac{1}{5},\; P(X = 0) = \frac{2}{5},$$

$$P\left(X \in \left\{-3, \frac{1}{2}\right\}\right) = 0,\; P(-4 \le X \le \frac{1}{2}) = 1.$$

b) Pokud $X = \mathrm{Mix}_w(U,V)$, kde U je diskrétní a V absolutně spojitá, potom $F_X = wF_U + (1-w)F_V$, přičemž w je dáno „součtem pravděpodobností u diskrétní části X“, tj. součtem skoků: $w = P(X \in \{-4; 0\}) = P(X = -4) + P(X = 0) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$. Tedy $(1-w) = \frac{2}{5}$. Aby platilo

$$\lim_{x \to \infty} F_V(x) = 1,$$

musíme z F_X odečíst „skoky“ a vynormovat vše konstantou $(1-w)^{-1} = \frac{5}{2}$. čili máme

$$F_V(x) = \begin{cases} \frac{5}{2} \cdot 0\,, & x < -4\,, \\ \frac{5}{2} \cdot \left(\frac{x+5}{5} - \frac{1}{5}\right)\,, & -4 \le x < -3\,, \\ \frac{5}{2} \cdot \left(\frac{2}{5} - \frac{1}{5}\right)\,, & -3 \le x < 0\,, \\ \frac{5}{2} \cdot \left(\frac{2x+4}{5} - \frac{3}{5}\right)\,, & 0 \le x < \frac{1}{2}\,, \\ \frac{5}{2} \cdot \left(1 - \frac{3}{5}\right)\,, & x \ge \frac{1}{2}\,. \end{cases}$$

Dostáváme

$$F_V(x) = \begin{cases} 0\,, & x < -4\,, \\ \frac{x+4}{2}\,, & -4 \le x < -3\,, \\ \frac{1}{2}\,, & -3 \le x < 0\,, \\ \frac{2x+1}{2}\,, & 0 \le x < \frac{1}{2}\,, \\ 1\,, & x \ge \frac{1}{2}\,, \end{cases}$$

odkud derivací skoro všude dostáváme hustotu

$$f_V(x) = \begin{cases} \frac{1}{2}\,, & -4 \le x < -3\,, \\ 1\,, & 0 \le x < \frac{1}{2}\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

Cvičení 10.19. *Alice a Bob hází kamenem na okno. Alice jej trefí s pravděpodobností $p_1 = 0.05$, Bob s pravděpodobností $p_2 = 0.08$, vzájemně nezávisle. Začíná Alice a střídají se. Jakmile se někdo strefí do okna, hra končí.*

a) Jaký je průměrný počet hodů kamenem ve hře?

b) V kterém kole hra průměrně skončí?

c) Jaký je rozptyl počtu kol?

Řešení. a) Nechť Z je náhodná veličina označující hod, ve kterém se okno rozbíjí. Její pravděpodobnostní funkce je rovna

$$p_X(k) = \begin{cases} \big((1-p_1)(1-p_2)\big)^{\frac{k-1}{2}} p_1\,, & k \text{ je liché}\,, \\ (1-p_1)^{\frac{k}{2}}(1-p_2)^{\frac{k}{2}-1} p_2\,, & k \text{ je sudé}\,, \end{cases} \quad k \in \mathbb{N}.$$

Potom střední hodnotu Z spočteme jako

$$\mathrm{E}Z = \sum_{k=1}^{\infty} k \cdot p_Z(k) = \sum_{k=0}^{\infty} (2k+1)\big((1-p_1)(1-p_2)\big)^{\frac{2k+1-1}{2}} p_1 + \sum_{k=1}^{\infty} (2k)(1-p_1)^{\frac{2k}{2}}(1-p_2)^{\frac{2k}{2}-1} p_2 =$$

$$= 2p_1 \sum_{k=0}^{\infty} k\big((1-p_1)(1-p_2)\big)^k + p_1 \sum_{k=0}^{\infty} \big((1-p_1)(1-p_2)\big)^k + \frac{2p_2}{1-p_2} \sum_{k=1}^{\infty} k\big((1-p_1)(1-p_2)\big)^k.$$

Označme si pro jednoduchost $t = (1-p_1)(1-p_2) = \frac{874}{1000}$. Potom máme

$$\mathrm{E}Z = \left(2p_1 + \frac{2p_2}{1-p_2}\right) \sum_{k=1}^{\infty} kt^k + p_1 \sum_{k=0} t^k = \left(2p_1 + \frac{2p_2}{1-p_2}\right) t \left(\sum_{k=0}^{\infty} t^k\right)' + \frac{p_1}{1-t} =$$

$$= \left(2p_1 + \frac{2p_2}{1-p_2}\right) t \left(\frac{1}{1-t}\right)' + \frac{p_1}{1-t} = \left(2p_1 + \frac{2p_2}{1-p_2}\right) \frac{t}{(1-t)^2} + \frac{p_1}{1-t},$$

což je po vyčíslení rovno

$$\mathrm{E}Z = 15.2\overline{7}.$$

b) Označme p pravděpodobnost, že během se jednoho kola okno rozbije. Platí $p = p_1 + (1-p_1)p_2 = 0.126$. Potom pokud Y je pořadí kola, ve kterém se okno rozbije, máme $Y \sim \mathrm{Geom}(p)$. Tedy $\mathrm{E}Y = \frac{1}{p} = 7.9365$.

c) Máme $Y \sim \mathrm{Geom}(p)$, tedy rozptyl je roven $\mathrm{D}Y = \frac{1-p}{p^2} = 55.0517$.

Cvičení 10.20. *Alice a Bob hrají hru, sestávající ze tří až pěti kol. Žádné kolo nemůže skončit remízou. Hráč vyhrává, vyhraje-li tři kola.*

a) Určete rozdělení pravděpodobnosti délky hry, jestliže Alice vyhraje jedno kolo s pravděpodobností p.

b) Určete střední hodnotu délky hry, jestliže Alice vyhraje jedno kolo s pravděpodobností $1/2$.

Řešení. a) Náhodná veličina udávající počet kol nabývá hodnot z množiny $\{3, 4, 5\}$ s pravděpodobností

$$\begin{aligned} P[X=3] &= \binom{2}{0}\left(p^3 + (1-p)^3\right), \quad P[X=4] = \binom{3}{1}\left(p^3\,(1-p) + p\,(1-p)^3\right), \\ P[X=5] &= \binom{4}{2}\left(p^3\,(1-p)^2 + p^2\,(1-p)^3\right). \end{aligned}$$

b) Pro $p = 1/2$ je střední hodnota rovna

$$\mathrm{E}X = \frac{6}{8} + \frac{24}{16} + \frac{60}{32} = \frac{33}{8}\,.$$

Cvičení 10.21. *Rozvodné závody dodávaly elektřinu, jejíž napětí ve voltech mělo normální rozdělení* $\mathrm{N}(230, 25)$. *Nyní se jim podařilo snížit rozptyl na* 10. *O kolik mohou zvýšit střední hodnotu při zachování horní meze, která je překročena jen s pravděpodobností* 10^{-4}*?*

Výsledky. Závody mohou zvýšit střední hodnotu napětí o zhruba 5.46 V.

Cvičení 10.22. *Spojitá náhodná veličina je frekvence v* Hz. *Jaký fyzikální rozměr má její rozptyl, směrodatná odchylka, medián, dále argumenty a výsledky distribuční a kvantilové funkce a hustoty?*

Řešení. Rozptyl Hz^2, směrodatná odchylka i medián Hz, distribuční funkce $\mathrm{Hz} \mapsto 1$, kvantilová funkce $1 \mapsto \mathrm{Hz}$, hustota $\mathrm{Hz} \mapsto \mathrm{Hz}^{-1} = \mathrm{s}$.

Cvičení 10.23. *Náhodná veličina X má distribuční funkci*

$$F_X(t) = \begin{cases} 0\,, & t < 0\,, \\ 1 - \exp(-2\,t)\,, & t \geq 0\,. \end{cases}$$

Popište rozdělení náhodné veličiny $Y = 2 - 2\,X$ a stanovte její střední hodnotu a rozptyl.

Řešení. Jedná se o exponenciální rozdělení s parametrem $\tau = 1/2$, $\mathrm{E}X = \tau = 1/2$, $\mathrm{D}X = \tau^2 = 1/4$,

$$\begin{aligned} f_X(t) = F'_X(t) &= \begin{cases} 0\,, & t < 0\,, \\ 2\,\exp(-2\,t)\,, & t \geq 0\,, \end{cases} \\ q_X(\alpha) = F_X^{-1}(\alpha) &= -\frac{1}{2}\,\ln(1-\alpha)\,. \end{aligned}$$

Změna znaménka:

$$\begin{aligned} F_{-X}(t) = 1 - F_X(-t) &= \begin{cases} \exp(2\,t)\,, & t < 0\,, \\ 0\,, & t \geq 0\,, \end{cases} \\ f_{-X}(t) = f_X(-t) &= \begin{cases} 2\,\exp(2\,t)\,, & t < 0\,, \\ 0\,, & t \geq 0\,, \end{cases} \\ q_{-X}(\alpha) = -q_X(1-\alpha) &= \frac{1}{2}\,\ln(\alpha)\,. \end{aligned}$$

Lineární zobrazení (nyní již násobíme $-X$ kladným číslem 2):

$$\begin{aligned} q_Y(\alpha) &= 2 + 2\,q_{-X}(\alpha) = 2 + \ln(\alpha)\,, \\ F_Y(t) &= q_Y^{-1}(\alpha) = F_{-X}\left(\frac{t}{2} - 1\right) = \begin{cases} \exp(t-2)\,, & t < 2\,, \\ 0\,, & t \geq 2\,, \end{cases} \\ f_Y(t) &= F'_Y(t) = \begin{cases} \exp(t-2)\,, & t < 2\,, \\ 0\,, & t \geq 2\,. \end{cases} \end{aligned}$$

$$\begin{aligned} \mathrm{E}Y &= 2 - 2\,\mathrm{E}X = 1\,, \\ \mathrm{D}Y &= 2^2\,\mathrm{D}X = 1\,. \end{aligned}$$

Cvičení 10.24. *Náhodná veličina X má alternativní rozdělení; nabývá hodnot $0, 1$ s pravděpodobností $1/2$. Náhodná veličina Y má rovnoměrné rozdělení na intervalu $\langle 0, 1\rangle$. Určete a znázorněte rozdělení náhodných veličin*

a) $2\,Y + 1$,

b) $\mathrm{Mix}_{2/3}(Y, X)$,

c) $X + Y$.

(Návod: X je směsí dvou konstantních náhodných veličin.)

Výsledky.

a) $2\,Y + 1$ má rozdělení $\mathrm{R}(1, 3)$.

b) $\mathrm{Mix}_{2/3}(Y, X)$ má distribuční funkci

$$F_{\mathrm{Mix}_{2/3}(Y,X)}(t) = \begin{cases} 0\,, & t < 0\,, \\ \dfrac{1}{6} + \dfrac{2}{3}\,t\,, & 0 \le t < 1\,, \\ 1\,, & t \ge 1\,. \end{cases}$$

c) $X + Y$ má rozdělení $\mathrm{R}(0, 2)$.

Cvičení 10.25. *Náhodná veličina X má binomické rozdělení* $\mathrm{Bi}(2, \frac{2}{3})$, *náhodná veličina Y má spojité rovnoměrné rozdělení* $\mathrm{R}(0, 1)$. *Popište a znázorněte rozdělení náhodných veličin*

a) $Y + \mathrm{E}X$,

b) $X - \mathrm{E}Y$,

c) $-2\,X$,

d) $-2\,Y$,

e) $\mathrm{Mix}_{1/3}(X, Y)$.

Výsledky. a) $Y + \mathrm{E}X$ má rozdělení $\mathrm{R}(\frac{2}{3}, \frac{5}{3})$.

b) $X - \mathrm{E}Y$ má pravděpodobnostní funkci v tabulce:

t	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$
$p_{X-\mathrm{E}Y}(t)$	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$

c) $-2\,X$ má pravděpodobnostní funkci v tabulce:

t	0	-2	-4
$p_{-2\,X}(t)$	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$

d) $-2\,Y$ má rozdělení $\mathrm{R}(-2, 0)$.

e) $\mathrm{Mix}_{1/3}(X, Y)$ má distribuční funkci

$$F_{\mathrm{Mix}_{1/3}(X,Y)}(t) = \begin{cases} 0\,, & t < 0\,, \\ \dfrac{1}{27} + \dfrac{2}{3}\,t\,, & 0 \le t < 1\,, \\ \dfrac{23}{27}\,, & 1 \le t < 2\,, \\ 1\,, & t \ge 2\,. \end{cases}$$

Cvičení 10.26. *Náhodná veličina X má hustotu*

$$f_X(t) = \begin{cases} c\,t\,, & t \in \langle 0,1 \rangle\,, \\ 0\,, & \textit{jinak}\,, \end{cases}$$

kde $c \in \mathbb{R}$. Vypočtěte střední hodnotu, určete a znázorněte distribuční funkce veličin $-X$ a X^2.

Výsledky. $\mathrm{E}X = 2/3$,

$$F_{-X}(t) = \begin{cases} 0\,, & t < -1 \\ 1 - t^2\,, & -1 \le t < 0\,, \\ 1\,, & t \ge 0\,. \end{cases}$$

X^2 má rozdělení $\mathrm{R}(0,1)$.

Cvičení 10.27. *Za první účast na zkoušce se platí 30 EUR, za každý opravný pokus $2\times$ více než za předešlý. Uchazeč má v každém pokusu pravděpodobnost úspěchu p. Na kolik ho v průměru zkouška přijde (v závislosti na p)?*

Řešení. Počet pokusů má (posunuté) geometrické rozdělení s pravděpodobnostní funkcí $p_X(k) = p\,(1-p)^{k-1}$, $k \in \{1,2,\ldots\}$. Střední hodnota úhrady je

$$30 \sum_{k=1}^{\infty} 2^{k-1}\, p\,(1-p)^{k-1} = 30\,p \sum_{k=1}^{\infty} (2\,(1-p))^{k-1}\,,$$

což je součet geometrické řady. Ta je pro $p \le 1/2$ divergentní (rovná ∞), pro $p > 1/2$ je střední hodnota úhrady

$$\frac{30\,p}{1 - 2\,(1-p)} = \frac{30\,p}{2\,p - 1}\,.$$

Cvičení 10.28. *Náhodná veličina X má diskrétní rozdělení s pravděpodobnostní funkcí $p_X(0) = \frac{1}{4}$, $p_X(1) = \frac{3}{4}$ a náhodná veličina Y má spojité rovnoměrné rozdělení na intervalu $\langle 0,1 \rangle$. Náhodná veličina Z je jejich směsí $\mathrm{Mix}_{\frac{1}{2}}(X,Y)$.*

a) Určete distribuční funkci F_Z.
b) Vypočtěte střední hodnotu $\mathrm{E}\,Z$.
c) Vypočtěte dolní kvartil, t.j. kvantil $q_Z(0.25)$.

Řešení. Pro požadovanou směs je $F_Z(u) = \frac{1}{2}\,F_X(u) + \frac{1}{2}\,F_Y(u)$. Pro jednotlivé složky dostaneme jejich distribuční funkce

$$F_X(u) = \begin{cases} 0, & u < 0, \\ \frac{1}{4}, & 0 \le u < 1, \\ 1, & u \ge 1, \end{cases} \qquad F_Y(u) = \begin{cases} 0, & u < 0, \\ u, & 0 \le u < 1, \\ 1, & u \ge 1. \end{cases}$$

a) Ze vzorce pro distribuční funkci je

$$F_Z(u) = \begin{cases} 0, & u < 0, \\ \frac{4u+1}{8}, & 0 \le u < 1, \\ 1, & u \ge 1. \end{cases}$$

b) Pro střední hodnotu platí obdobný vzorec

$$\mathrm{E}\,Z = \frac{1}{2}\,(\mathrm{E}\,X + \mathrm{E}\,Y)\,.$$

Pro střední hodnoty ve vzorci dostaneme:

$$\begin{aligned} \mathrm{E}\,X &= \sum_{u \in \mathbb{R}} u\,p_X(u) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{3}{4} = \frac{3}{4}\,; \\ \mathrm{E}\,Y &= \int_{-\infty}^{\infty} u\,F_Y'(u)\mathrm{d}u = \int_0^1 u\mathrm{d}u = \left(\frac{u^2}{2}\right)_0^1 = \frac{1}{2} \end{aligned}$$

a odtud

$$\mathrm{E}\,Z = \frac{1}{2}\,(\mathrm{E}\,X + \mathrm{E}\,Y) = \frac{1}{2}\left(\frac{3}{4} + \frac{1}{2}\right) = \frac{5}{8}\,.$$

c) Protože je $F_Z(0-) = 0$, $F_Z(0+) = \frac{1}{8}$, $F_Z(1-) = \frac{3}{8}$, $F_Z(1) = 1$ a $\frac{1}{8} = 0.125 < 0.25 < \frac{3}{8} = 0.375 \Longrightarrow 0 < q_Z(0.25) < 1$, tudíž

$$F_Z(u) = \frac{1}{4} \Longrightarrow \frac{4u+1}{8} = \frac{1}{4} \Longrightarrow 4u+1 = 2 \Longrightarrow u = \frac{1}{4}\,.$$

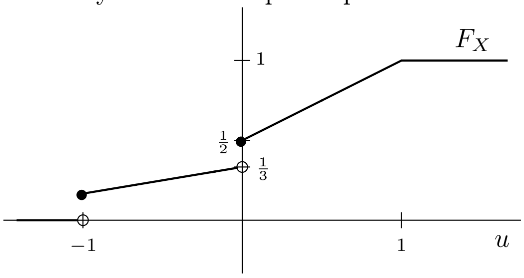
Je tedy $q_Z(0.25) = 0.25$.

Cvičení 10.29. *Náhodná veličina X má rozdělení určené distribuční funkcí F_X, kde*

$$F_X(u) = \begin{cases} 0, & -\infty < u < -1, \\ \frac{u+2}{6}, & -1 \le u < 0, \\ \frac{u+1}{2}, & 0 \le u < 1, \\ 1, & 1 \le u < \infty. \end{cases}$$

a) Vyjádřete náhodnou veličinu X jako směs dvou náhodných veličin, které mají spojité a diskrétní rozdělení.
b) Vypočtěte střední hodnotu $\mathrm{E}\,X$.
c) Vypočtěte horní kvartil, t.j. kvantil $q_X(0.75)$.

Řešení. Distribuční funkce má průběh znázorněný na obrázku, je nespojitá v bodech -1 a 0 a těchto hodnot nabývá s kladnou pravděpodobností.



$F_X(-1-) = 0,\ F_X(-1+) = \frac{1}{6},$

$F_X(0-) = \frac{1}{3},\ F_X(0+) = \frac{1}{2},$

$F_X(1-) = 1,\ F_X(1+) = 1,$

$P[X = -1] = \frac{1}{6},\ P[X = 0] = \frac{1}{6}.$

a) Jestliže označíme U náhodnou veličinu, která má diskrétní rozdělení a V tu, která má spojité rozdělení, pak ze skutečnosti
$P[X = 0 \vee X = -1] = P[X = -1] + P[X = 0] = \frac{1}{3}$
dostaneme, že

$$X = \mathrm{Mix}_{\frac{1}{3}}(U, V) \Longrightarrow F_X(u) = \frac{1}{3}\,F_U(u) + \frac{2}{3}\,F_V(u),\ E(X) = \frac{1}{3}\,E(U) + \frac{2}{3}\,E(V)\,.$$

Pro pravděpodobnostní funkci p_U náhodné veličiny U dostaneme

$$U:\ p_U(-1) = 3\,P(X = -1) = \frac{3}{6} = \frac{1}{2},\ p_U(0) = 3\,P(X = 0) = \frac{3}{6} = \frac{1}{2}\,.$$

Pro distribuční funkci F_V náhodné veličiny V dostaneme
$V:\ -1 \le u < 0:\ F_V(u) = \frac{3}{2}\left(F_X(u) - P[X = -1]\right) = \frac{3}{2}\left(\frac{u+2}{6} - \frac{1}{6}\right) = \frac{u+1}{4};$
$0 \le u < 1:\ F_V(u) = \frac{3}{2}\left(F_X(u) - P[X = -1 \vee X = 0]\right) = \frac{3}{2}\left(\frac{u+1}{2} - \frac{1}{3}\right) = \frac{3u+1}{4}.$
Rozdělení náhodné veličiny V můžeme popsat její hustotou f_V, pro kterou je

$$f_V(u) = \frac{3}{2}\,F_X'(u) \Longrightarrow f_V(u) = \frac{3}{2}\frac{1}{6} = \frac{1}{4},\ -1 < u < 0;\ f_V(u) = \frac{3}{2}\frac{1}{2} = \frac{3}{4},\ 0 < u < 1.$$

Je tedy

$$F_U(u) = \begin{cases} 0, & -\infty < u < -1, \\ \frac{1}{2}, & -1 \le u < 0, \\ 1, & 0 \le u < \infty, \end{cases} \qquad F_V(u) = \begin{cases} 0, & -\infty < u < -1, \\ \frac{u+1}{4}, & -1 \le u < 0, \\ \frac{3u+1}{4}, & 0 \le u < 1, \\ 1, & u \le 1 < \infty. \end{cases}$$

$$p_U(u) = \begin{cases} \frac{1}{2}, & u = -1, \\ \frac{1}{2}, & u = 0, \\ 0, & \text{jinde}, \end{cases} \qquad f_V(u) = \begin{cases} 0, & u < -1 \vee u > 1, \\ \frac{1}{4}, & -1 < u < 0, \\ \frac{3}{4}, & 0 < u < 1. \end{cases}$$

b) Pro střední hodnotu směsi je $\mathrm{E}\,X = \frac{1}{3}\,\mathrm{E}\,U + \frac{2}{3}\,\mathrm{E}\,V$.

Postupně dostaneme

$$
\begin{aligned}
\mathrm{E}\, U &= \sum_{u\in\mathbb{R}} u\, p_U(u) = (-1)\cdot\frac{1}{2} + 0\cdot\frac{1}{2} = -\frac{1}{2}\,,\\
\mathrm{E}\, V &= \int_{-\infty}^{\infty} u\, f_V(u)\mathrm{d}u = \int_{-1}^{0}\frac{u}{4}\mathrm{d}u + \int_{0}^{1}\frac{3u}{4}\mathrm{d}u = \left[\frac{u^2}{8}\right]_{u=-1}^{0} + \left[\frac{3u^2}{8}\right]_{u=0}^{1} = \\
&= -\frac{1}{8} + \frac{3}{8} = \frac{1}{4}\,,
\end{aligned}
$$

tedy

$$
\mathrm{E}\, X = \frac{1}{3}\left(\frac{-1}{2}\right) + \frac{2}{3}\cdot\frac{1}{4} = \frac{-1}{6} + \frac{1}{6} = 0\,.
$$

c) Je $F_X(0-) = 0$, $F_X(0+) = \frac{1}{6}$, $F_X(0-) = \frac{1}{3}$, $F_X(0+) = \frac{1}{2}$, $F_X(1) = 1$. Protože je $\frac{1}{2} < \frac{3}{4} < 1$, máme $0 < q_X(0.75) < 1$, a tedy

$$
\frac{3}{4} = F_X(u) \Longrightarrow \frac{3}{4} = \frac{u+1}{2} \Longrightarrow u = \frac{1}{2} \Longrightarrow q_X(0.75) = 0.5\,.
$$

Cvičení 10.30. *Náhodná veličina X má spojité rozdělení určené hustotou f_X, kde*

$$
f_X(u) = \begin{cases} 0, & u \in (-\infty,\, 1),\\ \frac{2}{u^3}, & u \in \langle 1,\, \infty). \end{cases}
$$

a) Určete distribuční funkci F_X.
b) Vypočtěte střední hodnotu a rozptyl náhodné veličiny X.
c) Vypočtěte pravděpodobnost $P[X > 2]$.

Řešení. a) Distribuční funkce:

$$
F_X(t) = \begin{cases} 0, & t \in (-\infty,\, 1),\\ \int_1^t \frac{2}{u^3}\, du = \left[\frac{-1}{u^2}\right]_1^t = 1 - \frac{1}{t^2}, & t \in \langle 1,\, \infty). \end{cases}
$$

b) Střední hodnota $E(X)$ a druhý moment $E(X^2)$:

$$
E(X) = \int_1^{\infty} u\cdot\frac{2}{u^3}\, du = \Big[-\frac{2}{u}\Big]_1^{\infty} = 2
$$

$$
E(X^2) = \int_1^{\infty} u^2\cdot\frac{2}{u^3}\, du = \int_1^{\infty}\frac{2}{u}\, du = \Big[2\ln u\Big]_1^{\infty} = \infty
$$

Tedy

$$
D(X) = E(X^2) - \big(E(X)\big)^2 = \infty\ .
$$

c)

$$
P[X > 2] = \int_2^{\infty}\frac{2}{u^3}\, du = \Big[\frac{-1}{u^2}\Big]_2^{\infty} = \frac{1}{4} = 0.25\,.
$$

Cvičení 10.31. *Najděte příklad nezáporné náhodné veličiny, která má střední hodnotu* 1 *a směrodatnou odchylku* 10*, nebo dokažte, že taková náhodná veličina neexistuje.*

Výsledky. Existuje. Např. vezměme náhodnou veličinu X, která má alternativní rozdělení s parametrem $q > 0$, takže $\mathrm{E}X = q$. Náhodná veličina $Y = \frac{1}{q}X$ má střední hodnotu 1 a rozptyl

$$
\mathrm{D}Y = \frac{1}{q^2}\,\mathrm{D}X = \frac{1-q}{q} = \frac{1}{q} - 1\,,
$$

který můžeme volbou q dosáhnout libovolně velký.

11 Náhodné vektory (vícerozměrné náhodné veličiny) a nezávislost náhodných veličin

Cvičení 11.1. *Náhodný vektor (X, Y) má diskrétní rozdělení určené pravděpodobnostní funkcí $p_{X,Y}(x, y)$, která je daná tabulkou*

$y \setminus x$	0	1	2
0	1/4	1/8	0
1	1/4	1/4	1/8

a) Jaká jsou jejich marginální rozdělení?

b) Určete kovarianční a korelační matici.

c) Jsou náhodné veličiny X a Y nezávislé? Zdůvodněte.

Řešení. a) Rozdělení náhodné veličiny X je

$$\begin{aligned} P[X=0] &= P[X=0, Y=0] + P[X=0, Y=1] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\ P[X=1] &= P[X=1, Y=0] + P[X=1, Y=1] = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}, \\ P[X=2] &= P[X=2, Y=0] + P[X=2, Y=1] = \frac{1}{8} + 0 = \frac{1}{8}. \end{aligned}$$

Rozdělení náhodné veličiny Y je analogicky

$$\begin{aligned} P[Y=0] &= \frac{1}{4} + \frac{1}{8} + 0 = \frac{3}{8}, \\ P[Y=1] &= \frac{1}{4} + \frac{1}{4} + \frac{1}{8} = \frac{5}{8}. \end{aligned}$$

b) Kovarianci vypočteme ze vztahu $\operatorname{cov}(X, Y) = \mathrm{E}XY - \mathrm{E}X \cdot \mathrm{E}Y$:

$$\begin{aligned} \mathrm{E}X &= 0 \cdot \frac{1}{2} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{1}{8} = \frac{5}{8}, \\ \mathrm{E}Y &= 0 \cdot \frac{3}{8} + 1 \cdot \frac{5}{8} = \frac{5}{8}, \\ \mathrm{E}XY &= 0 \cdot 0 \cdot \frac{1}{4} + 0 \cdot 1 \cdot \frac{1}{4} + 1 \cdot 0 \cdot \frac{1}{8} + 1 \cdot 1 \cdot \frac{1}{4} + 2 \cdot 0 \cdot 0 + 2 \cdot 1 \cdot \frac{1}{8} = \frac{1}{2}, \\ \operatorname{cov}(X, Y) &= \mathrm{E}XY - \mathrm{E}X \cdot \mathrm{E}Y = \frac{1}{2} - \frac{5}{8} \cdot \frac{5}{8} = \frac{7}{64}. \end{aligned}$$

Pro rozptyly dopočítáme

$$\begin{aligned} \mathrm{E}X^2 &= 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{1}{8} = \frac{7}{8}, \\ \mathrm{E}Y^2 &= 0^2 \cdot \frac{3}{8} + 1^2 \cdot \frac{5}{8} = \frac{5}{8}, \\ \mathrm{D}X &= \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{7}{8} - \left(\frac{5}{8}\right)^2 = \frac{31}{64}, \\ \mathrm{D}Y &= \mathrm{E}Y^2 - (\mathrm{E}Y)^2 = \frac{5}{8} - \left(\frac{5}{8}\right)^2 = \frac{15}{64}. \end{aligned}$$

Kovarianční matice je

$$\Sigma_{(X,Y)} = \begin{pmatrix} \frac{31}{64} & \frac{7}{64} \\ \frac{7}{64} & \frac{15}{64} \end{pmatrix}.$$

Korelaci X a Y spočteme jako

$$\varrho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathrm{D}X}\,\sqrt{\mathrm{D}Y}} = \frac{7/64}{\sqrt{31/64}\,\sqrt{15/64}} = \frac{7}{\sqrt{465}},$$

tudíž korelační matice je

$$\varrho_{(X,Y)} = \begin{pmatrix} 1 & \frac{7}{\sqrt{465}} \\ \frac{7}{\sqrt{465}} & 1 \end{pmatrix}.$$

c) Veličiny X a Y nejsou nezávislé už proto, že $\operatorname{cov}(X,Y) \neq 0$, nebo také např.

$$P[X=0, Y=0] = \frac{1}{4} \neq P[X=0]\,P[Y=0] = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16}\,.$$

(Pozn.: Aby byly nezávislé, muselo by platit $\forall i,j : P[X=i, Y=j] = P[X=i]\,P[Y=j]$.)

Cvičení 11.2. *Náhodný vektor (X,Y) má diskrétní rozdělení určené pravděpodobnostní funkcí $p_{X,Y}(x,y)$, která je daná tabulkou*

$x \backslash y$	-1	0	1
1	1/6	0	1/3
2	1/8	1/4	1/8

a) Vypočtěte pravděpodobnost $P[X > Y]$ a střední hodnotu $\mathrm{E}(X\,Y^2)$.

b) Rozhodněte, zda jsou náhodné veličiny X a Y nezávislé.

Řešení. a) Platí $P[X>Y] = 1 - P[X \leq Y]$ a podmínce $X \leq Y$ vyhovují z tabulky pouze hodnoty $X=1,\ Y=1$, tedy

$$P[X>Y] = 1 - p_{X,Y}(1,1) = 1 - \frac{1}{3} = \frac{2}{3}\,.$$

Střední hodnotu vypočteme pomocí vzorce

$$\mathrm{E}(X\,Y^2) = \sum_{(x,y)\in\mathbb{R}^2} x\,y^2\,p_{X,Y}(x,y) = 1\cdot\frac{1}{6} + 0\cdot 0 + 1\cdot\frac{1}{3} + 2\cdot\frac{1}{8} + 0\cdot\frac{1}{4} + 2\cdot\frac{1}{8} = 1\,.$$

b) Náhodné veličiny X a Y nejsou nezávislé, neboť je $p_{X,Y}(1,0) = 0$, a tudíž nemůže být splněna nutná a postačující podmínka pro nezávislost. Je nutně

$$p_{X,Y}(1,0) \neq p_X(1)\,p_Y(0)\,.$$

Cvičení 11.3. *Náhodný vektor (X,Y) má diskrétní rozdělení určené pravděpodobnostní funkcí $p_{X,Y}(x,y)$, která je daná tabulkou*

$y \backslash x$	1	2
0	1/3	1/3
1	0	1/3

Vypočtěte korelaci náhodných veličin X, Y.

Řešení. $EX = \frac{5}{3}$, $EY = \frac{1}{3}$, $DX = DY = \frac{2}{9}$, $E\,(X\,Y) = \frac{2}{3}$,

$$\varrho\,(X,Y) = \frac{E\,(X\,Y) - EX \cdot EY}{\sigma_X\,\sigma_Y} = \frac{1}{2}\,.$$

Cvičení 11.4. *Náhodný vektor (X,Y) má diskrétní rozdělení určené pravděpodobnostní funkcí $p_{X,Y}(x,y)$, která je daná tabulkou*

$y \backslash x$	-1	0	2
1	0.15	0.25	0
3	0.1	0.2	0.3

a) Vypočtěte pravděpodobnost $P[X < Y]$.

b) Určete marginální pravděpodobnostní funkci p_X a střední hodnotu $\mathrm{E}X$.

c) Jsou náhodné veličiny X a Y nezávislé?

Řešení. a) Podmínku $X < Y$ nesplňuje pouze dvojice $(2,1)$, je tedy

$$P[X < Y] = 1 - P[X \geq Y] = 1 - p_{X,Y}(2,1) = 1 - 0 = 1\,.$$

b) Marginální pravděpodobnostní funkci p_X vypočteme pomocí vzorce $p_X(u) = \sum_{v\in\mathbb{R}} p_{X,Y}(u,v)$. Dostaneme její hodnoty

x	-1	0	2
$p_X(x)$	0.25	0.45	0.3

Střední hodnotu $\mathrm{E}X$ vypočteme podle vzorce

$$\mathrm{E}X = \sum_{x\in\mathbb{R}} x\,p_X(x) = -1\cdot 0.25 + 0\cdot 0.45 + 2\cdot 0.3 = 0.35\,.$$

c) Protože je $p_{X,Y}(2,1) = 0$, nemůže být splněna podmínka nezávislosti $p_{X,Y}(x,y) = p_X(x)\cdot p_Y(y)$, tudíž náhodné veličiny X a Y nejsou nezávislé.

Cvičení 11.5. *Náhodný vektor (X,Y) má diskrétní rozdělení určené pravděpodobnostní funkcí $p_{X,Y}(x,y)$, která je daná tabulkou*

$y \backslash x$	0	1	2
-1	0.25	0.125	0.11
1	0.125	0.25	0.14

a) Vypočtěte kovarianci $\mathrm{cov}(X,Y)$.

b) Co můžete říci o nezávislosti náhodných veličin X a Y?

c) Určete pravděpodobnostní funkci p_Z náhodné veličiny $Z = X + Y$.

Řešení. a) Kovarianci vypočteme podle vzorce $\mathrm{E}XY - \mathrm{E}X\cdot\mathrm{E}Y$.

$$\begin{aligned}\mathrm{E}XY &= \sum_{(x,y)} x\,y\,p(x,y) = 1(-1)\cdot 0.125 + 1\cdot 1\cdot 0.25 + (-1)\cdot 2\cdot 0.11 + 1\cdot 2\cdot 0.14 = \\ &= -0.125 + 0.25 - 0.22 + 0.28 = 0.185\,,\end{aligned}$$

když jsme vynechali nulové sčítance. Obdobně

$$\begin{aligned}\mathrm{E}X &= \sum_{(x,y)} x\,p(x,y) = 1\cdot(0.125+0.25) + 2\cdot(0.11+0.14) = 0.875\,,\\ \mathrm{E}Y &= \sum_{(x,y)} y\,p(x,y) = -1\cdot(0.25+0.125+0.11) + 1\cdot(0.125+0.25+0.14) = 0.03\,.\end{aligned}$$

Tedy

$$\mathrm{cov}(X,Y) = 0.185 - 0.875\cdot 0.03 = 0.1588\,.$$

b) Protože je kovariance $\mathrm{cov}(X,Y) \neq 0$, můžeme říci, že náhodné veličiny nejsou nezávislé.

c) Možné hodnoty náhodné veličiny $Z = X + Y$ jsou -1, 0, 1, 2, 3 Pravděpodobnostní funkci dostaneme sečtením pravděpodobností z tabulky, které mají příslušnou hodnotu součtu x a y, tj.

z	-1	0	1	2	3
$p_Z(z)$	0.25	0.125	0.235	0.25	0.14

Cvičení 11.6. *Náhodné veličiny X a Y jsou nezávislé a mají diskrétní rozdělení s pravděpodobnostními funkcemi, které jsou zadány tabulkami*

x	-1	2
$p_X(x)$	0.3	0.7

y	0	1	3
$p_Y(y)$	0.2	0.45	0.35

a) Vypočtěte střední hodnotu $\mathrm{E}XY$.

b) Určete pravděpodobnostní funkci p_Z náhodné veličiny $Z = X + Y$.

c) Vypočtěte korelaci $\rho(X, Y)$.

Řešení. a) Protože jsou náhodné veličiny X a Y nezávislé, je $\mathrm{E}XY = \mathrm{E}X \cdot \mathrm{E}Y$, kde

$$
\begin{aligned}
\mathrm{E}X &= \sum_x x\, p_X(x) = -1 \cdot 0.3 + 2 \cdot 0.7 = 1.1\,, \\
\mathrm{E}Y &= \sum_y y\, p_Y(y) = 1 \cdot 0.45 + 3 \cdot 0.35 = 1.5\,,
\end{aligned}
$$

tedy $\mathrm{E}XY = 1.1 \cdot 1.5 = 1.65$.

b) Náhodná veličina Z nabývá hodnot $\{-1, 0, 2, 3, 5\}$, tedy

$$
\begin{aligned}
p_Z(-1) &= p_X(-1) \cdot p_Y(0) = 0.3 \cdot 0.2 = 0.06, \\
p_Z(0) &= p_X(-1) \cdot p_Y(1) = 0.3 \cdot 0.45 = 0.135, \\
p_Z(2) &= p_X(2) \cdot p_Y(0) + p_X(-1) \cdot p_Y(3) = 0.7 \cdot 0.2 + 0.3 \cdot 0.35 = 0.245, \\
p_Z(3) &= p_X(2) \cdot p_Y(1) = 0.7 \cdot 0.45 = 0.315, \\
p_Z(5) &= p_X(2) \cdot p_Y(3) = 0.7 \cdot 0.35 = 0.245\,.
\end{aligned}
$$

c) Protože jsou náhodné veličiny X a Y nezávislé, je korelace $\rho(X, Y) = 0$.

Cvičení 11.7. *Diskrétní náhodná veličina X má rovnoměrné rozdělení na množině $\{0, 1\}$ a spojitá náhodná veličina Y má rovnoměrné rozdělení na intervalu $[0, 1]$. Veličiny X, Y jsou nezávislé. Určete rozdělení veličiny $W = X + Y$ a veličiny $Z = XY$. (Návod: lze použít směs rozdělení, jehož jedna složka odpovídá situaci $X = 0$ a druhá $X = 1$.)*

Řešení. Náhodná veličina W má rovnoměrné rozdělení na intervalu $[0, 2]$. Náhodná veličina Z má distribuční funkci

$$
F_Z(t) = \begin{cases} 0\,, & t < 0\,, \\ 1/2\,, & t = 0\,, \\ 1/2 + t/2\,, & 0 \le t < 1\,, \\ 1\,, & \text{jinak}\,. \end{cases}
$$

Cvičení 11.8. *Sdružená hustota náhodných veličin X a Y je*

$$
f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{2} e^{-x-\frac{y}{2}}\,, & x > 0, y > 0\,, \\ 0\,, & \textit{jinak}\,. \end{cases}
$$

a) Jaká jsou jejich marginální rozdělení?

b) Jsou veličiny X a Y nezávislé?

c) Jak vypadá jejich korelační matice?

Řešení. a) Marginální hustoty jsou

$$
\begin{aligned}
f_X(x) &= \textstyle\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \\
&= \textstyle\int_0^{\infty} \frac{1}{2} e^{-x-\frac{y}{2}} dy = \frac{1}{2} e^{-x} \cdot [-2e^{-\frac{y}{2}}]_0^{\infty} = e^{-x} \text{ pro } x > 0 \text{ a } 0 \text{ jinak.} \\
f_Y(y) &= \textstyle\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \\
&= \textstyle\int_0^{\infty} \frac{1}{2} e^{-x-\frac{y}{2}} dx = \frac{1}{2} e^{-\frac{y}{2}} \cdot [-e^{-x}]_0^{\infty} = \frac{1}{2} e^{-\frac{y}{2}} \text{ pro } y > 0 \text{ a } 0 \text{ jinak.}
\end{aligned}
$$

b) Složky jsou nezávislé právě tehdy, když $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y), \forall x, y,$ což podle bodu 1. platí.

c) Z nezávislosti X, Y plyne okamžitě $\mathrm{cov}(X, Y) = 0 \Rightarrow \varrho(X, Y) = 0 \Rightarrow$

$$
\varrho_{(X,Y)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
$$

Cvičení 11.9. *Náhodný vektor má rovnoměrné rozdělení na trojúhelníku s vrcholy* $(0,0)$, $(1,0)$, $(1,1)$. *Popište a znázorněte distribuční funkce jeho složek (marginální rozdělení).*

Řešení. *1. postup:* Marginální hustoty jsou

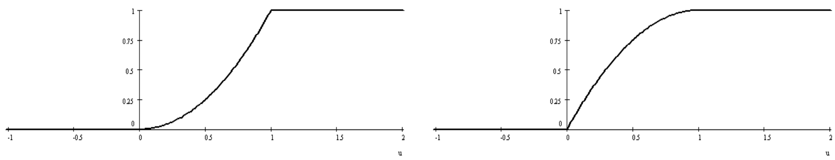
$$f_X(t) = \begin{cases} 2t\,, & 0 \le t \le 1\,, \\ 0 & \text{jinak,} \end{cases}$$

$$f_Y(t) = \begin{cases} 2(1-t)\,, & 0 \le t \le 1\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

distribuční funkce dostaneme jejich integrací:

$$F_X(t) = \int_{-\infty}^{u} f_X\,(t)\;\mathrm{d}t = \begin{cases} 0\,, & u < 0\,, \\ u^2\,, & 0 \le u \le 1\,, \\ 1\,, & u > 1\,, \end{cases}$$

$$F_Y(t) = \int_{-\infty}^{u} f_Y\,(t)\;\mathrm{d}t = \begin{cases} 0\,, & u < 0\,, \\ 2\,u - u^2\,, & 0 \le u \le 1\,, \\ 1\,, & u > 1\,. \end{cases}$$



Obrázek 5: Marginální distribuční funkce F_X (vlevo) a F_Y (vpravo).

2. postup: Distribuční funkce je podle definice dána poměrem obsahů ploch (vesměs se jedná o trojúhelníky nebo lichoběžníky, takže nepotřebujeme integrovat a vystačíme s geometrií ze základní školy); vždy je nutno dělit obsahem celého daného trojúhelníka, což je 1/2. Pro $0 \le u \le 1$ vychází

$$F_X\,(u) = \frac{\frac{u^2}{2}}{\frac{1}{2}} = u^2\,,$$

$$F_Y\,(u) = \frac{\frac{1}{2} - \frac{(1-u)^2}{2}}{\frac{1}{2}} = 2\,u - u^2\,.$$

Cvičení 11.10. *Náhodný vektor* (X,Y) *má sdruženou hustotu*

$$f_{X,Y}(x,y) = \begin{cases} y\,\mathrm{e}^{-y\,(x+1)}\,, & x > 0,\ y > 0\,, \\ 0\,, & \textit{jinak}\,. \end{cases}$$

Určete marginální rozdělení a jejich střední hodnoty.
Pomůcka:

$$\int t\,\mathrm{e}^{-t}\,\mathrm{d}t = -(t+1)\,\mathrm{e}^{-t} + c\,.$$

Řešení. Marginální hustotu veličiny X, resp. Y, dostaneme integrací hustoty $f_{X,Y}$ podle y, resp. x,

$$\begin{aligned} f_X(t) &= \int_0^{\infty} y\,\mathrm{e}^{-y\,(t+1)}\,\mathrm{d}y = 1/(t+1)^2\,, \\ f_Y(t) &= \int_0^{\infty} t\,\mathrm{e}^{-t\,(x+1)}\,\mathrm{d}x = \mathrm{e}^{-t}\,. \end{aligned}$$

Střední hodnota veličiny X je nekonečná

$$\mathrm{E}X = \int_0^{\infty} t/(t+1)^2\,\mathrm{d}t = [\ln(t+1) + 1/(t+1)]_{t=0}^{\infty} = \infty\,.$$

Střední hodnota veličiny Y je

$$\mathrm{E}Y = \int_0^{\infty} t\,\mathrm{e}^{-t}\,\mathrm{d}t = [-\mathrm{e}^{-t}]_{t=0}^{\infty} = 1\,.$$

Cvičení 11.11. *Náhodný vektor (X,Y) má rovnoměrné rozdělení na čtverci s rohy $[0,2]$, $[2,0]$, $[0,0]$ a $[2,2]$. Určete sdruženou distribuční funkci $F_{X,Y}$, sdruženou hustotu $f_{X,Y}$ a střední hodnotu náhodného vektoru $(X+3,2Y)$.*

Řešení. Protože (X,Y) má rovnoměrné rozdělení na čtverci o obsahu 4, sdružená hustota má tvar

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4}\,, & (x,y) \in [0,2]^2\,, \\ 0\,, & \text{jinak}\,. \end{cases}$$

Její integrací dostaneme sdruženou distribuční funkci

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_{X,Y}(u,v)\mathrm{d}v \right) \mathrm{d}u = \begin{cases} 0\,, & \min(x,y) < 0\,, \\ \frac{\min(x,2)\cdot\min(y,2)}{4}\,, & \min(x,y) \geq 0\,. \end{cases}$$

Střední hodnota náhodného vektoru $(X+3,2Y)$ je díky linearitě rovna $\mathrm{E}(X+3,2Y) = (\mathrm{E}X+3, 2\mathrm{E}Y)$. Střední hodnotu vypočteme pomocí marginálních hustot

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y = \begin{cases} \int_0^2 \frac{1}{4}\mathrm{d}y = \frac{1}{2}\,, & x \in [0,2]\,, \\ 0\,, & x \notin [0,2]\,, \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x = \begin{cases} \int_0^2 \frac{1}{4}\mathrm{d}x = \frac{1}{2}\,, & y \in [0,2]\,, \\ 0\,, & y \notin [0,2]\,. \end{cases}$$

Pak

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f_X(x)\ \mathrm{d}x = \int_0^2 \frac{x}{2}\ \mathrm{d}x = \left[\frac{x^2}{4}\right]_0^2 = 1,$$

$$\mathrm{E}Y = \int_{-\infty}^{\infty} y f_Y(y)\ \mathrm{d}y = \int_0^2 \frac{y}{2}\ \mathrm{d}y = \left[\frac{y^2}{4}\right]_0^2 = 1.$$

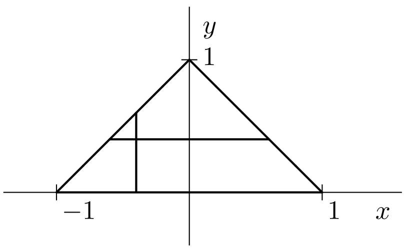
Tedy střední hodnota náhodného vektoru $(X+3,2Y)$ je rovna $(4,2)$.

Cvičení 11.12. *Náhodný vektor (X,Y) má rovnoměrné rozdělení v množině $A = \{(x,y) :\ y \geq 0,\ x+y \leq 1,\ y-x \leq 1\}$.*

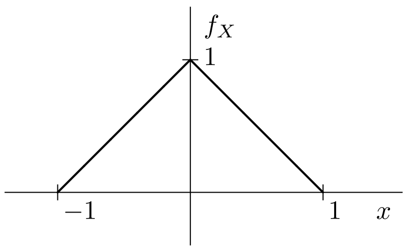
a) Vypočtěte kovarianci $\mathrm{cov}(X,Y)$.

b) Rozhodněte, zda jsou náhodné veličiny X a Y nezávislé.

Řešení. Náhodný vektor (X,Y) má rovnoměrné rozdělení, je tedy jeho sdružená hustota konstantní a je rovna převrácené hodnotě obsahu množiny A.



Obr. 1.



Obr. 2.

Sdružená hustota $f(x,y)$ je tedy dána vzorcem

$$f(x,y) = \begin{cases} 1\,, & (x,y) \in A\,, \\ 0\,, & (x,y) \notin A\,. \end{cases}$$

a) Jestliže použijeme symetrie množiny A vzhledem k ose y a sudosti funkce $f(x,y)$, dostaneme okamžitě $\mathrm{E}X = 0$ a $\mathrm{E}XY = 0$. Tedy kovariance $\mathrm{cov}(X,Y) = \mathrm{E}XY - \mathrm{E}X \cdot \mathrm{E}Y = 0$.

b) Náhodné veličiny nejsou nezávislé, neboť není sdružená hustota kladná na kartézském součinu intervalů (obdélníku) a nemůže tedy být součinem marginálních hustot.

Lze také místo výše uvedené úvahy provést následující výpočty:
Marginální hustotu f_X náhodné veličiny X vypočteme ze vzorce

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\,.$$

Ta bude nenulová pouze v intervalu $(-1,\,1)$ a výpočet musíme rozdělit pro kladné a záporné hodnoty proměnné x (viz obr. 1, svislý řez). Pak je

$$\begin{aligned} f_X(x) &= \int_0^{1+x} 1\,\mathrm{d}y = 1+x\,, \qquad x \in (-1,\,0)\,,\\ f_X(x) &= \int_0^{1-x} 1\,\mathrm{d}y = 1-x\,, \qquad x \in (0,\,1)\,. \end{aligned}$$

Obdobně vypočteme marginální hustotu f_Y náhodné veličiny Y ze vzorce

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\ \mathrm{d}x\,.$$

Ta bude nenulová pouze v intervalu $(0,1)$ a pro její hodnoty dostaneme (viz obr. 1, vodorovný řez)

$$f_Y(y) = \int_{y-1}^{1-y} 1\,\mathrm{d}x = 2\,(1-y)\,, \qquad y \in (0,\,1)\,.$$

Možnou nezávislost ověříme z rovnice

$$f(x,y) = f_X(x)\cdot f_Y(y) \Rightarrow 1 = (1+x)\cdot 2\,(1-y) \vee 1 = (1-x)\cdot 2\,(1-y)\,,$$

která neplatí na žádné části množiny A.

Pro kovarianci postupně vypočteme:

$$\begin{aligned} \mathrm{E}X &= \int_{-\infty}^{\infty} x\,f_X(x) = \int_{-1}^{0} x\,(1+x)\,\mathrm{d}x + \int_0^1 x\,(1-x)\,\mathrm{d}x = 0\,,\\ \mathrm{E}Y &= \int_{-\infty}^{\infty} y\,f_Y(y) = \int_0^1 2y(1-y)\,\mathrm{d}y = 2\int_0^1 y - y^2\,\mathrm{d}y = 2\left[\frac{y^2}{2} - \frac{y^3}{3}\right]_0^1 = \frac{1}{3}\,,\\ \mathrm{E}XY &= \int\!\!\int_{\mathbb{R}^2} x\,y\,f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int\!\!\int_A x\,y\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 y\left(\int_{y-1}^{1-y} x\,\mathrm{d}x\right)\mathrm{d}y = \\ &= \int_0^1 y\left[\frac{x^2}{2}\right]_{y-1}^{1-y}\mathrm{d}y = \int_0^1 0\,\mathrm{d}y = 0\,. \end{aligned}$$

Kovarianci vypočteme ze vzorce

$$\mathrm{cov}(X,Y) = \mathrm{E}XY - \mathrm{E}X\cdot\mathrm{E}Y = 0 - 0\cdot\frac{1}{3} = 0\,.$$

Cvičení 11.13. *Náhodný vektor (X,Y) má rovnoměrné rozdělení na množině $A = \{(u,v) \mid u^2+v^2 \le 4,\ u \ge 0,\ v \ge 0\}$. Určete*

a) kovarianci $\mathrm{cov}(X,Y)$;

b) zda jsou náhodné veličiny X a Y závislé či nezávislé.

Řešení. a) Sdružená hustota $f_{X,Y}$ je konstantní na množině A a její hodnota je rovna převrácené hodnotě obsahu množiny A. Je tedy

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\pi}\,, & (x,y) \in A\,,\\ 0\,, & (x,y) \notin A\,. \end{cases}$$

Potom je

$$\begin{aligned}
\mathrm{E}\,X &= \frac{1}{\pi}\iint_A x\mathrm{d}x\mathrm{d}y = \left|\begin{array}{ll} x=\varrho\cos\varphi, & 0<\varrho<2 \\ y=\varrho\sin\varphi, & 0\le\varphi\le\pi/2 \end{array}\right| = \\
&= \frac{1}{\pi}\int_0^{\pi/2}\left(\int_0^2 \varrho^2\cos\varphi\mathrm{d}\varrho\right)\mathrm{d}\varphi = \\
&= \frac{1}{\pi}\left[\frac{\varrho^3}{3}\right]_{\varrho=0}^2 [\sin\varphi]_{\varphi=0}^{\pi/2} = \frac{8}{3\pi} = \mathrm{E}\,Y\,; \qquad \text{(ze symetrie)} \\
\mathrm{E}\,XY &= \frac{1}{\pi}\iint_A xy\mathrm{d}x\mathrm{d}y = \left|\begin{array}{ll} x=\varrho\cos\varphi, & 0<\varrho<2 \\ y=\varrho\sin\varphi, & 0\le\varphi\le\pi/2 \end{array}\right| = \\
&= \frac{1}{\pi}\int_0^{\pi/2}\left(\int_0^2 \varrho^3\cos\varphi\sin\varphi\,\mathrm{d}\varrho\right)\mathrm{d}\varphi = \frac{1}{\pi}\left[\frac{\varrho^4}{4}\right]_{\varrho=0}^2\left[\frac{\sin^2\varphi}{2}\right]_{\varphi=0}^{\pi/2} = \frac{2}{\pi}\,.
\end{aligned}$$

Kovariance je rovna

$$\operatorname{cov}(X,Y) = \mathrm{E}\,XY - \mathrm{E}\,X\ \mathrm{E}\,Y = \frac{2}{\pi} - \frac{64}{9\pi^2} = \frac{1}{9\pi^2}(18\pi-64) \doteq -0.0839\,.$$

b) Náhodné veličiny X a Y nemohou být nezávislé, neboť je sdružená hustota $f_{X,Y}(x,y)$ kladná pouze v množině A, která není intervalem $[0,2]\times[0,2]$.

Alternativní postup bez transformací souřadnic:

Marginální hustoty:

$$\begin{aligned}
f_X(x) &= \int_0^{\sqrt{4-x^2}} \frac{1}{\pi}\mathrm{d}y = \frac{\sqrt{4-x^2}}{\pi}, \qquad x\in(0,\ 2),\ f_X(x)=0,\ x\notin(0,\ 2)\,, \\
f_Y(y) &= \int_0^{\sqrt{4-y^2}} \frac{1}{\pi}\mathrm{d}x = \frac{\sqrt{4-y^2}}{\pi}, \qquad y\in(0,\ 2),\ f_Y(y)=0,\ y\notin(0,\ 2)\,, \\
\mathrm{E}\,X &= \frac{1}{\pi}\int_0^2 x\sqrt{4-x^2}\mathrm{d}x = \frac{-1}{3\pi}\left[(4-x^2)^{3/2}\right]_{x=0}^2 = \frac{8}{3\pi} = \mathrm{E}\,Y \text{ (ze symetrie)}.
\end{aligned}$$

Zřejmě je

$$f_{X,Y}(x,y) = 0 < f_X(x)\cdot f_Y(y), \qquad (x,y)\in(0,2)^2\setminus A\,.$$

Cvičení 11.14. *V množině* $A=\{(x,y)\mid x^2+y^2\le 1,\ y\ge 0\}$ *volíme náhodně bod* (X,Y) *tak, že sdružená hustota je nepřímo úměrná vzdálenosti bodu* (X,Y) *od počátku.*

a) Určete sdruženou hustotu $f_{X,Y}$ *a pravděpodobnost* $P[X>Y]$.

b) Jsou náhodné veličiny X *a* Y *závislé či nezávislé?*

Řešení. a) Sdružená hustota $f_{X,Y}$ náhodného vektoru (X,Y) je dána vzorcem

$$f_{X,Y}(x,y) = \frac{a}{\sqrt{x^2+y^2}}\,,\ (x,y)\in A;\ f_{X,Y}(x,y)=0,\ (x,y)\notin A,$$

kde číslo a určíme z podmínky

$$\begin{aligned}
1 = \iint_A f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y &= \iint_A \frac{a}{\sqrt{x^2+y^2}}\mathrm{d}x\mathrm{d}y = \left|\begin{array}{ll} x=\varrho\cos\varphi, & 0<\varrho<1 \\ y=\varrho\sin\varphi, & 0\le\varphi\le\pi \end{array}\right| = \\
&= a\int_0^1\left(\int_0^\pi \mathrm{d}\varphi\right)\mathrm{d}\varrho = a\pi \Rightarrow a = \frac{1}{\pi}\,.
\end{aligned}$$

Je tedy

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{\pi\sqrt{x^2+y^2}}\,,\ (x,y)\in A;\ f_{X,Y}(x,y)=0,\ (x,y)\notin A\,, \\
P[X>Y] &= \iint_{x>y} \frac{1}{\pi\sqrt{x^2+y^2}}\mathrm{d}x\mathrm{d}y = \left|\begin{array}{ll} x=\varrho\cos\varphi, & 0<\varrho<1 \\ y=\varrho\sin\varphi, & 0\le\varphi\le\pi/4 \end{array}\right| = \\
&= \frac{1}{\pi}\int_0^1\left(\int_0^{\pi/4}\mathrm{d}\varphi\right)\mathrm{d}\varrho = \frac{1}{\pi}\frac{\pi}{4} = \frac{1}{4}\,.
\end{aligned}$$

b) Náhodné veličiny X a Y nejsou nezávislé, neboť je sdružená hustota $f_{X,Y}$ kladná na množině A (půlkruh), ale součin marginálních hustot f_X a f_Y je kladný na celém intervalu $(-1, 1) \times (0, 1)$, tudíž se nemohou rovnat.

Cvičení 11.15. *Pro náhodné veličiny X a Y platí*

$$\mathrm{E}X = -1\,, \qquad \mathrm{D}X = 3\,, \qquad \mathrm{E}Y = 5\,, \qquad \mathrm{D}Y = 4\,, \qquad \mathrm{cov}(X, Y) = -2.$$

Pro náhodné veličiny $U = 3\,X + 4\,Y - 1$ a $V = -2\,X + 2\,Y + 3$ určete kovarianci $\mathrm{cov}(U, V)$.

Řešení. Kovarianci vypočteme pomocí vzorce

$$\mathrm{cov}(U, V) = \mathrm{E}UV - \mathrm{E}U\,\mathrm{E}V\,.$$

Z linearity střední hodnoty dostaneme

$$\begin{aligned} \mathrm{E}U &= 2\,\mathrm{E}X + 4\,\mathrm{E}Y - 1 = -3 + 4 \cdot 5 - 1 = 16\,, \\ \mathrm{E}V &= -2\,\mathrm{E}X + 2\,\mathrm{E}Y + 3 = 2 + 10 + 3 = 15\,. \end{aligned}$$

Po roznásobení získáme vyjádření pro

$$UV = (3\,X + 4\,Y - 1)\,(-2\,X + 2\,Y + 3) = -6\,X^2 + 8\,Y^2 - 2\,X \cdot Y + 11\,X + 10\,Y - 3\,.$$

Ze vzorce pro rozptyl vypočteme

$$\mathrm{E}X^2 = \mathrm{D}X + (\mathrm{E}X)^2 = 3 + (-1)^2 = 4\,, \qquad \mathrm{E}Y^2 = \mathrm{D}Y + (\mathrm{E}Y)^2 = 4 + 5^2 = 29\,.$$

Ze vzorce pro kovarianci dostaneme

$$\mathrm{E}XY = \mathrm{cov}(X, Y) + \mathrm{E}X \cdot \mathrm{E}Y = -2 + (-1) \cdot 5 = -7\,.$$

Je pak

$$\mathrm{E}UV = -6 \cdot 4 + 8 \cdot 29 - 2 \cdot (-7) + 11(-1) + 10 \cdot 5 - 3 = 258\,.$$

Po dosazení dostaneme kovarianci

$$\mathrm{cov}(U, V) = \mathrm{E}UV - \mathrm{E}U \cdot \mathrm{E}V = 258 - 16 \cdot 15 = 18\,.$$

Cvičení 11.16. *Náhodné veličiny X, Y vyhovují vztahu*

$$\alpha\,X - \beta\,Y = \gamma\,,$$

kde α, β a γ jsou konstanty. Určete korelaci $\varrho(X, Y)$ a poměr směrodatných odchylek σ_X/σ_Y.

Řešení. Předpokládejme, že $\alpha \neq 0$, $\beta \neq 0$ a $\mathrm{E}X$ a $\mathrm{D}X$ existují. Potom

$$\begin{aligned} Y &= \frac{\alpha\,X - \gamma}{\beta}\,, \\ \mathrm{E}Y &= \frac{\alpha\,\mathrm{E}X - \gamma}{\beta}\,, \\ \sigma_Y^2 &= \left(\frac{\alpha}{\beta}\right)^2 \sigma_X^2 \end{aligned}$$

a pro korelaci $\varrho(X, Y)$ a poměr směrodatných odchylek dostáváme

$$\begin{aligned} \varrho(X, Y) &= \frac{\mathrm{E}XY - \mathrm{E}X\,\mathrm{E}Y}{\sigma_X\,\sigma_Y} = \frac{\frac{\alpha\,\mathrm{E}X^2 - \gamma\,\mathrm{E}X}{\beta} - \frac{\alpha\,(\mathrm{E}X)^2 - \gamma\,\mathrm{E}X}{\beta}}{|\frac{\alpha}{\beta}|\,\sigma_X^2} = \mathrm{sign}\,\frac{\alpha}{\beta}\,, \\ \frac{\sigma_Y}{\sigma_X} &= \left|\frac{\alpha}{\beta}\right|\,. \end{aligned}$$

Cvičení 11.17. *Náhodný vektor má kovarianční matici*

$$\begin{pmatrix} 4 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Najděte jeho korelační matici.

Řešení. Každý řádek i každý sloupec je potřeba vydělit příslušnou směrodatnou odchylkou, kterou nalezneme jako odmocninu z prvku na diagonále (=rozptylu), tj. 1. řádek i sloupec vydělíme 2, 2. řádek i sloupec vydělíme $\sqrt{2}$ a dostaneme

$$\begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & 1 \end{pmatrix}.$$

Cvičení 11.18. *Náhodný vektor (X, Y) má kovarianční matici $\Sigma_{(X,Y)} = \begin{pmatrix} 4 & -3 \\ -3 & 5 \end{pmatrix}$. Určete*

a) korelační matici $\varrho_{(X,Y)}$ náhodného vektoru (X, Y);

b) korelační matici $\varrho_{(U,V)}$ náhodného vektoru (U, V), kde $U = 2X + 1$ a $V = -3Y$.

Řešení. a) Z kovarianční matice dostaneme, že
$\mathrm{D}X = \mathrm{cov}(X, X) = 4,\ \mathrm{D}Y = \mathrm{cov}(Y, Y) = 5 \Rightarrow$

$$\varrho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{D}X\,\mathrm{D}Y}} = \frac{-3}{\sqrt{4 \cdot 5}} \doteq -0.6708\,.$$

Korelační matice náhodného vektoru (X, Y) je tedy

$$\varrho_{(X,Y)} \doteq \begin{pmatrix} 1 & -0.6708 \\ -0.6708 & 1 \end{pmatrix}.$$

b) Z vlastností rozptylu a kovariance dostaneme

$$\begin{aligned} \mathrm{cov}(U, U) &= \mathrm{D}U = \mathrm{D}(2X + 1) = 4\,\mathrm{D}X = 16\,, \\ \mathrm{cov}(V, V) &= \mathrm{D}V = \mathrm{D}(-3Y) = 9\,\mathrm{D}Y = 45\,, \\ \mathrm{cov}(U, V) &= \mathrm{cov}(2X + 1, -3Y) = 2 \cdot (-3) \cdot \mathrm{cov}(X, Y) = (-6) \cdot (-3) = 18\,, \\ \varrho(U, V) &= -\varrho(X, Y) \doteq 0.6708\,. \end{aligned}$$

Kovarianční a korelační matice náhodného vektoru (U, V) jsou

$$\Sigma_{(U,V)} = \begin{pmatrix} 16 & 18 \\ 18 & 45 \end{pmatrix}, \qquad \varrho_{(U,V)} \doteq \begin{pmatrix} 1 & 0.6708 \\ 0.6708 & 1 \end{pmatrix}.$$

Cvičení 11.19. *Náhodná veličina X má spojité rovnoměrné rozdělení v intervalu $(0,\ 2)$. Pro náhodnou veličinu $Y = X^2$ vypočtěte korelaci $\varrho(X, Y)$.*

Řešení. Hustota náhodné veličiny X je dána vzorcem

$$f_X(u) = \begin{cases} \frac{1}{2}, & u \in (0,\ 2), \\ 0, & u \notin (0,\ 2). \end{cases}$$

Potřebujeme určit $\mathrm{E}X$, $\mathrm{E}Y = \mathrm{E}X^2$, $\mathrm{E}(XY) = \mathrm{E}X^3$ a $\mathrm{E}Y^2 = \mathrm{E}X^4$; najdeme obecně pro $n \in \mathbb{N}$:

$$\begin{aligned} \mathrm{E}X^n &= \int_0^2 \frac{1}{2}\,u^n\,\mathrm{d}u = \left[\frac{u^{n+1}}{2(n+1)}\right]_0^2 = \frac{2^n}{n+1}\,, \\ \mathrm{E}X &= 1\,, \\ \mathrm{E}Y &= \mathrm{E}(X^2) = \frac{4}{3} \doteq 1.33\,, \\ \mathrm{E}XY &= \mathrm{E}(X^3) = \frac{8}{4} = 2\,, \\ \mathrm{E}(Y^2) &= \mathrm{E}(X^4) = \frac{16}{5} = 3.2\,, \\ \mathrm{D}X &= \mathrm{E}(X^2) - (\mathrm{E}X)^2 = \frac{4}{3} - 1 = \frac{1}{3} \doteq 0.33\,, \\ \mathrm{D}Y &= \mathrm{E}(Y^2) - (\mathrm{E}Y)^2 = \frac{16}{5} - \frac{16}{9} = \frac{64}{45} \doteq 1.42\,. \end{aligned}$$

Korelaci dostaneme dosazením do vzorce

$$\varrho(X, Y) = \frac{\mathrm{E}XY - \mathrm{E}X \cdot \mathrm{E}Y}{\sqrt{\mathrm{D}X\,\mathrm{D}Y}} = \frac{2 - 1 \cdot \frac{4}{3}}{\sqrt{\frac{1}{3} \cdot \frac{64}{45}}} = \frac{2 \cdot 3\sqrt{15}}{3 \cdot 8} = \frac{\sqrt{15}}{4} \doteq 0.9682\,.$$

12 Čebyševova nerovnost, centrální limitní věta

Cvičení 12.1. *Hodíme 420 krát pravidelnou šestistěnnou hrací kostkou a výsledky hodů sčítáme. Pomocí centrální limitní věty odhadněte pravděpodobnost, že součet bude ležet mezi čísly 1400 a 1550.*

Řešení. Pro $i = 1, \ldots, n$ (kde $n = 420$) si zavedeme diskrétní veličiny X_i = hodnota, která padne na kostce při i-tém hodu. Veličiny X_i považujeme za nezávislé. Střední hodnotu a rozptyl už máme spočítánu z předchozího příkladu:

$$EX_i = 3.5\,, \qquad DX_i = \frac{35}{12}\,.$$

Součet hodnot se vyjádří jako

$$X = \sum_{i=1}^{n} X_i\,.$$

Máme teď určit $P(1400 \le X \le 1550)$, což uděláme za pomoci centrální limitní věty použité na normovanou veličinu $\mathrm{norm}(X) = \frac{X - \mathrm{E}X}{\sqrt{\mathrm{D}X}}$. K tomu potřebujeme znát:

$$\begin{aligned} EX &= \sum_{i=1}^{n} \mathrm{E}X_i = 420 \cdot 3.5 = 1470\,, \\ \mathrm{D}X &= \sum_{i=1}^{n} \mathrm{D}X_i = 420 \cdot \frac{35}{12} = 1225 = (35)^2\,. \end{aligned}$$

Tedy

$$\mathrm{norm}(X) = \frac{X - \mathrm{E}X}{\sqrt{\mathrm{D}X}} = \frac{X - 1470}{35}$$

a můžeme psát (pomocí úprav nerovností):

$$\begin{aligned} P(1400 \le X \le 1550) &= P\Big(\frac{1400 - 1470}{35} \le \frac{X - 1470}{35} \le \frac{1550 - 1470}{35}\Big) = \\ &= P\Big(-2 \le \mathrm{norm}(X) \le \frac{16}{7}\Big) \doteq \Phi\Big(\frac{16}{7}\Big) - \Phi(-2) \doteq \\ &\doteq \Phi(2.2857) - 1 + \Phi(2) \doteq 0.98886 + 0.97725 - 1 = 0.96611\,. \end{aligned}$$

Cvičení 12.2. *Výška mužů (určitého věku) je náhodná veličina o střední hodnotě* 180 *cm a směrodatnou odchylkou* 10 *cm. Určete pravděpodobnost, že průměrná výška* $n = 20$ *mužů bude v intervalu* 175 *cm a* 185 *cm*

a) pomocí centrální limitní věty,

b) pomocí Čebyševovy nerovnosti.

Řešení. Výšku i-tého může (pro $i = 1, \ldots, 20$) si označíme jako X_i. Veličiny X_i považujeme za nezávislé, se střední hodnotou $\mathrm{E}X_i = 180$ cm a směrodatnou odchylkou $\sigma_i = \sqrt{DX_i} = 10$ cm.

Průměrná výška se pak vyjádří jako

$$\overline{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} X_i\,.$$

Zajímá nás $P(175 \le \overline{\boldsymbol{X}} \le 185)$. Budeme potřebovat:

$$\mathrm{E}\overline{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}X_i = 180$$

$$\mathrm{D}\overline{\boldsymbol{X}} = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{D}X_i = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = \frac{10^2}{n} = \frac{100}{20} = 5\,.$$

Takže

$$\mathrm{norm}(\overline{\boldsymbol{X}}) = \frac{\overline{\boldsymbol{X}} - \mathrm{E}\overline{\boldsymbol{X}}}{\sqrt{\mathrm{D}\overline{\boldsymbol{X}}}} = \frac{\overline{\boldsymbol{X}} - 180}{\sqrt{5}}\,.$$

a) Odhad pomocí centrální limitní věty - úpravami nerovností dostaneme:

$$P(175 \le \overline{\boldsymbol{X}} \le 185) = P\Big(\frac{175 - 180}{\sqrt{5}} \le \frac{\overline{\boldsymbol{X}} - 180}{\sqrt{5}} \le \frac{185 - 180}{\sqrt{5}}\Big) = P\Big(-\sqrt{5} \le \mathrm{norm}(\overline{\boldsymbol{X}}) \le \sqrt{5}\Big) \doteq$$

$$\doteq \Phi\Big(\sqrt{5}\Big) - \Phi\Big(-\sqrt{5}\Big) = 2 \cdot \Phi\Big(\sqrt{5}\Big) - 1 \doteq 2 \cdot \Phi\Big(2.2361\Big) - 1 \doteq 2 \cdot 0.98733 - 1 = 0.97466\,.$$

Tedy asi 97.5%.

b) Odhad pomocí Čebyševovy nerovnosti - použijeme tvar, který už máme:

$$P(175 \le \overline{\boldsymbol{X}} \le 185) = P\Big(\big|\mathrm{norm}(\overline{\boldsymbol{X}})\big| \le \sqrt{5}\Big) \ge 1 - \frac{1}{(\sqrt{5})^2} = \frac{4}{5} = 0.8$$

Dostáváme tedy sice spodní odhad 80%, zato přesný.

Cvičení 12.3. *Hmotnost jedné součástky v kilogramech je náhodná veličina s rozdělením o střední hodnotě 5 kg a rozptylu 9 kg^2. Každá součástka je zabalena do krabice o hmotnosti 2 kg. Určete pravděpodobnost, že auto, naložené $n = 150$ takovými (plnými) krabicemi bude přetížené, je-li maximální možné zatížení auta 1100 kg.*

Řešení. Pro $i = 1, \ldots, 150$ si zavedeme veličiny X_i = hmotnost součástky zabalené v krabici. Veličiny X_i považujeme za nezávislé s rozdělením se střední hodnotou $\mathrm{E}X_i = 5 + 2 = 7$ kg a rozptylem $\mathrm{D}X_i = 9$ kg^2.

Hmotnost nákladu auta pak bude

$$X = \sum_{i=1}^{n} X_i$$

což je opět veličina s rozdělením takovým, že

$$\mathrm{E}X = \sum_{i=1}^{n} \mathrm{E}X_i = n \cdot 7 = 1050$$

$$\mathrm{D}X = \sum_{i=1}^{n} \mathrm{D}X_i = n \cdot 9 = 1350$$

$$\sqrt{\mathrm{D}X} = \sqrt{1350} = 15\sqrt{6} \doteq 36.742\ .$$

Pravděpodobnost, že auto bude přetížené je tedy $P(X > 1100)$. Určíme ji za pomoci centrální limitní věty použité na normovanou veličinu

$$\mathrm{norm}(X) = \frac{X - E(X)}{\sqrt{D(X)}} = \frac{X - 1050}{15\sqrt{6}}$$

Pomocí úprav nerovností můžeme psát:

$$P(X > 1100) = P\Big(\frac{X - 1050}{15\sqrt{6}} > \frac{1100 - 1050}{15\sqrt{6}}\Big) = P\Big(\mathrm{norm}(X) > \frac{10}{3\sqrt{6}}\Big) \doteq$$

$$\doteq 1 - \Phi\Big(\frac{10\sqrt{6}}{3 \cdot 6}\Big) = 1 - \Phi\Big(\frac{5\sqrt{6}}{9}\Big) \doteq 1 - \Phi(1.3608) \doteq 1 - 0.91321 = 0.08679\ .$$

Pravděpodobnost, že auto bude přetížené, tedy bude asi 8,68%.

Cvičení 12.4. *Ryby mohou si vybrat ze 2 cest, z nichž jedna je správná (vede k potravě). Každá ryba nezávisle pozná správnou cestu s pravděpodobností $q = 0.6$. Jaká je pravděpodobnost, že „většinové hlasování“ v hejnu n ryb vybere správnou cestu?*

Řešení. Rozhodnutí jednotlivých ryb popisují nezávislé náhodné veličiny X_j, $j = 1, \ldots, n$ s alternativním rozdělením s parametrem $q = 0.6$. (Správnou cestu vyhodnocujeme jako 1, špatnou 0.) Z vlastností alternativního rozdělení

$$\mathrm{E}X_j = q\,, \qquad \mathrm{D}X_j = q\,(1 - q)\,.$$

Pro výběrový průměr

$$\mathrm{E}\overline{\boldsymbol{X}} = q\,, \qquad \mathrm{D}\overline{\boldsymbol{X}} = \frac{q\,(1 - q)}{n}\,,$$

jeho rozdělení pro velká n můžeme podle centrální limitní věty přibližně nahradit normálním rozdělením se stejnými parametry, tj. $\mathrm{N}\left(q, \frac{q\,(1-q)}{n}\right)$. Odchylku střední hodnoty od 50 %, $0.5 - q = 0.5 - 0.6 = -0.1$ budeme měřit směrodatnou odchylkou výběrového průměru

$$\sigma_{\overline{\boldsymbol{X}}} = \sqrt{\frac{q\,(1 - q)}{n}} = \sqrt{\frac{0.6 \cdot (1 - 0.6)}{n}} \doteq \frac{0.49}{\sqrt{n}}\,,$$

poměr $\frac{-0.1}{0.49}\sqrt{n}$ bude argumentem distribuční funkce normovaného normálního rozdělení. Pravděpodobnost, že se hejno rozhodne chybně, je

$$\begin{aligned} P\left[\overline{\boldsymbol{X}} \le 0.5\right] &\doteq F_{\mathrm{N}\left(q,\frac{q\,(1-q)}{n}\right)}(0.5-q) = \Phi\left(\frac{0.5-q}{\sigma_{\overline{\boldsymbol{X}}}}\right) \doteq \\ &\doteq \Phi\left(\frac{-0.1}{0.49}\sqrt{n}\right) \doteq \Phi\left(-0.204\,08\sqrt{n}\right). \end{aligned}$$

Pravděpodobnost správného rozhodnutí hejna je k ní doplňková,

$$\begin{aligned} P\left[\overline{\boldsymbol{X}} > 0.5\right] &\doteq 1 - F_{\mathrm{N}\left(q,\frac{q\,(1-q)}{n}\right)}(0.5-q) = 1 - \Phi\left(\frac{0.5-q}{\sigma_{\overline{\boldsymbol{X}}}}\right) = \Phi\left(\frac{q-0.5}{\sigma_{\overline{\boldsymbol{X}}}}\right) \doteq \\ &\doteq \Phi\left(\frac{0.1}{0.49}\sqrt{n}\right) \doteq \Phi\left(0.204\sqrt{n}\right). \end{aligned}$$

Číselné hodnoty pro několik hodnot n udává tabulka:

n	$0.204\sqrt{n}$	$P\left[\overline{\boldsymbol{X}} > 0.5\right]$
10	0.645	0.74
100	2.04	0.98
1000	6.45	$1 - 6 \cdot 10^{-11}$

Cvičení 12.5. *Ve vzorku je* 1 mg *uhlíku, tj. asi* $6 \cdot 10^{23} \cdot 10^{-3}/12 = 5 \cdot 10^{19}$ *atomů. Z nich je přibližně* $1/10^{12}$*, tj. asi* $5 \cdot 10^{7}$*, atomů radioaktivního izotopu* $C14$*. Určete symetrický* 95 %*-ní intervalový odhad počtu atomů, které se rozpadnou za* 1 *rok, tj. za* 1/5730 *poločasu rozpadu. Co o tom říká Čebyševova nerovnost?*

Řešení. Odhadujeme náhodnou veličinu X s rozdělením $\mathrm{Bi}\,(n,p)$, $n = 5 \cdot 10^7$, $p = 1 - 1/2^{1/5730} \doteq 1.2 \cdot 10^{-4}$ (=pravděpodobnost, že se atom v daném čase rozpadne),

$$\begin{aligned} \mathrm{E}X &= n\,p \doteq 6048\,, \\ \mathrm{D}X &= n\,p\,(1-p) \doteq 6047\,, \\ \sigma_X &= \sqrt{n\,p\,(1-p)} \doteq 78\,. \end{aligned}$$

Při aproximaci normálním rozdělením vyjdou meze $\mathrm{E}X \pm \sigma_X\,\Phi^{-1}(0.975) \doteq 6048 \pm 78 \cdot 1.96 \doteq 6048 \pm 153$, interval přibližně $\langle 5895, 6201\rangle$, relativní chyba zhruba $153/6048 \doteq 2.5\,\%$.

Exaktní výpočet z binomického rozdělení by byl pracný a vedl by k velmi podobným výsledkům.

Čebyševova nerovnost nezohledňuje znalost rozdělení (přibližně normální) a vede na intervalový odhad s tolerancí ε splňující nerovnost

$$\begin{aligned} \frac{\mathrm{D}X}{\varepsilon^2} &\le 0.05\,, \\ \varepsilon &\ge \sqrt{\frac{\mathrm{D}X}{0.05}} = \frac{\sigma_X}{\sqrt{0.05}} \doteq \frac{78}{\sqrt{0.05}} \doteq 349\,, \end{aligned}$$

meze $6048 \pm 349 = 6397$, interval přibližně $\langle 5699, 6397\rangle$, relativní chyba zhruba $349/6048 = 5.7\,\%$.

Cvičení 12.6. *Alice nabídla Bobovi sázku* 1 : 1000*, že nedokáže z* 500 *hodů mincí aspoň v* 60 % *hodit líc. Bob váhá, proto Alice navíc nabízí, že Bob má* 10 *pokusů (po* 500 *hodech) a stačí, když aspoň v jednom z nich uspěje. Kurs zůstává* 1 : 1000*. Má Bob sázku přijmout?*

Řešení. Je-li mince regulérní a $n = 500$ je počet pokusů, pak výběrový průměr má podle centrální limitní věty rozdělení přibližně $N\left(\frac{1}{2}, \frac{1}{4\,n}\right)$. Počet líců je $n\times$ větší, má tedy rozdělení přibližně $N\left(\frac{n}{2}, \frac{n}{4}\right)$. Pravděpodobnost, že Bob v jednom kole dosáhne o 10 % víc než polovinu líců, je

$$1 - \Phi\left(\frac{0.1\,n}{\sqrt{\frac{n}{4}}}\right) = 1 - \Phi\left(0.2\sqrt{n}\right) \doteq 1 - \Phi\,(4.472) \doteq 3.9 \times 10^{-6}\,.$$

Při 10 opakovaných pokusech se pravděpodobnost úspěchu zvýší méně než 10×, sázka zůstává pro Boba velmi nevýhodná.

Cvičení 12.8. *Letecká společnost prodává letenky a chce co nejvíce utržit. Letadlo má* 216 *míst, ale ví se, že zhruba* 5% *lidí se k odletu nedostaví. Kolik může společnost prodávat letenek na jeden let, chce-li držet pravděpodobnost, že k odletu se dostaví více než* 216 *lidí pod hladinou* $\alpha = 0.1$*?*

Řešení. Označíme n počet prodaných letenek, X_i, $i \in \{1, 2, ..., n\}$ náhodnou veličinu nabývající 0, pokud i-tý cestující nepřijde a 1, pokud přijde k odletu. Tedy X_i je náhodná veličina s alternativním rozdělením s parametrem $p = 0.95$. Čili platí $\mathrm{E}X_i = p$, $\mathrm{D}X_i = p(1-p)$ pro každé $i \in \{1, 2, ..., n\}$. Označme $X = \sum_{i=1}^{n} X_i$. Pro dostatečně velká n bude dle CLV platit, že veličina

$$\mathrm{norm}(X) = \frac{X - np}{\sqrt{np(1-p)}}$$

má téměř normované normální rozdělení, neboli $\mathrm{norm}(X) \sim N(0,1)$. Proto je možné pro každé $u \in \mathbb{R}$ psát
$P(\mathrm{norm}(X) < u) = F_{\mathrm{norm}(X)}(u) \doteq \Phi(u)$.
Chceme zjistit, jaké může být n, proto řešíme nerovnici

$$P\left(X > 216\right) < 0.1$$
$$P\left(\frac{X - np}{\sqrt{np(1-p)}} > \frac{216 - np}{\sqrt{np(1-p)}}\right) < 0.1$$

po dosazení $p = 0.95$ a U_n

$$P\left(\mathrm{norm}(X) > \frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) < 0.1$$

pronásobíme -1 a přičteme 1.

$$-P\left(\mathrm{norm}(X) > \frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) > -0.1$$
$$1 - P\left(\mathrm{norm}(X) > \frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) > 1 - 0.1$$
$$P\left(\mathrm{norm}(X) \le \frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) > 0.9$$
$$\Phi\left(\frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) > 0.9$$

Složíme s kvantilovou funkcí normovaného normálního rozdělení, tj. $q = \Phi^{-1}$. Ta je rostoucí, proto se nemění nerovnost.

$$\frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}} > q(0.9)$$

Tedy dostáváme

$$216 > q(0.9)\frac{\sqrt{19n}}{20} + \frac{19}{20}n,$$

což lze řešit buďto zkusmo (výraz napravo s každým n roste, takže to není tak těžké) nebo umocněním. Výraz nalevo v nerovnosti

$$216 - \frac{19}{20}n > q(0.9)\frac{\sqrt{19n}}{20}$$

bude kladný pouze pro $n < \frac{4320}{19} \doteq 227.36$ (tj. pro $n \le 227$), proto lze při umocnění nezměnit znaménko nerovnosti jen pro $n \le 227$. Zároveň pravá strana nerovnosti je vždy větší než 0, proto nerovnost pro

$n > 227$ platit nemůže. Pro $n \leq$ tedy po umocnění dostáváme

$$
\begin{aligned}
(216)^2 - \frac{2052}{5}n + \frac{361}{400}n^2 &> q^2(0.9)\frac{19n}{400} \\
\frac{361}{400}n^2 - \left(\frac{2052}{5} + q^2(0.9)\frac{19}{400}\right)n + (216)^2 &> 0 \\
D &= 64.083808229 \\
n_{1,2} &= \frac{410.478 \pm 8.005}{\frac{361}{200}} \\
n_1 &\doteq 231.86 \\
n_2 &\doteq 222.98.
\end{aligned}
$$

První kořen nevyhovuje podmínce $n \leq 227$, tedy lze říci, že nerovnici splňují pouze všechna $n \leq 222$. Společnost tedy může prodávat k letu 222 letenek a pravděpodobnost, že přijde příliš mnoho lidí, bude stále menší než 0.1.

Poznámka. Druhý kořen vyšel téměř rovný 223, proto naše řešení dává menší pravděpodobnost, že cestujících k odletu přijde příliš, než bylo požadovaných 0.1. Sledujme, kolik vychází CLV-odhady na tuto pravděpodobnost, pokud společnost prodá letenek postupně 222, 223 a 224. Dosadíme-li tato n do vzorce

$$
P\left(U_n > \frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) = 1 - P\left(U_n \leq \frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right) = 1 - \Phi\left(\frac{216 - \frac{19}{20}n}{\frac{\sqrt{19n}}{20}}\right),
$$

dostáváme postupně pravděpodobnosti

$$
\begin{aligned}
p_{222} &= 1 - \Phi\left(\frac{5.1}{3.247306}\right) = 0.058, \\
p_{223} &= 1 - \Phi\left(\frac{4.15}{3.254612}\right) = 0.1003, \\
p_{224} &= 1 - \Phi\left(\frac{3.2}{3.261901}\right) = 0.1635.
\end{aligned}
$$

Všimněte si, že přidáním jediné letenky vzroste pravděpodobnost z 6% na 10% a z 10% na 16%!!!

Cvičení 12.9. *Osoby A, B a C kandidují na prezidenta. Odhadněte, kolika lidí se musí průzkum týkat, aby pro každého kandidáta byla alespoň* 95% *pravděpodobnost, že statistická chyba volební preference nepřesáhne* 1%. *Předpokládejte, že volební preference potenciálního voliče nezávisí na preferencích ostatních voličů.*

Řešení. Budeme odhadovat pro každého kandidáta zvlášť. Preference pro kandidáta X má alternativní rozdělení Alt(p). Z centrální limitní věty dostaneme

$$
\frac{0.01}{\sqrt{p_X(1-p)}}\sqrt{n} \geq \Phi^{-1}(0.975)\,.
$$

Rozptyl nabývá maximální hodnoty pro $p = 1/2$. Pro $t = \Phi^{-1}(0.975) \doteq 2$ dostáváme (hrubý) odhad $n \geq 10000$.

Cvičení 12.10. *Házíme hrací kostkou a počítáme výskyt šestek. Určete, kolik musíme provést hodů, aby se relativní výskyt šestek (poměr počtu šestek ku počtu hodů) lišil od* $\frac{1}{6}$ *nejvýše o* $\varepsilon = 0.05$ *s pravděpodobností* $P^* = 0.95$*, a to*

a) výpočtem pomocí aproximace normálním rozdělením,

b) odhadem z Čebyševovy nerovnosti.

Řešení. Počet šestek v provedené sérii hodů je náhodná veličina X, která má binomické rozdělení Bin(n, p), kde n je počet hodů a $p = \frac{1}{6}$ je pravděpodobnost výskytu šestky při jednotlivých hodech. Pro parametry rozdělení dostaneme

$$
\mathrm{E}X = n\,p = \frac{n}{6}\,, \qquad \mathrm{D}X = n\,p\,(1-p) = \frac{5\,n}{36}\,.
$$

a) K výpočtu použijeme skutečnosti, že pro větší počet hodů má náhodná veličina X přibližně normální rozdělení $\mathrm{N}(\mu,\,\sigma^2)$, kde $\mu = \frac{n}{6}$ a $\sigma^2 = \frac{5\,n}{36}$. Je-li X počet šestek, pak podmínka z úlohy znamená

$$\left|\frac{X}{n} - \frac{1}{6}\right| < 0.05 \Rightarrow \left|X - \frac{n}{6}\right| < n \cdot 0.05\,.$$

Označíme-li F_X distribuční funkci náhodné veličiny X, dostaneme, že

$$\begin{aligned} 0.95 &= P(|X - \mathrm{E}(X)| < 0.05) = F_X(n\,(p + 0.05)) - F_X(n\,(p - 0.05)) = \\ &= \Phi\left(\frac{n \cdot 0.05}{\sqrt{\frac{5\,n}{36}}}\right) - \Phi\left(\frac{-n \cdot 0.05}{\sqrt{\frac{5\,n}{36}}}\right) = 2\,\Phi\left(\frac{0.3\sqrt{n}}{\sqrt{5}}\right) - 1 \Rightarrow \left(\frac{0.3\sqrt{n}}{\sqrt{5}}\right) = 0.975\,. \end{aligned}$$

Z tabulky kvantilů normálního rozdělení dostaneme, že

$$\frac{0.3\sqrt{n}}{\sqrt{5}} = \Phi^{-1}(0.975) \doteq 1.96 \Rightarrow \sqrt{n} = \frac{1.96 \cdot 2.236}{0.3} \doteq 14.609 \Rightarrow n \doteq 213.42\,.$$

Musíme tedy provést alespoň 214 hodů, abychom s požadovanou pravděpodobností dosáhli zadané tolerance pro relativní četnost šestek.

b) Odhad pro počet hodů můžeme získat i z Čebyševovy nerovnosti. Je pak

$$P\left(\left|\frac{X}{n} - \frac{1}{6}\right| < 0.05\right) \geq \frac{n\,p\,(1-p)}{n^2 \cdot 0.05^2} \geq 0.95 \Rightarrow \frac{p\,(1-p)}{n \cdot 0.05^2} < 0.05 \Rightarrow$$

$$n > \frac{5}{36 \cdot 0.05 \cdot 0.05^2} \doteq 1111.11\,.$$

K získání požadované tolerance výskytu šestek musíme provést 1112 hodů.

Cvičení 12.11. *Házíme 100-krát pravidelnou mincí a náhodná veličina X je počet rubů.*

a) Určete (přibližně) číslo ε tak, aby $P[|X - \mathrm{E}X| < \varepsilon] = 0.9$.

b) Pomocí Čebyševovy nerovnosti odhadněte pravděpodobnost $P[|X - \mathrm{E}X| > 10]$.

Řešení. Náhodná veličina X má binomické rozdělení Bi(100; 0.5). Je tedy $\mathrm{E}X = 100 \cdot 0.5 = 50$ a $\mathrm{D}X = 100 \cdot 0.5 \cdot (1 - 0.5) = 25$. Z centrální limitní věty vyplývá, že můžeme předpokládat pro náhodnou veličinu X přibližně normální rozdělení N(50; 25).

a) Číslo ε určíme z rovnice

$$\begin{aligned} 0.9 &= P[|X - \mathrm{E}X| < \varepsilon] = P[50 - \varepsilon < X < 50 + \varepsilon] = F_X(50 + \varepsilon) - F_X(50 - \varepsilon) = \\ 0.9 &= \Phi\left(\frac{\varepsilon}{5}\right) - \Phi\left(\frac{-\varepsilon}{5}\right) = 2\,\Phi\left(\frac{\varepsilon}{5}\right) - 1 \\ &\Rightarrow \Phi\left(\frac{\varepsilon}{5}\right) = 0.95 \Rightarrow \frac{\varepsilon}{5} = \Phi^{-1}(0.95) \Rightarrow \varepsilon \doteq 5 \cdot 1.645 = 8.225\,. \end{aligned}$$

b) Z Čebyševovy nerovnosti dostaneme

$$P[|X - \mathrm{E}X| > 10] \leq \frac{\mathrm{D}X}{10^2} \Rightarrow P[|X - 50| > 10] \leq \frac{25}{100} = 0.25\,.$$

Cvičení 12.12. *Standardní porce masa v řízku je* 150 *gramů. Směrodatná odchylka od této hodnoty je* 10 *gramů.*

a) Určete pravděpodobnost, že na 100 *porcí masa nebude spotřebováno více než* 15.3 *kilogramů masa.*

b) Pomocí Čebyševovy nerovnosti určete pravděpodobnost, že odchylka od standardní spotřeby na 100 *porcí není větší než* 0.8 kg.

Řešení. Máme $\mathrm{E}X_i = 150$, $\mathrm{D}X_i = 10^2$, $\mathrm{E}\sum_{i=1}^{100} X_i = 15\,000$, $\mathrm{D}\sum X_i^{100} = 10\,000$ a $\sqrt{\mathrm{D}\sum_{i=1}^{100} X_i} = 100$.

a) Podle CLV je

$$P\left(\sum_{i=1}^{100} X_i \leq 15\,300\right) = P\left(\frac{\sum X_i - 15\,000}{\sqrt{10\,000}} \leq \frac{15\,300 - 15\,000}{100}\right) = \Phi(3) = 0.99865.$$

b) z Čebyševovy nerovnosti je

$$P\left(\left|\sum_{i=1}^{100} X_i - \mathrm{E}\sum_{i=1}^{100} X_i\right| \le 800\right) \ge 1 - \frac{10\,000}{800^2} = 1 - \frac{1}{64} = \frac{63}{64}.$$

Cvičení 12.13. *Počet bodů ze zkouškové písemky je náhodná veličina pohybující se v rozmezí* $0 - 100$*, s průměrem* 53 *a rozptylem* 839*. Celkem* 200 *studentů psalo zkouškový test. Určete pravděpodobnost, že průměrný počet bodů u těchto studentů byl menší než* 50*, a uveďte za jakých předpokladů.*

Řešení. Označme X_i počet bodů i-tého studenta, $i = 1, 2, \ldots, 200$, a dále $X = \sum_{i=1}^{200} X_i$. Hledáme pravděpodobnost

$$P\left(\frac{\sum_{i=1}^{200} X_i}{200} < 50\right) = P(X < 10^4) = P\left(\frac{X - 200 \cdot 53}{\sqrt{200}\sqrt{839}} < \frac{10^4 - 200 \cdot 53}{\sqrt{200}\sqrt{839}}\right).$$

Pokud jsou veličiny X_i nezávislé (což nejsou např., když studenti opisují), platí podle CLV

$$\mathrm{norm}(X) = \frac{X - 10600}{\sqrt{200}\sqrt{839}} \sim \mathrm{N}(0,1).$$

Tedy lze psát

$$P\left(\mathrm{norm}(X) < \frac{-600}{409.63}\right) = \Phi(-1.46) = 1 - \Phi(1.46) = 1 - 0.93 = 0.07.$$

Cvičení 12.14. *Krevní tlak má v populaci střední hodnotu* $\mathrm{E}X = 138$ *mmHg a směrodatnou odchylku* $\sigma_X = 20$ *mmHg. Určete rozmezí, v němž se nachází aspoň* $\beta = 80\,\%$ *populace. Uveďte použité předpoklady.*

Řešení. Čebyševova nerovnost (protože neznáme typ rozdělení) zaručuje

$$P\left[|X - \mathrm{E}X| \ge \varepsilon\right] \le \frac{\sigma_X^2}{\varepsilon^2}\,.$$

Položíme

$$\begin{aligned}\frac{\sigma_X^2}{\varepsilon^2} &\le 1 - \beta = 0.2,\\ \varepsilon &\ge \frac{\sigma_X}{\sqrt{0.2}} = \sigma_X\sqrt{5} \doteq 45\,.\end{aligned}$$

Hledané meze jsou $\mathrm{E}X \pm \sigma_X\sqrt{5} \doteq 138 \pm 45 = (93, 183)$.

Část II
Základy matematické statistiky

13 Intervalové odhady charakteristik rozdělení

13.1 Intervalové odhady normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$

13.1.1 Odhad střední hodnoty

Cvičení 13.1. *Soubor dat* (75, 85, 58, 72, 70, 75) *je náhodným výběrem z normálního rozdělení* $\mathrm{N}(\mu,\ \sigma^2)$. *Stanovte* 95% *interval spolehlivosti pro střední hodnotu* μ.

Řešení. K určení intervalového odhadu použijeme statistiku

$$T = \frac{\overline{\boldsymbol{X}} - \mu}{S_{\overline{\boldsymbol{X}}}} = \frac{\overline{\boldsymbol{X}} - \mu}{S_{\boldsymbol{X}}} \sqrt{n} \quad \Rightarrow \quad \overline{\boldsymbol{X}} - \frac{S_{\boldsymbol{X}}}{\sqrt{n}}\, q_{\mathrm{t}(5)}(0.975) \le \mu \le \overline{\boldsymbol{X}} + \frac{S_{\boldsymbol{X}}}{\sqrt{n}}\, q_{\mathrm{t}(5)}(0.975)\,,$$

kde $\overline{\boldsymbol{X}}$ je výběrový průměr, $S^2_{\boldsymbol{X}}$ výběrový rozptyl a n je rozsah výběru. Ze zadání dostaneme $n = 6$ a

$$\overline{\boldsymbol{X}} = \frac{1}{6} \sum_{i=1}^{6} x_i = \frac{435}{6} = 72.5\,,$$

$$s^2_{\boldsymbol{x}} = \frac{1}{5}\left(\sum_{i=1}^{6} x_i^2 - 6 \cdot (\overline{\boldsymbol{X}})^2\right) = \frac{385.5}{5} \doteq 77.1\,, \qquad s_{\boldsymbol{x}} \doteq 8.781\,.$$

Z tabulek kvantilů studentova rozdělení dostaneme $q_{\mathrm{t}(5)}(0.975) \doteq 2.57$, a tedy

$$63.29 \le \mu \le 81.71\,.$$

Cvičení 13.2. *Opakovaná měření stejné náhodné veličiny dala následující výsledky:*

17	37	70	34	64	52	72	21	47	32	94	62	71	34	43

Jakou hodnotu h *překročí střední hodnota s pravděpodobností nejvýše* $\alpha = 1\,\%$? *Uveďte použité předpoklady.*

Řešení. $n = 15$, $\bar{\boldsymbol{x}} = 50$, $s_{\boldsymbol{x}} = 21,7$,

$$h \ge \bar{\boldsymbol{x}} + \frac{s_{\boldsymbol{x}}}{\sqrt{n}}\, q_{t(n-1)}(1-\alpha) \ \doteq 50 + \frac{21.7}{\sqrt{15}} \cdot 2.625 = 64.7.$$

Předpoklady: Chyby jednotlivých měření mají stejné normální rozdělení a jsou nezávislé (bývá splněno).

13.1.2 Odhad rozptylu a směrodatné odchylky

Cvičení 13.3. *Opakovaná měření stejné koncentrace látky vedla k následujícím výsledkům:* $(0.2, 0.23, 0.21, 0.16, 0.18, 0.19, 0.14, 0.18, 0.21)$. *Najděte symetrické oboustranné* 90%*-ní odhady střední hodnoty, rozptylu a směrodatné odchylky.*

Řešení. Odhadujeme parametry náhodné veličiny X z realizace rozsahu $n = 9$, jejíž statistiky jsou realizace výběrového průměru $\overline{\boldsymbol{X}} \doteq 0.189$, realizace výběrového rozptylu $s_x^2 \doteq 7.6 \cdot 10^{-4}$, realizace výběrové směrodatné odchylky $s_x = \sqrt{s_x^2} \doteq 2.76 \cdot 10^{-2}$. Intervalový odhad střední hodnoty:

$$\begin{aligned}
&\left\langle \overline{\boldsymbol{X}} - \frac{s_x}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(0.95), \overline{\boldsymbol{X}} + \frac{s_x}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(0.95) \right\rangle \doteq \\
&\doteq \left\langle 0.189 - \frac{2.76 \cdot 10^{-2}}{3}\, \underbrace{q_{\mathrm{t}(8)}(0.95)}_{1.86}, 0.189 + \frac{2.76 \cdot 10^{-2}}{3}\, q_{\mathrm{t}(8)}(0.95) \right\rangle \doteq \\
&\doteq \langle 0.172, 0.206 \rangle\ .
\end{aligned}$$

Intervalový odhad rozptylu:

$$\begin{aligned}
&\left\langle \frac{(n-1)\,s_x^2}{q_{\chi^2(n-1)}\,(0.95)}, \frac{(n-1)\,s_x^2}{q_{\chi^2(n-1)}\,(0.05)} \right\rangle \doteq \\
&\doteq \left\langle \frac{8\cdot 7.6\cdot 10^{-4}}{\underbrace{q_{\chi^2(8)}\,(0.95)}_{15.51}}, \frac{8\cdot 7.6\cdot 10^{-4}}{\underbrace{q_{\chi^2(8)}\,(0.05)}_{2.73}} \right\rangle \doteq \\
&\doteq \left\langle 3.9\cdot 10^{-4},\, 2.2\cdot 10^{-3} \right\rangle .
\end{aligned}$$

Intervalový odhad směrodatné odchylky (odmocnina z předchozího):

$$\begin{aligned}
&\left\langle \sqrt{\frac{(n-1)\,s_x^2}{q_{\chi^2(n-1)}\,(0.95)}}, \sqrt{\frac{(n-1)\,s_x^2}{q_{\chi^2(n-1)}\,(0.05)}} \right\rangle \doteq \\
&\doteq \left\langle \sqrt{3.9\cdot 10^{-4}},\, \sqrt{2.2\cdot 10^{-3}} \right\rangle \doteq \left\langle 1.97\cdot 10^{-2}. 4.7\cdot 10^{-2} \right\rangle .
\end{aligned}$$

Všimněte si, že intervalové odhady výběrového rozptylu, resp. směrodatné odchylky nejsou symetrické kolem jejich bodových odhadů $s_x^2 \doteq 7.6\cdot 10^{-4}$, resp. $s_x \doteq 2.76\cdot 10^{-2}$.

Cvičení 13.4. *Soubor* (70, 84, 89, 70, 74, 70) *je náhodným výběrem z normálního rozdělení* $\mathrm{N}(\mu, \sigma)$. *Určete* 95% *interval spolehlivosti pro rozptyl* σ^2.

Řešení. Interval spolehlivosti pro rozptyl určíme ze statistiky

$$Y = (n-1)\,\frac{S_{\boldsymbol{X}}^2}{\sigma^2}\,,$$

má rozdělení $\chi^2(n-1)$, kde n je rozsah výběru a $S_{\boldsymbol{X}}^2$ je výběrový rozptyl. Ze zadání je $n = 6$,

$$\overline{\boldsymbol{X}} = \frac{1}{6}\sum_{i=1}^{6} x_i = \frac{457}{6} \doteq 76.17\,, \quad s_{\boldsymbol{x}}^2 = \frac{1}{5}\left(\sum_{i=1}^{6} x_i^2 - 6\cdot(\overline{\boldsymbol{X}})^2\right) \doteq \frac{341.787}{5} \doteq 68.358\,.$$

Pro 95% pravděpodobnost je

$$q_{\chi^2(5)}(0.025) = 0.831 \le Y \le q_{\chi^2(5)}(0.975) \Rightarrow$$

$$\frac{341.787}{12.83} \doteq 26.64 \le \sigma^2 \le \frac{341.787}{0.831} \doteq 411.296\,.$$

Cvičení 13.5. *V souboru* $y = (72, 65, 75, 87, 72, 73, 70, 83, 85)$ *jsou uvedeny váhy ve skupině studentů. Za předpokladu, že se jedná o náhodný výběr z náhodné veličiny* Y*, která má normální rozdělení, stanovte* 95% *horní jednostranný interval spolehlivosti pro rozptyl* σ_Y^2*, tedy mez* σ_0^2*, pro kterou* $P[\sigma_Y^2 \le \sigma_0^2] = 0.95$.

Řešení. Ke stanovení intervalového odhadu použijeme skutečnosti, že náhodná veličina (statistika)

$$V = \frac{(n-1)\,S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$$

má χ^2-rozdělení s $n-1$ stupni volnosti. Z 95% jednostranného intervalu spolehlivosti pro její hodnoty získáme hledaný odhad. Pro zadaný výběr dostaneme

$$n = 9, \qquad \bar{\boldsymbol{y}} = \frac{682}{9} \doteq 75.778, \qquad \sum_{i=1}^{9} y_i^2 = 52\,130, \qquad s_y^2 = 56.194.$$

Z tabulek kvantilů odečteme pro rozdělení $\chi^2(8)$ kvantil $q_{\chi^2(8)}(0.05) \doteq 2.73$. Pro realizaci dostaneme interval spolehlivosti

$$2.73 \le \frac{8\cdot 56.194}{\sigma_Y^2} \qquad \Rightarrow \qquad \sigma_Y^2 \le \sigma_0^2 := \frac{449.552}{2.73} \doteq 164.67\,.$$

Tedy $\sigma_Y^2 \in (-\infty, 164.67\rangle$ s pravděpodobností 0.95.

Cvičení 13.6. *Deset opakovaných měření obsahu alkoholu ve vzorku krve má průměrnou hodnotu* $\overline{\boldsymbol{X}} =$ 0.15 *promile (alkoholu v krvi) a směrodatnou odchylku* $s_{\overline{\boldsymbol{X}}} = 0.01$ *promile. Jakou hodnotu* σ_0 *překročí chyba metody (t.j. směrodatná odchylka) s pravděpodobností nejvýše* $\alpha = 1$ %*? Uveďte použité předpoklady.*

Řešení. U veličiny

$$Y = \text{„}\textit{obsah alkoholu v krvi (v promile)}\text{“}$$

budeme předpokládat normální rozdělení $\mathrm{N}(\mu, \sigma^2)$. Naše veličina

$$X = \text{„}\textit{chyba měření obsahu alkoholu v krvi (v promile)}\text{“}$$

bude tedy určena jako $X = Y - \mu$ s rozdělením $\mathrm{N}(0, \sigma^2)$. Jednotlivá měření považujeme za nezávislá. Hledáme horní $1 - \alpha = 99\%$ intervalový odhad pro parametr rozptylu σ^2, který bude tvaru $(0, \sigma_0^2\rangle$.

Pro statistiku

$$T = \frac{(n-1)\, S_{\boldsymbol{X}}^2}{\sigma^2}$$

s χ^2-rozdělením s $n - 1 = 9$ stupni volnosti dostáváme, že s pravděpodobností $1 - \alpha = 99$ % pro realizaci t veličiny T platí, že:

$$q_{\chi^2(n-1)}(\alpha) \le t = \frac{(n-1)\, s_{\boldsymbol{x}}^2}{\sigma^2}\,,$$

takže

$$\sigma^2 \le \frac{(n-1)\, s_{\boldsymbol{x}}^2}{q_{\chi^2(n-1)}(\alpha)} =: \sigma_0^2$$

a hledaná hranice je

$$\sigma_0 = s_{\boldsymbol{x}} \cdot \sqrt{\tfrac{n-1}{q_{\chi^2(n-1)}(\alpha)}} = 0.01 \cdot \sqrt{\frac{9}{q_{\chi^2(9)}(0.01)}} \doteq \frac{0.03}{\sqrt{2.0879}} \doteq 0.02076\,.$$

Předpoklady: Chyby jednotlivých měření mají stejné normální rozdělení a jsou nezávislé (bývá splněno).

14 Odhad parametrů (metoda momentů, metoda maximální věrohodnosti)

14.1 Odhady diskrétních rozdělení

Cvičení 14.1. *Počet kazů na tabulkách skla se řídí Poissonovým rozdělením. Bylo pozorováno*

17 *tabulek*	*bez kazu*
4 *tabulky*	*s* 1 *kazem*
1 *tabulka*	*s* 2 *kazy*
2 *tabulky*	*s* 3 *kazy*
1 *tabulka*	*s* 5 *kazy.*

Metodou maximální věrohodnosti určete parametr λ tohoto Poissonova rozdělení.

Řešení. Máme realizaci náhodného výběru $(x_1, \ldots, x_n)$.

Pro náhodnou veličinu s rozdělením $\mathrm{Po}(\lambda)$ je $P[X = k] = \dfrac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$. Věrohodnost je

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}\,\mathrm{e}^{-\lambda} = \left(\frac{\lambda^0}{0!}\,\mathrm{e}^{-\lambda}\right)^{17} \left(\frac{\lambda^1}{1!}\,\mathrm{e}^{-\lambda}\right)^{4} \left(\frac{\lambda^2}{2!}\,\mathrm{e}^{-\lambda}\right)^{1} \left(\frac{\lambda^3}{3!}\,\mathrm{e}^{-\lambda}\right)^{2} \left(\frac{\lambda^5}{5!}\,\mathrm{e}^{-\lambda}\right)^{1} \\ &= \frac{\lambda^{17}}{8640}\,\mathrm{e}^{-25\lambda}\,, \end{aligned}$$

logaritmus věrohodnosti

$$l(\lambda) = \ln L(\lambda) = 17 \ln \lambda - 25\lambda - \ln 8640$$

a derivace

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{17}{\lambda} - 25\,.$$

Řešením je

$$\frac{17}{\widehat{\lambda}} - 25 = 0 \qquad \Longrightarrow \qquad \widehat{\lambda} = \frac{17}{25}\,.$$

Cvičení 14.2. *Kostka A je správná, kostka B dává výsledky s pravděpodobnostmi v tabulce. Házeli jsme jednou z nich, četnosti výsledků udává tabulka. Odhadněte, kterou kostkou se házelo.*

hodnota	1	2	3	4	5	6
p_A	1/6	1/6	1/6	1/6	1/6	1/6
p_B	1/10	1/10	1/10	1/10	1/10	1/2
četnost	10	12	15	10	12	25

Řešení. Věrohodnost pro první kostku je

$$(1/6)^{84} \doteq 4.318 \cdot 10^{-66}\,,$$

pro druhou

$$(1/10)^{10+12+15+10+12} \cdot (1/2)^{25} \doteq 2.98 \cdot 10^{-67}\,.$$

Věrohodnější je odhad, že se házelo kostkou A.

Lze také testovat shodu s teoretickým rozdělením obdobně jako v χ^2-testu a vybrat variantu s nižší hodnotou kritéria; postup je pracnější a vede ke stejnému závěru.

Cvičení 14.3. *V urně je mnoho hracích kostek, z nichž některé jsou správné, některé falešné. Na falešných padá šestka s pravděpodobností $1/2$, zbývající čísla mají stejnou pravděpodobnost. Opakovaně jsme vytáhli kostku, hodili jí a vrátili ji zpět. Četnost výsledků udává tabulka:*

hodnota	1	2	3	4	5	6
četnost	18	20	12	15	10	25

Odhadněte, kolik procent kostek je falešných.

Řešení. Podíl falešných kostek označme $p \in \langle 0,1\rangle$.

Metoda momentů:

Střední hodnota výsledku pro správnou kostku je 3.5, pro falešnou 4.5, pro směs s koeficientem p vychází $3.5\,(1-p)+4.5\,p = 3.5+p$.

Realizace výběrového průměru je 3.54.

Srovnáním těchto dvou hodnot vyjde odhad $\widehat{p} = 0.04 \in \langle 0,1\rangle$, což vyhovuje zadání.

Metoda maximální věrohodnosti:

Ve směsi rozdělení má šestka pravděpodobnost $\frac{1}{6}\,(1-p)+\frac{1}{2}\,p = \frac{1+2\,p}{6}$ a padla 25×, ostatní čísla $\frac{1}{6}\,(1-p)+\frac{1}{10}\,p = \frac{5-2\,p}{30}$ a padla 75× (není třeba mezi nimi rozlišovat).

$$L(p) = \left(\frac{5-2\,p}{30}\right)^{75} \cdot \left(\frac{1+2\,p}{6}\right)^{25},$$
$$\ell(p) = 75\,\ln(5-2\,p)+25\,\ln(1+2\,p)-75\,\ln 30-25\,\ln 6\,.$$

Maximum nastává pro $\widehat{p}$ takové, že

$$\frac{\partial}{\partial \widehat{p}}\ell(\widehat{p}) = \frac{-150}{5-2\,\widehat{p}}+\frac{50}{1+2\,\widehat{p}} = 0\,,$$
$$\widehat{p} = \frac{1}{4} \in \langle 0,1\rangle\,.$$

Cvičení 14.4. *Náhodná veličina X je směsí náhodných veličin Y, Z, jejichž pravděpodobnostní funkce jsou dány tabulkou:*

hodnota	1	2	3	4
p_Y	0.4	0.4	0.1	0.1
p_Z	0.1	0.1	0.4	0.4
pozorovaná četnost	12	13	9	6

Poslední řádek udává četnosti hodnot v realizaci náhodného výběru s rozdělením, které má náhodná veličina X. Odhadněte z nich neznámý koeficient směsi w.

Řešení. Metoda momentů:

$$\begin{aligned}
\mathrm{E}X &= w\,\mathrm{E}Y+(1-w)\,\mathrm{E}Z = \\
&= w\,(0.4\cdot 1+0.4\cdot 2+0.1\cdot 3+0.1\cdot 4)+(1-w)\,(0.1\cdot 1+0.1\cdot 2+0.4\cdot 3+0.4\cdot 4) = \\
&= 1.9\,w+3.1\,(1-w) = 3.1-1.2\,w = \frac{12\cdot 1+13\cdot 2+9\cdot 3+6\cdot 4}{12+13+9+6} = \frac{89}{40} = 2.225\,,\\
w &= \frac{35}{48} \doteq 0.729\,17\,.
\end{aligned}$$

Vyhovuje zadání.

Metoda maximální věrohodnosti:

$0.4\,w+0.1\,(1-w) = 0.3\,w+0.1$, $0.1\,w+0.4\,(1-w) = 0.4-0.3w\,.$

hodnota	1	2	3	4
p_X	$0.1+0.3\,w$	$0.1+0.3\,w$	$0.4-0.3w$	$0.4-0.3w$
pozorovaná četnost	12	13	9	6

$$\begin{aligned}
L(w) &= (0.1+0.3\,w)^{12+13}\cdot(0.4-0.3w)^{9+6} = (0.1+0.3\,w)^{25}\,(0.4-0.3w)^{15}\,,\\
\ell(w) &= \ln(L(w)) = 25\,\ln(0.1+0.3\,w)+15\,\ln(0.4-0.3w)\,,\\
\ell'(w) &= \frac{7.5}{0.1+0.3\,w}-\frac{4.5}{0.4-0.3\,w} = 0\,,\\
w &= \frac{17}{24} \doteq 0.708\,33\,.
\end{aligned}$$

Cvičení 14.5. *Dvě diskrétní náhodné veličiny X, Y mají pravděpodobnostní funkce dané tabulkou. Odhadněte koeficient c směsi $Z = \mathrm{Mix}_c(X,Y)$ z četností jejích realizací uvedených v tabulce.*

hodnota	1	2	3	4
p_X	0.1	0.2	0.2	0.5
p_Y	0.5	0.2	0.2	0.1
četnost	30	20	15	35

Řešení. Metoda momentů:

$$\begin{aligned}
\mathrm{E}X &= 0.1 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.5 = 3.1\,, \\
\mathrm{E}Y &= 0.5 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.1 = 1.9\,, \\
\mathrm{E}Z &= c\,\mathrm{E}X + (1-c)\,\mathrm{E}Y = 1.9 + 1.2\,c\,, \\
\bar{z} &= \frac{30 + 2 \cdot 20 + 3 \cdot 15 + 4 \cdot 35}{100} = \frac{51}{20} = 2.55 = 1.9 + 1.2\,\widehat{c}\,, \\
\widehat{c} &= \frac{2.55 - 1.9}{1.2} \doteq 0.5417\,.
\end{aligned}$$

Metoda maximální věrohodnosti:

hodnota	1	2	3	4
p_Z	$0.5 - 0.4\,c$	0.2	0.2	$0.1 + 0.4\,c$

$$\begin{aligned}
L(c) &= (0.5 - 0.4\,c)^{30} \cdot 0.2^{20+15} \cdot (0.1 + 0.4\,c)^{35}\,, \\
\ell(c) &= \ln L(c) = 30\,\ln(0.5 - 0.4\,c) + 35\,\ln 0.2 + 35\,\ln(0.1 + 0.4\,c)\,, \\
0 &= \frac{\partial \ell(\widehat{c})}{\partial \widehat{c}} = 30\,\frac{-0.4}{0.5 - 0.4\,\widehat{c}} + 35\,\frac{0.4}{0.1 + 0.4\,\widehat{c}}\,, \\
\widehat{c} &= \frac{29}{52} \doteq 0.5577\,.
\end{aligned}$$

Cvičení 14.6. *Náhodná veličina může nabývat hodnot* $0, 1, 2$. *Její rozdělení, závislé na parametrech* p, q, *a četnost hodnot v realizaci uvádí tabulka:*

hodnota	0	1	2
teoretická pravděpodobnost	p	q	q^2
pozorovaná četnost	2	12	6

Odhadněte parametry p, q.

Řešení. Parametry jsou vázány podmínkou $p = 1 - q - q^2$.

Metoda momentů: $\mu_X = q + 2q^2$, $m_X = \frac{1 \cdot 12 + 2 \cdot 6}{2 + 12 + 6} = \frac{6}{5}$, $\mu_X = m_X$.

$\Longrightarrow q_1 = -\frac{1}{20}\sqrt{265} - \frac{1}{4} = -1.063\,9$ (nevyhovuje), $q_2 = \frac{1}{20}\sqrt{265} - \frac{1}{4} = 0.563\,94$ (vyhovuje), $p = 1 - q_2 - q_2^2 = 0.118\,03$.

Metoda maximální věrohodnosti: $L(q) = 2\ln p + 12\ln q + 6\ln q^2 = 2\ln(1 - q - q^2) + 24\ln q$, $\frac{\partial L}{\partial q}(q) = \frac{24}{q} + 2\frac{-2q-1}{1-q^2-q} = 0 \Longrightarrow q_1 = -\frac{3}{2}$(nevyhovuje), $q_2 = \frac{4}{7} = 0.571\,43$ (vyhovuje), $p = 1 - q_2 - q_2^2 = \frac{5}{49} = 0.102\,04$.

Cvičení 14.7. *Náhodná veličina* X *nabývá hodnot s pravděpodobnostmi dle tabulky, kde* c, q *jsou reálné parametry rozdělení. Z četností hodnot v náhodném výběru, uvedených v tabulce, odhadněte pravděpodobnosti všech hodnot.*

hodnota i	1	2	3
pravděpodobnost $p_X(i)$	$c - q$	c	$c + q$
četnost n_i	10	10	5

Řešení. Protože součet pravděpodobností všech hodnot je 1, musí být $c = 1/3$, zbývá odhadnout parametr q.

Metoda maximální věrohodnosti:

$$\begin{aligned}
\ell(q) &= 10\,\ln\left(\frac{1}{3} - q\right) + 10\,\ln\frac{1}{3} + 5\,\ln\left(\frac{1}{3} + q\right)\,, \\
0 = \ell'(q) &= \frac{-10}{\frac{1}{3} - q} + \frac{5}{\frac{1}{3} + q}\,, \\
q &= -\frac{1}{9}\,.
\end{aligned}$$

Odhady pravděpodobností hodnot $1, 2, 3$ jsou po řadě $4/9, 1/3, 2/9$.

Metoda momentů:
Střední hodnota je

$$\mathrm{E}X = \frac{1}{3} - q + 2 \cdot \frac{1}{3} + 3 \cdot \left(\frac{1}{3} + q\right) = 2 + 2\,q\,,$$

její odhad z realizace

$$\frac{1}{n} \sum_i i\, n_i = \frac{1}{25}\,(10 + 2 \cdot 10 + 3 \cdot 5) = \frac{9}{5}\,,$$

rovnice

$$2 + 2\,q = \frac{9}{5}$$

má řešení $q = -1/10$. Odhady pravděpodobností hodnot $1, 2, 3$ jsou po řadě

$$\frac{1}{3} + \frac{1}{10},\ \frac{1}{3},\ \frac{1}{3} - \frac{1}{10},$$

což vyhovuje zadání.

Cvičení 14.8. *Házíme nesymetrickou mincí. Padl $12\times$ rub a $8\times$ líc. Chceme odhadnout pravděpodobnost q, se kterou padá rub. Použijte metodu maximální věrohodnosti.*

Výsledky. $q = 3/5$.

Cvičení 14.9. *Náhodné veličiny X, Y a Z mají rozdělení daná tabulkou*

veličina \ *hodnota*	1	2	3	4
X	$1/2$	$1/4$	$1/8$	$1/8$
Y	$1/3$	$1/3$	$1/6$	$1/6$
Z	$2/5$	$3/10$	$1/5$	$1/10$

Z realizace

hodnota	1	2	3	4
četnost	43	30	15	12

rozhodněte, který z modelů (daných rozděleními veličin X, Y, Z) je nejlepší.

Řešení. Náhodné veličiny X, Y, Z dávají na dané realizaci po řadě věrohodnosti

$$\begin{aligned}(1/2)^{43} * (1/4)^{30} * (1/8)^{15} * (1/8)^{12} &\doteq 4.08 \cdot 10^{-56}\,,\\ (1/3)^{43} * (1/3)^{30} * (1/6)^{15} * (1/6)^{12} &\doteq 1.45 \cdot 10^{-56}\,,\\ (2/5)^{43} * (3/10)^{30} * (1/5)^{15} * (1/10)^{12} &\doteq 5.22 \cdot 10^{-56}\,,\end{aligned}$$

největší je pro Z.

Cvičení 14.10. *Posuďte, který z pravděpodobnostních modelů v tabulce nejlépe odpovídá pozorovaným četnostem známek:*

známka	1	2	3	4
pravděpodobnost dle modelu A	$1/4$	$1/4$	$1/4$	$1/4$
pravděpodobnost dle modelu B	$1/6$	$1/6$	$1/3$	$1/3$
pravděpodobnost dle modelu C	$1/8$	$1/8$	$1/4$	$1/2$
četnost	20	27	70	99

Řešení. Na dané realizaci dostáváme po řadě věrohodnosti

$$\begin{aligned}(1/4)^{20} * (1/4)^{27} * (1/4)^{70} * (1/4)^{99} &\doteq 9.02 \cdot 10^{-131}\,,\\ (1/6)^{20} * (1/6)^{27} * (1/3)^{70} * (1/3)^{99} &\doteq 5.21 \cdot 10^{-118}\,,\\ (1/8)^{20} * (1/8)^{27} * (1/4)^{70} * (1/2)^{99} &\doteq 4.06 \cdot 10^{-115}\,,\end{aligned}$$

největší je pro model C.

Cvičení 14.11. *Gen se vyskytuje ve 4 variantách A, B, C, D. Model předpokládá, že B se vyskytuje 3× častěji než A a D 3× častěji než C. Odhadněte jejich pravděpodobnosti na základě zjištěných četností v tabulce.*

varianta	*A*	*B*	*C*	*D*
četnost	10	15	15	40

Výsledky. $\frac{5}{64}, \frac{15}{64}, \frac{11}{64}, \frac{33}{64}$.

Cvičení 14.12. *V osudí jsou dva druhy kostek, na prvních jsou čísla* $1, \ldots, 6$*, na druhých pouze* $1, 3, 5$*, u obou druhů jsou všechny možné výsledky stejně pravděpodobné. Vytáhli jsme* 20 *kostek a jednou jimi hodili; četnost výsledků udává tabulka. Odhadněte, kolik z těchto* 20 *kostek bylo prvního druhu.*

a)

hodnota	1	2	3	4	5	6
četnost	5	2	4	2	6	1

b)

hodnota	1	2	3	4	5	6
četnost	3	4	4	4	2	3

Výsledky. a) 10, b) 20.

Cvičení 14.13. *Náhodná veličina nabývá výsledky* $1, 2, 3$*. Tabulka uvádí jejich pravděpodobnosti a pozorované četnosti. Odhadněte parametry* a, b*.*

hodnota	1	2	3
teoretická pravděpodobnost	$a+b$	$a+2b$	$a+3b$
četnost	10	10	20

Výsledky. $a = b = 1/9$.

14.2 Odhady spojitých rozdělení

Cvičení 14.14. *Doba do poruchy daného přístroje má exponenciální rozdělení. Od začátku roku bylo zjištěno, že stroj se porouchal postupně za* 20 *dní,* 37.5 *dní,* 28 *dní,* 10.5 *dní a* 54 *dní. Metodou maximální věrohodnosti určete parametr* w *tohoto exponenciálního rozdělení.*

Řešení. Předpokládáme hustotu $f(x) = w\,\mathrm{e}^{-wx}$ pro $x > 0$. Věrohodnostní funkce je

$$\Lambda(w) = \prod_{i=1}^{n} w\,\mathrm{e}^{-wx_i} = w\,\mathrm{e}^{-w\cdot 20} w\,\mathrm{e}^{-w\cdot 37.5} w\,\mathrm{e}^{-w\cdot 28} w\,\mathrm{e}^{-w\cdot 10.5} w\,\mathrm{e}^{-w\cdot 54} = w^5\,\mathrm{e}^{-w\cdot 150}\,.$$

Logaritmicko-věrohodnostní funkce je

$$\lambda(w) = \ln \Lambda(w) = n \ln w - w \sum_{i=1}^{n} x_i = 5 \ln w - 150\,w\,,$$

Její derivace

$$\lambda'(w) = \frac{n}{w} - \sum_{i=1}^{n} x_i = \frac{5}{w} - 150\,.$$

Nulová derivace a maximum věrohodnosti je pro

$$\widehat{w} = \frac{1}{\bar{\boldsymbol{x}}} = \frac{1}{30}\,.$$

Cvičení 14.15. *Exponenciální rozdělení má pro* $t \geq 0$ *hustotu*

$$f_X(t) = w\,\mathrm{e}^{-w\,t}\,.$$

Na základě náhodného výběru

0.42600148; 0.87249437; 0.14744007; 0.01452009; 1.44667089

najděte maximálně věrohodný odhad parametru w*.*

Výsledky. 1.72.

Cvičení 14.16. *Předpokládáme, že náhodná veličina X má posunuté exponenciální rozdělení s hustotou*

$$f_X(t) = \begin{cases} \dfrac{1}{\tau}\,\exp\left(-\dfrac{t-T}{\tau}\right), & t \ge T,\\ 0 & \textit{jinak,} \end{cases}$$

kde $\tau > 0$. Z realizace $x = (2, 3, 8, 4, 10, 3, 5)$ odhadněte parametry T, τ.

Řešení. Metoda maximální věrohodnosti:

$$\lambda(T,\tau) = \ln \Lambda(T,\tau) = \ln\left(\prod_{i=1}^{n} \frac{1}{\tau}\exp\left(-\frac{x_i - T}{\tau}\right)\right) = -n\,\ln\tau - \frac{1}{\tau}\sum_{i=1}^{n} x_i + \frac{1}{\tau}\,n\,T\,,$$

pokud $T \le \min\limits_i x_i$ (jinak 0). To je rostoucí funkce T, takže $\widehat{T} = \min\limits_i x_i$.

$$0 = \frac{\partial\lambda}{\partial\tau}(\widehat{T},\widehat{\tau}) = -\frac{n}{\widehat{\tau}} + \frac{1}{\widehat{\tau}^2}\underbrace{\sum_{i=1}^{n} x_i}_{n\,\bar{\boldsymbol{x}}} - \frac{1}{\widehat{\tau}^2}\,n\,\widehat{T}\,,$$

$$\widehat{\tau} = \bar{\boldsymbol{x}} - \widehat{T} = \bar{\boldsymbol{x}} - \min_i x_i\,.$$

V našem případě $\widehat{T} = 2$, $\widehat{\tau} = 5 - 2 = 3$.

Metoda momentů:

$$\begin{aligned} \mathrm{E}X &= \int_{\mathbb{R}} t\,f_X(t)\,\mathrm{d}t = \int_T^\infty \frac{t}{\tau}\exp\left(-\frac{t-T}{\tau}\right)\mathrm{d}t = (-t-\tau)\exp\left(-\frac{t-T}{\tau}\right)\Big|_{t=T}^{\infty} = \\ &= T + \tau\,,\\ \mathrm{E}X^2 &= \int_{\mathbb{R}} t^2\,f_X(t)\,\mathrm{d}t = \int_T^\infty \frac{t^2}{\tau}\exp\left(-\frac{t-T}{\tau}\right)\mathrm{d}t = \\ &= \left(-t^2 - 2\,\tau\,t - 2\,\tau^2\right)\exp\left(-\frac{t-T}{\tau}\right)\Big|_{t=T}^{\infty} = \\ &= T^2 + 2\,\tau\,T + 2\,\tau^2 = (T+\tau)^2 + \tau^2 = (\mathrm{E}X)^2 + \tau^2\,. \end{aligned}$$

K těmto výsledkům lze dojít bez integrování, neboť $X = Y + T$, kde T je konstanta a Y je náhodná veličina s exponenciálním rozdělením, $\mathrm{E}Y = \tau$, $\sigma_Y^2 = \tau^2$; $\mathrm{E}X = \mathrm{E}Y + T$, $\sigma_X^2 = \sigma_Y^2$, $\mathrm{E}X^2 = \mathrm{E}X^2 + \sigma_X^2$.

V našem případě

$$m_X = \frac{1}{n}\sum_{i=1}^{n} x_i = 5\,, \qquad m_{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{227}{7} \doteq 32.429\,,$$

soustava rovnic

$$\begin{aligned} \widehat{T} + \widehat{\tau} &= m_X = 5\,,\\ m_X^2 + \widehat{\tau}^2 &= m_{X^2} = \frac{227}{7}\,, \end{aligned}$$

má kladné řešení $\widehat{\tau} = \frac{2\sqrt{91}}{7} \doteq 2.725\,5$, $\widehat{T} \doteq 2.274\,5$, které ovšem neodpovídá zadání, neboť $\widehat{T} > x_1 = 2$, takže nalezený model nepřipouští pozorovanou hodnotu x_1 (ta by měla nulovou hustotu pravděpodobnosti).

Cvičení 14.17. *a) Určete parametr α tak, aby funkce $f(t)$*

$$f(t) = \begin{cases} 0\,, & t < 0\,,\\ \alpha\,t\,\mathrm{e}^{-wt}\,, & 0 \le t \end{cases}$$

byla hustotou pravděpodobnosti.

b) Pomocí metody maximální věrohodnosti odhadněte parametr w pro hustotu z bodu a) na základě realizace náhodného výběru $(x_1, x_2, \ldots, x_n)$.

Řešení. a) Protože funkce f je hustotou pravděpodobnosti, platí

$$1 = \int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = \int_{0}^{+\infty} \alpha\, t\,\mathrm{e}^{-wt}\,\mathrm{d}t = -\frac{\alpha}{w}\left[t\,\mathrm{e}^{-wt}\right]_{t=0}^{\infty} + \frac{\alpha}{w}\int_{0}^{+\infty} \mathrm{e}^{-wt}\,\mathrm{d}t =$$
$$= -\frac{\alpha}{w^2}\left[\mathrm{e}^{-wt}\right]_{t=0}^{\infty} = \frac{\alpha}{w^2}\,.$$

Parametr $\alpha = w^2$.

b) Věrohodnostní funkce má tvar

$$\Lambda(w) = \prod_{i=1}^{n} w^2\, x_i\,\mathrm{e}^{-wx_i} = w^{2\,n}\,\mathrm{e}^{-w\sum x_i}\prod_{i=1}^{n} x_i\,.$$

Po zlogaritmování a derivaci dostaneme

$$(\ln\Lambda(w))' = \left(2\,n\,\ln w - w\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\ln x_i\right)' = \frac{2\,n}{w} - \sum_{i=1}^{n} x_i\,.$$

Tento výraz položíme roven nule a parametr vyjde

$$\widehat{w} = \frac{2\,n}{\sum_{i=1}^{n} x_i} = \frac{2}{\bar{\boldsymbol{x}}}\,.$$

Cvičení 14.18. *Pomocí metody maximální věrohodnosti i metody momentů odhadněte parametr a pro hustotu*

$$f_X(t) = \begin{cases} 0\,, & t < a\,,\\ \mathrm{e}^{a-t}\,, & a \le t\,. \end{cases}$$

Úlohu řešte obecně pro realizaci $(x_1, x_2, \dots, x_n)$ *a také pro konkrétní realizaci* $(1, 2, 2, 2, 3, 3, 4)$.

Řešení. Metoda maximální věrohodnosti:
Naším cílem je maximalizovat funkci

$$\Lambda(a) = \prod_{i=1}^{n} \mathrm{e}^{a-x_i} = \mathrm{e}^{n\,a}\cdot\mathrm{e}^{-\sum x_i}\,.$$

Tato funkce nabývá maxima pro největší přípustnou hodnotu parametru
$\widehat{a} = \min\{x_1, \dots x_n\}$, resp. $\widehat{a} = 1$.

Metoda momentů:

$$m_1 = \int_{-\infty}^{\infty} t\,f_X(t)\,\mathrm{d}t = \int_{a}^{\infty} t\,\mathrm{e}^{a-t}\,\mathrm{d}t = \left[-t\,\mathrm{e}^{a-t}\right]_{t=a}^{\infty} + \int_{a}^{\infty}\mathrm{e}^{a-t}\,\mathrm{d}t =$$
$$= \left[-t\,\mathrm{e}^{a-t}\right]_{t=a}^{\infty} + \left[-\mathrm{e}^{a-t}\right]_{t=a}^{\infty} = a + 1\,.$$

Odtud pro parametr $\widehat{a}$ dostaneme $\widehat{a} = \overline{x} - 1 = \frac{\sum x_i}{n} - 1$, resp. $\widehat{a} = \frac{17}{7} - 1 \doteq 1.43$, což nedává smysl.

Cvičení 14.19. *Datový soubor* $\mathbf{x} = (-4,\,-3,\,-2,\,-1.5,\,0.5,\,1,\,2.5,\,3)$ *je realizací náhodné veličiny* X*, která má spojité rovnoměrné rozdělení v intervalu* $\langle -h,\,h\rangle$*. Metodou momentů určete odhad parametru* h*. (Ověřte, zda odhad odpovídá zadání.)*

Řešení. Rozsah souboru je $n = 8$. Náhodná veličina X má střední hodnotu rovnou nule, proto musíme použít další momenty (jen pro představu: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = -\frac{3.5}{8} = -0.4375$). Dostaneme $\mathrm{E}(X^2) = \int_{-h}^{h}\frac{x^2}{2h}\,dx = \left[\frac{x^3}{6h}\right]_{-h}^{h} = \frac{h^2}{3}$. Odhad druhého momentu je

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{47.75}{8} = 5.96875\,.$$

Odhad parametru získáme jako řešení rovnice

$$\frac{h^2}{3} = \mathrm{E}(X^2) = m_2 = \frac{47.75}{8} \quad\Longrightarrow\quad h = \sqrt{17.90625} \doteq 4.2316\,.$$

Protože všechny hodnoty ze souboru leží v intervalu $\langle -h,\,h\rangle = \langle -4.2316,\,4.2316\rangle$, můžeme nalezenou hodnotu h tudíž považovat za hledaný odhad parametru rozdělení.

Cvičení 14.20. *Za předpokladu, že soubor* $\mathbf{x} = (114, 105, 118, 119, 117, 108)$ *je realizací náhodného výběru ze spojitého rovnoměrného rozdělení v intervalu* $\langle \mu - h, \mu + h \rangle$*, odhadněte metodou momentů parametry rozdělení.*

Řešení. Hledáme odhady pro dva neznámé parametry μ a h. Použijeme tudíž rovnosti prvních dvou momentů. Pro náhodnou veličinu X, která má rovnoměrné rozdělení, dostaneme $\mathrm{E}X = \mu$, $\mathrm{D}X = \frac{h^2}{3}$, $\mathrm{E}X^2 = (\mathrm{E}X)^2 + \mathrm{D}X = \mu^2 + \frac{h^2}{3}$. Odhady parametrů získáme jako řešení soustavy rovnic

$$\widehat{\mathrm{E}X} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad \widehat{\mathrm{E}X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2\,.$$

Po dosazení hodnot ze souboru dostaneme

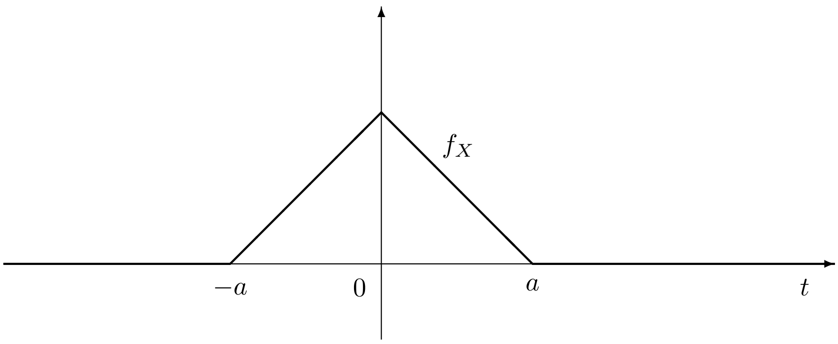
$$\overline{x} = \frac{1}{6}\cdot 681 = 113.5\,, \qquad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{6}\cdot 77\,459 \doteq 12\,909.83\,.$$

Soustava pro odhady hledaných parametrů má tvar

$$\mu = 113.5\,, \qquad \mu^2 + \frac{h^2}{3} \doteq 12909.83 \quad \Longrightarrow \quad h \doteq 9.096\,.$$

Protože je $\mu - h \doteq 104.4$ a $\mu + h \doteq 122.6$, leží všechny hodnoty ze souboru v intervalu $\langle \mu - h, \mu + h\rangle$. Nalezené hodnoty můžeme tudíž považovat za hledané odhady parametrů rozdělení.

Cvičení 14.21. *Předpokládáme, že náhodná veličina* X *má (po částech lineární) hustotu dle obrázku.*



Na základě realizace

a) $\mathbf{x} = (-2, 1, 1)$

b) $\mathbf{x} = (-1, 1, 2)$

odhadněte parametr $a > 0$.

Výsledky. $a = 3$, a to pro obě varianty, neboť hustota závisí jen na absolutní hodnotě dat.
(Druhý kořen kvadratické rovnice, 4/3, nevyhovuje požadavku $a > 2$.)

15 Testy střední hodnoty a rozptylu

15.1 Testy střední hodnoty normálního rozdělení

15.1.1 Při *známém* rozptylu σ^2

Cvičení 15.1. *Zařízení váží součástky s chybou, která má normální rozdělení o neznámé střední hodnotě a směrodatné odchylce* $\sigma = 0.66$ g. *Otestujte na hladině významnosti* 5 %, *zda je možné, aby měření nebylo zatíženo systematickou chybou (tj. střední hodnota chyb byla nulová), pokud při* $n = 9$ *kontrolních měřeních byly naměřeny tyto chyby (v gramech):*

$$0.3,\ 0.4,\ -0.8,\ 0.1,\ -1.3,\ -1.1,\ -0.6,\ 0.2,\ -0.5.$$

Řešení. Realizace výběrového průměru vychází $\bar{\boldsymbol{x}} = -3.3/9 \doteq -0.367$ g. Směrodatná odchylka výběrového průměru je $\sigma/\sqrt{n} = 0.22$ g. Testovací statistiku

$$\frac{\bar{\boldsymbol{x}}}{\sigma}\sqrt{n} \doteq \frac{-0.367}{0.22} \doteq 1.67$$

porovnáme s kvantilem $\Phi^{-1}(0.025) = -\Phi^{-1}(0.975) \doteq -1.96$ a nulovou hypotézu (že zařízení nemá systematickou chybu) nezamítáme.

Cvičení 15.2. *Teploměrem, o jehož chybě předpokládáme, že má normální rozdělení se směrodatnou odchylkou* $\sigma = 3°$*, jsme provedli* 30 *měření stejné teploty. Průměrný výsledek byl* 101°*. Otestujte na hladině významnosti* 5 %*, zda teplota nepřesahuje* 100°*.*

Výsledky. Hodnotu kritéria 1.83 porovnáme s $\Phi^{-1}(0.95) \doteq 1.64$ a nulovou hypotézu zamítáme.

15.1.2 Při *neznámém* rozptylu

Cvičení 15.3. *Testujte na hladině významnosti* $\alpha = 0.1$*, jestli následující normálně rozdělená data mají střední hodnotu* $\mu = 5$*:* $\boldsymbol{x} = (6.1, 1.2, 3.4, 8.1, 5.1, 6, 4.7, 4)$.

Řešení. Předpokládáme, že $\boldsymbol{x}$ je realizací náhodného výběru $(X_1, ..., X_8)$ z rozdělení $\mathrm{N}(\mu_X, \sigma_X^2)$, kde ani jeden z parametrů není znám. Testujeme hypotézu

$$H_0\colon \mu_X = \mu$$

proti hypotéze

$$H_1\colon \mu_X \neq \mu.$$

Používáme testovací statistiku

$$T = \frac{\overline{\boldsymbol{X}} - \mu}{S_X} \sqrt{n},$$

kde $n = 8$, $\overline{\boldsymbol{X}} = 4.825$, $\mu = 5$ a $S_X \doteq 2.059$, která má za platnosti H_0 Studentovo t-rozdělení se 7 stupni volnosti. Provádíme oboustranný test, proto $|T|$ porovnáváme s kvantilem $q_{\mathrm{t}(7)}(1 - \frac{\alpha}{2}) \doteq 1.89$. Platí $|t| \doteq 0.24 < 1.89 \doteq q_{\mathrm{t}(7)}(0.95)$, proto na hladině významnosti *nezamítáme* hypotézu, že střední hodnota tohoto rozdělení je rovna 5.

Cvičení 15.4. *Výrobce tvrdí, že spotřeba jím vyráběného automobilu je* 6 l/100 km*. Průměrná spotřeba u* 49 *uživatelů ale byla* 6.4 l/100 km*. Dále byla naměřena výběrová směrodatná odchylka* 1.6 l/100 km*. Testujte na hladině* 5 %*, zda měl výrobce pravdu, a uveďte použité předpoklady.*

Řešení. Předpoklady: spotřeba má normální rozdělení se střední hodnotou μ, měření odpovídají náhodnému výběru.

Za nulovou, resp. alternativní hypotézu si zvolíme
H_0: spotřeba = 6 l/100 km,
H_1: spotřeba ≠ 6 l/100 km.
Za platnosti H_0 má náhodná veličina

$$\frac{\overline{\boldsymbol{X}}_n - 6}{1.6}\sqrt{49}$$

rozdělení t(48). Vypočteme proto hodnotu

$$t = \frac{6.4 - 6}{1.6} \cdot 7 = 1.75\,.$$

Jelikož $|t| < q_{\mathrm{t}(48)}(0.975) \doteq 2.011$, hypotézu H_0 ve prospěch H_1 nezamítáme.
Pokud za nulovou, resp. alternativní, hypotézu zvolíme
H_0': spotřeba ≤ 6 l/100 km,
H_1': spotřeba > 6 l/100 km,
(záleží na přesné formulaci zadání), pak hodnotu $t = 1.75$ porovnáme s kvantilem $q_{\mathrm{t}(48)}(0.95) \doteq 1.68$, neboť alternativě tentokrát odpovídá „horních 5 %". A jelikož $T_0 > q_{\mathrm{t}(48)}(0.95) \doteq 1.68$, hypotézu H_0' ve prospěch H_1' zamítáme. (Předpoklad normálního rozdělení není adekvátní, ale přibližně jej lze použít.)

Cvičení 15.5. *Pacientovi byly naměřeny následující hodnoty systolického krevního tlaku* [torr]*:* 132*,* 135*,* 140*,* 153*,* 120*,* 148*,* 125*. Posuďte na hladině významnosti* 5 %*, zda střední hodnota odpovídá normálu, kterým je* 120 torr*. Uveďte použité předpoklady.*

Řešení. Výběrový průměr 136.1, výběrová směrodatná odchylka $\frac{5}{21}\sqrt{21}\sqrt{118} \doteq 11.85$, kritérium $\frac{136.1-120}{11.85}\sqrt{7} \doteq 3.6$. Nulovou hypotézu, vzhledem k hodnotám kvantilů Studentova rozdělení $q_{\mathrm{t}(6)}(0.025) \doteq -2.45$ a $q_{\mathrm{t}(6)}(0.975) \doteq 2.45$, zamítáme.

Cvičení 15.6. *Ve skupině studentů byly zjištěny výšky* (187, 181, 177, 168, 186, 185) *v* cm*. Za předpokladu, že se jedná o náhodný výběr z normálního rozdělení* $\mathrm{N}(\mu, \sigma^2)$*, testujte na hladině významnosti* $\alpha = 5$ % *hypotézu*

$H_0\colon\ \mu \geq 181$ *proti alternativě* $H_1\colon\ \mu < 181$.

Řešení. Pro testování hypotézy zvolíme jednovýběrový t-test, kde je testovací statistika

$$T = \frac{\overline{\boldsymbol{X}} - \mu}{S} \sqrt{6} \sim \mathrm{t}(5)$$

a kritický obor je $\{T \mid T < q_{\mathrm{t}(5)}(0.05) \doteq -2.02\}$. Rozsah souboru je $n = 6$, výběrový průměr $\overline{\boldsymbol{X}} = \frac{542}{3} \doteq 180.667$, výběrový rozptyl $S_X^2 = \frac{2352}{5 \cdot 9} \doteq 52.267$ a výběrová směrodatná odchylka $S_X \doteq 7.23$.

Po dosazení dostaneme realizaci statistiky

$$t = \frac{180.667 - 181}{7.23} \sqrt{6} = -0.113\,.$$

Protože je

$$t = -0.113 > q_T(0.05) \doteq -2.02\,,$$

hypotézu H_0 nezamítáme.

Cvičení 15.7. *Provádíme průzkum, jaký skutečný objem piva točí v nejmenované restauraci. Zakoupeno bylo $n = 10$ piv a jejich objem byl (v litrech):*

$$0.510,\ 0.462,\ 0.491,\ 0.466,\ 0.451,\ 0.503,\ 0.475,\ 0.487,\ 0.512,\ 0.505\,.$$

Předpokládejte, že natočený objem piva se řídí normálním rozdělením a jednotlivá měření jsou nezávislá.

a) Pro zvolenou hladinu $\alpha = 5$ % odhadněte (intervalově) střední hodnotu objemu natočeného piva.

b) Na hladině α otestujte hypotézu, že dostaneme natočeno alespoň 0.5 litru.

Řešení. V obou případech máme neznámý rozptyl. Spočítáme $\overline{\boldsymbol{x}} = 0.4862$, $s_{\boldsymbol{x}}^2 \doteq 0.0004722$ a $s_{\boldsymbol{x}} \doteq 0.02173$.

a) Kvantil Studentova rozdělení je $q_{t(n-1)}(1 - \frac{\alpha}{2}) = q_{t(9)}(0.975) \doteq 2.26$.

   Intervalový odhad střední hodnoty je

$$\left\langle \overline{\boldsymbol{x}} - \frac{s_{\boldsymbol{x}}}{\sqrt{n}}\, q_{t(n-1)}\Big(1 - \frac{\alpha}{2}\Big), \overline{\boldsymbol{x}} + \frac{s_{\boldsymbol{x}}}{\sqrt{n}}\, q_{t(n-1)}\Big(1 - \frac{\alpha}{2}\Big) \right\rangle \doteq \langle 0.4707, 0.5017 \rangle\,,$$

   protože $\frac{s_{\boldsymbol{x}}}{\sqrt{n}}\, q_{t(n-1)}(1 - \frac{\alpha}{2}) \doteq 0.01553$.

b) Nulová hypotéza $\mathbf{H}_0 : \mu \geq 0.5 = c$,
   alternativní hypotéza: $\mathbf{H}_1 : \mu < 0.5$.

   Kvantil Studentova rozdělení je $q_{t(n-1)}(1 - \alpha) = q_{t(9)}(0,95) \doteq 1.83$. Pro $T = \frac{\overline{\boldsymbol{X}} - c}{S_{\boldsymbol{x}}} \sqrt{n}$ je $t \doteq -2.008 < -1.83 \doteq -q_{t(n-1)}(1 - \alpha)$, a tedy hypotézu zamítáme.

Cvičení 15.8. *Předpokládáme, že následující data pocházejí z normálního rozdělení:*

$$28, 26, 37, 29, 27, 29, 26, 24, 43, 38, 32, 24, 26, 29, 34, 23, 18.$$

Otestujte na hladině významnosti $\alpha = 5$ % hypotézu, že toto rozdělení má střední hodnotu (právě) $\mathrm{E}X = 33$. *Uveďte použité předpoklady.*

Řešení. $n = 17$, $\bar{\boldsymbol{x}} = 29$, $s_{\boldsymbol{x}} = 6.19$,

$$t = \frac{\bar{\boldsymbol{x}} - \mathrm{E}X}{s_{\boldsymbol{x}}} \sqrt{n} \doteq \frac{29 - 33}{6.19} \cdot \sqrt{17} = -2.66$$

porovnáme s kvantily $q_{t(n-1)}(1 - \alpha/2) = 2.12$ a $q_{t(n-1)}(\alpha/2) = -q_{t(n-1)}(1 - \alpha/2) = -2.12$, tj. $-2.66 < -2.12$ a hypotézu zamítáme.

Předpoklad: Náhodné veličiny jsou nezávislé.

Cvičení 15.9. *Alice a Bob hrají hru, sestávající ze čtyř až sedmi kol. Žádné kolo nemůže skončit remízou. Hráč vyhrává, vyhraje-li čtyři kola. V následující tabulce jsou zaznamenány délky tisíce po sobě jdoucích her.*

počet kol	4	5	6	7
četnost	121	245	321	313

Za předpokladu, že Alice a Bob mají stejnou šanci na vítězství v jednom kole, je střední hodnota délky hry 5.8125. *Na hladině významnosti* $\alpha = 5\,\%$ *otestujte, zda je tomu tak.*

Řešení. Spočítáme realizaci výběrového průměru $\bar{\boldsymbol{x}} = 5.826$ a výběrového rozptylu $s^2_{\boldsymbol{x}} = 1.01273$. Spočítáme realizaci statistiky t, porovnáme s kvantilem normálního rozdělení

$$|t| = \frac{5.826 - 5.8125}{\sqrt{1.01273}}\sqrt{999} \doteq 0.425 < \Phi^{-1}(0.975) \doteq 1.96$$

a hypotézu nezamítáme. *(Použití normálního rozdělení místo Studentova je možné díky velkému rozsahu výběru.)*

Cvičení 15.10. *Náhodná veličina X je odpor rezistoru v ohmech. Posuďte na hladině významnosti* 5 % *hypotézu, že střední hodnota náhodné veličiny X je* 100 Ω, *jestliže z* 20 *nezávisle vybraných rezistorů vyšla realizace výběrového průměru* 101 Ω *a realizace výběrové směrodatné odchylky* 2 Ω.

Výsledky. Hodnotu kritéria 2.37 porovnáme s $q_{\mathrm{t}(19)}(0.975) \doteq 2.09$ a nulovou hypotézu zamítáme.

Cvičení 15.11. *Z* 10 *měření krevního tlaku u jednoho pacienta jsme obdrželi výběrový průměr* 150 torr *a výběrovou směrodatnou odchylku* 20 torr. *Rozhodněte na hladině významnosti* 5 %, *zda je střední hodnota krevního tlaku nejvýše* 140 torr. *Za jakých předpokladů výsledek platí?*

Výsledky. Hodnotu kritéria 1.58 porovnáme s $q_{\mathrm{t}(9)}(0.95) \doteq 1.83$ a nulovou hypotézu nezamítáme.

Cvičení 15.12. *Voltmetr vykázal následující četnosti chyb měření. Otestujte na hladině významnosti* 1 % *hypotézu, že má nulovou stálou chybu. Diskutujte použité předpoklady.*

chyba [mV]	−0.2	−0.1	0	0.1	0.2	0.3
četnost chyby	2	2	10	10	5	1

Výsledky. Hodnotu kritéria 2.66 porovnáme s $q_{\mathrm{t}(29)}(0.995) \doteq 2.76$ a nulovou hypotézu nezamítáme.

15.2 Testy rozptylu normálního rozdělení

Cvičení 15.13. *Soubor dat* (65, 71, 97, 72, 75) *je náhodným výběrem z normálního rozdělení* $\mathrm{N}(\mu, \sigma^2)$. *Na hladině významnosti $\alpha = 5$ % testujte hypotézu o rozptylu $H_0\colon \sigma^2 = 120$ proti alternativě $H_1\colon \sigma^2 \neq 120$.*

Řešení. K řešení úlohy použijeme jednovýběrový F-test. Náhodný výběr má 5 hodnot a rozptyl rozdělení odhadneme pomocí výběrového rozptylu,

$$\overline{\boldsymbol{X}} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{380}{5} = 76\,,$$

$$(n-1)\,S^2 = \sum_{i=1}^{5} x_i^2 - 5\,(\overline{\boldsymbol{X}})^2 = 29\,484 - 5\cdot 5\,776 \doteq 4.151\,.$$

Použijeme skutečnost, že statistika

$$T = \frac{(n-1)\,S^2}{\sigma_0^2}$$

má rozdělení $\chi^2(4)$. Po dosazení dostaneme její hodnotu

$$t = \frac{4\cdot 151}{120} \doteq 5.03\,.$$

Porovnáme s kvantily $q_{\chi^2(4)}(0.025) \doteq 0.484$, $q_{\chi^2(4)}(0.975) \doteq 11.14$ a hypotézu H_0 nezamítáme.

Cvičení 15.14. *Do laboratoře bylo odesláno* 5 *stejných vzorků krve ke stanovení obsahu alkoholu. Výsledky byly:* 0.8, 1, 0.6, 1.4, 0.9 *promile. Posuďte na hladině významnosti* 5 %, *zda směrodatná odchylka měření je nejvýše* 0.1 *promile. Uveďte použité předpoklady.*

Řešení. Z výběrového rozptylu vypočítáme testovací statistiku

$$t = \frac{(n-1)\,s^2_{\boldsymbol{x}}}{\mathrm{D}X} \doteq \frac{4\cdot 0.088}{0.1^2} \doteq 35.2\,,$$

kterou porovnáme s kvantilem $q_{\chi^2(n-1)}(1-\alpha) = q_{\chi^2(4)}(0.95) \doteq 9.49$, hypotézu zamítáme. Vycházíme z předpokladu, že chyby jednotlivých měření jsou nezávislé a mají všechny stejné normální rozdělení; potom má testovací statistika rozdělení $\chi^2(n-1)$.

Cvičení 15.15. *Z 10 měření stejného napětí nám vyšla výběrová směrodatná odchylka voltmetru 3 mV. Posuďte na hladině významnosti 5 %, zda směrodatná odchylka voltmetru je nejvýše 2 mV, jak uvádí výrobce. Uveďte použité předpoklady.*

Řešení. Z výběrového rozptylu vypočítáme testovací statistiku

$$t = \frac{(n-1)\, s_{\boldsymbol{x}}^2}{\mathrm{D}X} = \frac{81}{4} = 20.25\,,$$

kterou porovnáme s kvantilem $q_{\chi^2(n-1)}(1-\alpha) \doteq 16.92$, hypotézu zamítáme. Vycházíme z předpokladu, že chyby jednotlivých měření jsou nezávislé a mají všechny stejné normální rozdělení; potom má testovací statistika rozdělení $\chi^2(n-1)$.

Cvičení 15.16. *Generátor náhodných čísel s normovaným normálním rozdělením dal následující výsledky:*
$-0.503,\ 0.811,\ 1.078,\ -0.501,\ 0.562,\ -1.032,\ 0.152,\ 0.859,\ -0.156,\ 2.213\,.$
Posuďte na hladině významnosti 5 %, zda data odpovídají předpokládanému rozptylu.

Řešení. Výběrový průměr je 0.3483, součet kvadrátů 9, 387373. Předpokládaný rozptyl je 1, výběrový rozptyl je 0.9082, pro oboustranný odhad použijeme testovací statistiku

$$\frac{(n-1)\, S_X^2}{\mathrm{D}X} \doteq 9 \cdot 0.9082 \doteq 8.174\,,$$

která má rozdělení χ^2 s 9 stupni volnosti, porovnáme s kvantily $q_{\chi^2(9)}(0.025) \doteq 2.7$, $q_{\chi^2(9)}(0.975) \doteq 19.02$ a nulovou hypotézu nezamítáme.

Cvičení 15.17. *Předpokládáme, že následující data pocházejí z normálního rozdělení:*

9	13	11	5	15	12	11	13	11	5	9	9	7

Otestujte na hladině významnosti $\alpha = 5$ % hypotézu, že toto rozdělení má směrodatnou odchylku (právě) $\sigma_X = 2.5$. Uveďte použité předpoklady.

Řešení. $n = 13,\ s_{\boldsymbol{x}} = 3.055$,

$$t = \frac{(n-1)\, s_{\boldsymbol{x}}^2}{\sigma_X^2} \doteq \frac{12 \cdot 3.055^2}{2.5^2} = 17.92.$$

porovnáme s kvantily $q_{\chi^2(n-1)}(\alpha/2) = 4.4$ a $q_{\chi^2(n-1)}(1-\alpha/2) = 23.34$, tj. $4,4 < 17.92 < 23.34$ a hypotézu nezamítáme.

Předpoklad: Náhodné veličiny jsou nezávislé.

Cvičení 15.18. *Předpokládáme, že následující data pocházejí z normálního rozdělení:*

7	9	13	15	16	6	10	4	4	10	18	6	12	9	16	5

Otestujte na hladině významnosti $\alpha = 5$ % hypotézu, že toto rozdělení má směrodatnou odchylku (právě) $\sigma_X = 3.5$. Uveďte použité předpoklady.

Řešení. $n = 16,\ s_{\boldsymbol{x}} = 4.5753$,

$$t = \frac{(n-1)\, s_{\boldsymbol{x}}^2}{\sigma_X^2} \doteq \frac{15 \cdot 4.5753^2}{3.5^2} = 25.6.$$

porovnáme s kvantily $q_{\chi^2(n-1)}(\alpha/2) = 6.91$ a $q_{\chi^2(n-1)}(1-\alpha/2) = 28.85$ a hypotézu nezamítáme.

Předpoklad: Náhodné veličiny jsou nezávislé.

Cvičení 15.19. *Otestujte na hladině významnosti 5 %, zda směrodatná odchylka obsahu kakaové masy v čokoládě může být menší nebo rovna udávané 3 (a uveďte předpoklady!), pokud jsme u pěti náhodně vybraných kusů čokolád naměřili hodnoty*

$$25,\ 28,\ 20,\ 23,\ 26.$$

Výsledky. Hodnotu kritéria 4.13 porovnáme s $q_{\chi^2(4)}(0.95) \doteq 9.49$ a nulovou hypotézu nezamítáme.

15.3 Porovnání dvou normálních rozdělení

15.3.1 Test rozptylů dvou normálních rozdělení

Cvičení 15.20. *Oštěpařky Anežka a Bára provedly po řadě* 9 *a* 7 *hodů. Výsledky v metrech jsou zaznamenány v tabulce. Na hladině významnosti* 5 % *rozhodněte, zda Bára má rozptyl výkonů nejvýše takový jako Anežka.*

Anežka	45	52	48	60	55	51	59	47	50
Bára	60	69	75	49	53	50	47		

Řešení. Porovnáme realizace výběrových rozptylů $s_{\boldsymbol{a}}^2 \doteq 27.1$, $s_{\boldsymbol{b}}^2 \doteq 117$,

$$\frac{s_{\boldsymbol{a}}^2}{s_{\boldsymbol{b}}^2} \doteq 0.231\,,$$

srovnáme s kvantilem F-rozdělení

$$q_{\mathrm{F}(8,6)}(0.05) = \frac{1}{q_{\mathrm{F}(6,8)}(0.95)} \doteq \frac{1}{3.58} \doteq 0.28$$

a hypotézu zamítáme, tj. prohlašujeme, že Anežka má menší rozptyl než Bára.

Mohli jsme místo toho opačný poměr

$$\frac{s_{\boldsymbol{b}}^2}{s_{\boldsymbol{a}}^2} \doteq 4.33$$

srovnat s kvantilem $q_{\mathrm{F}(6,8)}(0.95) \doteq 3.58$ (se stejným výsledkem).

Cvičení 15.21. *Jeden vzorek byl rozdělen na mnoho stejných částí a zaslán opakovaně k měření dvěma laboratořím. Výsledky jsou v tabulce. Posuďte na hladině významnosti* 5 %*, zda rozptyl jejich výsledků je stejný. Uveďte použité předpoklady.*

1. laboratoř	10.1	10.3	11.1	9.7	10.4	10.8	10.4
2. laboratoř	9.8	9.6	11.3	9.3	10.5	10.7	10.2

Výsledky. Hodnotu kritéra 0.431, resp. $1/0.431 \doteq 2.32$, srovnáme s kvantily $q_{\mathrm{F}(6,6)}(0.975) \doteq 5.82$, $q_{\mathrm{F}(6,6)}(0.025) = 1/q_{\mathrm{F}(6,6)}(0.975) \doteq 1/5.82 \doteq 0.172$, hypotézu nezamítáme.

Další úlohy na testy rozptylů dvou rozdělení jsou v kapitole 15.3.3.

15.3.2 Testy středních hodnot dvou normálních rozdělení se známým rozptylem σ^2

Cvičení 15.22. *Stejný rezistor (jehož odpor považujeme za dostatečně stálý) jsme měřili dvěma ohmmetry se směrodatnou odchylkou* $\sigma = 0.01\ \Omega$*, a to každým a) jednou, b)* $5\times$*. Realizace výběrového průměru měření byly* $\bar{\boldsymbol{x}} = 3.002\ \Omega$ *pro první ohmmetr,* $\bar{\boldsymbol{y}} = 3.020\ \Omega$ *pro druhý. Posuďte na hladině významnosti* 1 %*, zda je soustavný rozdíl mezi údaji obou ohmmetrů.*

Řešení. Je-li n rozsah výběrů (zde u obou stejný), pak kritériem je

$$\frac{|\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}|}{\sigma} \sqrt{\frac{n}{2}}\,,$$

což vychází a) 1.27, b) 2.95. Porovnáme s kvantilem $\Phi^{-1}(0.995) \doteq 2.576$ a nulovou hypotézu zamítáme pouze v případě b).

Cvičení 15.23. *Teploměrem jsme měřili teplotu na dvou místech, jednu* $5\times$*, druhou* $10\times$*. Vyšly nám realizace výběrového průměru* 38.1° *a* 38.7°*. Předpokládáme, že rozhodující vliv na chybu měření má směrodatná odchylka teploměru* $\sigma = 0.5^\circ$*. Posuďte na hladině významnosti* 5 %*, zda je druhá teplota vyšší.*

Výsledky. Nulová hypotéza je, že první teplota je menší nebo rovna druhé (alternativa, že je vyšší). Hodnotu kritéria

$$\frac{|\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}|}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \doteq 2.19$$

porovnáme s kvantilem $\Phi^{-1}(0.95) \doteq 1.645$ a nulovou hypotézu zamítáme.

15.3.3 Testy středních hodnot dvou normálních rozdělení se (stejným) neznámým rozptylem

Cvičení 15.24. *U testovací skupiny* 20 *pacientů, kterým byl podáván lék na snížení krevního tlaku, byla naměřena realizace výběrového průměru* 140 torr*, realizace výběrové směrodatné odchylky* 20 torr*. U srovnávací skupiny* 50 *pacientů, kterým lék nebyl podáván, byl naměřena realizace výběrového průměru* 150 torr*, realizace výběrové směrodatné odchylky* 15 torr*. Posuďte, zda je tím prokázána účinnost léku na hladině významnosti* 1 %*. Uveďte použité předpoklady.*

Řešení. $\bar{\boldsymbol{x}} = 140,\ s_{\boldsymbol{x}} = 20,\ m = 20,$
$\bar{\boldsymbol{y}} = 150,\ s_{\boldsymbol{y}} = 15,\ n = 50,$
$H_0' : s_{\boldsymbol{x}}^2 = s_{\boldsymbol{y}}^2, \qquad H_1' : s_{\boldsymbol{x}}^2 \neq s_{\boldsymbol{y}}^2$
$\frac{s_{\boldsymbol{x}}^2}{s_{\boldsymbol{y}}^2} = \frac{16}{9} \doteq 1.778$ porovnáme s

$$q_{F(19,49)}(0.995) \doteq 2.47, \qquad q_{F(19,49)}(0.005) = \frac{1}{q_{F(49,19)}(0.995)} \doteq \frac{1}{2.96} \doteq 0.338,$$

hypotézu H_0' o rovnosti rozptylů nezamítáme.

Odhad rozptylu a směrodatné odchylky:

$$s^2 = \frac{(m-1)\ s_{\boldsymbol{x}}^2 + (n-1)\ s_{\boldsymbol{y}}^2}{m+n-2} \doteq 273.9, \qquad s \doteq \sqrt{273.9} \doteq 16.55,$$

$H_0 : \mathrm{E}X \geq \mathrm{E}Y, \qquad H_1 : \mathrm{E}X < \mathrm{E}Y$

$$t = \frac{\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \doteq \frac{-10}{16.55\sqrt{\frac{1}{20} + \frac{1}{50}}} \doteq -2.284$$

porovnáme s $q_{t(68)}(0.01) = -q_{t(68)}(0.99) \doteq -2.38$ a hypotézu, že lék nesnižuje krevní tlak, nezamítáme na hladině významnosti 1 %. (Mohli bychom ji zamítnout na hladině významnosti 5 %, pro tu je $q_{t(68)}(0.05) = -q_{t(68)}(0.95) \doteq -1.66$.)

Předpoklady: normální rozdělení (stejné uvnitř každého souboru), nezávislost, stejné rozptyly.

Cvičení 15.25. *Stejnou veličinu jsme měřili dvěma metodami, každou* $10\times$*. Výsledky shrnuje následující tabulka.*

metoda \ *parametr*	*výběrový průměr*	*výběrová směrodatná odchylka*
1.	20	3
2.	21	5

Posuďte na hladině významnosti 5 %*, zda lze považovat obě metody za stejně přesné a jejich střední hodnoty za stejné.*

Výsledky. Testovací kritérium pro rovnost rozptylů $25/9 \doteq 2.78$, resp. $9/25 = 0.36$, srovnáme s kvantily $q_{\mathrm{F}(9,9)}(0.975) \doteq 3.2$, $q_{\mathrm{F}(9,9)}(0.025) = 1/q_{\mathrm{F}(9,9)}(0.975) \doteq 1/3.2 \doteq 0.31$, rovnost rozptylů nezamítáme.

Testovací kritérium pro rovnost středních hodnot vychází

$$\sqrt{\frac{10}{3^2 + 5^2}} \doteq 0.54\,,$$

porovnáme s $q_{\mathrm{t}(18)}(0.975) \doteq 2.1$, rovnost středních hodnot nezamítáme.

Cvičení 15.26. *V řetězcích A a B jsme koupili* 11 *balíčků cukru a jejich zvážením jsme dospěli k těmto hodnotám:*

parametr \ *řetězec*	*A*	*B*
výběrový průměr	0.951 kg	0.912 kg
výběrový rozptyl	0.021 kg^2	0.067 kg^2
výběrová směrodatná odchylka	0.144 kg	0.258 kg

Je možné na základě těchto dat zamítnout na hladině významnosti 5 % *hypotézu, že střední hodnoty hmotnosti balíčků cukru v těchto dvou řetězcích jsou stejné?*

Výsledky. Testovací kritérium pro rovnost rozptylů 3.19, srovnáme s $q_{\mathrm{F}(10,10)}(0.975) \doteq 3.72$, rovnost rozptylů nezamítáme.

Testovací kritérium pro rovnost středních hodnot vychází 0.42 porovnáme s $q_{t(20)}(0.975) \doteq 2.09$, rovnost středních hodnot nezamítáme.

Cvičení 15.27. *Máme dva postupy, jimiž z vody filtrujeme dusičnany. Realizovali jsme* 7 *měření při postupu A a* 5 *měření při postupu B. Obsah dusičnanů (v* mg /100 ml*) po filtraci udává následující tabulka. Rozhodněte, zda rozdíl v účinnosti filtrace můžeme považovat za statisticky významný.*

postup A	3	2.1	0.9	4.1	1.6	2.3	3.1
postup B	3	5	4	2.4	1.8		

Výsledky. $\bar{\boldsymbol{a}} = 2.442$, $s_{\boldsymbol{a}} = 1.058$, $\bar{\boldsymbol{b}} = 3.24$, $s_{\boldsymbol{b}} = 1.276$; testovací statistika $t = 1.185$ nedovoluje zamítnout ani v oboustranném, ani v jednostranném testu na hladině významnosti 5 %.

15.4 Testy středních hodnot dvou normálních rozdělení – párový test

15.4.1 Pro *známý* rozptyl σ^2

Cvičení 15.28. *Různá napětí jsme měřili vždy současně dvěma voltmetry se směrodatnou odchylkou* $\sigma = 10$ mV. *Naměřili jsme následující dvojice hodnot* [V]*:*

X	3	3.5	4	4.5	5	4.5	4	3.5	3	2
Y	3.005	3.506	4.011	4.512	5.019	4.518	4.010	3.508	2.999	1.995

Posuďte na hladině významnosti 1 % *hypotézu, že střední hodnoty údajů voltmetrů jsou stejné.*

Řešení. Testovací statistiku

$$t = \frac{|\bar{X} - \bar{Y}|}{\sigma}\sqrt{\frac{n}{2}} = \frac{0.0083}{0.01}\sqrt{5} \doteq 1.86$$

porovnáme s kvantilem normovaného normálního rozdělení $\Phi^{-1}(0.995) \doteq 2.576$ a hypotézu nezamítáme.

Cvičení 15.29. *Dvěma stejnými teploměry jsme měřili teploty současně na dvou různých stanovištích. Předpokládáme, že pro chybu měření je rozhodující jejich směrodatná odchylka* $\sigma = 2$°C. *(Teploměry jsou zkalibrovány, takže jejich systematickou odchylku považujeme za zanedbatelnou.) Naměřené hodnoty jsou v tabulce. Posuďte na hladině významnosti* 5 %*, zda teplota na druhém stanovišti je vyšší než na prvním.*

1. teploměr	10.5	12.1	13.3	15.2	14.8	11.6	10.9	8.9
2. teploměr	11.6	13.3	16.6	17.8	16.0	14.3	12.8	11.2

Řešení. Nemůžeme testovat ostrou nerovnost, ale místo ní posoudíme nulovou hypotézu, že teplota na prvním stanovišti je větší nebo rovná teplotě na druhém. Realizace výběrov0ho průměru rozdílu teplot mezi prvním a druhým stanovištěm je $\bar{\boldsymbol{\delta}} = -2.0125$. Podle nulové hypotézy by měla být nezáporná, testovací statistiku

$$\frac{\bar{\boldsymbol{\delta}}}{\sigma}\sqrt{\frac{n}{2}} = \frac{-2.0375}{2}\sqrt{4} \doteq -2.0125$$

porovnáme s kvantilem $\Phi^{-1}(0.05) \doteq -1.96$ a hypotézu zamítáme. Teplota na prvním stanovišti není větší nebo rovná teplotě na druhém, tedy na druhém je větší.

15.4.2 Pro *neznámý* rozptyl

Cvičení 15.30. *U* 8 *praváků jsme změřili délku prostředníčku na pravé a levé ruce, hodnoty v milimetrech uvádí tabulka.*

pravá	84	76	89	85	80	69	58	68
levá	81	74	90	84	77	67	59	70

Na hladině významnosti 5 % *posuďte hypotézu, že praváci mají delší prostředníček na pravé ruce, a uveďte předpoklady.*

Řešení. Označme X_i a Y_i po řadě délky prostředníčků u praváků z náhodného výběru na pravé a levé ruce (s realizacemi x_i, y_i v tabulce), $i \in \{1, ..., 8\}$. Předpokládáme, že náhodné veličiny $\varDelta_i = X_i - Y_i$ mají stejné normální rozdělení $\mathrm{N}(\mathrm{E}\varDelta, \mathrm{D}\varDelta)$. Nemůžeme testovat nulovou hypotézu, že $\mathrm{E}\varDelta > 0$ (aniž bychom specifikovali minimální dovolenou hodnotu rozdílu). Místo toho můžeme testovat nulovou hypotézu, že $\mathrm{E}\varDelta \leq 0$; pokud ji zamítneme, můžeme tvrdit, že praváci mají delší prostředníček na pravé ruce. Naše hypotézy jsou:

$$\begin{aligned} H_0: &\quad \mathrm{E}\varDelta \leq 0\,, \\ H_1: &\quad \mathrm{E}\varDelta > 0\,. \end{aligned}$$

Vypočteme realizace výběrového průměru a rozptylu:

$$\bar{\boldsymbol{\delta}} = 0.875\,, \quad s_{\boldsymbol{\delta}}^2 = \frac{1}{7}\sum_{i=1}^{8}(\delta_i - \bar{\boldsymbol{\delta}})^2 \doteq 3.8393\,.$$

Testujeme realizaci statistiky

$$t = \frac{\bar{\boldsymbol{\delta}}}{s_{\boldsymbol{\delta}}}\sqrt{8} \doteq 1.26$$

na Studentovo t-rozdělení se 7 stupni volnosti. Zamítací kritérium je $t > q_{\mathrm{t}(7)}(0.95) \doteq 1.89$, neboť jde o jednostranný test. Proto na hladině významnosti 5 % nezamítáme hypotézu, že mají praváci pravý prostředníček nejvýše tak dlouhý jako levý.

Cvičení 15.31. *U dvou benzínových stanic byly vždy v tutéž dobu sledovány ceny benzínu, výsledky jsou v tabulce:*

X	32.50	32.20	31.30	30.60	29.20	27.60	27.20	26.90	25.90	25.90	23.90	23.90
Y	32.70	32.30	31.50	30.60	29.30	27.70	27.40	26.70	26.50	25.50	24.90	23.50

Posuďte na hladině významnosti 5 % *hypotézu, že benzín u stanice* X *není levnější. Uveďte použité předpoklady.*

Řešení. Rozdíly cen jsou
$\boldsymbol{\delta} = (-0.2, -0.1, -0.2, 0, -0.1, -0.1, -0.2, 0.2, -0.6, 0.4, -1, 0.4)$,
$n = 12$, $\bar{\boldsymbol{\delta}} = -0.125$, $s_{\boldsymbol{\delta}}^2 \doteq 0.153$, $s_{\boldsymbol{\delta}} \doteq 0.391$,

$$t = \frac{\bar{\boldsymbol{\delta}}}{s_{\boldsymbol{\delta}}}\sqrt{n} \doteq -1.107\,,$$

porovnáme s kvantilem $q_{t(11)}(0.05) = -q_{t(11)}(0.95) \doteq -1.80$ a nulovou hypotézu nezamítáme.
Předpoklady pro párový test: střední hodnoty náhodných veličin v obou výběrech kolísají stejně, odchylky od nich mají normální rozdělení a jsou nezávislé.

Cvičení 15.32. *Má se rozhodnout, zda se u automobilů dané značky při seřízení geometrie vozu sjíždějí obě přední pneumatiky stejně rychle. Bylo proto vybráno 8 vozů a po jisté době změřeno, o kolik* mm *se sjely jejich pravé a levé pneumatiky. Výsledky udává tabulka.*

pravá	1.1	2	0.9	2.1	1.6	1.7	1.8	1.4
levá	1.6	2.3	1.3	1.9	1.9	1.6	2	1.9

Výsledky. $\bar{\boldsymbol{\delta}} = -0.2375$, $s_{\boldsymbol{\delta}} \doteq 0.262$, hodnota kritéria $-2.57 < -2.36 \doteq q_{\mathrm{t}(7)}(0.025)$, hypotézu zamítáme.

Cvičení 15.33. *Testujeme vliv kouření na zdraví jedince. U* 41 *párů dvojčat ženského pohlaví, kde jedno z dvojčat je nekuřačka a druhé je kuřačka, byla změřena hustota kostní hmoty bederní páteře. Výsledky měření pro nekuřačky označme* N_i*, výsledky měření pro kuřačky označme* K_i *a nakonec položme* $X_i = N_i - K_i$*. V následující tabulce jsou naměřené hodnoty pro průměr hustoty kostí, směrodatné odchylky, průměr rozdílu hustoty kostí dvojčat a tomu odpovídající směrodatné odchylky*

$\bar{\boldsymbol{n}}$	$\bar{\boldsymbol{k}}$	$\bar{\boldsymbol{x}}$	$s_{\boldsymbol{n}}$	$s_{\boldsymbol{k}}$	$s_{\boldsymbol{x}}$
0.795	0.759	0.036	0.1281	0.1345	0.0896

Otestujte na hladině významnosti 5 % *hypotézu, že kouření neškodí zdraví (nesnižuje hustotu kostí).*

Výsledky. Hodnota kritéria $\frac{0.036}{0.0896}\sqrt{41} \doteq 2.57 > 1.68 \doteq q_{\mathrm{t}(40)}(0.95)$, hypotézu zamítáme (dokonce i na hladině významnosti 1 %, které odpovídá kvantil $q_{\mathrm{t}(40)}(0.99) \doteq 2.42$). *(Ostatní zadané hodnoty nebyly potřeba.)*

16 χ^2-test dobré shody

16.1 χ^2-test dobré shody rozdělení a modelu

Cvičení 16.1. *Účastníci konference budou ubytováni ve čtyřpatrovém penziónu s 12 pokoji, v každém patře jsou tři pokoje se dvěma lůžky. Každý z $n = 20$ účastníků poslal organizátorům nezávisle svůj požadavek čísla pokoje, kde by chtěl být ubytovaný. Čísla byla následující*

$$8, 12, 5, 4, 3, 5, 6, 12, 11, 2, 6, 4, 2, 12, 11, 9, 6, 7, 9, 9.$$

Otestujte na hladině významnosti $\alpha = 5$ % hypotézu
H_0: rozdělení účastníků do pater je rovnoměrné
proti alternativě
H_1: rozdělení účastníků do pater není rovnoměrné.

patro	1	2	3	4
čísla pokojů	$1-3$	$4-6$	$7-9$	$10-12$

Řešení. K rozhodování použijeme χ^2-test dobré shody. Setřídíme data do skupin a vypočteme empirické četnosti, které zapíšeme spolu s teoretickými četnostmi do tabulky

i	1	2	3	4
čísla pokojů	$1-3$	$4-6$	$7-9$	$10-12$
n_i	3	7	5	5
np_i	5	5	5	5

Ze zadání dostaneme $n = 20$ počet dat, $k = 4$ počet tříd a $p_i = 0.25$, $1 \leq i \leq 4$, pro rovnoměrné rozdělení. Dosadíme do vzorce pro statistiku testu, která má v tomto případě rozdělení přibližně $\chi^2(3)$.

$$\sum_{i=1}^{4} \frac{(n_i - np_i)^2}{np_i} = \frac{1}{5}\,(2^2 + 2^2) = 1.6\,,$$

porovnáme s kvantilem $q_{\chi^2(3)}(0.95) = 7.81$ a hypotézu H_0 nezamítáme.

Cvičení 16.2. *Realizací náhodného výběru jsme dostali následující četnosti hodnot:*

hodnota	0	1	2	3	4	5
pozorovaná četnost	2	7	15	12	3	1

Posuďte na hladině významnosti 5 % *hypotézu, že výběr pochází z binomického rozdělení* $\mathrm{Bi}\,(5, q)$*, kde q neznáme.*

Řešení. Odhad q metodou momentů: $\mathrm{E}X = 5\,q = \bar{\boldsymbol{x}} = 2.25$, $q = 0.45$. Stejný výsledek dává i metoda maximální věrohodnosti, viz [6, str. 180]. Pravděpodobnostní funkce je $p_k = p_X(k) = \binom{5}{k}\,q^k\,(1-q)^{5-k}$.

hodnota k	0	1	2	3	4	5
pozorovaná četnost n_k	2	7	15	12	3	1
teoretická četnost $n\,p_k$	2.013	8.236	13.476	11.026	4.511	0.738

Pro $k \in \{0, 5\}$ vychází teoretická četnost příliš malá, musíme sdružit třídy:

hodnota k	$0-1$	2	3	$4-5$
pozorovaná četnost n_k	9	15	12	4
teoretická četnost $n\,p_k$	10.2	13.5	11	5.2
příspěvek ke kritériu $\dfrac{(n_k - n\,p_k)^2}{n\,p_k}$	0.15	0.17	0.09	0.3

Hodnotu kritéria 0.71 porovnáme s kvantilem $q_{\chi^2(2)}\,(0.95) \doteq 5.99$ a hypotézu nezamítáme.

Cvičení 16.3. *Firma má tři pobočky. Dva roky bylo sledováno, která z nich zaznamenala nejvyšší měsíční výnos. Bylo zjištěno, že nejvýnosnější byla první pobočka desetkrát, druhá šestkrát a třetí osmkrát. Je možné říct, že první pobočka je nejvýnosnější dvakrát častěji druhá pobočka a přitom druhá pobočka je nejvýnosnější stejně často jako třetí pobočka? Testujte na hladině* 5 %.

Řešení. Označme p_i pravděpodobnost, že nejvýnosnější pobočkou měsíce bude i-tá pobočka a X_i počet měsíců, kdu byla i-tá pobočka nejvýnosnější pobočkou měsíce. Za nulovou, resp. alternativní, hypotézu si zvolíme
$H_0\colon p_1 = \frac{1}{2}, p_2 = p_3 = \frac{1}{4}$,
H_1: hodnoty jsou jiné.
Za platnosti H_0 má náhodná veličina

$$\sum_{i=1}^{3} \frac{(X_i - 24 \cdot p_i)^2}{24 \cdot p_i}$$

rozdělení χ^2_2. Vypočteme proto hodnotu

$$\chi^2 = \frac{(10-12)^2}{12} + \frac{(6-6)^2}{6} + \frac{(8-6)^2}{6} = 1.$$

Jelikož $\chi^2 < \chi^2_{2;0.95} = 5.992$, hypotézu H_0 ve prospěch H_1 nezamítáme.

Cvičení 16.4. *Pro soubor dat* (22, 26, 19, 18, 17, 19, 20, 26, 25, 16, 20, 18, 16, 26, 25, 23, 20, 21, 23, 23)*, která jsou bodovými hodnoceními testu v předmětu A0B01PSI, testujte na hladině významnosti* $\alpha = 5$ % *hypotézu*
H_0: *výběr je z rovnoměrného rozdělení rozsahu* $\langle 15, 26\rangle$
proti alternativě
H_1: *rozdělení není rovnoměrné.*
(Volte stejnou šířku třídy $d = 3$.*)*

Řešení. Setřídíme data do skupin a vypočteme empirické četnosti, které zapíšeme spolu s teoretickými četnostmi do tabulky:

i	1	2	3	4
rozsah	$15-17$	$18-20$	$21-23$	$24-26$
n_i	3	7	5	5
$n\,p_i$	5	5	5	5

Ze zadání dostaneme $n = 20$ počet dat, $k = 4$ počet tříd a $p_i = 0.25$, $1 \le i \le 4$, pro rovnoměrné rozdělení. Dosadíme do vzorce pro statistiku testu, která má v tomto případě rozdělení $\chi^2(3)$.

$$\sum_{i=1}^{4} \frac{(n_i - n\,p_i)^2}{n\,p_i} = \frac{1}{5}\,(2^2 + 2^2) = 1.6$$

porovnáme s kvantily $q_{\chi^2(3)}(0.95) = 7.81$, resp. $q_{\chi^2(3)}(0.025) = 0.216$ a $q_{\chi^2(3)}(0.975) = 9.35$, a hypotézu H_0 nezamítáme.

Cvičení 16.5. *Náhodná veličina* X *nabývá hodnot z množiny* $\{1,2,3,4,5,6\}$ *s pravděpodobností*

$$P[X = i] = \gamma\, i^2\,.$$

Na hladině významnosti $\alpha = 5$ % *otestujte hypotézu, že následující data pocházejí z tohoto rozdělení.*

hodnota	1	2	3	4	5	6
četnost	10	20	100	140	270	330

Řešení. Aby P byla pravděpodobnost, musí platit $\gamma = 1/91$.

hodnota	1	2	3	4	5	6
četnost	10	20	100	140	270	330
teor. pr.	1/91	4/91	9/91	16/91	25/91	36/91
teor. čet.	9.56	38.24	86.04	152.97	239.01	344.18
χ^2	0.02	8.7	2.26	1.1	4.02	0.58

$$t = \sum_{k=1}^{6} \frac{(n_i - n\,p_i)^2}{n\,p_i} \doteq 0.02 + 8.7 + 2.26 + 1.1 + 4.02 + 0.58 = 16.68\,.$$

Statistiku porovnáme s kvantilem $q_{\chi^2(6-1)}(0.95) \doteq 11.07$ (případně s kvantily $q_{\chi^2(6-1)}(0.975) \doteq 12.83$ a $q_{\chi^2(6-1)}(0.025) \doteq 0.83$) a hypotézu zamítáme.

Cvičení 16.6. *Na hladinách významnosti* 5 %, *resp.* 1 %, *otestujte hypotézu, že následující data pocházejí z binomického rozdělení* Bi(3, 5/6).

hodnota	0	1	2	3
četnost	3	15	35	47

Řešení.

hodnota	0	1	2	3
n_i	3	15	35	47
p_i	0.005	0.07	0.347	0.578
np_i	0.5	7	34.7	57.8

Hodnota 0 má příliš malou teoretickou četnost, sloučíme ji s hodnotou 1:

hodnota	0-1	2	3
n_i	18	35	47
p_i	0.075	0.347	0.578
np_i	7.5	34.7	57.8
χ_i^2	14.7	0.003	2.02

Hodnotu kritéria 16.7 porovnáme s $q_{\chi^2(2)}(1-\alpha) \doteq 5.99$, resp. 9.21, pro $\alpha = 5$ %, resp. 1 %, a hypotézu zamítáme.

Cvičení 16.7. *Počty narozených dětí v České republice během let* 2006 *až* 2011 *(zaokrouhleny na tisíce) jsou uvedeny v tabulce. Testujte na hladině významnosti* 5 % *hypotézu, že mají data rovnoměrné rozdělení. Posuďte předpoklady a použitelnost metody.*

rok	2006	2007	2008	2009	2010	2011
četnost	106 000	115 000	120 000	118 000	117 000	107 000

Řešení. Použijeme χ^2-test dobré shody. Máme $k = 6$ tříd a celkový počet je $n = 683\,000$. Předpokládáme rovnoměrné rozdělení, tedy pravděpodobnosti příslušnosti k jednotlivým třídám jsou stejné, a to $p_i = 1/6,\ 1 \le i \le 6$, $n\,p_i \doteq 113\,833$. Statistika testu je dána vzorcem

$$T = \sum_{i=1}^{6} \frac{(n_i - np_i)^2}{np_i}$$

a má přibližně rozdělení χ^2 s 5 stupni volnosti. Po dosazení získáme její realizaci

$$\begin{aligned} t &= \frac{1}{113\,833}\left(7\,833^2 + 1\,167^2 + 6\,167^2 + 4\,167^2 + 3\,167^2 + (-6\,833)^2\right) \doteq \\ &\doteq \frac{10^6}{113\,833}(61.356 + 1.362 + 38.032 + 17.363 + 10.029 + 46.69) = \\ &= \frac{174.833 \cdot 10^6}{113\,833} \doteq 1\,536\,. \end{aligned}$$

Porovnáme s kvantilem $q_{\chi^2(5)}(0.95) \doteq 11.07$ rozdělení $\chi^2(5)$. Protože je $t = 1\,536 \gg 11.07$, hypotézu, že data jsou výběrem z rovnoměrného rozdělení, zamítáme, a to na velmi vysoké hladině významnosti. Je to tím, že data neodpovídala přesně rovnoměrnému rozdělení a bylo jich mnoho. Na výsledku nic nemění skutečnost, že se předpokládanému jen asymptoticky blížíme a že data byla zaokrouhlena.

Cvičení 16.8. *Po dobu jednoho měsíce jsme měřili počet aut, který projel mezi 11:59 a 12:00 křižovatkou na Karlově náměstí. Jednotlivé četnosti jsou uvedeny v tabulce.*

počet aut	4	5	6	7	8	9	10	11	12	13	14	15
četnost	3	3	1	3	2	2	3	5	3	1	3	1

Ověřte na hladině $\alpha = 5$ % *hypotézu, že se jedná o geometrické rozdělení se střední hodnotou* 9.

Řešení. Pro střední hodnotu 9 je parametr $q = 0.9$. Příspěvky jednotlivých pozorování jsou v následující tabulce, přičemž kvantilová funkce má nejvýše hodnotu $q_{\chi^2(11)}(0.95) = 19.68$ (kdybychom neslučovali do méně skupin). Stačí se například podívat na hodnotu pro počet aut $i = 11$ a hypotézu na hladině 5 % zamítáme.

i	4	5	6	7	8	9
n_i	3	3	1	3	2	2
p_i	0.066	0.059	0.053	0.048	0.043	0.039
np_i	1.968	1.771	1.594	1.435	1.291	1.162
t_i	2.097	5.884	12.174	21.584	34.850	52.854

i	10	11	12	13	14	15
n_i	3	5	3	1	3	1
p_i	0.035	0.031	0.028	0.025	0.023	0.021
np_i	1.046	0.941	0.847	0.763	0.686	0.618
t_i	76.645	107.469	146.801	196.385	258.274	334.888

Cvičení 16.9. *Chceme zjistit, zda si jistý druh ptáka buduje hnízda rovnoměrně po krajině. K tomu jsme rozdělili testovací region na* 6 *souvislých částí, jejichž rozlohy v* km^2 *jsou uvedeny v tabulce. Tabulka udává i počet hnízd nalezených v dané části regionu. Za hladinu významnosti považujte* $\alpha = 5$ %.

oblast	*A*	*B*	*C*	*D*	*E*	*F*
rozloha	*5*	*10*	*10*	*5*	*15*	*15*
počet hnízd	*14*	*22*	*28*	*12*	*40*	*34*

Řešení. Využijeme test dobré shody. Počet hnízd v regionu je $n = 150$. Označíme X_i, $i \in \{1, ..., n\}$, veličinu nabývající hodnot $A, ..., F$, v závislosti na tom, v které části regionu leží i-té hnízdo. Dále označme $p_j = P[X_i = j]$, $j \in \{A, ..., F\}$ (neboť předpokládáme, že X_i jsou stejně rozdělené). Můžeme tedy doplnit tabulku následovně:

oblast	A	B	C	D	E	F
rozloha	5	10	10	5	15	15
počet hnízd n_j	14	22	28	12	40	34
teoretická pravděpodobnost p_j	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{3}{12}$	$\frac{3}{12}$
teoretický počet hnízd np_j	12.5	25	25	12.5	37.5	37.5
příspěvek ke kritériu $\frac{(n_j-np_j)^2}{np_j}$	0.18	0.36	0.36	0.02	0.167	0.327

Testovací statistika

$$T = \sum_{j=A}^{F} \frac{(n_j - np_j)^2}{np_j}$$

má za předpokladu rovnoměrného rozdělení hnízd v regionu χ^2-rozdělení o 5 stupních volnosti. Z našich realizací získáme $t \doteq 1.413$, což porovnáme s tabulkovou hodnotou kvantilu $q_{\chi^2(5)}(1-\alpha) \doteq 11.1 > t$. Na hladině významnosti 5 % tedy nemůžeme zamítnout, že si pták buduje hnízda rovnoměrně po krajině.

Cvičení 16.10. *Sportovec* 25× *prohrál (*0 *bodů),* 118× *remizoval (*1 *bod) a* 123× *vyhrál (*2 *body). Posuďte na hladině významnosti* 5 %*, zda tato data vyhovují binomickému rozdělení* $\mathrm{Bi}(2, q)$*, kde* $q \in \langle 0, 1 \rangle$ *je neznámý parametr.*

Výsledky. Metoda mometů vede k odhadu $q \doteq 0.684$, což je polovina realizace výběrového průměru, metoda maximální věrohodnosti vede ke stejnému odhadu. Hodnotu kritéria 0.188 porovnáme s kvantilem $q_{\chi^2(1)}(0.95) \doteq 3.84$ a nulovou hypotézu nezamítáme.

Cvičení 16.11. *Posuďte na hladině významnosti* 5 %*, zda data v tabulce odpovídají následujícímu pravděpodobnostnímu modelu: Každý rok je přijímán stejný počet studentů (*1200*), z každého ročníku do dalšího postoupí* 80 %*, ostatní fakultu opustí.*

ročník	1	2	3	4	5
počet studentů	1200	860	650	530	450

Výsledky. *(Jedná se o oříznuté geometrické rozdělení.)* Hodnotu kritéria 43.6 porovnáme s kvantilem $q_{\chi^2(3)}(0.95) \doteq 7.81$ a nulovou hypotézu zamítáme. K tomu stačil např. příspěvek 18.1 pro 3. ročník.

Cvičení 16.12. *Sledujeme, zda při bombardování Londýna padaly bomby rovnoměrně po celém Londýně. K tomu jsme si Londýn rozdělili na 7 stejně velkých oblastí. Počet bomb dopadlých do jednotlivých oblastí udává tabulka. Posuďte na hladině významnosti $\alpha = 5$ %.*

oblast	1	2	3	4	5	6	7	*Celkem*
počet bomb	25	18	32	20	38	16	26	175

Výsledky. Hodnotu kritéria 14.96 porovnáme s kvantilem $q_{\chi^2(6)}(0.95) \doteq 12.59$ a nulovou hypotézu zamítáme.

Cvičení 16.13. *Úspěšnost u zkoušek ve vztahu k počtu přítomných studentů udává tabulka:*

počet přítomných	16	28	36	50	45	43	50	46	37	24
počet úspěšných	8	22	26	17	26	16	15	17	15	6

Otestujte na hladině významnosti 5 % hypotézu, že pravděpodobnost úspěchu byla u všech zkouškových termínů stejná. (Data jsou z předmětu MVT v zimním semestru 2010/11.)

Výsledky. Hodnotu kritéria 22.14 porovnáme s kvantilem $q_{\chi^2(9)}(0.95) \doteq 16.92$ a nulovou hypotézu zamítáme. Otázkou zůstává, zda je to způsobeno rozdíly v obtížnosti písemek, v přípravě studentů přihlášených na různé termíny, nebo něčím jiným.

Cvičení 16.14. *Tabulka uvádí, kolik z respondentů odpovědělo v průzkumu na otázku kladně, v závislosti na vzdělání. Máme důvod se domnívat, že odpověď závisí na vzdělání?*

ukončené vzdělání	*počet respondentů*	*počet kladných odpovědí*
žádné	5	1
základní	195	10
střední	450	14
vyšší střední	150	10
vysokoškolské	200	15
celkem	1000	50

Výsledky. První skupina má malou teoretickou četnost, musíme ji sdružit s jinou, nejlépe následující („žádné nebo základní"). Hodnotu kritéria 6.64 porovnáme s kvantilem $q_{\chi^2(3)}(0.95) \doteq 7.81$ a nulovou hypotézu nezamítáme.

16.2 χ^2-test dobré shody dvou rozdělení a nezávislosti dvou rozdělení

Cvičení 16.15. *U 120 osob byla pozorována výše jejich platu a schopnost splácet úvěr. Naměřeny byly následující sdružené četnosti:*

plat \ *schopnost splácet*	*špatná*	*dobrá*
nízký	10	20
střední	15	45
vysoký	5	25

Jsou vlastnosti plat a schopnost splácet úvěr nezávislé? Testujte na hladině 5 %.

Řešení. Za nulovou, resp. alternativní, hypotézu si zvolíme
H_0: veličiny jsou nezávislé,
H_1: veličiny nejsou nezávislé.
Za platnosti H_0 má náhodná veličina

$$\sum_{i=1}^{3}\sum_{j=1}^{2}\frac{\left(n_{ij}-\frac{n_{i.}\cdot n_{.j}}{n}\right)^2}{\frac{n_{i.}\cdot n_{.j}}{n}}$$

rozdělení $\chi^2_{(3-1)(2-1)} = \chi^2_2$. Doplníme proto tabulku o marginální četnosti

plat \ schopnost splácet	špatná	dobrá	celkem
nízký	10	20	30
střední	15	45	60
vysoký	5	25	30
celkem	30	90	120

a vypočteme hodnotu

$$\chi^2 = \frac{(10 - \frac{30\cdot 30}{120})^2}{\frac{30\cdot 30}{120}} + \frac{(20 - \frac{30\cdot 90}{120})^2}{\frac{30\cdot 90}{120}} + \frac{(15 - \frac{60\cdot 30}{120})^2}{\frac{60\cdot 30}{120}} + \frac{(45 - \frac{60\cdot 90}{120})^2}{\frac{60\cdot 90}{120}} + \frac{(5 - \frac{30\cdot 30}{120})^2}{\frac{30\cdot 30}{120}} + \frac{(25 - \frac{30\cdot 90}{120})^2}{\frac{30\cdot 90}{120}} \doteq 2.22.$$

Jelikož $\chi^2 = 2.22 < \chi^2_{0.95,2} = 5.992$, nezamítáme hypotézu H_0 ve prospěch H_1.

Cvičení 16.16. *Na 100 lidech byla pozorována barva očí a vlasů. Data jsou shrnuta v tabulce. Na hladině 5 % testujte hypotézu o nezávislosti barvy očí a vlasů.*

vlasy \ oči	*tmavé*	*světlé*
modré	10	20
šedé	10	10
hnědé	40	10

Řešení. Použijeme χ^2-test nezávislosti 2 znaků. Kritická hodnota je $q_{\chi^2(2)}(0.95) \doteq 5.992$. Potřebujeme jednotlivé marginální četnosti: $n_{1\cdot} = 30$, $n_{2\cdot} = 20$, $n_{3\cdot} = 50$, $n_{\cdot 1} = 60$, $n_{\cdot 2} = 40$. Hodnota testovací statistiky je

$$t = \frac{(10 - \frac{30\cdot 60}{100})^2}{\frac{30\cdot 60}{100}} + \cdots \doteq 18.$$

Nulovou hypotézu o nezávislosti proto zamítáme.

Cvičení 16.17. *Chceme zkoumat, zda u zločinců hmotnost souvisí s jejich rozumovými schopnostmi. Posuďte na základě dat v tabulce (počty vězňů v kategoriích).*

rozumové schopnosti \ hmotnost	≤ 75 kg	> 75 kg	*celkem*
normální	272	124	396
snížené	82	15	97
celkem	354	139	493

Řešení. Pokud bychom ponechali marginální rozdělení a předpokládali nezávislé náhodné veličiny, dostali bychom teoretické četnosti dle následující tabulky:

rozumové schopnosti \ hmotnost	≤ 75 kg	> 75 kg	celkem
normální	284.3	111.7	396
snížené	69.7	27.3	97
celkem	354	139	493

Příspěvky jednotlivých položek ke kritériu uvádí tabulka:

hmotnost / rozumové schopnosti	≤ 75 kg	> 75 kg
normální	0.54	1.37
snížené	2.19	**5.58**

Jejich součet, 9.67, je hodnotou kritéria, kterou porovnáme s kvantilem $t_{\chi^2(1)}(0.95) \doteq 3.84$ a nulovou hypotézu zamítáme. K tomu nám stačila i samotná hodnota v pravém dolním rohu poslední tabulky.

17 Test korelace dvou výběrů z normálních rozdělení

Cvičení 17.1. *V tabulce jsou uvedeny zaokrouhlené hodnoty HDP (v* 10^9 *Kč) ve 3. a 4. čtvrtletí příslušných let. Testujte hypotézu o korelovanosti na hladině významnosti* 5 %.

rok	2007	2008	2009	2010	2011
3. čtvrtletí	935	997	939	959	965
4. čtvrtletí	978	991	978	986	996

Řešení. Označíme X, resp. Y náhodné veličiny, které odpovídají 3., resp. 4. čtvrtletí. Koeficient korelace $\varrho(X,Y)$ odhadneme pomocí výběrového koeficientu korelace $R_{\boldsymbol{X},\boldsymbol{Y}}$, resp. jeho realizace $r_{\boldsymbol{x},\boldsymbol{y}}$,

$$R_{\boldsymbol{X},\boldsymbol{Y}} = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})(Y_i - \overline{\boldsymbol{Y}})}{\sqrt{\left(\sum\limits_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})^2\right)\left(\sum\limits_{i=1}^{n}(Y_i - \overline{\boldsymbol{Y}})^2\right)}}\,.$$

Rozsah souborů $n = 5$ a pro daná data dostaneme

$$\sum_{i=1}^{5} x_i = 4\,795, \qquad \sum_{i=1}^{5} y_i = 4\,929, \qquad \sum_{i=1}^{5} x_i y_i = 4\,727\,513,$$

$$\sum_{i=1}^{5} x_i^2 = 4\,600\,861, \qquad \sum_{i=1}^{5} y_i^2 = 4\,859\,261.$$

Po dosazení do vzorce pro realizaci výběrového koeficientu korelace dostaneme

$$r_{\boldsymbol{x},\boldsymbol{y}} = \frac{n\sum\limits_{i=1}^{5} x_i y_i - \left(\sum\limits_{i=1}^{5} x_i\right)\left(\sum\limits_{i=1}^{5} y_i\right)}{\sqrt{\left(n\sum\limits_{i=1}^{5} x_i^2 - \left(\sum\limits_{i=1}^{5} x_i\right)^2\right)\left(n\sum\limits_{i=1}^{5} y_i^2 - \left(\sum\limits_{i=1}^{5} y_i\right)^2\right)}} = \frac{3\,010}{\sqrt{12\,280 \cdot 1\,264}} \doteq 0.764\,.$$

Statistiku

$$T = \frac{R_{\boldsymbol{X},\boldsymbol{Y}}\sqrt{n-2}}{\sqrt{1 - R^2_{\boldsymbol{X},\boldsymbol{Y}}}}$$

testujeme na t-rozdělení o $n - 2 = 3$ stupních volnosti. Pro její realizaci dostaneme hodnotu

$$t = \frac{r_{\boldsymbol{x},\boldsymbol{y}}\sqrt{n-2}}{\sqrt{1 - r^2_{\boldsymbol{x},\boldsymbol{y}}}} \doteq \frac{0.764\sqrt{3}}{\sqrt{1 - 0.764^2}} \doteq 2.051.$$

Kritický obor pro test je $|t| > q_{t(3)}(0.975) \doteq 3.18$. Protože $|t| \doteq 2.051 < q_{t(3)}(0.975) \doteq 3.18$, hypotézu o nekorelovanosti nezamítáme.

Cvičení 17.2. *Pro realizace* $\boldsymbol{x} = (22, 15, 30, 27, 29)$ *a* $\boldsymbol{y} = (10, 6, 8, 4, 8)$ *náhodných výběrů z veličin* X, Y *testujte na hladině významnosti* $\alpha = 5$ % *jejich korelovanost.*

Řešení. Testujeme hypotézu o koeficientu korelace $\varrho(X,Y)$ mezi náhodnými veličinami X a Y,
H_0: $\varrho(X,Y)=0$, náhodné veličiny X a Y jsou nekorelované
proti hypotéze
H_1: $\varrho(X,Y)\neq 0$, náhodné veličiny X a Y jsou korelované.

K testování použijeme výběrový koeficient korelace $R_{\boldsymbol{x},\boldsymbol{y}}$ a testovou statistiku $T=\frac{R_{\boldsymbol{x},\boldsymbol{y}}\sqrt{n-2}}{\sqrt{1-r^2(X,Y)}}$, která má Studentovo rozdělení $\mathrm{t}(n-2)$, kde n je rozsah výběrů. Realizaci výběrového koeficientu korelace vypočteme ze vzorce

$$r_{\boldsymbol{x},\boldsymbol{y}}=\frac{n\sum\limits_{i=1}^{n}x_iy_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{\sqrt{\left(n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2\right)\left(n\sum\limits_{i=1}^{n}y_i^2-\left(\sum\limits_{i=1}^{n}y_i\right)^2\right)}}\,.$$

Je $n=5$,

$$\sum_{i=1}^{n}x_i=123,\ \sum_{i=1}^{n}y_i=36,\ \sum_{i=1}^{n}x_i^2=3\,179,\ \sum_{i=1}^{n}y_i^2=280,\ \sum_{i=1}^{n}x_iy_i=890.$$

Po dosazení hodnot dostaneme

$$r_{\boldsymbol{x},\boldsymbol{y}}\doteq 0.078,\quad t\doteq 0.1354.$$

Z tabulek nalezneme kvantil $q_{\mathrm{t}(3)}(0.975)$ rozdělení $\mathrm{t}(3)$, který má hodnotu $q_{\mathrm{t}(3)}(0.975)\doteq 3.18$. Protože je $T=0.1354<q_{\mathrm{t}(3)}(0.975)\doteq 3.18$, hypotézu H_0 nezamítáme.

Cvičení 17.3. *Na vzorku 50 pacientů byla zjištěna korelace -0.4 mezi tělesnou hmotností a věkem, kterého se dožili. Otestujte na hladině významnosti 5 % hypotézu, že zvýšená hmotnost nezkracuje život.*

Řešení.

$$t=\frac{r_{\boldsymbol{x},\boldsymbol{y}}\sqrt{n-2}}{\sqrt{1-r^2_{\boldsymbol{x},\boldsymbol{y}}}}=\frac{-0.4\sqrt{50-2}}{\sqrt{1-0.4^2}}\doteq -3.02\,,$$

porovnáme s $q_{t(48)}(0.05)=-q_{t(48)}(0.95)\doteq -1.68$ a hypotézu zamítáme na hladině významnosti 5 %.

Cvičení 17.4. *Pro náhodné výběry $X=(7,9,4,3,2,4)$ a $Y=(6,5,6,4,0,7)$, které jsou výsledky z 1., resp. 2. testu, testujte na hladině významnosti $\alpha=5$ % hypotézu*
H_0: *náhodné veličiny jsou nekorelované*
proti alternativě
H_1: *náhodné veličiny jsou korelované.*

Za jaké podmínky můžete tímto způsobem testovat závislost či nezávislost náhodných veličin?

Řešení. K testování hypotézy použijeme realizaci výběrového koeficientu korelace

$$r_{\boldsymbol{x},\boldsymbol{y}}=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{\left(n\sum_{i=1}^{n}x_i^2-(n\,\bar{\boldsymbol{x}})^2\right)\left(n\sum_{i=1}^{n}y_i^2-(n\,\bar{\boldsymbol{y}})^2\right)}}\,,$$

kde $n=6$ je počet dat v souborech. Po dosazení postupně dostáváme:

$$\begin{aligned}
n\,\bar{\boldsymbol{x}} &= \sum_{i=1}^{6}x_i=29\,,\qquad n\,\bar{\boldsymbol{y}}=\sum_{i=1}^{6}y_i=28\,,\\
n\sum_{i=1}^{6}x_i\,y_i &- \left(\sum_{i-1}^{6}x_i\right)\left(\sum_{i=1}^{6}y_i\right)=6\cdot 151-29\cdot 28=94\,,\\
\sum_{i=1}^{6}x_i^2 &= 175\,,\qquad \sum_{i=1}^{6}y_i^2=162\,,\\
n\sum_{i=1}^{6}x_i^2 &= 162-\left(\sum_{i=1}^{6}x_i\right)^2=6\cdot 175-29^2=209\,,\\
n\sum_{i=1}^{6}y_i^2 &= 162-\left(\sum_{i=1}^{6}y_i\right)^2=6\cdot 162-28^2=188\,.
\end{aligned}$$

Potom dostaneme realizaci výběrového koeficientu korelace

$$r_{\boldsymbol{xy}}=\frac{94}{\sqrt{209\cdot 188}}\doteq 0.4742$$

a odtud získáme hodnotu statistiky

$$t = \frac{r_{\boldsymbol{xy}}\sqrt{n-2}}{\sqrt{1-r_{\boldsymbol{xy}}^2}} \doteq \frac{0.4742 \cdot 2}{\sqrt{1-0.4742^2}} \doteq 1.077\,.$$

Testujeme na Studentovo rozdělení t(4) a kritický obor určíme pomocí tabulek kvantilů, kde nalezneme $q_{\mathrm{t}(4)}(0.975) = 2.78$, tedy

$$W = \{t : |t| > 2.78\}\,.$$

Protože hodnota statistiky t neleží v kritickém oboru, testovanou hypotézu H_0 nezamítáme.

Pro nezávislost je nutné, aby rozdělení byla normální. Testujeme jen lineární závislost, jiné druhy závislosti tento test nemusí prokázat.

Cvičení 17.5. *Testujte na hladině významnosti* 1 %*, zda veličiny výška a váha novorozeněte jsou korelované, pokud jsme u* 100 *náhodně vybraných dětí zjistili průměrnou váhu* 3.42 kg *a výšku* 51 cm*, výběrovou směrodatnou odchylku váhy* 0.5 kg *a výšky* 4 cm*. Označíme-li po řadě váhu a výšku i-tého novorozence* x_i*, resp.* y_i*, je součet* $\sum_{i=1}^{100} x_i y_i$ *roven* 17557 kg cm.

Řešení. Dle [6, Věta 12.4.2]

$$r_{\boldsymbol{x},\boldsymbol{y}} = \frac{n}{n-1}\,\frac{\frac{1}{n}\sum\limits_{j=1}^{n} x_j\,y_j - \bar{\boldsymbol{x}}\,\bar{\boldsymbol{y}}}{s_{\boldsymbol{x}}\,s_{\boldsymbol{y}}} = \frac{100}{99}\,\frac{\frac{1}{100}\,17557 - 3.42\cdot 51}{0.5\cdot 4} \doteq 0.58\,.$$

Hodnotu testovací statistiky

$$t = \frac{r_{\boldsymbol{x},\boldsymbol{y}}\sqrt{n-2}}{\sqrt{1-r_{\boldsymbol{x},\boldsymbol{y}}^2}} = \frac{0.58\sqrt{98}}{\sqrt{1-0.58^2}} \doteq 7.06\,.$$

porovnáme s kvantilem $q_{\mathrm{t}(98)}(0.995) \doteq 2.63$ a hypotézu, že veličiny jsou nekorelované, zamítáme.

18 Znaménkový test

Cvičení 18.1. *Zpráva z tisku: „O globálním oteplování nemůže být pochyb, protože z posledních* 10 *let bylo* 9 *teplotně nadprůměrných.“ Posuďte statistickou významnost tohoto pozorování.*

Řešení. Argument je velmi přesvědčivý, nicméně je potřeba ho posoudit podle statistických kritérií. Nulová hypotéza budiž, že průměrná teplota ve sledovaném období není vyšší než dlouhodobý průměr, alternativa, že je vyšší. Jelikož nemáme dostatek dat pro odhad rozdělení a dokonce ani nemáme k dispozici jednotlivé odchylky, jen jejich znaménka, nabízí se pouze znaménkový test.

Předpokládáme, že dlouhodobý průměr je mediánem rozdělení, z něhož byl pořízen tento náhodný výběr. Pak kladné i záporné odchylky mají stejnou pravděpodobnost 1/2. (Nulové odchylky nezahrnujeme do souboru.) Za předpokladu nulové hypotézy se počet záporných (stejně jako kladných) odchylek řídí binomickým rozdělením Bi(10, 1/2). Pravděpodobnost, že počet záporných odchylek bude 1 nebo nižší, je

$$\sum_{k=0}^{1} p_{\mathrm{Bi}(10,1/2)}(k) = \sum_{k=0}^{1}\binom{10}{k}\frac{1}{2^{10}} = (1+10)\,\frac{1}{2^{10}} \doteq 0.011\,.$$

Toto je dosažená významnost, která dovoluje zamítnout nulovou hypotézu na hladině významnosti 5 %, ale nikoli 1 %. *(Zde jsme součet kombinačních čísel spočítali snadno, pro velký rozsah výběru bychom použili náhradu normálním rozdělením.)*

Ze zprávy nevyplývá, zda nekladný výsledek byl nulový nebo záporný, předpokládali jsme záporný. Kdyby byl nulový, ignorovali bychom ho a měli 9 údajů, všechny kladné, což vede na dosaženou významnost $1/2^9 \doteq 0.002$, tedy ještě významnější výsledek.

Kdybychom v nulové hypotéze předpokládali rovnost místo nerovnosti (dlouhodobý průměr je mediánem rozdělení) s alternativou, že tyto hodnoty jsou různé, dostali bychom oboustranný test. Pak bychom posuzovali pravděpodobnost, že počet *méně častých* odchylek bude 1 nebo nižší, tj. počet záporných (stejně jako kladných) odchylek bude v kritickém oboru $\{0, 1, 9, 10\}$. Ta je

$$2\sum_{k=0}^{1} p_{\mathrm{Bi}(10,1/2)}(k) = 2\sum_{k=0}^{1}\binom{10}{k}\frac{1}{2^{10}} = 2\,(1+10)\,\frac{1}{2^{10}} \doteq 0.022\,,$$

což opět dovoluje zamítnout nulovou hypotézu na hladině významnosti 5 %, ale nikoli 1 %. (Výsledek v oboustranném testu je méně významný než v jednostranném.)

Test má ale metodické chyby. Především vyhodnocuje měřená data dodatečně. Správně bychom měli formulovat hypotézy předem, stanovit, že budeme provádět pozorování následujících 10 let a vyhodnotíme znaménkovým testem např. na hladině významnosti 5 %. Dodatečné vyhodnocení existujících dat umožňuje volbu začátku, což snižuje významnost testu. Kdybychom např. posunuli začátek o rok a ten se ukázal teplotně podprůměrný, zjistili bychom jen 8 teplotně nadprůměrných let z 10 a dosažená významnost

$$\sum_{k=0}^{2} p_{\mathrm{Bi}(10,1/2)}(k) = \sum_{k=0}^{2} \binom{10}{k} \frac{1}{2^{10}} = (1 + 10 + 45)\,\frac{1}{2^{10}} \doteq 0.055$$

by nedovolila zamítnout nulovou hypotézu ani na hladině významnosti 5 %.

Také délka testu byla stanovena dodatečně, nikoli předem. Takto se seriózně argumentovat nedá.

Cvičení 18.2. *Zařízení váží součástky s chybou, jejíž rozdělení není známo. Otestujte na hladině významnosti* 5 %*, zda je možné, aby měření nebylo zatíženo systematickou chybou (tj. medián chyb byl nulový), pokud při* 9 *kontrolních měřeních byly naměřeny tyto chyby (v gramech):*

$$0.3,\ 0.4,\ -0.8,\ 0.1,\ -1.3,\ -1.1,\ -0.6,\ -0.2,\ -0.5$$

Výsledky. Z 9 hodnot jsou 3 záporné a 6 kladných, dosažená významnost při jednostranném testu 0.25 nedovoluje zamítnout nulovou hypotézu.

Cvičení 18.3. *Chceme zjistit, jestli v daném podniku nalévají systematicky pod míru. Z* 30 *piv, která jsme si za poslední měsíc nechali natočit, bylo* 15 *piv natočeno pod rysku,* 4 *k rysce a* 11 *piv nad rysku. Jaký je náš závěr?*

Výsledky. Počítáme jen 26 položek, rozdělených na 15 a 11, dosažená významnost při jednostranném testu 0.28 nedovoluje zamítnout nulovou hypotézu, že správná hodnota je mediánem tohoto rozdělení.

19 Příklady pro opakování

V této kapitole jsou příklady, které se tematicky váží k více kapitolám, takže nemohly být zařazeny jen k jednomu tématu. Mohou sloužit k opakování látky a procvičení souvislostí.

Cvičení 19.1. *Nezávislé náhodné veličiny X, Y, Z mají po řadě rozdělení* $\mathrm{N}(2,3)$, $\mathrm{N}(5,1)$, $\mathrm{N}(0,1)$*. Určete*

a) rozdělení náhodné veličiny $X - Y$,

b) střední hodnotu náhodné veličiny $X \cdot Y$,

c) rozdělení náhodné veličiny Z^2.

Výsledky.

a) $X - Y$ má rozdělení $\mathrm{N}(-3,4)$,

b) $\mathrm{E}(X \cdot Y) = 10$,

c) Z^2 má rozdělení $\chi^2(1)$ (dle definice).

Cvičení 19.2. *Nezávislé náhodné veličiny X, Y, Z mají po řadě rozdělení* $\mathrm{N}(2,3)$, $\mathrm{N}(0,1)$, $\mathrm{N}(0,1)$*. Určete*

a) rozdělení náhodné veličiny $X + Y$,

b) střední hodnotu směsi náhodných veličin $\mathrm{Mix}_{1/2}(X, Y)$,

c) rozdělení náhodné veličiny $Y^2 + Z^2$.

Výsledky.

a) $X + Y$ má rozdělení $\mathrm{N}(2,4)$,

b) $\mathrm{E}\,\mathrm{Mix}_{1/2}(X, Y) = 1$,

c) $Y^2 + Z^2$ má rozdělení $\chi^2(2)$ (dle definice).

Cvičení 19.3. *Na desce jsou kruhové kapky. Jejich plošný obsah v* mm^2 *má rozdělení χ^2 s* 1 *stupněm volnosti. Jaké je rozdělení a medián jejich obvodu?*

Výsledky. Obvod má stejné rozdělení jako náhodná veličina $|S|$, kde S má $\mathrm{N}(0, 4\,\pi)$, medián je $2\,\pi\,\Phi^{-1}(3/4) \doteq 2.39$ mm.

Cvičení 19.4. *Jaké jsou vztahy mezi nezávislostí a nekorelovaností náhodných veličin? Uveďte jeden příklad u každé kombinace těchto vlastností, která může nastat.*

Řešení. Nezávislé náhodné veličiny jsou nekorelované, příkladů je mnoho. Příklady ostatních případů:

Závislé a korelované: $X = Y$ libovolné kromě konstantních.

Závislé a nekorelované: (X, Y) nabývá hodnot $(-2, -1)\,, (-1, 1)\,, (1, 1)\,, (2, -1)$ s pravděpodobnostmi 1/4. Pak $\mathrm{E}X = \mathrm{E}Y = \mathrm{E}(X\,Y) = 0$,

$$P[X = 2, Y = 1] = 0 \neq P[X = 2] \cdot P[Y = 1] = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}\,.$$

Cvičení 19.5. *Podmínka nekorelovanosti náhodných veličin je tvaru rovnosti dvou reálných čísel, což, jak známo, je velmi neobvyklý případ. Co z toho vyplývá pro nekorelovanost náhodných veličin?*

Řešení. Pokud jsou náhodné veličiny závislé, je pravděpodobnost, že vyjdou nekorelované, velmi malá (typicky nulová); nemůžeme to však vyhodnotit, takže nanejvýš můžeme vyvrátit hypotézu, že jsou nekorelované.

Pokud jsou však nezávislé (což není tak neobvyklé), pak nekorelovanost vychází z podstaty pokusu a platí přesně.

Důsledkem je, že dostatečně přesný (rozsáhlý) test na nekorelovanost odhalí závislost náhodných veličin s vysokou pravděpodobností, ačkoli jistotu nedává ani teoreticky přesná nekorelovanost.

Cvičení 19.6. *Nechť X je náhodná veličina, která má spojité rovnoměrné rozdělení v intervalu $(0,\,1)$. Položme $Y = X^2$.*

a) Určete korelační matici náhodného vektoru $(X,\,Y)$.

b) Jsou náhodné veličiny X a Y závislé či nezávislé?

Řešení. a) Hustota f_X náhodné veličiny X je konstantní v intervalu $(0, 1)$ a je rovna převrácené hodnotě délky intervalu, tedy

$$f_X(x) = \begin{cases} 1, & x \in (0,\,1), \\ 0, & x \notin (0,\,1). \end{cases}$$

Potom je

$$\begin{aligned}
\mathrm{E}\,X &= \int_0^1 x \mathrm{d}x = \left[\frac{x^2}{2}\right]_{x=0}^1 = \frac{1}{2}\,, \\
\mathrm{E}\,Y &= \mathrm{E}(X^2) = \int_0^1 x^2 \mathrm{d}x = \left[\frac{x^3}{3}\right]_{x=0}^1 = \frac{1}{3}\,, \\
\mathrm{E}(XY) &= \mathrm{E}(X^3) = \int_0^1 x^3 \mathrm{d}x = \left[\frac{x^4}{4}\right]_{x=0}^1 = \frac{1}{4}\,, \\
\mathrm{E}(Y^2) &= \mathrm{E}(X^4) = \int_0^1 x^4 \mathrm{d}x = \left[\frac{x^5}{5}\right]_{x=0}^1 = \frac{1}{5}\,.
\end{aligned}$$

Odtud dostaneme

$$\begin{aligned}
\mathrm{D}\,X &= \mathrm{E}(X^2) - (\mathrm{E}\,X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\,, \\
\mathrm{D}\,Y &= \mathrm{E}(Y^2) - (\mathrm{E}\,Y)^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}\,.
\end{aligned}$$

Koeficient korelace je

$$\varrho(X, Y) = \frac{\mathrm{E}(XY) - \mathrm{E}\,X\;\mathrm{E}\,Y}{\sqrt{\mathrm{D}\,X\;\mathrm{D}\,Y}} = \frac{\frac{1}{4} - \frac{1}{2}\,\frac{1}{3}}{\sqrt{\frac{1}{12}\,\frac{4}{45}}} = \frac{\sqrt{15}}{4} \doteq 0.9682\,.$$

Uvážíme-li, že $\varrho(X, X) = \varrho(Y, Y) = 1$, je korelační matice náhodného vektoru (X, Y) rovna

$$\left(\begin{array}{c|c} 1 & \frac{\sqrt{15}}{4} \\ \frac{\sqrt{15}}{4} & 1 \end{array}\right).$$

b) Náhodné veličiny X a Y jsou závislé, neboť je $Y = X^2$. Tato skutečnost je potvrzena i tím, že $\varrho(X, Y) \neq 0$.

Cvičení 19.7. *Při realizaci náhodného vektoru (X, Y) byly pozorovány uvedené četnosti výskytu hodnot. Odhadněte:*

a) sdruženou pravděpodobnostní funkci $p_{X,Y}$;

b) marginální pravděpodobnostní funkce p_X a p_Y;

c) střední hodnotu $\mathrm{E}X$.

$y \backslash x$	-1	0	1
0	3	5	7
2	1	4	12

Řešení. a) Všech hodnot je celkem $n = 3+5+7+1+4+12 = 32$. Jestliže označíme n_{ij} četnost výskytu hodnoty (x_i, y_j), pak odhadem hodnoty sdružené pravděpodobnostní funkce je $p_{X,Y}(x_i, y_j) = \frac{n_{ij}}{n}$. Příslušné hodnoty sepíšeme do tabulky

$y \backslash x$	-1	0	1
0	$\frac{3}{32} = 0.09375$	$\frac{5}{32} = 0.15625$	$\frac{7}{32} = 0.21875$
2	$\frac{1}{32} = 0.03125$	$\frac{4}{32} = 0.125$	$\frac{12}{32} = 0.375$

b) (Maximálně věrohodné) odhady hodnot marginálních pravděpodobnostních funkcí dostaneme obdobně jako v případě sdružené pravděpodobnostní funkce. Sečteme příslušné četnosti nebo sečteme hodnoty sdružené pravděpodobnostní funkce po řádcích či sloupcích.

$y \backslash x$	-1	0	1	p_Y
0	$\frac{3}{32} = 0.09375$	$\frac{5}{32} = 0.15625$	$\frac{7}{32} = 0.21875$	$\frac{15}{32} \doteq 0.4688$
2	$\frac{1}{32} = 0.03125$	$\frac{4}{32} = 0.125$	$\frac{12}{32} = 0.375$	$\frac{17}{32} \doteq 0.5312$
p_X	$\frac{4}{32} = 0.125$	$\frac{9}{32} \doteq 0.2813$	$\frac{19}{32} \doteq 0.5937$	1

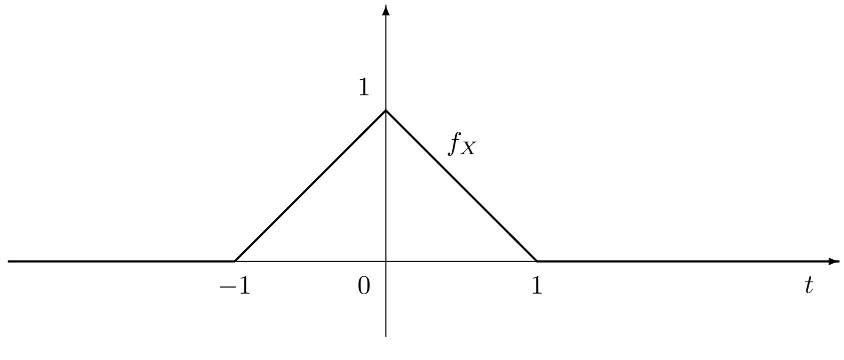
c) Střední hodnotu $\mathrm{E}X$ odhadneme jako realizaci výběrového průměru. Dostaneme

$$\overline{x} = -1 \cdot \frac{4}{32} + 0 \cdot \frac{9}{32} + 1 \cdot \frac{19}{32} = \frac{15}{32} = 0.46875\,.$$

Cvičení 19.8. *Náhodná veličina X má rozdělení s hustotou*

$$f_X(t) = \begin{cases} 1 - |t|, & t \in (-1, 1), \\ 0 & \textit{jinde.} \end{cases}$$

(viz obrázek). Určete distribuční funkci F_X. Dále spočtěte střední hodnotu a rozptyl veličiny X a její kovarianci s náhodnou veličinou $Y = X^2 - 1$.



Výsledky.

$$F_X(t)=\begin{cases}0, & t<-1,\\ \dfrac{1}{2}\,(t+1)^2, & -1\le t<0,\\ 1-\dfrac{1}{2}\,(t-1)^2, & 0\le t<1,\\ 1, & t\ge 1,\end{cases}$$

$\mathrm{E}X=0$, $\mathrm{D}X=1/6$, $\mathrm{cov}(X,Y)=0$.

Cvičení 19.9. *Směrodatná odchylka měřicího přístroje je* 36*, systematická chyba je nulová. Kolik je třeba provést měření, aby s pravděpodobností alespoň* 0.9 *chyba alespoň jednoho měření v absolutní hodnotě nepřekročila* 7.2*?*

Řešení. Označme jev A_i, že chyba měření X_i je v intervalu $[-7.2, 7.2)$; to nastává s pravděpodobností

$$\begin{aligned} P(A_i) &= P(-7.2\le X_i\le 7.2)=P(-7.2/36\le X_i/36\le 7.2/36)\\ &= \Phi(0.2)-\Phi(-0.2)=2\,\Phi(0.2)-1\doteq 0.1586\,. \end{aligned}$$

Hledané n pak vyhovuje vztahu

$$0.9\le P\left(\bigvee_{i=1}^{n} A_i\right)=1-P\left(\bigwedge_{i=1}^{n}\bar{A}_i\right)=1-(0.8414)^n\,.$$

Řešením je $n>13$.

Cvičení 19.10. *Alice střílí třikrát nezávisle na cíl, pravděpodobnost zásahu v jednom pokusu je $p_A=1/3$. Bob střílí na cíl nezávisle dvakrát a pravděpodobnost zásahu v jednom pokusu je $p_B=1/2$. Pomocí entropie určete, který střelec má v průměru méně náhodný počet zásahů cíle.*

Výsledky. Entropie výsledku Alice je

$$\frac{7}{3}\,\log_2 3-2\doteq 1.7\,,$$

Boba $3/2=1.5$, tj. menší (a méně náhodný výsledek).

Cvičení 19.11. *Pokud generátor náhodných čísel je nedokonalý (dává některé výsledky s vyšší pravděpodobností než jiné), typickým technickým řešením je zpětná vazba, která to kompenzuje. Posuďte možnost uplatnění tohoto principu.*

Řešení. Takový generátor by nebyl dobrý, jeho výsledky by byly závislé. Pokud např. náhodou vyjdou 3 stejné výsledky za sebou, má být pravděpodobnost opakování téhož výsledku v dalším pokusu stále stejná, ale zde by se snížila.

Cvičení 19.12. *Vysvětlete rozdíly mezi následujícími pojmy:*

- *střední hodnota,*
- *výběrový průměr,*
- *realizace výběrového průměru.*

Řešení. Střední hodnota nemusí existovat. Pokud existuje, je to číslo, které nám může zůstat utajeno; projevuje se pouze zprostředkovaně v realizacích náhodné veličiny a je limitou některých odhadů. Výběrový průměr je náhodná veličina vypočítaná z náhodného výběru, na rozdíl od střední hodnoty vždy existuje (pro numerické náhodné veličiny). Pokud původní rozdělení má rozptyl, je výběrový průměr nestranným konzistentním odhadem střední hodnoty, takže k ní v jistém smyslu konverguje pro rozsah výběru jdoucí do nekonečna. Realizace výběrového průměru je číslo získané z realizace náhodného výběru, sloužící k (realizaci) odhadu neznámé střední hodnoty.

Cvičení 19.13. *Předpokládejme, že politická strana má volební preference* 3 %*. Jaká je pravděpodobnost, že v průzkumu odhad jejích preferencí dosáhne aspoň* 5 %*, je-li rozsah výběru a)* 250*, b)* 500*?*

Výsledky. a) 3.2 %, b) 0.44 %.

Cvičení 19.14. *Na stejném místě měříme teplotu dvěma nezávislými teploměry se směrodatnými odchylkami* 2°C*. Ukazují* 3°C*, resp.* 2.5°C*. Jaké je riziko, že mrzne? Uveďte použité předpoklady.*

Řešení. Aritmetický průměr obou údajů je 2.75°C se směrodatnou odchylkou $\sqrt{2} \doteq 1.41$°C,

$$\Phi\left(\frac{-2.75}{\sqrt{2}}\right) \doteq 1 - \Phi\,(1.944\,54) \doteq 1 - 0.974 = 0.026\,.$$

Cvičení 19.15. *Náhodná veličina X je počet dětí ve školním věku v jedné rodině. Předpokládáme, že má Poissonovo rozdělení s parametrem $\lambda = 0.8$, tj.*

$$p_X(k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,, \qquad k \in \{0, 1, 2, \ldots\}\,,$$
$$\mathrm{E}X = \lambda\,, \qquad \mathrm{D}X = \lambda\,.$$

Ve městě bydlí $n = 10\,000$ rodin. Jaký počet míst ve školách bude postačovat s pravděpodobností aspoň 95 %? (Předpokládáme, že všechny děti chodí do školy v obci, ve které bydlí.) Uveďte použité předpoklady.

Řešení. *1. postup:* Použijeme centrální limitní větu; počet dětí má přibližně normální rozdělení $\mathrm{N}(n\,\lambda, n\,\lambda) = \mathrm{N}(8\,000, 8\,000)$. Výsledkem je kvantil

$$q_{\mathrm{N}(n\,\lambda,n\,\lambda)}\,(0.95) = n\,\lambda + \sqrt{n\,\lambda}\,\Phi^{-1}\,(0.95) \doteq 8\,000 + \sqrt{8\,000}\,1.645 \doteq 8147.13\,.$$

Zaokrouhlíme nahoru; potřebujeme aspoň 8148 míst.

2. postup: Součet nezávislých Poissonových rozdělení má Poissonovo rozdělení, zde s parametrem $n\,\lambda = 8\,000$. Pro intervalový odhad je nahradíme normálním rozdělením $\mathrm{N}(8\,000, 8\,000)$, další postup je stejný.

Předpokládáme nezávislost počtu dětí v jednotlivých rodinách. Existence rozptylu je zaručena předpoklady. Dále považujeme počet rodin za dostatečně velký na to, abychom mohli zanedbat chybu v náhradě výsledného (Poissonova) rozdělení normálním.

Cvičení 19.16. *Po $2/3$ dní neprší. V ostatní dny má srážkový úhrn v* mm *přibližně logaritmickonormální rozdělení* $\mathrm{LN}(0, 2.5)$*, tj. rozdělení náhodné veličiny tvaru $X = \exp(Y)$, kde Y má rozdělení* $\mathrm{N}(0, 2.5)$*. Její hustota je*

$$f_X(u) = \begin{cases} \frac{1}{u\,\sqrt{5\,\pi}}\exp\left(-\frac{(\ln u)^2}{5}\right), & u > 0,\\ 0 & \textit{jinak.}\end{cases}$$

Odhadněte, jak velký denní úhrn srážek je překročen $1\times$ za 100 let.

Řešení. Označme p podmíněnou pravděpodobnost, že v den, který je deštivý, spadnou zmíněné extrémní srážky. Počet takových dní za století se řídí binomickým rozdělením $\mathrm{Bi}(n, p)$, kde $n = 365.25 \cdot 100/3 = 12175$ je počet všech deštivých dní za století. Podle zadání je střední hodnota tohoto rozdělení $n\,p = 1$, tedy $p = 1/n \doteq 8.2 \cdot 10^{-5}$. Tomu odpovídá $q_Y(1 - p) = \sqrt{2.5}\,\Phi^{-1}(1 - p) \doteq 1.58 \cdot 3.8 \doteq 6$ a denní úhrn srážek

$$q_X(1 - p) = \exp(q_Y(1 - p)) \doteq \exp(6) \doteq 407 \text{ mm}\,.$$

Cvičení 19.17. *Profesor chodí na přednášky s malým zpožděním. Zjistil, že studenti chtějí statisticky vyhodnotit toto zpoždění. Napadl ho trik: na poslední přednášku přijde hodně pozdě, čímž zvýší rozptyl a zpoždění nevyjde statisticky významné. Má tato strategie naději na úspěch? Zdůvodněte. Jaké testy mohou studenti zvolit pro svoji hypotézu?*

Výsledky. Studenti mohou použít test střední hodnoty dle kapitoly 15.1.2. Trik se zvýšením rozptylu může někdy fungovat jako v Cvičení 10.31, záleží na rozsahu použitelných hodnot, který zde není neomezený. Kdyby studenti použili robustní odhad, který ignoruje vychýlené hodnoty *(outliers)*, trik by nefungoval.

Statistické tabulky

(Dle [6]; viz též `http://www.statsoft.com/textbook/sttable.html`.)

Poznámka 1. Pro účely testů hypotéz na hladině významnosti α používáme obvykle kvantily $q_X(\alpha)$, $q_X(1-\alpha)$, $q_X(\alpha/2)$, $q_X(1-\alpha/2)$, v závislosti na tvaru nulové hypotézy. V tabulkách bývají uvedeny jen některé, pokud další lze snadno dopočítat. Často bývají ve statistických tabulkách uvedeny *kritické hodnoty*. Kritickou hodnotou testu na hladině významnosti α může někdy být kvantil $q_X(1-\alpha/2)$, což hrozí nedorozuměním. Proto je potřeba vždy před použitím neznámých tabulek ověřit význam údajů.

Tabulka 11: Kvantilová funkce normovaného normálního rozdělení

α	$\Phi^{-1}(\alpha)$	α	$\Phi^{-1}(\alpha)$
0.500	0.000	0.99550	2.612
0.550	0.126	0.99600	2.652
0.600	0.253	0.99650	2.697
0.650	0.385	0.99700	2.748
0.700	0.524	0.99750	2.807
0.750	0.674	0.99800	2.878
0.800	0.842	0.99850	2.968
0.850	1.036	0.99900	3.090
0.900	1.282	0.99950	3.290
0.950	1.645	0.99955	3.320
0.955	1.695	0.99960	3.353
0.960	1.751	0.99965	3.389
0.965	1.812	0.99970	3.432
0.970	1.881	0.99975	3.481
0.975	1.960	0.99980	3.540
0.980	2.054	0.99985	3.616
0.985	2.170	0.99990	3.719
0.990	2.326	0.99995	3.891
0.995	2.576	0.99999	4.265

Tabulka 12: Distribuční funkce normovaného normálního rozdělení

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0.00	0.5000	0.76	0.7764	1.52	0.9357	2.28	0.98870
0.02	0.5080	0.78	0.7823	1.54	0.9382	2.30	0.98928
0.04	0.5160	0.80	0.7881	1.56	0.9406	2.32	0.98983
0.06	0.5239	0.82	0.7939	1.58	0.9429	2.34	0.99036
0.08	0.5319	0.84	0.7995	1.60	0.9452	2.36	0.99086
0.10	0.5398	0.86	0.8051	1.62	0.9474	2.38	0.99134
0.12	0.5478	0.88	0.8106	1.64	0.9495	2.40	0.99180
0.14	0.5557	0.90	0.8159	1.66	0.9515	2.42	0.99224
0.16	0.5636	0.92	0.8212	1.68	0.9535	2.44	0.99266
0.18	0.5714	0.94	0.8264	1.70	0.9554	2.46	0.99305
0.20	0.5793	0.96	0.8315	1.72	0.9573	2.48	0.99343
0.22	0.5871	0.98	0.8365	1.74	0.9591	2.50	0.99379
0.24	0.5948	1.00	0.8413	1.76	0.9608	2.52	0.99413
0.26	0.6026	1.02	0.8461	1.78	0.9625	2.54	0.99446
0.28	0.6103	1.04	0.8508	1.80	0.9641	2.56	0.99477
0.30	0.6179	1.06	0.8554	1.82	0.9656	2.58	0.99506
0.32	0.6255	1.08	0.8599	1.84	0.9671	2.60	0.99534
0.34	0.6331	1.10	0.8643	1.86	0.9686	2.62	0.99560
0.36	0.6406	1.12	0.8686	1.88	0.9699	2.64	0.99585
0.38	0.6480	1.14	0.8729	1.90	0.9713	2.66	0.99609
0.40	0.6554	1.16	0.8770	1.92	0.9726	2.68	0.99632
0.42	0.6628	1.18	0.8810	1.94	0.9738	2.70	0.99653
0.44	0.6700	1.20	0.8849	1.96	0.9750	2.72	0.99674
0.46	0.6772	1.22	0.8888	1.98	0.9761	2.74	0.99693
0.48	0.6844	1.24	0.8925	2.00	0.9772	2.76	0.99711
0.50	0.6915	1.26	0.8962	2.02	0.9783	2.78	0.99728
0.52	0.6985	1.28	0.8997	2.04	0.9793	2.80	0.99744
0.54	0.7054	1.30	0.9032	2.06	0.9803	2.82	0.99760
0.56	0.7123	1.32	0.9066	2.08	0.9812	2.84	0.99774
0.58	0.7190	1.34	0.9099	2.10	0.9821	2.86	0.99788
0.60	0.7257	1.36	0.9131	2.12	0.9830	2.88	0.99801
0.62	0.7324	1.38	0.9162	2.14	0.9838	2.90	0.99813
0.64	0.7389	1.40	0.9192	2.16	0.9846	2.92	0.99825
0.66	0.7454	1.42	0.9222	2.18	0.9854	2.94	0.99836
0.68	0.7517	1.44	0.9251	2.20	0.9861	2.96	0.99846
0.70	0.7580	1.46	0.9279	2.22	0.9868	2.98	0.99856
0.72	0.7642	1.48	0.9306	2.24	0.9875	3.00	0.99865
0.74	0.7704	1.50	0.9332	2.26	0.9881	3.02	0.99874

Tabulka 13: Kvantily χ^2-rozdělení $q_{\chi^2(\eta)}(\alpha)$ (η = počet stupňů volnosti)

α / η	0.005	0.01	0.025	0.05	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.000039	0.000157	0.0010	0.0039	3.84	5.02	6.63	7.88	10.83	12.12
2	0.010	0.020	0.051	0.103	5.99	7.38	9.21	10.60	13.82	15.20
3	0.072	0.115	0.216	0.352	7.81	9.35	11.34	12.84	16.27	17.73
4	0.207	0.297	0.484	0.711	9.49	11.14	13.28	14.86	18.47	20.00
5	0.412	0.554	0.831	1.145	11.07	12.83	15.09	16.75	20.51	22.11
6	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46	24.10
7	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32	26.02
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12	27.87
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88	29.67
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59	31.42
11	2.60	3.05	3.82	4.57	19.68	21.92	24.73	26.76	31.26	33.14
12	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91	34.82
13	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82	34.53	36.48
14	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32	36.12	38.11
15	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80	37.70	39.72
16	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27	39.25	41.31
17	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72	40.79	42.88
18	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16	42.31	44.43
19	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58	43.82	45.97
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00	45.31	47.50
21	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.40	46.80	49.01
22	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80	48.27	50.51
23	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18	49.73	52.00
24	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56	51.18	53.48
25	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93	52.62	54.95
26	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29	54.05	56.41
27	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.65	55.48	57.86
28	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99	56.89	59.30
29	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34	58.30	60.73
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67	59.70	62.16
31	14.46	15.66	17.54	19.28	44.99	48.23	52.19	55.00	61.10	63.58
32	15.13	16.36	18.29	20.07	46.19	49.48	53.49	56.33	62.49	64.99
33	15.82	17.07	19.05	20.87	47.40	50.73	54.78	57.65	63.87	66.40
34	16.50	17.79	19.81	21.66	48.60	51.97	56.06	58.96	65.25	67.80
35	17.19	18.51	20.57	22.47	49.80	53.20	57.34	60.27	66.62	69.20
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77	73.40	76.10
45	24.31	25.90	28.37	30.61	61.66	65.41	69.96	73.17	80.08	82.87
50	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49	86.66	89.56
55	31.73	33.57	36.40	38.96	73.31	77.38	82.29	85.75	93.17	96.16
60	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95	99.61	102.70
65	39.38	41.44	44.60	47.45	84.82	89.18	94.42	98.10	105.99	109.16
70	43.28	45.44	48.76	51.74	90.53	95.02	100.43	104.21	112.32	115.58
75	47.21	49.48	52.94	56.05	96.22	100.84	106.39	110.29	118.60	121.94
80	51.17	53.54	57.15	60.39	101.88	106.63	112.33	116.32	124.84	128.26
85	55.17	57.63	61.39	64.75	107.52	112.39	118.24	122.32	131.04	134.54
90	59.20	61.75	65.65	69.13	113.15	118.14	124.12	128.30	137.21	140.78
95	63.25	65.90	69.92	73.52	118.75	123.86	129.97	134.25	143.34	146.99
100	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17	149.45	153.16

Tabulka 14: Kvantily t-rozdělení $q_{t(\eta)}(\alpha)$ (η = počet stupňů volnosti)

α / η	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
1	6.31	12.7	31.8	63.7	127.3	318.3	636.6
2	2.92	4.30	6.96	9.92	14.1	22.3	31.6
3	2.35	3.18	4.54	5.84	7.5	10.2	12.9
4	2.13	2.78	3.75	4.60	5.60	7.17	8.61
5	2.02	2.57	3.36	4.03	4.77	5.89	6.87
6	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	1.89	2.36	3.00	3.50	4.03	4.79	5.41
8	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	1.78	2.18	2.68	3.05	3.43	3.93	4.32
13	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	1.75	2.12	2.58	2.92	3.25	3.69	4.01
17	1.74	2.11	2.57	2.90	3.22	3.65	3.97
18	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	1.72	2.09	2.53	2.85	3.15	3.55	3.85
21	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	1.72	2.07	2.51	2.82	3.12	3.50	3.79
23	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	1.71	2.06	2.49	2.80	3.09	3.47	3.75
25	1.71	2.06	2.49	2.79	3.08	3.45	3.73
26	1.71	2.06	2.48	2.78	3.07	3.43	3.71
27	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	1.70	2.05	2.46	2.76	3.04	3.40	3.66
30	1.70	2.04	2.46	2.75	3.03	3.39	3.65
35	1.69	2.03	2.44	2.72	3.00	3.34	3.59
40	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	1.67	2.00	2.39	2.66	2.91	3.23	3.46
80	1.66	1.99	2.37	2.64	2.89	3.20	3.42
100	1.66	1.98	2.36	2.63	2.87	3.17	3.39
∞	1.64	1.96	2.33	2.58	2.81	3.09	3.29

Tabulka 15: 0.95-kvantily F-rozdělení $q_{\mathrm{F}(\xi,\eta)}(0.95)$ (ξ = počet stupňů volnosti čitatele, η = počet stupňů volnosti jmenovatele)

η \ ξ	2	4	6	8	10	12	14	16	18	20
2	19.0	19.2	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4
4	6.94	6.39	6.16	6.04	5.96	5.91	5.87	5.84	5.82	5.80
6	5.14	4.53	4.28	4.15	4.06	4.00	3.96	3.92	3.90	3.87
8	4.46	3.84	3.58	3.44	3.35	3.28	3.24	3.20	3.17	3.15
10	4.10	3.48	3.22	3.07	2.98	2.91	2.86	2.83	2.80	2.77
12	3.89	3.26	3.00	2.85	2.75	2.69	2.64	2.60	2.57	2.54
14	3.74	3.11	2.85	2.70	2.60	2.53	2.48	2.44	2.41	2.39
16	3.63	3.01	2.74	2.59	2.49	2.42	2.37	2.33	2.30	2.28
18	3.55	2.93	2.66	2.51	2.41	2.34	2.29	2.25	2.22	2.19
20	3.49	2.87	2.60	2.45	2.35	2.28	2.22	2.18	2.15	2.12
25	3.39	2.76	2.49	2.34	2.24	2.16	2.11	2.07	2.04	2.01
30	3.32	2.69	2.42	2.27	2.16	2.09	2.04	1.99	1.96	1.93
40	3.23	2.61	2.34	2.18	2.08	2.00	1.95	1.90	1.87	1.84
50	3.18	2.56	2.29	2.13	2.03	1.95	1.89	1.85	1.81	1.78
60	3.15	2.53	2.25	2.10	1.99	1.92	1.86	1.82	1.78	1.75
80	3.11	2.49	2.21	2.06	1.95	1.88	1.82	1.77	1.73	1.70
100	3.09	2.46	2.19	2.03	1.93	1.85	1.79	1.75	1.71	1.68
150	3.06	2.43	2.16	2.00	1.89	1.82	1.76	1.71	1.67	1.64
200	3.04	2.42	2.14	1.98	1.88	1.80	1.74	1.69	1.66	1.62
∞	3.00	2.37	2.10	1.94	1.83	1.75	1.69	1.64	1.60	1.57

η \ ξ	25	30	40	50	60	80	100	150	200	∞
2	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
4	5.77	5.75	5.72	5.70	5.69	5.67	5.66	5.65	5.65	5.63
6	3.83	3.81	3.77	3.75	3.74	3.72	3.71	3.70	3.69	3.67
8	3.11	3.08	3.04	3.02	3.01	2.99	2.97	2.96	2.95	2.93
10	2.73	2.70	2.66	2.64	2.62	2.60	2.59	2.57	2.56	2.54
12	2.50	2.47	2.43	2.40	2.38	2.36	2.35	2.33	2.32	2.30
14	2.34	2.31	2.27	2.24	2.22	2.20	2.19	2.17	2.16	2.13
16	2.23	2.19	2.15	2.12	2.11	2.08	2.07	2.05	2.04	2.01
18	2.14	2.11	2.06	2.04	2.02	1.99	1.98	1.96	1.95	1.92
20	2.07	2.04	1.99	1.97	1.95	1.92	1.91	1.89	1.88	1.84
25	1.96	1.92	1.87	1.84	1.82	1.80	1.78	1.76	1.75	1.71
30	1.88	1.84	1.79	1.76	1.74	1.71	1.70	1.67	1.66	1.62
40	1.78	1.74	1.69	1.66	1.64	1.61	1.59	1.56	1.55	1.51
50	1.73	1.69	1.63	1.60	1.58	1.54	1.52	1.50	1.48	1.44
60	1.69	1.65	1.59	1.56	1.53	1.50	1.48	1.45	1.44	1.39
80	1.64	1.60	1.54	1.51	1.48	1.45	1.43	1.39	1.38	1.32
100	1.62	1.57	1.52	1.48	1.45	1.41	1.39	1.36	1.34	1.28
150	1.58	1.54	1.48	1.44	1.41	1.37	1.34	1.31	1.29	1.22
200	1.56	1.52	1.46	1.41	1.39	1.35	1.32	1.28	1.26	1.19
∞	1.51	1.46	1.39	1.35	1.32	1.27	1.24	1.20	1.17	1.00

Tabulka 16: 0.975-kvantily F-rozdělení $q_{F(\xi,\eta)}(0.975)$ (ξ = počet stupňů volnosti čitatele, η = počet stupňů volnosti jmenovatele)

η \ ξ	2	4	6	8	10	12	14	16	18	20
2	39.0	39.2	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4
4	10.65	9.60	9.20	8.98	8.84	8.75	8.68	8.63	8.59	8.56
6	7.26	6.23	5.82	5.60	5.46	5.37	5.30	5.24	5.20	5.17
8	6.06	5.05	4.65	4.43	4.30	4.20	4.13	4.08	4.03	4.00
10	5.46	4.47	4.07	3.85	3.72	3.62	3.55	3.50	3.45	3.42
12	5.10	4.12	3.73	3.51	3.37	3.28	3.21	3.15	3.11	3.07
14	4.86	3.89	3.50	3.29	3.15	3.05	2.98	2.92	2.88	2.84
16	4.69	3.73	3.34	3.12	2.99	2.89	2.82	2.76	2.72	2.68
18	4.56	3.61	3.22	3.01	2.87	2.77	2.70	2.64	2.60	2.56
20	4.46	3.51	3.13	2.91	2.77	2.68	2.60	2.55	2.50	2.46
25	4.29	3.35	2.97	2.75	2.61	2.51	2.44	2.38	2.34	2.30
30	4.18	3.25	2.87	2.65	2.51	2.41	2.34	2.28	2.23	2.20
40	4.05	3.13	2.74	2.53	2.39	2.29	2.21	2.15	2.11	2.07
50	3.97	3.05	2.67	2.46	2.32	2.22	2.14	2.08	2.03	1.99
60	3.93	3.01	2.63	2.41	2.27	2.17	2.09	2.03	1.98	1.94
80	3.86	2.95	2.57	2.35	2.21	2.11	2.03	1.97	1.92	1.88
100	3.83	2.92	2.54	2.32	2.18	2.08	2.00	1.94	1.89	1.85
150	3.78	2.87	2.49	2.28	2.13	2.03	1.95	1.89	1.84	1.80
200	3.76	2.85	2.47	2.26	2.11	2.01	1.93	1.87	1.82	1.78
∞	3.69	2.79	2.41	2.19	2.05	1.94	1.87	1.80	1.75	1.71

η \ ξ	25	30	40	50	60	80	100	150	200	∞
2	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5
4	8.50	8.46	8.41	8.38	8.36	8.33	8.32	8.30	8.29	8.26
6	5.11	5.07	5.01	4.98	4.96	4.93	4.92	4.89	4.88	4.85
8	3.94	3.89	3.84	3.81	3.78	3.76	3.74	3.72	3.70	3.67
10	3.35	3.31	3.26	3.22	3.20	3.17	3.15	3.13	3.12	3.08
12	3.01	2.96	2.91	2.87	2.85	2.82	2.80	2.78	2.76	2.72
14	2.78	2.73	2.67	2.64	2.61	2.58	2.56	2.54	2.53	2.49
16	2.61	2.57	2.51	2.47	2.45	2.42	2.40	2.37	2.36	2.32
18	2.49	2.44	2.38	2.35	2.32	2.29	2.27	2.24	2.23	2.19
20	2.40	2.35	2.29	2.25	2.22	2.19	2.17	2.14	2.13	2.09
25	2.23	2.18	2.12	2.08	2.05	2.02	2.00	1.97	1.95	1.91
30	2.12	2.07	2.01	1.97	1.94	1.90	1.88	1.85	1.84	1.79
40	1.99	1.94	1.88	1.83	1.80	1.76	1.74	1.71	1.69	1.64
50	1.92	1.87	1.80	1.75	1.72	1.68	1.66	1.62	1.60	1.55
60	1.87	1.82	1.74	1.70	1.67	1.63	1.60	1.56	1.54	1.48
80	1.81	1.75	1.68	1.63	1.60	1.55	1.53	1.49	1.47	1.40
100	1.77	1.71	1.64	1.59	1.56	1.51	1.48	1.44	1.42	1.35
150	1.72	1.67	1.59	1.54	1.50	1.45	1.42	1.38	1.35	1.27
200	1.70	1.64	1.56	1.51	1.47	1.42	1.39	1.35	1.32	1.23
∞	1.63	1.57	1.48	1.43	1.39	1.33	1.30	1.24	1.21	1.00

Tabulka 17: 0.99-kvantily F-rozdělení $q_{\mathrm{F}(\xi,\eta)}(0.99)$ (ξ = počet stupňů volnosti čitatele, η = počet stupňů volnosti jmenovatele)

η \ ξ	2	4	6	8	10	12	14	16	18	20
2	99.0	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4
4	18.00	15.98	15.21	14.80	14.55	14.37	14.25	14.15	14.08	14.02
6	10.92	9.15	8.47	8.10	7.87	7.72	7.60	7.52	7.45	7.40
8	8.65	7.01	6.37	6.03	5.81	5.67	5.56	5.48	5.41	5.36
10	7.56	5.99	5.39	5.06	4.85	4.71	4.60	4.52	4.46	4.41
12	6.93	5.41	4.82	4.50	4.30	4.16	4.05	3.97	3.91	3.86
14	6.51	5.04	4.46	4.14	3.94	3.80	3.70	3.62	3.56	3.51
16	6.23	4.77	4.20	3.89	3.69	3.55	3.45	3.37	3.31	3.26
18	6.01	4.58	4.01	3.71	3.51	3.37	3.27	3.19	3.13	3.08
20	5.85	4.43	3.87	3.56	3.37	3.23	3.13	3.05	2.99	2.94
25	5.57	4.18	3.63	3.32	3.13	2.99	2.89	2.81	2.75	2.70
30	5.39	4.02	3.47	3.17	2.98	2.84	2.74	2.66	2.60	2.55
40	5.18	3.83	3.29	2.99	2.80	2.66	2.56	2.48	2.42	2.37
50	5.06	3.72	3.19	2.89	2.70	2.56	2.46	2.38	2.32	2.27
60	4.98	3.65	3.12	2.82	2.63	2.50	2.39	2.31	2.25	2.20
80	4.88	3.56	3.04	2.74	2.55	2.42	2.31	2.23	2.17	2.12
100	4.82	3.51	2.99	2.69	2.50	2.37	2.27	2.19	2.12	2.07
150	4.75	3.45	2.92	2.63	2.44	2.31	2.20	2.12	2.06	2.00
200	4.71	3.41	2.89	2.60	2.41	2.27	2.17	2.09	2.03	1.97
∞	4.61	3.32	2.80	2.51	2.32	2.18	2.08	2.00	1.93	1.88

η \ ξ	25	30	40	50	60	80	100	150	200	∞
2	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
4	13.91	13.84	13.75	13.69	13.65	13.61	13.58	13.54	13.52	13.46
6	7.30	7.23	7.14	7.09	7.06	7.01	6.99	6.95	6.93	6.88
8	5.26	5.20	5.12	5.07	5.03	4.99	4.96	4.93	4.91	4.86
10	4.31	4.25	4.17	4.12	4.08	4.04	4.01	3.98	3.96	3.91
12	3.76	3.70	3.62	3.57	3.54	3.49	3.47	3.43	3.41	3.36
14	3.41	3.35	3.27	3.22	3.18	3.14	3.11	3.08	3.06	3.00
16	3.16	3.10	3.02	2.97	2.93	2.89	2.86	2.83	2.81	2.75
18	2.98	2.92	2.84	2.78	2.75	2.70	2.68	2.64	2.62	2.57
20	2.84	2.78	2.69	2.64	2.61	2.56	2.54	2.50	2.48	2.42
25	2.60	2.54	2.45	2.40	2.36	2.32	2.29	2.25	2.23	2.17
30	2.45	2.39	2.30	2.25	2.21	2.16	2.13	2.09	2.07	2.01
40	2.27	2.20	2.11	2.06	2.02	1.97	1.94	1.90	1.87	1.80
50	2.17	2.10	2.01	1.95	1.91	1.86	1.82	1.78	1.76	1.68
60	2.10	2.03	1.94	1.88	1.84	1.78	1.75	1.70	1.68	1.60
80	2.01	1.94	1.85	1.79	1.75	1.69	1.65	1.61	1.58	1.49
100	1.97	1.89	1.80	1.74	1.69	1.63	1.60	1.55	1.52	1.43
150	1.90	1.83	1.73	1.66	1.62	1.56	1.52	1.46	1.43	1.33
200	1.87	1.79	1.69	1.63	1.58	1.52	1.48	1.42	1.39	1.28
∞	1.77	1.70	1.59	1.52	1.47	1.40	1.36	1.29	1.25	1.00

Tabulka 18: 0.995-kvantily F-rozdělení $q_{F(\xi,\eta)}(0.995)$ (ξ = počet stupňů volnosti čitatele, η = počet stupňů volnosti jmenovatele)

η \ ξ	2	4	6	8	10	12	14	16	18	20
2	199.0	199.2	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4
4	26.28	23.15	21.98	21.35	20.97	20.70	20.51	20.37	20.26	20.17
6	14.54	12.03	11.07	10.57	10.25	10.03	9.88	9.76	9.66	9.59
8	11.04	8.81	7.95	7.50	7.21	7.01	6.87	6.76	6.68	6.61
10	9.43	7.34	6.54	6.12	5.85	5.66	5.53	5.42	5.34	5.27
12	8.51	6.52	5.76	5.35	5.09	4.91	4.77	4.67	4.59	4.53
14	7.92	6.00	5.26	4.86	4.60	4.43	4.30	4.20	4.12	4.06
16	7.51	5.64	4.91	4.52	4.27	4.10	3.97	3.87	3.80	3.73
18	7.21	5.37	4.66	4.28	4.03	3.86	3.73	3.64	3.56	3.50
20	6.99	5.17	4.47	4.09	3.85	3.68	3.55	3.46	3.38	3.32
25	6.60	4.84	4.15	3.78	3.54	3.37	3.25	3.15	3.08	3.01
30	6.35	4.62	3.95	3.58	3.34	3.18	3.06	2.96	2.89	2.82
40	6.07	4.37	3.71	3.35	3.12	2.95	2.83	2.74	2.66	2.60
50	5.90	4.23	3.58	3.22	2.99	2.82	2.70	2.61	2.53	2.47
60	5.79	4.14	3.49	3.13	2.90	2.74	2.62	2.53	2.45	2.39
80	5.67	4.03	3.39	3.03	2.80	2.64	2.52	2.43	2.35	2.29
100	5.59	3.96	3.33	2.97	2.74	2.58	2.46	2.37	2.29	2.23
150	5.49	3.88	3.25	2.89	2.67	2.51	2.38	2.29	2.21	2.15
200	5.44	3.84	3.21	2.86	2.63	2.47	2.35	2.25	2.18	2.11
∞	5.30	3.72	3.09	2.74	2.52	2.36	2.24	2.14	2.06	2.00

η \ ξ	25	30	40	50	60	80	100	150	200	∞
2	199.4	199.5	199.5	199.5	199.5	199.5	199.5	199.5	199.5	199.5
4	20.00	19.89	19.75	19.67	19.61	19.54	19.50	19.44	19.41	19.32
6	9.45	9.36	9.24	9.17	9.12	9.06	9.03	8.98	8.95	8.88
8	6.48	6.40	6.29	6.22	6.18	6.12	6.09	6.04	6.02	5.95
10	5.15	5.07	4.97	4.90	4.86	4.80	4.77	4.73	4.71	4.64
12	4.41	4.33	4.23	4.17	4.12	4.07	4.04	3.99	3.97	3.90
14	3.94	3.86	3.76	3.70	3.66	3.60	3.57	3.53	3.50	3.44
16	3.62	3.54	3.44	3.37	3.33	3.28	3.25	3.20	3.18	3.11
18	3.38	3.30	3.20	3.14	3.10	3.04	3.01	2.96	2.94	2.87
20	3.20	3.12	3.02	2.96	2.92	2.86	2.83	2.78	2.76	2.69
25	2.90	2.82	2.72	2.65	2.61	2.55	2.52	2.47	2.45	2.38
30	2.71	2.63	2.52	2.46	2.42	2.36	2.32	2.28	2.25	2.18
40	2.48	2.40	2.30	2.23	2.18	2.12	2.09	2.04	2.01	1.93
50	2.35	2.27	2.16	2.10	2.05	1.99	1.95	1.90	1.87	1.79
60	2.27	2.19	2.08	2.01	1.96	1.90	1.86	1.81	1.78	1.69
80	2.17	2.08	1.97	1.90	1.85	1.79	1.75	1.69	1.66	1.56
100	2.11	2.02	1.91	1.84	1.79	1.72	1.68	1.62	1.59	1.49
150	2.03	1.94	1.83	1.76	1.70	1.63	1.59	1.53	1.49	1.37
200	1.99	1.91	1.79	1.71	1.66	1.59	1.54	1.48	1.44	1.31
∞	1.88	1.79	1.67	1.59	1.53	1.45	1.40	1.32	1.28	1.00

Literatura

[1] Anděl, J.: Základy matematické statistiky. Matfyzpress, Praha, 2007.

[2] Dupač, V., Hušková, M.: Pravděpodobnost a matematická statistika. Karolinum, Praha, 1999.

[3] Likeš, J., Machek, J.: Matematická statistika. 2. vydání, SNTL, Praha, 1988.

[4] Mood, A.M., Graybill, F.A., Boes, D.C.: Introduction to the Theory of Statistics. 3rd ed., McGraw-Hill, 1974.

[5] Nagy, I.: Pravděpodobnost a matematická statistika. Cvičení. Skriptum FD ČVUT, Praha, 2002.

[6] Navara, M.: *Pravděpodobnost a matematická statistika.* Skriptum ČVUT, Praha, 2007.

[7] Něničková, A.: Matematická statistika — cvičení. Skriptum ČVUT, Praha, 1990.

[8] Novovičová, J.: *Pravděpodobnost a matematická statistika.* Skriptum FD ČVUT, Praha, 2002.

[9] Papoulis, A.: Probability and Statistics, Prentice-Hall, 1990.

[10] Rogalewicz, V.: Pravděpodobnost a statistika pro inženýry. Skriptum FEL ČVUT, 2. vydání, Praha, 2000.

[11] Wasserman, L.: All of Statistics. A Concise Course in Statistical Inference. Springer, 2004.

[12] Zvára, K., Štěpán, J.: Pravděpodobnost a matematická statistika (2. vydání). Matfyzpress, MFF UK, Praha, 2002.