PRAVDĚPODOBNOST A MATEMATICKÁ STATISTIKA

Mirko Navara

2007
České vysoké učení technické v Praze
Nakladatelství ČVUT

Lektor: doc. Vladimír Rogalewicz, CSc.

Nakladatelství ČVUT upozorňuje autory na dodržování autorských práv.
Za jazykovou a věcnou správnost obsahu díla odpovídá autor. Text neprošel jazykovou ani redakční úpravou.

© Mirko Navara, 2007
ISBN 978-80-01-03795-9

Předmluva

Tento text je určen pro studenty technických vysokých škol, zejména elektrotechnických fakult, jako úvod do teorie pravděpodobnosti a metod matematické statistiky. Zvláštní důraz je kladen na otázky použitelnosti jednotlivých metod a interpretaci jejich výsledků, spíše než na podrobné odvození, které lze najít v literatuře.

Pokud je na začátku kapitoly citována použitá literatura, označuje pramen, z jehož přístupu jsem nejvíc čerpal a který doporučuji pro další studium (není-li řečeno jinak).

Skriptum je psáno tak, aby umožnilo různé průchody podle výběru látky přednášejícím. Lze tedy jen s minimálními úpravami vynechat některé obtížnější kapitoly, zejména Kolmogorovův pravděpodobnostní model, konstrukce pravděpodobností, entropii, charakteristickou funkci náhodné veličiny a mnohé statistické metody. Přesto jsou uvedeny pro ty, kdo by chtěli lépe pochopit souvislosti nebo nastudovat příslušné téma bez nutnosti přechodu na jinou terminologii a značení z jiných učebnic. Výjimkami jsou směs náhodných veličin a kvantilová funkce, která se prolíná celým textem, neboť je použita i ke značení kvantilů.

U příkladů cituji použité prameny jen někdy, často se jedná o všeobecně používané úlohy. Hvězdičkami * jsou označeny příklady a cvičení, které látku nejen procvičují, ale i doplňují o témata, která by neměla být při studiu přehlédnuta a mohou na ně být odkazy v dalším textu.

Při četbě pramenů jsem ve většině z nich našel chyby včetně chybně řešených příkladů. Statistika mě nutí očekávat, že ani toto skriptum nebude výjimkou. K tomu účelu bude podporováno na stránce `http://math.feld.cvut.cz/skripta/pms.html`, kde čtenáři najdou errata, případné dodatky i kontakt na autora či správce stránky. Tímto prosím o hlášení nalezených chyb.

Skriptum bylo vytvořeno v prostředí *Scientific Workplace* (s výpočetní podporou *MuPAD*) a vysázeno v systému LaTeX s využitím šablony nakladatelství Springer. Jelikož se jedná o mezinárodní prostředí, prosím čtenáře o shovívavost při zjištěných nedostatcích české sazby (nicméně o nahlášení závad). Statistické tabulky byly vygenerovány tabulkovým procesorem *MS Excel*. Na tvorbě obrázků pomocí programů *xfig*, *Maple*® a *Scientific Workplace/MuPAD* se velmi ochotně podíleli Prof. Ing. Václav Hlaváč, CSc. a Ing. Milan Petrík. Některé výpočty (zejména s vysokými mocninami a kombinačními čísly) byly prováděny v systému *Maple*® díky podpoře projektu MŠMT 1N04075 „Realizace interaktivně informačního portálu pro vědecko-technické aplikace". Podmínky pro práci na skriptu a inspiraci poskytl výzkumný záměr MSM 6840770013 „Rozhodování a řízení pro průmyslovou výrobu III".

K podstatnému vylepšení textu přispělo různými způsoby mnoho lidí, zejména Prof. RNDr. J. Štěpán, DrSc., Doc. Vladimír Rogalewicz, CSc., Prof. Ing. Václav Hlaváč, CSc., Doc. RNDr. Eduard Krajník, CSs., RNDr. Petr Olšák, RNDr. Petr Savický, CSc., Doc. RNDr. Jaroslav Tišer, CSc., Doc. RNDr. Josef Tkadlec, CSc., Doc. RNDr. Peter Volauf, CSc., Ing. Tomáš Kroupa, Ph.D., Ing. Vít Zýka, Ph.D., Ing. Milan Petrík, Ing. Václav Gerla, Lukáš Přívozník, Jakub Břach a další. Jejich pomoci si vážím a tímto za ni děkuji.

Praha, 13. srpna 2007 *Mirko Navara*

Obsah

Seznam obrázků

Seznam tabulek

Přehled značení

Odchylky od českého standardu

V následujících případech je použit anglický standard místo českého; podobný trend však lze pozorovat i v jiných současných českých učebnicích, např. [Nagy 2002, Novovičová 2002].

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	$\exp \Omega = \{A \mid A \subseteq \Omega\}$	potenční množina (=množina všech podmnožin) množiny
$.\restriction.$	$X\restriction_{\Omega_j}$	zúžení funkce na podmnožinu jejího definičního oboru
$*$	$(f_X * f_Y)(u)$	konvoluce

Základní pojmy teorie pravděpodobnosti

Ω	Ω	množina všech elementárních jevů
$\mathbf{1}, \Omega$	$P(\mathbf{1}) = P(\Omega) = 1$	jev jistý
$\mathbf{0}, \emptyset$	$P(\mathbf{0}) = P(\emptyset) = 0$	jev nemožný
$P(.)$	$P(A)$	pravděpodobnost jevu
$P[.]$	$P[X \leq 0]$	pravděpodobnost jevu (pouze v kontextu pravděpodobnosti hodnot náhodných veličin)
$P(.\|.)$	$P(A\|B)$	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*)
Φ	$\Phi = F_{\mathrm{N}(0,1)}$	distribuční funkce *normovaného normálního* rozdělení
Mix	$\mathrm{Mix}_w(X, Y)$	směs dvou náhodných veličin
Mix	$\mathrm{Mix}_{(w_1,\ldots,w_n)}(X_1, \ldots, X_n)$	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
$\mathrm{cov}(.,.)$	$\mathrm{cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y$	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
Po	$\mathrm{Po}(w)$	Poissonovo
R	$\mathrm{R}(a,b)$	spojité rovnoměrné
N	$\mathrm{N}(\mu,\sigma^2)$	normální (=Gaussovo)
N(0,1)	$\mathrm{N}(0,1)$	normované normální
LN	$\mathrm{LN}(\mu,\sigma^2)$	logaritmicko-normální
Ex	$\mathrm{Ex}(w)$	exponenciální
N	$\mathrm{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$	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
$\bullet$	$X \bullet Y = \mathrm{E}(XY)$	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
$S_.^2, S^2$	$S_{\boldsymbol{X}}^2, S^2$	výběrový rozptyl
$s_.^2, s^2$	$s_{\boldsymbol{x}}^2, s^2$	realizace výběrového rozptylu
$\widehat{\Sigma}_.^2$	$\widehat{\Sigma}_{\boldsymbol{X}}^2$	výběrový 2. centrální moment (alternativní odhad rozptylu)
$\widehat{\sigma}_.^2$	$\widehat{\sigma}_{\boldsymbol{x}}^2$	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	$H_0 : \mathrm{E}X \leq c$	nulová hypotéza
H_1	$H_1 : \mathrm{E}X > c$	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

1. Základní pojmy teorie pravděpodobnosti

1.1 Kdy potřebujeme pravděpodobnostní popis?

Řekneme, že systém je **náhodný**, jestliže jeho budoucí chování nedovedeme s jistotou předpovědět. Je to obvykle tím, že příčiny nějakých jevů neznáme, nebo jsou tak složité, že je nedovedeme vyhodnotit (např. pro nedostatečnou znalost počátečních podmínek). Účelné chování v náhodném systému vyžaduje připravit se na ty události, které nejspíše nastanou. Např. na letní dovolené v Čechách budeme asi spíše potřebovat sluneční brýle než teplé rukavice, a podle toho se na ni vybavíme.

Pro *pravděpodobnostní (=stochastickou) neurčitost* je typické, že je časově omezená. Pravděpodobnostní popis nahrazuje nedostatek našich znalostí *před* provedením náhodného pokusu. Až pokus proběhne, bude jasné, který z možných výsledků nastal a který ne, neurčitost zmizí. Teorie pravděpodobnosti se zabývá jen takovými jevy.

Omezujeme se na takové náhodné pokusy, které lze opakovat za stejných, nebo aspoň velmi podobných podmínek. (To je ještě více potřeba ve statistice.) Jen zkušenost z opakování pokusu nám může být užitečná pro rozhodování o budoucích akcích. Řídké jevy takto studovat nelze a je velmi obtížné (byť žádoucí) rozšířit použitelnost pravděpodobnostních a statistických metod na náhodné pokusy, jejichž opakovatelnost je omezená.

Poznámka 1.1.1: Často se pravděpodobnost demonstruje na příkladech z hazardních her. Pro teorii pravděpodobnosti je typické studium jevů, o nichž lze uzavírat sázky. Máme totiž splněny dva předpoklady: Před pokusem není výsledek znám, po pokusu je však nezpochybnitelný.

Můžeme například studovat náhodné jevy, jako je čas našeho příchodu na zastávku autobusu, čas příjezdu autobusu a zda autobus stihneme. Můžeme studovat otázku, zda bude zítra dálnice D1 průjezdná, nikoli však, zda bude „průjezdná bez obtíží", protože to není výrok (nelze ho za všech výsledků pokusu jednoznačně vyhodnotit v termínech „pravdivý-nepravdivý").

Poznámka 1.1.2: Setkáváme se i s jinými typy neurčitosti, které nedovolují obvyklý pravděpodobnostní popis, ač se o to mnozí pokoušejí. Pravděpodobnostní popis dovoluje např. usuzovat o výsledku sportovního zápasu, ale nikoli o tom, zda naše družstvo hrálo dobře nebo zda podalo lepší výkon než minule. Tato informace je *vágní*, nemáme na ni jednoznačnou odpověď ani poté, co jsme pokus provedli. Takový výsledek může být splněn více či méně, aniž bychom mohli říci, jak je pravděpodobné, že náš tým hrál dobře. K tomu bychom totiž potřebovali náhodný pokus, který na tuto otázku dává náhodnou, ale jednoznačnou odpověď. Přesto i takové informace nám něco říkají a mohou být užitečné při dalším rozhodování. Jejich popis si vyžaduje více než dvě pravděpodobnostní hodnoty. Tímto zobecněním se zabývá *fuzzy logika*. Vzhledem k odlišnému původu a vlastnostem popisované neurčitosti nedovoluje náhradu pravděpodobnostním popisem.

Poznámka 1.1.3: V kvantových systémech se projevuje *kvantová neurčitost*. Ta je zobecněním pravděpodobnostní neurčitosti a sdílí s ní základní vlastnosti: popisuje systém před pokusem, zatímco po pokusu máme jednoznačné odpovědi na některé otázky, které jsme testovali. Rozdíl oproti klasické teorii pravděpodobnosti je v tom, že na některé otázky nemáme odpověď, neboť testováním jsme změnili systém a přišli o možnost jiných testů, které v původním stavu bylo možno provést.

Vymezili jsme aspoň rámcově oblast zkoumání, abychom mohli přikročit k matematickému popisu.

Motivační příklady

Příklad 1.1.4 (cena losu): Za účast ve statistickém průzkumu dostala Alice los. Jeden los z 1000 vyhrává 1000 €, ostatní nic. Bob její los ztratil ještě před losováním. Jakou jí má zaplatit adekvátní náhradu? Hodnota losu *po losování* je 0 nebo 1000 € v závislosti na tom, zda šlo o vyhrávající los. My však nevíme, zda na ztracený los připadla výhra.

Adekvátní hodnota losu před losováním je $\frac{1}{1000}\,1000 = 1\,€$, což je průměrná hodnota po losování. Obvyklá prodejní cena takového losu bývá 2 €, přesto se losy kupují. Vysvětlení jsou spíše psychologická. Vnímání peněz není lineární, malou ztrátu nepocítíme, velkou výhru ano. Lidé jsou také optimističtí; mnohý si myslí, že právě on bude tím jedním z tisíce. (Málokdy na to pomyslí, když usedá za volant auta a může být jedním z tisíce, který havaruje.)

Poznámka 1.1.5: Hodnota losu před losováním nemusí být důležitá jen pro jeho prodej, ale i pro obezřetnost při skladování a manipulaci. Např. v roce 2006 zůstala na Slovensku nevyzvednutá výhra téměř 1 000 000 €. O příčině lze jen spekulovat. Je možné, že majitel doklad ztratil nebo na něj zapomněl. Kdyby věděl, že jde o poukázku na velkou výhru, zřejmě by s ní nakládal mnohem pečlivěji, avšak před losováním pro to nebyl důvod.

Průměrná cena rozhoduje i o účasti v anketách apod., které jsou odměňovány tak, že vylosovaný účastník dostane velkou výhru, ostatní nic.

Příklad 1.1.6 (rozložení podnikatelského rizika): Otevřeme-li si v létě stánek se slunečními brýlemi, riskujeme velkou ztrátu v případě špatného počasí. Opatrnější bude prodávat navíc např. deštníky, v tom případě jistě nějaké tržby budou, ale vzdáváme se i naděje na velký relativní zisk v případě dobrého počasí. Stejné úvahy se dělají i při velkých investicích, a to nejen při tvorbě portfolia.

Příklad 1.1.7 (riziko pojištění): Cena pojištění se stanovuje v závislosti na pravděpodobnosti pojistné události a na odhadu výšky vyplacené částky.

Příklad 1.1.8 (doba trvání cesty): Každý den si klademe otázku, kdy musíme vyrazit do školy, aby riziko pozdního příchodu bylo menší než jistá „přijatelná" mez. Zde nám postačí hrubý odhad, ale pokud jde o plánování návaznosti operací např. u stavby mrakodrapu, pak je podrobná analýza nutná, jde o složitější formu obdobného problému.

Poznámka 1.1.9 (minumum historie): Za počátky teorie pravděpodobnosti jsou považovány práce B. Pascala (1623–1662), P. de Fermata (1601–1655) a Ch. Huygense (1629–1695), kteří studovali počítání s pravděpodobnostmi a zavedli pojem střední hodnoty. Klíčový byl příspěvek J. Bernoulliho (1654–1705), který mj. dokázal zákon velkých čísel, a A. de Moivrea (1667–1754), který za omezujících předpokladů dokázal centrální limitní větu (a to dříve, než se narodil K.F. Gauss (1777–1855), podle nějž je limitní rozdělení pojmenováno!). Díky nim se studium pravděpodobnosti stalo součástí matematiky. P.S. Laplace (1749–1827) zformuloval v r. 1812 matematický model pravděpodobnosti, který přes své nedostatky posloužil při odvození řady dodnes používaných metod. K vývoji významně přispěli zejména S.D. Poisson (1781-1840), K.F. Gauss (1777–1855) a P.L. Čebyšev (1821–1894). A.A. Markov (1856–1922) položil základy teorie náhodných procesů jako posloupností náhodných pokusů. Přesto byl matematický aparát shledáván nedostatečným a jeho oprava byla jedním z hlavních úkolů matematiky, které pro 20. století vytyčil D. Hilbert (1862–1943). Teprve r. 1933 A.N. Kolmogorov (1903–1987), v návaznosti na práce E. Borela (1871–1956) z teorie míry, podal uspokojivou axiomatizaci pravděpodobnosti. Jeho přístup je dnes standardní.

1.2 Laplaceův model pravděpodobnosti

Intuitivní chápání pravděpodobnosti zformuloval P.S. Laplace v r. 1812. I když je dnes přežité, splnilo svou roli ve vývoji a lze na něm snadno demonstrovat některé základní pojmy, které

později upřesníme. (Někdy se hovoří o **klasické definici pravděpodobnosti**, ale tento pojem se dnes vyskytuje i v jiných významech.)

Mějme náhodný pokus, který má $n \in \mathbb{N}$ různých, vzájemně se vylučujících výsledků, které jsou *stejně možné*. Pak pravděpodobnost jevu, který nastává právě při k z těchto výsledků, je k/n.

Zřejmě je problém definovat, co znamená, že výsledky jsou „stejně možné". Rádi bychom řekli „stejně pravděpodobné" (což podle této definice jsou), jenže pak bychom dostali definici kruhem, neboť pravděpodobnost je to, co chceme definovat. Taková konstrukce je v matematice nepřijatelná. Trvalo však více než jedno století, než se ji podařilo korektně nahradit.

V Laplaceově pojetí náhodného pokusu všech n „stejně možných" výsledků nazýváme **elementární jevy** (angl. *elementary event, simple event*). Množinu všech elementárních jevů označme Ω. Každý **jev** (angl. *event*) pozorovatelný v tomto pokusu je popsán podmnožinou $A \subseteq \Omega$ těch elementárních jevů, při nichž příslušný jev nastává.

Úmluva. *Nadále budeme jevy ztotožňovat s příslušnými množinami elementárních jevů a používat pro ně množinové operace (místo výrokových).*

Speciálně **jev jistý** (angl. *the certain event, sure event*), který vždy nastane, je reprezentován množinou Ω, značíme jej též **1**; **jev nemožný** (angl. *the impossible event*), který nikdy nenastane, je reprezentován prázdnou množinou $\emptyset$, značíme jej též **0**. Pro jevy $A, B \subseteq \Omega$ je jejich **konjunkce** („and") reprezentována průnikem $A \cap B$ a **disjunkce** („or") sjednocením $A \cup B$. **Jev opačný** k A (neboli **negace** jevu A, tj. jev, který nastává, právě když A nenastává), je reprezentován množinovým doplňkem $\bar{A} = \Omega \setminus A$. **Jev složený** z jevů $A_1, \ldots, A_k$ je jakýkoli jev, který lze vyjádřit pomocí jevů $A_1, \ldots, A_k$ a spojek výrokové logiky (konjunkce, disjunkce a negace, přičemž např. disjunkci nepotřebujeme), speciálně i jev jistý $A_1 \cup \bar{A}_1$, a jev nemožný $A_1 \cap \bar{A}_1$. Pokud jev B nastává vždy, když nastane A, tj. platí-li implikace $A \implies B$, pak v množinové reprezentaci tomu odpovídá inkluze $A \subseteq B$. Jestliže jev A nastává při k elementárních jevech, pak jeho pravděpodobnost je $P(A) = k/n$. Všechny jevy pozorovatelné v náhodném pokusu tvoří **jevové pole**; zde je jím množina $\exp \Omega$ všech podmnožin množiny Ω.

Náhodná veličina (angl. *random variable*) je libovolná funkce $X\colon \Omega \to \mathbb{R}$. Její **střední hodnota** (angl. *mean, expected value, expectation*) je

$$\mathrm{E}X = \frac{1}{n} \sum_{\omega \in \Omega} X(\omega) \,, \tag{1.1}$$

kde $n = |\Omega|$ je počet prvků množiny Ω. Lze ji interpretovat následovně: Je-li hodnota náhodné veličiny hodnotou výhry ve hře, pak střední hodnota je spravedlivá cena za účast ve hře. (Pojmy náhodné veličiny a střední hodnoty později ještě zobecníme.)

Někdy můžeme na stejnou pravděpodobnost elementárních jevů usuzovat na základě symetrie – není důvod, proč by některý z nich měl mít jinou pravděpodobnost než ostatní.

Příklad 1.2.1: Pro jeden hod kostkou tvoří množinu všech elementárních jevů všech 6 možných výsledků, $\Omega = \{1, \ldots, 6\}$ (označujeme je odpovídajícími čísly). Každý elementární jev má stejnou pravděpodobnost $1/6$ a každý jev $A \subseteq \Omega$ má pravděpodobnost $|A|/6$, závislou jen na počtu prvků množiny A (tj. na počtu elementárních jevů, při nichž nastává jev A).

Příklad 1.2.2: Hodíme dvěma kostkami a za výsledek pokusu považujeme součet čísel na obou kostkách. To je náhodná veličina Y, která může nabývat 11 hodnot $2, 3, \ldots, 12$. Tyto výsledky však nejsou stejně pravděpodobné a jejich pravděpodobnost není $1/11$. Správné bude rozlišovat obě kostky a pořadí čísel na nich. Pak dostaneme $6^2 = 36$ elementárních jevů, kterými jsou uspořádané dvojice čísel $1, \ldots, 6$. Např. výsledku 9 odpovídá čtyřprvková množina elementárních jevů $\{(3,6), (4,5), (5,4), (6,3)\}$, výsledku 4 množina $\{(1,3), (2,2), (3,1)\}$, výsledku 12 množina $\{(6,6)\}$. Pravděpodobnosti výsledků uvádí tabulka:

výsledek	2	3	4	5	6	7	8	9	10	11	12
počet elementárních jevů	1	2	3	4	5	6	5	4	3	2	1
pravděpodobnost	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Střední hodnota je

$$EY = \frac{1}{36}2 + \frac{1}{18}3 + \frac{1}{12}4 + \frac{1}{9}5 + \frac{5}{36}6 + \frac{1}{6}7 + \frac{5}{36}8 + \frac{1}{9}9 + \frac{1}{12}10 + \frac{1}{18}11 + \frac{1}{36}12 = 7,$$

což nepřekvapí, neboť je to dvojnásobek střední hodnoty výsledku hodu jednou kostkou,

$$\frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}.$$

Na problém narazíme, pokud není zřejmé, jak najít stejně pravděpodobné elementární jevy. Jen někdy je možné vystačit s upraveným Laplaceovým modelem, jako v následujícím příkladu.

Příklad 1.2.3: Příklad 1.2.1 lze pozměnit tak, že kostka je nepravidelná, např. šestka padá s pravděpodobností 1/2, ostatní čísla se stejnými pravděpodobnostmi 1/10. Můžeme si představit následující náhodný pokus, který dává stejný výsledek: hrací kostka má 10 stěn, všechny stejně pravděpodobné, na 5 z nich je šestka, na zbývajících čísla $1, \ldots, 5$. Pravděpodobnost se změní na

$$P^*(A) = \begin{cases} \frac{|A|}{10} & \text{pro } 6 \notin A, \\ \frac{|A|+4}{10} & \text{pro } 6 \in A. \end{cases} \tag{1.2}$$

Předchozí postup zjevně selhává, pokud by např. pravděpodobnost šestky měla být dána iracionálním číslem.

Poznámka 1.2.4: Původní hrací kostky byly zvířecí záprstní kůstky a jednotlivé výsledky nebyly zdaleka stejně pravděpodobné; odhad skutečné pravděpodobnosti se promítal do pravidel her.

Příklad 1.2.5: Navrhněte, jak pro dané $n \in \mathbb{N}$, $n \geq 3$, uspořádat náhodný pokus, který má n stejně pravděpodobných výsledků.

Řešení: Místo kostky bychom mohli házet pravidelným n-bokým hranolem, dostatečně dlouhým, aby nezůstal ležet na podstavě. (Můžeme kvůli tomu konce zaoblit.) Tato myšlenka nalezne uplatnění i v počítačových generátorech náhodných čísel. Pro $n = 2k$, $k \in \mathbb{N}$, bychom také mohli použít dva pravidelné k-boké jehlany spojené podstavami. □

Ještě vážnější námitkou proti Laplaceově definici pravděpodobnosti je existence úloh, které mají nekonečně mnoho možných výsledků.

Příklad 1.2.6: Kdybychom místo kostkou či hranolem házeli dlouhým válcem, bylo by teoreticky nekonečně mnoho poloh, v nichž se může zastavit.

Příklad 1.2.7: Ve hře „kámen-nůžky-papír" se má vybrat jeden ze dvou hráčů tak, aby měli oba stejnou šanci. To se obvykle dělá tak, že oba současně zvolí svou strategii, a ta se v jednom kole vyhodnotí podle tabulky (V=výhra hráče, který volil řádek, P=prohra, R=remíza). V případě remízy se hraje další kolo podle stejných pravidel, dokud se nerozhodne.

	kámen	nůžky	papír
kámen	R	V	P
nůžky	P	R	V
papír	V	P	R

Takový postup je sice spravedlivý, ale může trvat libovolně dlouho; neexistuje konečný počet kol, který by zaručoval rozhodnutí. Všech možných průběhů hry je nekonečně mnoho.

Příklad 1.2.8 (ruinování rulety): Máme hru, při níž hráč vsadí zvolenou částku, s pravděpodobností 1/2 o ni přijde, s pravděpodobností 1/2 vyhraje dvojnásobek (tj. dostane peníze zpět a k tomu stejně peněz, kolik vsadil). Hráč zvolí následující strategii „zaručené výhry". V prvním kole vsadí 1 €. Pokud vyhraje, skončí (s výhrou 1 €). Pokud prohraje, vsadí v dalším kole vždy dvojnásobek předchozí částky. Tak pokračuje do první výhry, po ní končí. Odnáší si výhru 1 €, neboť v kole k vsadil 2^{k-1} €, vydělal 2^{k-1} € a z toho uhradil prohry z předchozích kol, které činily

$$\sum_{i=0}^{k-2} 2^i = 2^{k-1} - 1\,€\,.$$

Hráč má malou, ale zajištěnou výhru. Kromě toho může hru opakovat. Úloha vznikla zjednodušením a malou úpravou pravidel rulety, podle tohoto postupu lze opravdu hrát. Máme tedy stroj na peníze, přestože o každém kole lze jednoduše ukázat, že průměrná výhra v něm je nulová. Kde je chyba?

Laplaceova definice pravděpodobnosti nám tento paradox neumožňuje rozřešit, neboť není omezen počet kol. Tím pádem ani nelze najít konečný počet stejně pravděpodobných výsledků, postihujících všechny možné situace. K řešení budeme potřebovat jiný popis.

Lze si rovněž představit analogové měření veličiny, při němž výsledek je vyjádřen reálným číslem. Pak máme dokonce nespočetně mnoho možných výsledků.

Poznámka 1.2.9: Námitka, že omezená přesnost dovoluje rozlišit jen konečně mnoho hodnot, naráží na problém, že by tato přesnost musela být předem známa. Vždy však máme možnost použít přesnější měřicí metodu a umožnit více výsledků. Kromě toho popis veličin reálnými čísly bývá jednodušší než pomocí kvantovaných veličin s konečně mnoha hodnotami.

Rozšíření Laplaceova modelu pravděpodobnosti

Příklad 1.2.10: Místo hrací kostky házíme krabičkou od zápalek, jejíž strany jsou nestejně dlouhé. Jaká je pravděpodobnost možných výsledků?

Některé z námitek proti Laplaceově definici pravděpodobnosti lze řešit, pokud připustíme, že *elementární jevy nemusí být stejně pravděpodobné*. Takto lze popsat např. hod libovolně nepravidelnou kostkou z předchozího příkladu. Odstraní se tím i omezení hodnot pravděpodobnosti na racionální čísla. Přináší to ale i problémy.

Především ztrácíme návod, jak pravděpodobnost stanovit. Víme, že je to funkce, která elementárním (i všem od nich odvozeným) jevům přiřazuje čísla z intervalu $\langle 0, 1\rangle$ a splňuje jisté podmínky (které zformulujeme později). Takových funkcí však může být mnoho a my nemáme návod, jak z nich vybrat tu pravou. Např. při házení nepravidelnou kostkou známe všechny možné výsledky a tušíme, že mají různé pravděpodobnosti, ale nevíme jaké. Ukáže se, že tato nevýhoda je neodstranitelná a bude nás provázet celou učebnicí. Vlastně je důvodem pro vznik statistiky, která se v podstatě zabývá tím, jak k danému opakovatelnému pokusu vybrat vhodný pravděpodobnostní model.

Dále lze namítnout, že nemáme jak definovat pravděpodobnost a vyhnout se definici kruhem. V kapitole 1.7 se dočkáme řešení v podobě Kolmogorovova modelu pravděpodobnosti, které je jen zčásti uspokojivé. Pravděpodobnost nadefinujeme jako *jakoukoli* funkci splňující stanovené podmínky. Ty nám umožní s ní dále pracovat, aniž bychom se starali, zda tato pravděpodobnost odpovídá našemu experimentu, nebo nějakému jinému. (To je v matematice obvyklý přístup, že pojmy jako např. *lineární prostor* jsou zavedeny podle vlastností, které splňují, a abstrahují od jejich konkrétní podoby.)

Hlavní problém ale je, že ani rozšíření Laplaceova modelu pravděpodobnosti nedovoluje popsat náhodné pokusy s nekonečně mnoha výsledky. Určitá šance na zobecnění by byla, pokud je elementárních jevů spočetně mnoho. Nicméně výsledky řady pokusů jsou přirozeně popsány

např. reálnými čísly nebo body v prostoru, kterých je nespočetně mnoho. Zde předchozí přístup zcela selhává. Několik takových úloh připomeneme v kapitole 1.6.

Ukázali jsme popis náhodného pokusu pomocí *konečně mnoha stejně pravděpodobných* elementárních jevů. Ten byl po staletí východiskem netriviálních základních poznatků z této disciplíny a umožňuje jednoduchou představu o významu základních pojmů včetně náhodné veličiny a střední hodnoty. Jeho možnosti jsou omezené a později jej zobecníme. Tento model zůstane i nadále použitelný pro řadu náhodných pokusů a závěry z něj učiněné zůstávají v platnosti i pro pozdější přístupy, je ovšem třeba kontrolovat předpoklady, z nichž vychází.

Ukázali jsme, že Laplacův model si vyžaduje zásadnější ústupky než to, že elementární jevy nemusí být stejně pravděpodobné. K jeho revizi přistoupíme v kapitole 1.7.

Cvičení 1.2.11: Zkuste upravit příklad 1.2.7 tak, aby byl předem znám maximální počet kol.

Řešení: Zakážeme jeden znak, např. papír. Zbylé vyhodnocujeme tak, že jeden z hráčů vyhrává při rovnosti znaků, druhý při nerovnosti. Dostáváme následující tabulku a rozhodnutí padne hned v prvním kole. Matematická varianta téhož principu: Hráči volí celá čísla; první vyhrává, je-li součet lichý, druhý při sudém součtu.

	kámen	nůžky
kámen	P	V
nůžky	V	P

□

Cvičení 1.2.12: Zkuste upravit příklad 1.2.7 tak, abychom spravedlivě vybrali jednoho ze tří hráčů v předem známém maximálním počtu kol.

Řešení: Použijeme všechny tři znaky, přiřadíme jim čísla $0, 1, 2$. Volby všech tří hráčů sečteme a vezmeme zbytek po dělení třemi. Každému hráči je předem přidělena jedna zbytková třída, při níž je vybrán. Takto každý z hráčů má stejnou možnost ovlivnit výsledek. Pravděpodobnost, že bude vybrán, je $9/27 = 1/3$. Rozhodnutí padne hned v prvním kole. Cvičení 1.2.11 bylo totéž pro zbytkové třídy při dělení dvěma. □

Cvičení 1.2.13: Zkuste upravit příklad 1.2.7 tak, abychom spravedlivě vybrali jednoho ze dvou hráčů v předem známém maximálním počtu kol. Přitom každému hráči ponecháme 3 možné strategie.

Řešení: Tato úloha nemá řešení. Počet všech elementárních jevů je 3^k, kde k je maximální počet kol. (Pokud hra skončí po menším než maximálním počtu kol, můžeme ji prodloužit na stejný počet kol, přičemž další kola již neovlivní známý výsledek.) Při lichém počtu elementárních jevů neexistuje jev s pravděpodobností $1/2$. S rostoucím počtem kol lze znevýhodnění některého hráče pouze snížit, nikoli odstranit. □

Cvičení 1.2.14 (vyhrávající kostky): Máme na vybranou 3 hrací kostky K, L, M, na nichž jsou napsána čísla dle následující tabulky; všechna jsou stejně pravděpodobná. Alice si vybere jednu kostku, pak si Bob vybere jednu ze zbývajících kostek. Oba jednou hodí vybranou kostkou; komu padne vyšší číslo, vyhrává (rovnost nemůže nastat).

K	3	3	3	3	3	3
L	1	1	4	4	4	4
M	2	2	2	2	5	5

Kterou kostku si má Alice, resp. Bob vybrat, aby měli co nejvyšší naději na výhru?

Řešení: Kostka K dává konstantní výsledek, který proti kostce L dává šanci na výhru $1/3$, proti kostce M $2/3$, což by svědčilo pro preferenci kostky L. Avšak L proti M vyhrává pouze při výsledku $(4, 2)$, který nastává s pravděpodobností $(2/3)^2 = 4/9 < 1/2$. Tabulka uvádí pravděpodobnost výhry kostky v příslušném řádku nad kostkou ve sloupci:

	K	L	M
K		$\frac{1}{3}$	$\frac{2}{3}$
L	$\frac{2}{3}$		$\frac{4}{9}$
M	$\frac{1}{3}$	$\frac{5}{9}$	

Žádná z kostek není lepší než zbývající dvě. Při libovolné volbě Alice si Bob může vybrat kostku, s níž má větší naději na výhru. (Stejně jako ve hře „kámen-nůžky-papír", zde však hráči volí své strategie postupně a výsledek je náhodný.) Bob je tedy zvýhodněn a při správné strategii ve většině případů vyhrává. Alice v průměru prohrává, nejméně tehdy, když si vybere kostku L; pak Bob vybere M a vyhrává s pravděpodobností 5/9. □

Cvičení 1.2.15: Najděte jiná čísla na kostkách, se kterými by příklad 1.2.14 fungoval podobně.

Řešení: Např. na kostce K by mohla být čísla $0, 0, 3, 3, 6, 6$. □

Cvičení 1.2.16: Jak se změní strategie v příkladu 1.2.14, pokud si Bob může vybrat kostku až ve chvíli, kdy Alice hodila (a je znám výsledek)?

Řešení: V tomto případě znalost výsledku Alice neumožňuje zlepšit Bobovu strategii. □

Cvičení 1.2.17: Změňme pravidla hry z příkladu 1.2.14 tak, že Alice si kostku vybírá, ale Bob ze zbylých jen losuje (s pravděpodobností 1/2). Jakou kostku si má Alice vybrat?

Řešení: Alice si vybere kostku L a *vyhrává* s pravděpodobností

$$\frac{\frac{2}{3}+\frac{4}{9}}{2} = \frac{5}{9} > \frac{1}{2}.$$

□

1.3 Kombinatorické pojmy a vzorce

(Dle [Zvára, Štěpán 2002].)

Zde zopakujeme základní situace používané při modelování náhodných pokusů. Poslouží nám „urnový model". V urně je n rozlišitelných objektů (např. očíslovaných koulí). Vystačíme zde s Laplaceovým modelem pravděpodobnosti, v němž předpokládáme, že každý objekt v urně má stejnou pravděpodobnost, že bude vytažen. Postupně vytáhneme k objektů; otázka zní, kolik je možných (posloupností) výsledků. Záleží ovšem na tom, zda tažené objekty vracíme zpět do urny (pak se jedná o **výběr s vracením**, angl. *sampling with replacement*), nebo nikoli (pak se jedná o **výběr bez vracení**, angl. *sampling without replacement*). Dále záleží na tom, zda ve výsledcích rozlišujeme pořadí tažených objektů, nebo pouze, které objekty byly (někdy) taženy. Pokud nám *záleží na pořadí*, pak hovoříme **uspořádaném výběru** a o **variacích**. Pokud *nerozlišujeme pořadí*, v němž byly objekty taženy, pak hovoříme **neuspořádaném výběru** a o **kombinacích**. Podle toho rozlišíme 4 základní situace (zde řazené od nejjednodušší k nejsložitější z hlediska matematického popisu):

Uspořádaný výběr s vracením je popsán **variacemi s opakováním**. Rozlišujeme pořadí tahů a objekty mohou být taženy opakovaně. Počet způsobů, kterými lze takto vybrat k objektů z n, je

$$n^k,$$

neboť při každém z k tahů máme n možností. Všech n^k možných výsledků má v Laplaceově modelu stejnou pravděpodobnost $1/n^k$.

Uspořádaný výběr bez vracení je popsán **variacemi bez opakování**. V prvním tahu máme n možností, ve druhém už jen $n-1$ atd. Počet způsobů, kterými lze takto vybrat k objektů z n, je

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

Alternativní značení: $V(n,k)$, $V_k(n)$. Speciálně pro $k = n$ (vytáhneme postupně všechny objekty) dostáváme počet všech **permutací** neboli **pořadí** (**bez opakování**) n prvků:

$$n \cdot (n-1) \cdot \ldots \cdot 1 = n!.$$

Alternativní značení: $P(n)$, P_n. V Laplaceově modelu mají všechny variace s opakováním stejnou pravděpodobnost $\frac{(n-k)!}{n!}$, všechny permutace pravděpodobnost $\frac{1}{n!}$.

Neuspořádaný výběr bez vracení je popsán **kombinacemi bez opakování**. Ve srovnání s variacemi bez opakování je jich méně v poměru $k!$, což je počet všech pořadí tažených objektů, neboť na pořadí nezáleží. Vždy $k!$ různých variací bez opakování odpovídá jedné kombinaci bez opakování. Počet způsobů, kterými lze takto vybrat k objektů z n, je

$$\frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k},$$

což je binomický koeficient (kombinační číslo). Alternativní značení: $C(n,k)$, $C_k(n)$. Všechny kombinace bez opakování mají v Laplaceově modelu stejnou pravděpodobnost $1/\binom{n}{k}$.

Kombinace bez opakování lze též motivovat následovně: Provedeme výběr bez vracení, a poté všechny tažené (očíslované) objekty seřadíme podle velikosti; tím se ztratí informace o pořadí, v němž byly taženy. Tedy $\binom{n}{k}$ je počet *rostoucích* posloupností k čísel z množiny $\{1,\ldots,n\}$.

Neuspořádaný výběr s vracením je popsán **kombinacemi s opakováním**. Počet způsobů, kterými lze takto vybrat k objektů z n, je

$$\binom{n+k-1}{k}.$$

Důkaz tohoto výsledku podáme později. Alternativní značení: $C'(n,k)$, $C'_k(n)$.

Kombinace s opakováním lze též motivovat následovně: Provedeme výběr bez vracení, a poté všechny tažené (očíslované) objekty seřadíme podle velikosti. Tedy $\binom{n+k-1}{k}$ je počet *neklesajících* posloupností k čísel z množiny $\{1,\ldots,n\}$.

Nyní zdůvodníme vzorec pro počet kombinací s opakováním.Všechny neklesající posloupnosti délky k můžeme vytvořit tak, že napíšeme za sebe všechna čísla $1,\ldots,n$ a za každé tolik čárek, kolikrát se v kombinaci vyskytuje, např. pro $n = 4$ zapíšeme posloupnost $(1,1,2,4,4,4)$ jako

1 | |
2 |
3
4 | | |

nebo (1 | | 2 | 3 4 | | |). (První znak je vždy 1.) Víme, jak jdou čísla za sebou, takže je beze ztráty informace můžeme všechna nahradit stejným znakem, např. tečkou: (. | | . | . . | | |). Každé kombinaci s opakováním odpovídá posloupnost n teček a k čárek. Ta začíná vždy tečkou, ze zbývajících $n+k-1$ znaků je k čárek, což lze zajistit $\binom{n+k-1}{k}$ způsoby.

Každé kombinaci s opakováním pak může odpovídat jiný počet variací s opakováním (které jsou stejně pravděpodobné).

Příklad 1.3.1: Hodíme dvěma hracími kostkami (jako v příkladu 1.2.2), ale výsledek uspořádáme vzestupně. Pak např. výsledek $(1,1)$ má pravděpodobnost $1/36$, ale výsledek $(1,2)$ má pravděpodobnost $2\times$ větší, neboť před uspořádáním mohl být $(1,2)$ nebo $(2,1)$.

Popíšeme postup, kterým lze spočítat variace s opakováním odpovídající jedné kombinaci s opakováním. Předpokládejme, že máme posloupnost délky k, tvořenou hodnotami $1,\ldots,n$, přičemž hodnota j se v ní opakuje k_j-krát, $\sum_{j=1}^{n} k_j = k$. Otázka je, kolik *různých* posloupností lze vytvořit z původní posloupnosti změnou pořadí prvků. (Opakované hodnoty nejsou rozlišitelné.) Dostáváme počet **permutací** (neboli **pořadí**) **s opakováním**:

$$\frac{k!}{k_1! \cdot \ldots \cdot k_n!}\,.$$

Tento výsledek vychází z úvahy, že k rozlišitelných prvků bychom mohli seřadit $k!$ způsoby, avšak nejsme schopni rozlišit změny pořadí stejných prvků, kterých je v jednotlivých třídách postupně $k_1,\ldots,k_n$. Právě tolik (stejně pravděpodobných) variací s opakováním odpovídá jedné kombinaci s opakováním; tím je určena její pravděpodobnost

$$\frac{k!}{k_1! \cdot \ldots \cdot k_n! \cdot n^k}\,,$$

která ovšem závisí na číslech $k_1,\ldots,k_n$, tedy na tom, kolikrát se v ní která číslice opakuje. Proto – na rozdíl od všech předchozích případů – kombinace s opakováním *nejsou stejně pravděpodobné* (ani v Laplaceově modelu).

Speciálně pro $n = 2$ dostáváme počet permutací s opakováním

$$\frac{k!}{k_1! \cdot k_2!} = \frac{k!}{k_1! \cdot (k-k_1)!} = \binom{k}{k!}\,,$$

což je počet kombinací bez opakování (ovšem k_1-prvkových z k prvků, zatímco permutace s opakováním byly z $n = 2$ prvků).

Tvrzení 1.3.2: *Počet všech podmnožin m-prvkové množiny je 2^m.*

Důkaz: Každý z m prvků původní množiny do podmnožiny buď patří, nebo nepatří; to jsou dvě možnosti, výsledkem je počet variací s opakováním ze dvou prvků („patří"/„nepatří"). □

Tvrzení 1.3.3: *Počet všech k-prvkových podmnožin m-prvkové množiny je $\binom{m}{k}$.*

Důkaz: Každý z m prvků původní množiny do podmnožiny buď patří, nebo nepatří, avšak patřit jich tam musí právě k; výsledkem je počet k-prvkových kombinací bez opakování z m prvků. □

Důsledek 1.3.4:

$$\sum_{k=0}^{m} \binom{m}{k} = 2^m\,.$$

Důkaz: Obě strany vyjadřují počet podmnožin m-prvkové množiny; na levé straně jsou počítány zvlášť podle počtu prvků. Plyne to také z binomické věty:

$$2^m = (1+1)^m = \sum_{k=0}^{m} \binom{m}{k} 1^k\, 1^{m-k} = \sum_{k=0}^{m} \binom{m}{k}\,,$$

ale při jejím důkazu bývají naopak použity zde uvedené kombinatorické výsledky. □

Tvrzení 1.3.5:

$$\binom{m}{k} = \binom{m-1}{k} + \binom{m-1}{k-1}\,.$$

Důkaz: Levá strana je počet všech k-prvkových podmnožin m-prvkové množiny. Z nich těch, do kterých nepatří první prvek, je $\binom{m-1}{k}$ (všech k prvků vybíráme z $m-1$ zbývajících). Těch podmnožin, do kterých patří první prvek, je $\binom{m-1}{k-1}$ (dalších $k-1$ prvků vybíráme z $m-1$ zbývajících). □

Následující příklad porovnává všechny čtyři druhy výběrů.

Příklad 1.3.6: ** Alice a Bob žijí ve státě, který má $n = 10^7$ obyvatel. Proběhne tam statistický průzkum, do kterého 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: Následující úvaha je společná pro všechny typy výběrů, bez ohledu na to, jaká funkce v popisuje počet možností. Celkový počet možností je $v(n,k)$. Spíše než ty možnosti, v nichž je vybrána Alice, je snazší spočítat všechny ostatní, těch je $v(n-1,k)$. Možností, při nichž je vybrána Alice, je $v(n,k) - v(n-1,k)$. Počet případů, kdy není vybrána Alice ani Bob, je $v(n-2,k)$. Ten budeme potřebovat v úloze (b). Ve výrazu $v(n,k) - 2v(n-1,k)$ jsme od celkového počtu případů odečetli počet výběrů bez Alice a počet výběrů bez Boba, takže jsme dvakrát odečetli výběry bez Alice i Boba, a musíme je jednou přičíst. Proto počet možností, kdy je vybrána Alice i Bob, je

$$v(n,k) - 2v(n-1,k) + v(n-2,k) .$$

Jsou-li všechny možnosti stejně pravděpodobné (tj. vždy kromě neuspořádaného výběru s vracením), vyjdou následující pravděpodobnosti (A, resp. B, značí jev, že byla vybrána Alice, resp. Bob):

(a)

$$P(A) = P(B) = 1 - P(\bar{A}) = 1 - \frac{v(n-1,k)}{v(n,k)} ,$$

(b)

$$P(A \cap B) = 1 - P(\bar{A}) - P(\bar{B}) + P(\bar{A} \cap \bar{B}) = 1 - 2\frac{v(n-1,k)}{v(n,k)} + \frac{v(n-2,k)}{v(n,k)} .$$

Zkoumané jevy $A, B, A \cap B$ nezávisí na tom, zda jde o výběr uspořádaný nebo neuspořádaný, jejich pravděpodobnosti na tom nemohou záviset.

Případ (c) má nenulovou pravděpodobnost pouze pro výběry s vracením.

Druhy výběrů rozlišíme indexy u v, P.

Uspořádaný výběr bez vracení:

$$v_{ub}(n,k) = \frac{n!}{(n-k)!} \doteq 6.73 \cdot 10^{69\,997} ,$$

$$P_{ub}(A) = 1 - \frac{v_{ub}(n-1,k)}{v_{ub}(n,k)} = 1 - \frac{(n-1)!}{(n-1-k)!}\frac{(n-k)!}{n!} = 1 - \frac{n-k}{n} = \frac{k}{n} = 0.001 ,$$

$$P_{ub}(\bar{A} \cap \bar{B}) = \frac{(n-2)!}{(n-2-k)!}\frac{(n-k)!}{n!} = \frac{(n-k)(n-k-1)}{n(n-1)} \doteq 0.998\,001 ,$$

$$P_{ub}(A \cap B) = 1 - 2\frac{n-k}{n} + \frac{(n-k)(n-k-1)}{n(n-1)} = \frac{k(k-1)}{n(n-1)} \doteq 9.999 \cdot 10^{-7} .$$

Poslední výsledek potvrzuje i úvaha, že Alice bude vybrána s pravděpodobností $\frac{k}{n}$, a poté ze zbývajících obyvatel Bob do zbytku výběru s pravděpodobností $\frac{k-1}{n-1}$.

Neuspořádaný výběr bez vracení:

$$v_{nb}(n,k) = \binom{n}{k} = \frac{v_{ub}(n,k)}{k!} \doteq 2.365 \cdot 10^{34\,338} ,$$

pravděpodobnosti jsou stejné jako pro uspořádaný výběr bez vracení.

Uspořádaný výběr s vracením:

$$v_{us}(n,k) = n^k = 10^{70\,000}\,,$$
$$P_{us}(A) = 1 - \left(\frac{n-1}{n}\right)^k \doteq 9.995 \cdot 10^{-4}\,,$$
$$P_{us}(A \cap B) = 1 - 2\left(\frac{n-1}{n}\right)^k + \left(\frac{n-2}{n}\right)^k \doteq 9.989 \cdot 10^{-7}\,.$$

(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 $k\, v_{us}(n-1,k-1) = k\,(n-1)^{k-1}$. Pravděpodobnost, že Alice bude vybrána více než jednou, je

$$1 - \frac{v_{us}(n-1,k)}{v_{us}(n,k)} - \frac{k\, v_{us}(n-1,k-1)}{v_{us}(n,k)} = 1 - \left(\frac{n-1}{n}\right)^k - \frac{k\,(n-1)^{k-1}}{n^k} \doteq 4.996 \cdot 10^{-7}\,.$$

Neuspořádaný výběr s vracením: Všech možností je

$$v_{ns}(n,k) = \binom{n+k-1}{k} \doteq 5.203 \cdot 10^{34342}\,,$$

ale *nejsou stejně pravděpodobné.* Počty možností bychom mohli vypočítat jako v obecném případě, ale pravděpodobnosti z nich nelze snadno určit. Pravděpodobnosti jsou stejné jako pro uspořádaný výběr s vracením. □

Poznámka 1.3.7: Předchozí příklad má vadu. Respondenty v průzkumu nemohou být nemluvňata, lidé v komatu apod. Nevybíráme tedy ze všech obyvatel (jejichž počet bývá znám), ale z menší cílové skupiny, jejíž rozsah není snadné stanovit.

Poznámka 1.3.8: Příklad 1.3.6 ukazuje velmi důležitou skutečnost: Pravděpodobnosti, že nějaké objekty budou vybrány, nezávisí na tom, zda uvažujeme uspořádaný nebo neuspořádaný výběr. Proto se v učebnicích obvykle rozlišují pouze výběry s vracením a bez vracení a o uspořádanosti se nehovoří. Činíme tak i v tomto skriptu. Pro výpočet pravděpodobností se používá ze dvou ekvivalentních možností ten model, který je jednodušší, což bývá *neuspořádaný výběr bez vracení* nebo *uspořádaný model s vracením.*

Když bude rozsah výběru malý ve srovnání s počtem všech objektů, lze předpokládat, že pravděpodobnost opakovaného výběru téhož objektu bude zanedbatelná a rozdíl mezi výběry bez vracení a s vracením bude nepodstatný:

Věta 1.3.9: *Pro pevně dané $k \in \mathbb{N}$ a pro $n \to \infty$ ($n \in \mathbb{N}$) se poměr počtů variací (resp. kombinací) bez opakování a s opakováním blíží jedné, tj.*

$$\lim_{n\to\infty} \frac{n!}{(n-k)!\,n^k} = 1\,,$$
$$\lim_{n\to\infty} \frac{\binom{n}{k}}{\binom{n+k-1}{k}} = 1\,.$$

Důkaz:

$$\frac{n!}{(n-k)!\,n^k} = \frac{n\,(n-1)\cdots(n-(k-1))}{n^k} = 1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \to 1\,,$$
$$\frac{\binom{n}{k}}{\binom{n+k-1}{k}} = \frac{n\,(n-1)\cdots(n-(k-1))}{(n+(k-1))\cdots(n+1)\,n} = \frac{1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{\left(1+\frac{k-1}{n}\right)\cdots\left(1+\frac{1}{n}\right)1} \to 1\,,$$

kde důležité je, že počet činitelů k je konstantní. □

Důsledek 1.3.10: *Pro $n \gg k$ můžeme počet variací (resp. kombinací) s opakováním počítat podle přibližných vzorců*

$$\frac{n!}{(n-k)!} \doteq n^k ,$$
$$\binom{n}{k} \doteq \frac{n^k}{k!} .$$

Ve statistice míváme velké výběry z ještě mnohem většího počtu objektů, takže i stanovení počtu všech možností může být zdlouhavé. *(Nevěřící čtenář si může zkusit přepočítat příklad 1.3.6.)* Např. výpočet

$$\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!\,(n-k)!}$$

dle posledního vzorce vyžaduje $2\,k-2$ nebo $2\,(n-k)-2$ operací. Důsledek 1.3.10 dovoluje počítat variace jako mocninu pomocí logaritmů. Faktoriál v druhém vzorci má složitost lineární v k, ale jsou známé aproximace, např. Stirlingův vzorec (C.11), takže můžeme docílit nižší složitost téměř nezávislou na argumentech. Stirlingův vzorec ovšem můžeme použít i přímo pro definiční vztah.

Zopakovali jsme si kombinatorické vzorce ze střední školy. Spolu s nimi jsme ukázali čtyři základní typy výběrů, uspořádaných a neuspořádaných, s vracením a bez vracení. Zdůvodnili jsme, že v obvyklých úlohách (kde výsledek záleží jen na tom, jaké objekty byly taženy, nikoli v jakém pořadí) můžeme pro výpočet pravděpodobnosti použít dle potřeby uspořádaný nebo neuspořádaný výběr. Pravděpodobnost výskytu ve výběru to neovlivní. Tím se můžeme zejména vyhnout uspořádanému výběru bez vracení, v němž jako jediném nejsou všechny možnosti stejně pravděpodobné, takže jej nelze řešit Laplaceovým modelem pravděpodobnosti (pouze rozšířeným).

Navíc, pokud má výběr mnohem méně prvků než je všech možných objektů, i rozdíl mezi výběrem s vracením a bez vracení může být nepodstatný. To se bude hodit pro volbu co nejjednoduššího popisu zejména ve statistice.

Příklad 1.3.11: [Němčíková 1990, Př. 2.24] * 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. Hledaný jev nastává pro ty výběry, v nichž je k vadných výrobků. Z K vadných vybereme k, to lze provést $\binom{K}{k}$ způsoby, a z $M-K$ dobrých vybereme $m-k$, to lze $\binom{M-K}{m-k}$ způsoby, celkový počet možností je $\binom{K}{k}\binom{M-K}{m-k}$. Výsledná pravděpodobnost je

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

Odvodili jsme tzv. **hypergeometrické rozdělení** (viz dodatek B.1). Stejným rozdělením se řídí např. počet zjištěných výskytů onemocnění ve vzorku populace. □

Cvičení 1.3.12 (riziko reklamace dodávky): Předpokládáme, že mezi výrobky je $p = 0.001$ zmetků, rozložených náhodně (nezávisle). Jaké je riziko, že v dodávce $n = 1000$ výrobků bude zmetků více než (a) 1 (b) 2. Jak by se situace změnila, kdyby se zmetkovitost zvýšila na $p = 0.002$? Napřed si zkuste odhadnout výsledek.

Řešení: Jedná se o binomické rozdělení $\mathrm{Bi}(n, p)$, zajímají nás z něj jen pravděpodobnosti nejnižších výsledků:

počet zmetků	pravděpodobnost obecně	pravděpodobnost číselně
0	$(1-p)^n$	0.367 70
1	$n\,p\,(1-p)^{n-1}$	0.368 06
2	$\frac{n\,(n-1)}{2}\,p^2\,(1-p)^{n-2}$	0.184 03

(a) Nebudeme sčítat pravděpodobnosti počtu zmetků od 2 do 1000, nýbrž od jednotky odečteme pravděpodobnost opačného jevu (žádný nebo jeden zmetek):

$$1 - (1-p)^n - np\,(1-p)^{n-1} \doteq 0.264\,24\,.$$

(b) Odečítáme navíc pravděpodobnost 2 zmetků:

$$1 - (1-p)^n - np\,(1-p)^{n-1} - \frac{n\,(n-1)}{2}\,p^2\,(1-p)^{n-2} \doteq 8.020\,9 \cdot 10^{-2}\,.$$

Vzorce platí obecně; pro $p = 0.002$ jsou přibližné číselné hodnoty (a) 0.594 27, (b) 0.323 32. □

Cvičení 1.3.13 (jak vyrábět čipy): Předpokládáme, že při výrobě paměťových prvků je pravděpodobnost vadného prvku $p = 2^{-27} \doteq 7.450\,6 \cdot 10^{-9}$ a jsou rozloženy náhodně (nezávisle).

(a) Jaká je pravděpodobnost, že na čipu s $n = 2^{29}$ paměťovými prvky jsou všechny dobré?
(b) Jak se pravděpodobnost změní, když napřed vyrobíme 2^{30} prvků v 16 sekcích po $m = 2^{26}$ a pak nám stačí z nich vybrat 8 dobrých?

Napřed si zkuste odhadnout výsledek.

Řešení: (a) $(1-p)^n \doteq \mathrm{e}^{-np} = \mathrm{e}^{-4} \doteq 1.831\,6 \cdot 10^{-2}$
(b) $q = (1-p)^m \doteq \mathrm{e}^{-mp} = \mathrm{e}^{-\frac{1}{2}} \doteq 0.606\,53$, pomocí $q = \exp(-mp)$ vychází

$$\sum_{i=8}^{16} \binom{16}{i} q^i\,(1-q)^{16-i} \doteq 0.869\,73\,.$$

Předpoklad nezávislosti je problematický a čísla v této úloze jsou fiktivní. Nicméně jsme ukázali princip, který umožnil pokrok ve výrobě pamětí za cenu dodatečné operace, při níž se vyberou bezporuchové sekce. □

Cvičení 1.3.14: Kufr má číselný zámek s 3 číslicemi 0–9. Zapomněli jsme kód, ale víme, že v něm byla jednotka a že dvě číslice byly stejné. Kolik je možností?

Cvičení 1.3.15: Kufr má dva číselné zámky, každý s 3 číslicemi 0–9. Jaký počet (vhodně zvolených) pokusů zaručí, že otevřeme oba zámky? Jak by se výsledek změnil, kdyby místo toho byl jediný zámek s 6 číslicemi?

1.4 Vlastnosti pravděpodobnosti

Shrňme si vlastnosti, které by pravděpodobnost měla mít nejen v Laplaceově modelu (na němž je lze snadno ukázat), ale i v jeho zobecněních.

Příklad 1.4.1: Alice: „Bobe, jak je pravděpodobné, že tu zkoušku oba uděláme a budeme mít o víkendu volno?“ Bob: „Řekl bych 70 %.“ Alice: „Ale včera jsi tvrdil, že ji na 90 % uděláš a že já mám stejnou šanci.“ Bob: „No a?“ Alice: „To není možné, a jestli to nevidíš, tak tu zkoušku asi neuděláš.“ Zdůvodněte!

Dva jevy nazýváme **neslučitelné** (též **vzájemně se vylučující**), jestliže nemohou nastat současně. Jsou tedy reprezentovány disjunktními množinami, proto se nazývají také **disjunktní**. Pro množinu jevů řekneme, že jsou **po dvou neslučitelné**, jestliže každé dva z nich jsou neslučitelné.

Následující tvrzení shrnuje nejdůležitější vlastnosti pravděpodobnosti. V Laplaceově modelu vyplývají jednoduše z definice. Později v obecnějším přístupu zůstanou v platnosti, ovšem jejich důkaz ze slabších předpokladů bude o něco obtížnější.

Tvrzení 1.4.2: *Pravděpodobnost P splňuje pro libovolné jevy A, B následující vlastnosti:*

1. $P(A) \in \langle 0, 1 \rangle$.
2. $P(\mathbf{0}) = 0$, $P(\mathbf{1}) = 1$.
3. $P(\bar{A}) = 1 - P(A)$.
4. $A \subseteq B \implies P(A) \leq P(B)$.
5. $A \subseteq B \implies P(B \setminus A) = P(B) - P(A)$.
6. $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$.
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Důkaz: Dokážeme pouze poslední vztah. Vyjádříme $A \cup B$ jako sjednocení neslučitelných jevů, $A \cup B = A \cup (B \setminus (A \cap B))$. Odtud

$$P(A \cup B) = P(A \cup (B \setminus (A \cap B))) = P(A) + P(B \setminus (A \cap B)) = P(A) + P(B) - P(A \cap B).$$

□

Nyní vyřešíme úvodní příklad:

Řešení (příkladu 1.4.1): Označíme-li A, resp. B, jev, že Alice, resp. Bob, udělá zkoušku, pak podle právě dokázaného vztahu

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.9 + 0.9 - 0.7 = 1.1\,,$$

což nelze. □

Zobecněním předchozí úvahy dostaneme z podmínky

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq 1$$

tzv. **Benferroniho nerovnost**

$$P(A \cap B) \geq P(A) + P(B) - 1\,.$$

Definice 1.4.3: ***Úplný systém jevů*** *je množina jevů $\{B_j \mid j \in I\}$, které jsou po dvou neslučitelné a jejichž sjednocením je jev jistý.* $\bigcup_{j \in I} B_j = \mathbf{1}$.

Úplný systém jevů má tu vlastnost, že vždy nastane právě jeden z nich. Úplný systém *dvou* jevů musí nutně být tvaru $\{C, \bar{C}\}$, kde C je nějaký jev.

Příklad 1.4.4: Pro výsledky zkoušky by mohl být úplný systém jevů $\{B_1, \ldots, B_5\}$, kde B_j znamená známku j ($j = 1, \ldots, 4$) a B_5 znamená, že žádná známka nebyla udělena (např. student se omluvil nebo zkouška se vůbec nekonala).

Tvrzení 1.4.5: *Nechť $\{B_1, \ldots, B_n\}$ je úplný systém jevů. Pak pro pravděpodobnost jevu A platí*

$$P(A) = \sum_{j=1}^{n} P(A \cap B_j)\,.$$

Speciálně pro $A = \mathbf{1}$

$$\sum_{j=1}^{n} P(B_j) = 1\,,$$

pro úplný systém dvou jevů, $\{B, \bar{B}\}$:

$$P(A) = P(A \cap B) + P\left(A \cap \bar{B}\right)\,. \tag{1.3}$$

Uvedli jsme základní vlastnosti pravděpodobnosti, které v Laplaceově modelu přirozeně vycházejí, ale zůstanou v platnosti i v jeho zobecněních. Nadále je budeme hojně používat, proto by čtenář neměl pokračovat bez jejich osvojení.

Cvičení 1.4.6: [Hsu 1996] Náhodný pokus má 3 možné výsledky a, b, c. Najděte jejich pravděpodobnosti, víte-li, že $P(\{a, c\}) = 0.75$, $P(\{b, c\}) = 0.75$.

Řešení: $P(a) = 0.4$, $P(b) = 0.25$, $P(c) = 0.35$.

Cvičení 1.4.7: [Hsu 1996] Najděte pravděpodobnost jevu $A \cup B \cup C$, jestliže $P(A) = P(B) = 1/4$, $P(C) = 1/3$, $P(A \cap B) = 1/8$, $P(A \cap C) = 1/6$, $P(B \cap C) = 0$.

Řešení: 13/24.

1.5 Nezávislé jevy

Nezávislost dvou jevů

Příklad 1.5.1: Mnohdy spolu dva náhodné jevy „nesouvisí". Například při hodu dvěma kostkami lze předpokládat, že výsledky na sobě nezávisí; vzájemné působení kostek (gravitační, magnetické apod.) je příliš malé na to, aby ovlivnilo výsledek, zvláště, budou-li kostky např. v různých městech.

Jinak tomu bude např. u cen ropy na burzách (i v různých městech). Ty se navzájem ovlivňují, navíc se všechny mění podle společných příčin.

Někdy není jasné, zda jsou jevy nezávislé, např. výsledky dvou fotbalových zápasů. Stejně tak jevy, že dva cestující stihnou lety z různých letišť, se zdají na první pohled nezávislé, ale nemusí tomu tak být; jedna příčina může způsobit zpoždění všech letů a oběma cestujícím pomoci nezmeškat let.

Nezávislé jevy představují důležitou výjimku z pravidla „vše souvisí se vším", která nám umožní radikální zjednodušení popisu. Proto potřebujeme nezávislost jevů matematicky charakterizovat. Jak se ukáže, podaří se to jen zčásti.

Příklad 1.5.2 (nezávislé zdroje): Reportéři se snaží zprávy ověřovat z „nezávislých zdrojů". Jestliže tutéž informaci potvrdí ne jeden, ale dva zdroje, z nichž každý je na 90 % pravdivý, pak se tím zvýší pravděpodobnost, že informace je pravdivá. To ovšem jen za předpokladu „nezávislosti", který by byl porušen, kdyby např. oba zdroje převzaly tutéž zprávu od stejného člověka. Jaká je pravděpodobnost, že shodná informace ze dvou *nezávislých* zdrojů je pravdivá?

Definice 1.5.3: *Dva jevy A, B jsou* ***nezávislé*** *(angl.* independent*), jestliže $P(A \cap B) = P(A) \cdot P(B)$. V opačném případě nazýváme jevy A, B* ***závislé****.*

Poznámka 1.5.4: I když je zvykem říkat, že jeden jev nezávisí na druhém, matematická nezávislost jevů je symetrická relace; oba jevy mají stejnou roli.

Poznámka 1.5.5: Chtěli bychom vyjádřit, že mezi nezávislými jevy není žádná souvislost. Bohužel neexistuje matematické kritérium, které by to vystihovalo. Můžeme prokázat závislost, nikoli však nezávislost (podobně jako nezávislost svědka u soudu). Proto matematický pojem nezávislosti je slabší – spokojíme se s tím, že pravděpodobnosti složených výrazů vycházejí takové, jako kdyby mezi jevy žádná souvislost nebyla. Nemůžeme například vyloučit, že výsledky jevů jsou generovány jedním společným procesem, jehož pravděpodobnosti jsou nastaveny právě tak, že se nám jevy jeví nezávislé. S podobnou situací jsme se již setkali v příkladu 1.7.9: neslučitelnost jevů nelze poznat z jejich pravděpodobností. Podobně „tu pravou" nezávislost, např. z příkladu 1.5.2, nedovedeme formulovat pomocí pravděpodobností. Matematický pojem nezávislosti je proto slabší. I v tom je zásadní rozdíl mezi nezávislostí a neslučitelností.

Tvrzení 1.5.6: *Jsou-li jevy A, B nezávislé, pak jsou nezávislé také jevy $A, \bar{B}$ (a též dvojice jevů $\bar{A}, B$ a $\bar{A}, \bar{B}$).*

Důkaz:

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A) \cdot (1 - P(B)) = P(A) \cdot P(\bar{B}) .$$

Podobně postupujeme v ostatních případech. □

Důsledek 1.5.7: *Jsou-li jevy A, B nezávislé, pak jejich pravděpodobnosti jednoznačně určují pravděpodobnosti všech výrazů složených z A, B. Speciálně*

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B) . \tag{1.4}$$

Důkaz: Každý složený jev lze vyjádřit v úplné disjunktivní normální formě jako disjunkci některých z následujících navzájem neslučitelných jevů: $A \cap B$, $A \cap \bar{B}$, $\bar{A} \cap B$, $\bar{A} \cap \bar{B}$. Jejich pravděpodobnosti jsou díky nezávislosti známy. Speciálně

$$\begin{aligned} P(A \cup B) &= P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B) = \\ &= P(A) \cdot P(B) + P(A) \cdot (1 - P(B)) + (1 - P(A)) \cdot P(B) = \\ &= P(A) + P(B) - P(A) \cdot P(B) . \end{aligned}$$

Stejný výsledek lze dostat přímo z de Morganova zákona a nezávislosti jevů $\bar{A}, \bar{B}$:

$$\begin{aligned} P(A \cup B) &= P\left(\overline{\bar{A} \cap \bar{B}}\right) = 1 - P(\bar{A} \cap \bar{B}) = \\ &= 1 - P(\bar{A}) \cdot P(\bar{B}) = 1 - (1 - P(A)) \cdot (1 - P(B)) = \\ &= P(A) + P(B) - P(A) \cdot P(B) . \end{aligned}$$

□

Příklad 1.5.8: K chlazení reaktoru slouží dvě čerpadla, každé funguje s pravděpodobností 0.9999. Jaká je pravděpodobnost, že funguje aspoň jedno (což stačí pro provoz reaktoru), jestliže jejich poruchy jsou nezávislé?

Řešení: Podle (1.4)

$$0.9999 + 0.9999 - 0.9999^2 = 0.999\,999\,99 .$$

Riziko chyby z této příčiny je 10^{-8}. Problematický je však předpoklad nezávislosti; některé zdroje poruch působí na čerpadla nezávisle, ale jiné mohou být společné. □

Tvrzení 1.5.9: *Dva jevy, z nichž jeden má pravděpodobnost 0 nebo 1, jsou nezávislé.*

Důkaz: Nechť $P(B) = 0$. Pak $P(A \cap B) = 0 = P(A) \cdot P(B)$. Pro $P(B) = 1$ můžeme použít předchozí výsledek a tvrzení 1.5.6. □

Řešení (příkladu 1.5.2): Příklad s nezávislými zdroji nabízí jednoduché, avšak chybné řešení: Ze zpráv prvního zdroje je 1/10 nepravdivých, druhý nezávislý zdroj sníží riziko chyby opět 10×, tedy zpráva potvrzená ze dvou nezávislých zdrojů je s pravděpodobností 1 % nepravdivá, s pravděpodobností 99 % pravdivá. To je obdobný výpočet jako v příkladu 1.5.8: $0.9+0.9-0.9^2 = 0.99$. Bohužel, tento zjednodušující pohled nerespektuje skutečnost, že zdroje se k některým otázkám *nemusí vyjádřit*. Také nezohledňuje, kolik nepravdivých informací ke zdrojům přichází (a jen některé z nich jsou přeneseny dál). Nemáme tedy dostatečné podklady pro přesný závěr. I přesto je však zřejmé, že potvrzení informace jednoho zdroje druhým, nezávislým zdrojem zvýší pravděpodobnost, že informace je pravdivá; bez předpokladu nezávislosti by to tak být nemuselo.

Předchozí úlohu by bylo možno zpřesnit tak, aby uvedený výpočet byl v pořádku. Např. klademe otázku, na kterou jsou možné pouze odpovědi ano/ne (např. zda president dorazil na plánovanou návštěvu). Oba zdroje znají pravdu a sdělí ji, ale jejich informace může být při přenosu zkreslena tak, že s pravděpodobností 10 % je mylná. Pak při přijetí dvou stejných odpovědí je pravděpodobnost jejich pravdivosti skutečně 99 %. Složitější případy se naučíme řešit v kapitole o podmíněné pravděpodobnosti. □

Příklad 1.5.10: Pokud bychom v příkladu 1.4.1 předpokládali, že výsledky Alice a Boba jsou nezávislé, vyšla by pravděpodobnost, že zkoušku udělají oba, $0.9 \cdot 0.9 = 0.81$. Tento předpoklad však nemusí být splněn, i když nebudou opisovat; pokud se připravovali společně, asi budou mít podobné názory na různou obtížnost jednotlivých témat. Bude-li v testu otázka, která jednomu připadá snadná, bude tak asi připadat i druhému. V uvedeném příkladu ovšem Bobův odhad 70 % měl zásadní vadu, protože tato hodnota není možná ani pro závislé jevy.

Příklad 1.5.11: [Zvára, Štěpán 2002] ** 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. (Tj. mají k dispozici náhodný generátor, který je špatný.) Protože neznají řešení cvičení 1.2.11 (podle něhož by minci ani nepotřebovali), dohodnou se, že hodí mincí dvakrát. Alice vyhrává, pokud padnou stejné výsledky, Bob při různých výsledcích. Kdo z nich má větší naději vyhrát?

Řešení: Nechť líc padá s pravděpodobností $1/2 + \varepsilon$, rub s pravděpodobností $1/2 - \varepsilon$, kde $\varepsilon \in (-1/2, 1/2)$. Můžeme předpokládat, že oba hody jsou nezávislé. V tom případě pravděpodobnost, že padne 2× líc, resp. rub, je $(1/2 + \varepsilon)^2$, resp. $(1/2 - \varepsilon)^2$. Alice vyhrává s pravděpodobností

$$\left(\tfrac{1}{2} + \varepsilon\right)^2 + \left(\tfrac{1}{2} - \varepsilon\right)^2 = \tfrac{1}{2} + 2\varepsilon^2 > \tfrac{1}{2}\,.$$

Pro kontrolu, Bob vyhrává při dvou stejně pravděpodobných výsledcích, a to s pravděpodobností

$$2\,\left(\tfrac{1}{2} + \varepsilon\right)\,\left(\tfrac{1}{2} - \varepsilon\right) = \tfrac{1}{2} - 2\varepsilon^2 < \tfrac{1}{2}\,,$$

což potvrzuje předchozí výsledek.

Na tomto příkladu lze ukázat více skutečností. Navzdory jednoduchosti je k pravděpodobnostnímu popisu potřeba rozšířený Laplaceův model (který čtenář jistě snadno najde), nevystačili bychom s původním Laplaceovým. Alice vyhrává častěji bez ohledu na to, která strana mince padá s vyšší pravděpodobností (pouze při rovnosti je $\varepsilon = 0$ a šance jsou vyrovnané). Pravděpodobnost se však od $1/2$ liší málo, o $h(\varepsilon) = 2\varepsilon^2$ namísto původního ε. Je-li např. $\varepsilon = 0.01$, pak Alice vyhrává s pravděpodobností $1/2 + 2\varepsilon^2 = 0.5002$. Tím jsme získali návod, jak udělat *ze špatného náhodného generátoru lepší.* □

Příklad 1.5.12: ** Zkuste vylepšit předchozí příklad tak, aby pravděpodobnost výhry byla ještě bližší $1/2$.

Řešení: Budeme nutně potřebovat více než dva hody. Možné výsledky můžeme vyhodnotit např. tak, že při sudém počtu líců vyhrává Alice, při lichém Bob. Např. 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 vyjde

$$\left(\tfrac{1}{2} - \varepsilon\right)^4 + 6\,\left(\tfrac{1}{2} - \varepsilon\right)^2\,\left(\tfrac{1}{2} + \varepsilon\right)^2 + \left(\tfrac{1}{2} + \varepsilon\right)^4 = \tfrac{1}{2} + 8\varepsilon^4 = 0.500\,000\,08\,.$$

Je možné také uvažovat tak, že postup z příkladu 1.5.11 aplikujeme na jeho vlastní výsledek, tj. dvojici hodů dvakrát opakujeme a rozhodneme podle toho, zda výsledky byly stejné; výsledná pravděpodobnost se liší od $1/2$ o $h(h(\varepsilon)) = 8\varepsilon^4$. To je ekvivalentní s prvně uvedeným postupem (pro libovolný počet hodů rovný mocnině dvou); i rozhodovací kritérium je stejné.

Uvažujme ještě velmi špatnou minci, u níž líc padá s pravděpodobností 0.9. Provedeme-li např. $2^5 = 32$ pokusů, pak pravděpodobnost, že počet líců je sudý, se liší od $1/2$ o číslo, získané z $\varepsilon = 0.4$ pomocí 5× opakované aplikace funkce h (odvozené v příkladu 1.5.11),

$$h\,(h\,(h\,(h\,(h\,(\varepsilon))))) \doteq 3.96 \cdot 10^{-4}\,.$$

Tento postup dovoluje vytvořit *z velmi špatného náhodného generátoru libovolně dobrý* (byť pomalejší). To se využívá v generátorech náhodných čísel. □

Cvičení 1.5.13: Data z osobního počítače zálohujeme na síťovém disku jednou týdně. Riziko, že během týdne přijdeme o data, je pro osobní počítač 0.3 % a pro síťový disk 0.1 %. Jaké je riziko, že přijdeme o data i zálohu během jednoho týdne, resp. během 156 týdnů (přibližně 3 let)? (Řešte za předpokladu nezávislosti, ale zvažte jeho adekvátnost.)

Řešení: Každý týden podstupujeme riziko $0.003 \cdot 0.001 = 3 \cdot 10^{-6}$. Za 156 týdnů to je

$$1 - \left(1 - 3 \cdot 10^{-6}\right)^{156} \doteq 4.7 \cdot 10^{-4}.$$

Nezávislost lze dobře předpokládat pro poruchy disků (kromě např. požáru budovy, v níž jsou oba disky), ale společným rizikem může být např. počítačový virus. Pro zohlednění závislosti v našem odhadu rizika však obvykle nemáme potřebná data. □

Cvičení 1.5.14: [Zvára, Štěpán 2002] Náhodně vybereme (např. losováním) k z n lidí (mezi nimiž je Alice a Bob) tak, že každá kombinace má stejnou pravděpodobnost $1/\binom{n}{k}$, kde $2 \leq k < n$. (Jedná se o neuspořádaný výběr bez opakování.) Jev A, resp. B znamená, že mezi vybranými je Alice, resp. Bob. Jsou jevy A, B nezávislé?

Řešení: Alici lze vybrat tolika způsoby, kolika lze vybrat $k-1$ lidí ze zbývajících $n-1$, takže pravděpodobnost jejího výběru (stejně jako Bobova) je

$$P(A) = P(B) = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{(n-1)!\,(n-k)!\,k!}{(n-k)!\,(k-1)!\,n!} = \frac{k}{n},$$

což nepřekvapí. Alici současně s Bobem lze vybrat tolika způsoby, kolika lze vybrat $k-2$ lidí ze zbývajících $n-2$, tj. s pravděpodobností

$$P(A \cap B) = \frac{\binom{n-2}{k-2}}{\binom{n}{k}} = \frac{(n-2)!\,(n-k)!\,k!}{(n-k)!\,(k-2)!\,n!} = \frac{k\,(k-1)}{n\,(n-1)} < \left(\frac{k}{n}\right)^2 = P(A) \cdot P(B),$$

takže jevy A, B nejsou nezávislé. (Je-li vybrána Alice, snižuje se tím pravděpodobnost, že bude vybrán i Bob.) □

Cvičení 1.5.15: Mohou být jevy $A, \bar{A}$ (resp. A, A) nezávislé?

Řešení: Z definice dostáváme $P(A \cap \bar{A}) = P(A) \cdot P(\bar{A})$, po úpravě $P(\mathbf{0}) = 0 = P(A) \cdot (1 - P(A))$. To je kvadratická rovnice pro $P(A)$ s řešeními $0, 1$. Možné to tedy je, a to právě tehdy, když $P(A) \in \{0, 1\}$. Stejný výsledek dostáváme i v druhém případě, jako řešení rovnice $P(A) = P(A) \cdot P(A)$. □

Nezávislost více jevů

Zobecnění nezávislosti na více než dva jevy přináší nový problém. Nabízí se následující jednoduchá možnost:

Definice 1.5.16: *Jevy $A_1, \ldots, A_n$ se nazývají* ***po dvou nezávislé****, jestliže každé dva z nich jsou nezávislé.*

Následující příklad ukazuje, že je to slabá podmínka, která nestačí pro vyjádření všech důležitých vlastností jevů, které spolu nesouvisí.

Příklad 1.5.17: Pro hod dvěma mincemi rozlišíme jevy

- A: na první minci padl líc,
- B: na druhé minci padl líc,
- C: na mincích padly různé výsledky.

Jevy A, B jsou zjevně nezávislé; snadno ověříme, že jevy A, B, C jsou po dvou nezávislé:

$$P(A) = P(B) = P(C) = \frac{1}{2},$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}.$$

Nicméně konjunkce všech tří jevů je nemožná, $P(A \cap B \cap C) = P(0) = 0$, ačkoli pro „opravdu nezávislé" jevy bychom ji očekávali s pravděpodobností $P(A) \cdot P(B) \cdot P(C) = 1/8$. Vidíme, že mezi (po dvou nezávislými) jevy A, B, C existuje závislost, která se však projeví až při studiu více než dvou jevů. Pomohlo by přidat požadavek $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$, avšak nesmíme vynechat podmínky $P(A \cap B) = P(A) \cdot P(B)$ atd.

Definice 1.5.18: *Konečná množina jevů $\mathcal{M}$ se nazývá* ***nezávislá****, jestliže*

$$P\Big(\bigcap_{A \in \mathcal{K}} A\Big) = \prod_{A \in \mathcal{K}} P(A) \tag{1.5}$$

pro všechny podmnožiny $\mathcal{K} \subseteq \mathcal{M}$. V tom případě též říkáme, že jevy z množiny $\mathcal{M}$ jsou ***nezávislé****, nebo že $\mathcal{M}$ je množina nezávislých jevů.*

Tvrzení 1.5.6 má následující zobecnění:

Tvrzení 1.5.19: *Je-li $\mathcal{M} = \{A_1, \ldots, A_n\}$ nezávislá množina jevů, pak je nezávislá i každá množina $\{B_1, \ldots, B_n\}$, kde $B_j \in \{A_j, \bar{A}_j\}$, $j = 1, \ldots, n$, a též každá její podmnožina.*

Důkaz: Důkaz je analogický tvrzení 1.5.6, např.

$$P\Big(\bigcap_{j \le n-1} A_j \cap \bar{A}_n\Big) = P\Big(\bigcap_{j \le n-1} A_j\Big) - P\Big(\bigcap_{j \le n} A_j\Big) = \prod_{j \le n-1} P(A_j) - \prod_{j \le n} P(A_j) =$$
$$= (1 - P(A_n)) \prod_{j \le n-1} P(A_j) = P(\bar{A}_n) \prod_{j \le n-1} P(A_j).$$

□

Poznámka 1.5.20: Připomeňme, že v předchozím tvrzení pro každé j můžeme do nezávislé množiny vybrat buď A_j, nebo $\bar{A}_j$, ale nikoli oba tyto jevy.

Důsledek 1.5.21: *Jsou-li jevy $A_1, \ldots, A_n$ nezávislé, pak jejich pravděpodobnosti jednoznačně určují pravděpodobnosti všech výrazů složených z $A_1, \ldots, A_n$.*

Důkaz: Obdobný důsledku 1.5.7. □

Někdy budeme potřebovat zobecnění nezávislosti na nekonečně mnoho jevů. Následující definice se vyhne možným problémům s existencí nekonečných součinů.

Definice 1.5.22: *Nekonečná množina jevů se nazývá* ***nezávislá****, pokud každá její konečná podmnožina je nezávislá, tj. pokud (1.5) platí pro každou* konečnou *podmnožinu $\mathcal{K}$.*

Příklad 1.5.23: Kolik rovnic je potřeba otestovat, abychom ověřili nezávislost n jevů?

Řešení: Rovnost definující nezávislost je potřeba ověřit pro všechny podmnožiny n-prvkové množiny, těch je 2^n. Z nich je $n + 1$ triviálně splněno, a to n pro jednoprvkové množiny, tvaru $P(A) = P(A)$, a 1 pro prázdnou množinu, $P(1) = 1$ (kde součin prázdné množiny činitelů pokládáme rovný neutrálnímu prvku 1). Zbývá ověřit $2^n - n - 1$ rovnic. □

Příklad 1.5.24: Kolik reálných čísel potřebujeme k popisu pravděpodobností všech složených jevů vytvořených z n *nezávislých* jevů?

Řešení: Stačí znát n pravděpodobností výchozích jevů $A_1, \ldots, A_n$. Z nich určíme

$$P(B_1 \cap \ldots \cap B_n) = \prod_{j \leq n} P(B_j)$$

pro všechna $B_j \in \{A_j, \bar{A}_j\}$, $j = 1, \ldots, n$. Každý složený jev lze vyjádřit jako disjunktní sjednocení jevů tohoto tvaru, tedy jeho pravděpodobnost lze určit z předchozích. □

Případná nezávislost jevů dovoluje značnou úsporu při popisu pravděpodobností náhodného pokusu. To je ještě důležitější při *hledání* tohoto popisu, neboť se data obvykle nepostačují k odhadu velkého počtu parametrů. Proto nezávislost jevů využijeme všude, kde to bude možné.

Poznámka 1.5.25: Přesné splnění podmínek nezávislosti může být vzácné. (Jsou ve tvaru rovností reálných čísel.) V praxi se proto s tímto předpokladem pracuje často i u jevů, u nichž nezávislost je splněna jen přibližně; výsledné závěry jsou pak zatíženy další chybou a platí také jen přibližně. Nicméně je to často jediný způsob, jak dojít aspoň k nějakým použitelným závěrům. Musíme však uvážit adekvátnost předpokladu nezávislosti a být si vědomi důsledků vyplývajících z jeho porušení.

Ukázali jsme si, jak se pravděpodobnostní popis zjednoduší, pokud víme, že některé jevy jsou nezávislé. V tom případě nám k popisu stačí méně parametrů. Nesmíme zapomínat, že jevy po dvou nezávislé ještě nemusí být nezávislé.

Cvičení 1.5.26: Na zajíce vystřelí současně tři střelci, jejichž pravděpodobnosti zásahu jsou $a = 0.2$, $b = 0.8$ a $c = 0.9$. Jaká je pravděpodobnost, že aspoň jedna z ran zajíce zasáhne?

Řešení: Považujeme-li jevy za nezávislé, pak pravděpodobnost, že zajíc nebude zasažen, je

$$(1-a)(1-b)(1-c) = 0.016\,.$$

Pravděpodobnost aspoň jednoho zásahu je

$$1 - (1-a)(1-b)(1-c) = 0.984\,.$$

□

Cvičení 1.5.27: Útočná raketa je zasažena obrannou raketou s pravděpodobností $1/2$. Kolik obranných raket je třeba vyslat, aby pravděpodobnost zásahu byla aspoň (a) 90 %, (b) 99 %.

Řešení: Za předpokladu nezávislosti je pravděpodobnost zásahu k obrannými raketami $1 - 1/2^k$, z toho vyplývají meze (a) $k \geq \log_2(10) \doteq 3.32$, $k \geq 4$, (b) $k \geq \log_2(100) \doteq 6.64$, $k \geq 7$. □

Cvičení 1.5.28: Předpokládejme, že náhodné jevy A, B, C jsou nezávislé a mají po řadě pravděpodobnosti 0.1, 0.3, 0.4. Určete pravděpodobnosti jevů $A \cup (B \cap C)$, $A \cup (\bar{A} \cap B \cap C)$, $(A \cup B) \cap (\bar{A} \cup C)$.

Řešení: Jevy A, $B \cap C$ jsou nezávislé, s využitím důsledku 1.5.7 dostáváme

$$\begin{aligned} P(A \cup (B \cap C)) &= P(A) + P(B \cap C) - P(A) \cdot P(B \cap C) = \\ &= P(A) + P(B) \cdot P(C) - P(A) \cdot P(B) \cdot P(C) = \\ &= 0.1 + 0.3 \cdot 0.4 - 0.1 \cdot 0.3 \cdot 0.4 = 0.208\,. \end{aligned}$$

Jevy A, $\bar{A} \cap B \cap C$ nejsou nezávislé, ale jsou neslučitelné,

$$\begin{aligned} P\left(A \cup (\bar{A} \cap B \cap C)\right) &= P(A) + P(\bar{A}) \cdot P(B) \cdot P(C) = \\ &= P(A) + (1 - P(A)) \cdot P(B) \cdot P(C) = 0.1 + 0.9 \cdot 0.3 \cdot 0.4 = 0.208\,. \end{aligned}$$

Jedná se ovšem o výraz sémanticky ekvivalentní s předchozím, podle zákona absorpce $A \cup (\bar{A} \cap X) = A \cup X$.

Jev $(A \cup B) \cap (\bar{A} \cup C)$ si převedeme do disjunktivní normální formy $(A \cap C) \cup (\bar{A} \cap B)$, kde jevy v závorkách jsou neslučitelné,

$$\begin{aligned} P\left((A \cap C) \cup (\bar{A} \cap B)\right) &= P(A \cap C) + P(\bar{A} \cap B) = \\ &= P(A) \cdot P(C) + P(\bar{A}) \cdot P(B) = \\ &= 0.1 \cdot 0.4 + 0.9 \cdot 0.3 = 0.31\,. \end{aligned}$$

Ve všech případech je možné vyjít z úplné disjunktivní normální formy, což by zde vedlo na poněkud složitější výpočet. □

Cvičení 1.5.29: K chlazení reaktoru slouží tři čerpadla, každé funguje s pravděpodobností 0.9999. Jaká je pravděpodobnost, že fungují aspoň dvě, resp. aspoň jedno (což stačí pro provoz reaktoru), jestliže jejich poruchy jsou nezávislé? (Srovnej s příkladem 1.5.8.)

Řešení: Booleovskou funkci „aspoň dva ze tří" lze pro jevy A, B, C vyjádřit např. $(A \cap B) \cup (A \cap C) \cup (B \cap C)$. V tomto výrazu však nejsou jednotlivé konjunkce neslučitelné. Proto jej pro výpočet pravděpodobnosti upravíme do ekvivalentního tvaru $(A \cap B) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C)$, který je disjunkcí neslučitelných výrazů. Dostáváme pravděpodobnost

$$0.9999^2 + 2 \cdot 0.9999^2 \cdot (1 - 0.9999) \doteq 0.999\,999\,97\,.$$

Podle očekávání je riziko chyby $3 \cdot 10^{-8}$ přibližně 3× větší než v příkladu 1.5.8.

Podobně můžeme postupovat pro funkci „aspoň jeden ze tří". Zde je ale jednodušší přejít k negacím a popsat ji vzorcem $\overline{\bar{A} \cap \bar{B} \cap \bar{C}}$, kde máme konjunkci nezávislých jevů. Riziko chyby vychází $0.0001^3 = 1 \cdot 10^{-12}$ a pravděpodobnost, že aspoň jedno čerpadlo funguje, je $1 - 1 \cdot 10^{-12} = 0.999\,999\,999\,999$.

Předpoklad nezávislosti je opět problematický. *(Číslo 0.9999 je fiktivní, ale úloha má reálný podklad; všechny tři uvedené konfigurace se u jaderných elektráren užívají.)* □

Cvičení 1.5.30: [Hsu 1996] Hodíme dvěma kostkami a rozlišíme následující jevy:

A: součet je 7,

B: součet je 6,

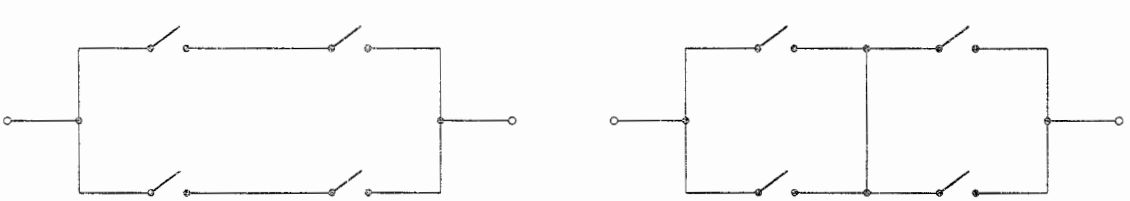
C: výsledek na první kostce je 4.

Jsou některé z jevů A, B, C nezávislé?

Řešení: Jevy A, C jsou nezávislé: $P(A) = P(C) = 1/6$, $P(A \cap C) = 1/36$.

Cvičení 1.5.31: [Hsu 1996] Určete pravděpodobnosti nezávislých jevů A, B, jestliže $P(A \cap B) = 0.16$, $P(A \cup B) = 0.64$.

Cvičení 1.5.32: Čtyři spínače v zabezpečovacím zařízení jsou spojeny (a) po dvou sériově a pak paralelně, (b) po dvou paralelně a pak sériově, viz obr. 1.1. Sepnou nezávisle s pravděpodobností $p \in (0, 1)$. Pro které zapojení je větší pravděpodobnost, že celá soustava bude pouštět proud?



Obrázek 1.1. Dvě zapojení spínačů

Řešení: (a) $P(A) = 1 - \left(1 - p^2\right)^2 = p^2\left(2 - p^2\right)$.

(b) $P(B) = \left(1 - (1-p)^2\right)^2 = p^2\,(2-p)^2$.

$P(B) - P(A) = p^2\left((2-p)^2 - \left(2 - p^2\right)\right) = 2p^2\,(p-1)^2 \geq 0$.

Cvičení 1.5.33: Dokažte, že jsou-li jevy A, B, C nezávislé, pak i jevy $A, B \cap C$ jsou nezávislé. Ukažte (např. pomocí příkladu 1.5.17), že nestačí předpoklad, že jevy A, B, C jsou *po dvou* nezávislé.

1.6 Geometrická pravděpodobnost

Typickým příkladem s nekonečně mnoha výsledky jsou geometrické úlohy, vedoucí na tzv. **geometrickou pravděpodobnost**.

Příklad 1.6.1: Jaká je pravděpodobnost, že náhodně vybraný bod čtverce leží v kruhu do čtverce vepsaném?

Příklad 1.6.2 (Buffonova úloha, angl. *Buffon's needle problem*): * Na linkovaný papír hodíme jehlu, jejíž délka je rovna a-násobku vzdálenosti mezi linkami, $a \leq 1$. Jaká je pravděpodobnost, že jehla protne nějakou linku?

V takovýchto úlohách předpokládáme, že výběr bodů je „rovnoměrný" v tom smyslu, že pravděpodobnost, že bod padne do určitého útvaru (v prostoru možných výsledků), je úměrná obsahu tohoto útvaru, nezáleží na jeho poloze, orientaci apod. Pak řešením je poměr obsahů ploch odpovídajících studovanému jevu a jevu jistému. Tím trochu předbíháme; korektní zavedení potřebných pojmů, hlavně „rovnoměrnosti", přinese až kapitola 1.7. Zvolený postup výkladu ale odpovídá historickému vývoji disciplíny, kde geometrická pravděpodobnost sehrála důležitou motivační roli před formulací korektního matematického popisu.

Řešení (příkladu 1.6.1): Za uvedených předpokladů je pravděpodobnost dána poměrem obsahu kruhu a čtverce (kde strana se rovná průměru), což je $\pi/4 \doteq 0.785$. □

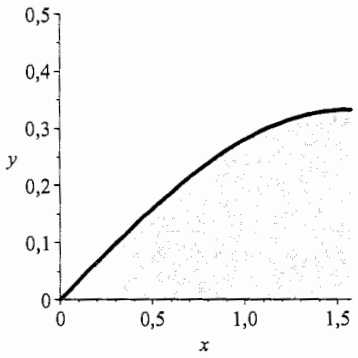
Řešení (příkladu 1.6.2): Výsledek můžeme vyjádřit v závislosti na dvou náhodných veličinách: úhlu mezi linkou a jehlou $x \in \langle 0, \pi/2\rangle$ a vzdálenosti středu jehly od nejbližší linky $y \in \langle 0, 1/2\rangle$ (za jednotku bereme vzdálenost mezi linkami). Předpokládáme, že tyto náhodné veličiny jsou nezávislé a mají rovnoměrná rozdělení na příslušných intervalech. (Mohli jsme uvažovat znaménka a zavést náhodné veličiny jako orientované, ale došli bychom ke stejným závěrům. Také maximální hodnoty se nezdají být správně ošetřeny; např. pro $y \in (0, 1/2)$ připadají na každou linku dvě přímky se vzdáleností od linek y, zatímco pro $y \in \{0, 1/2\}$ jen jedna přímka, ale tento případ nastává s nulovou pravděpodobností.) 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 $y < \frac{1}{2}\,a \sin x$ (viz obr. 1.2),

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

(Nezáleží na tom, zda použijeme ostrou nebo neostrou nerovnost, oba případy se liší s nulovou pravděpodobností.) Hledaná pravděpodobnost je poměr obsahů množin A a Ω, přičemž A je plocha pod křivkou, jejíž integrací dostaneme obsah, a Ω je obdélník:

$$P(A) = \frac{1}{\frac{1}{2}\,\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{1}{2}\,a \sin x \,\mathrm{d}x = \frac{2}{\pi}\,a \doteq 0.636\,619\,772\,a\,. \tag{1.6}$$

Jelikož tuto pravděpodobnost lze odhadnout na základě experimentu (tzv. *metodou Monte Carlo*), máme návod, jak lze (při známé délce jehly, např. rovné vzdálenosti mezi linkami) experimentálně změřit iracionální číslo π. To mohli udělat již staří Řekové, ale k interpretaci výsledku pokusu jim chyběly znalosti zde použitých vzorců, které pochopitelně vyžadovaly i rozsáhlé znalosti o čísle π. Podobně mohli použít jednodušší příklad 1.6.1. □



Obrázek 1.2. Oblast vyhovující Buffonově úloze pro $a = 2/3$

Pomocí Buffonovy úlohy 1.6.2 lze naopak při známém π touto metodou měřit délku jehly, což není dobrý nápad, nicméně je východiskem pro užitečnější výsledek:

Příklad 1.6.3: [Zvára, Štěpán 2002] Pozměňme Buffonovu úlohu 1.6.2 tak, že místo jehly házíme *konvexní* n-úhelník, jehož *průměr* (maximální vzdálenost dvou bodů) je menší než vzdálenost linek. Jaká je pravděpodobnost, že protne některou linku?

Řešení: Předpoklady zajišťují, že n-úhelník protíná nejvýše jednu linku. Nastane právě jedna z následujících možností:

- n-úhelník neprotíná žádnou linku,
- některý vrchol leží na lince,
- právě dvě (různé) strany n-úhelníka protínají jednu linku.

Druhá možnost má nulovou pravděpodobnost a dále ji neuvažujeme. Označme A_j jev, že j-tá strana protíná nějakou linku, a A jev, že n-úhelník protíná nějakou linku. Pak

$$A = \bigcup_{j=1}^{n-1} \bigcup_{k=j+1}^{n} A_j \cap A_k \,,$$

což je sjednocení disjunktních jevů.

$$P(A) = \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} P(A_j \cap A_k) = \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} P(A_j \cap A_k) \,.$$

Pro každé j známe dle (1.6) pravděpodobnost

$$P(A_j) = \sum_{k \in M_j} P(A_j \cap A_k) = \frac{2}{\pi} a_j \,,$$

kde a_j je délka j-té strany a $M_j = \{1, \ldots, n\} \setminus \{j\}$. Po dosazení

$$P(A) = \frac{1}{2} \sum_{j=1}^{n} \sum_{k \in M_j} P(A_j \cap A_k) = \frac{1}{2} \sum_{j=1}^{n} \frac{2}{\pi} a_j = \frac{1}{\pi} \sum_{j=1}^{n} a_j \,,$$

kde poslední suma je obvod n-úhelníka.

Dostali jsme metodu, která dovoluje odhadnout obvod *konvexního* mnohoúhelníka. Protože libovolnou konvexní množinu lze vyjádřit jako limitu mnohoúhelníků (a uvidíme, že pravděpodobnost se limitou zachová), lze takto odhadnout obvod libovolné *konvexní* množiny, jejíž obvod je konečný (stačí zajistit, že vzdálenost linek je větší než průměr množiny). To už může být nástroj k řešení obtížné úlohy. Bohužel tento postup nelze použít na nekonvexní množiny. □

Příklad 1.6.4 (odhad plochy pevniny): Chceme odhadnout plochu, kterou zabírá pevnina na Zemi. Vzhledem ke komplikovanému tvaru je to velmi obtížné. Můžeme uvažovat o následující snazší alternativě: Povrch Země lze odhadnout poměrně snadno, vyjdeme z povrchu koule či elipsoidu a provedeme drobné korekce na skutečný tvar. Pak volíme náhodně body na Zemi a zkoušíme, zda jsou na pevnině nebo na moři. Podíl plochy pevniny k povrchu Země je pravděpodobnost, že náhodně vybraný bod na Zemi leží na pevnině (je-li výběr bodů prováděn "rovnoměrně"; je ovšem obtížné definovat, co to znamená). Při velkém počtu bodů lze očekávat, že odhadneme poměr plochy pevniny a moří. Později se naučíme stanovit přesnost této metody.

Příklad 1.6.5: Náhodně vybereme dva body K, L na kružnici o jednotkovém poloměru se středem S. Jaká je pravděpodobnost, že jejich vzdálenost (přímá, měřená v rovině) je delší než nějaká mez, např. $\sqrt{3}$ (=délka strany vepsaného rovnostranného trojúhelníka)?

Řešení: Vzdálenost závisí jen na úhlu $\angle KSL$, ten je náhodně vybrán z intervalu $\langle 0, \pi \rangle$. Vzdálenost je větší než $\sqrt{3}$, právě když $\angle KSL > 2/3\,\pi$ (rovnost nastává, právě když K, L jsou dva z vrcholů vepsaného rovnostranného trojúhelníka). Hledaná pravděpodobnost vychází 1/3. □

Podle Laplaceovy definice může pravděpodobnost nabývat pouze racionálních hodnot. To se jeví jako neopodstatněné omezení. V příkladech 1.6.1 a 1.6.2 vyšla pravděpodobnost iracionální (a lze z ní vypočítat π).

Kromě toho je problém, jak vůbec formulovat podmínky pokusu, např. že „všechny body mají stejnou pravděpodobnost, že budou vybrány." Jelikož je jich nekonečně mnoho, tato pravděpodobnost nemůže být kladná, musí tedy být nulová. To nám však neumožňuje řešit takové úlohy prostým poměrem počtu příznivých případů k počtu všech případů. Ukážeme, že v některých úlohách na geometrickou pravděpodobnost není jasné, jaký poměr obsahů má být řešením a nedostaneme jednoznačný výsledek.

Příklad 1.6.6 (Bertrandův paradox): [Rogalewicz 2000] Jaká je pravděpodobnost, že délka tětivy kružnice o jednotkovém poloměru je delší než $\sqrt{3}$?

Řešení: 1. postup: Náhodně vybereme střed tětivy jako libovolný bod uvnitř kružnice. Délka tětivy bude větší než $\sqrt{3}$, právě když střed bude ležet v kružnici vepsané příslušnému rovnostrannému trojúhelníku. Ta má poloměr 1/2 a obsah 4× menší než původní kružnice, takže hledaná pravděpodobnost vychází 1/4.

2. postup: Náhodně vybereme střed tětivy pomocí vzdálenosti od středu kružnice (z intervalu $\langle 0, 1 \rangle$) a úhlu od středu (z intervalu $\langle 0, 2\,\pi \rangle$). Jev nezávisí na úhlu, pouze na vzdálenosti, ta má být menší než 1/2, což nastane s pravděpodobností 1/2.

3. postup: Jedno řešení už jsme měli v příkladu 1.6.5; výsledek byl 1/3.

Shrnutí: Různé postupy vedou k odlišným výsledkům, úloha není korektně zadaná. Není totiž jasné, jakým způsobem se má náhodně vybrat tětiva. Všechny postupy jsou správné, ale každý popisuje jinou úlohu s jiným rozdělením pravděpodobností tětiv. □

Rozdíl mezi příklady 1.6.5 a 1.6.6 je v tom, že v prvním jsme neváhali, co to znamená náhodně vybrat dva *body* kružnice. Pro každý oblouk je pravděpodobnost, že vybraný bod do něj padne, úměrná délce oblouku (tedy i jeho středovému úhlu). Pokud však úkolem bylo náhodně vybrat *tětivu*, pak není zřejmé, zda ji máme určit koncovými body jako dříve, nebo jiným, stejně přirozeným způsobem. Pokud čtenář i po tomto příkladu věří, že Bertrandův paradox vznikl z neopatrnosti, které se lze vyhnout jako v příkladu 1.6.5, možná znejistí v následující úloze:

Příklad 1.6.7: Jaká je pravděpodobnost, že vzdálenost dvou náhodně vybraných bodů *elipsy* je větší než stanovená mez?

Řešení: Podobně jako v příkladu 1.6.5 chceme náhodně zvolit body elipsy. Zde opravdu není zřejmé, co by mělo být „rovnoměrné" rozdělení. Má být pravděpodobnost, že bude vybrán bod z daného oblouku elipsy, úměrná středovému úhlu, nebo délce oblouku? U kružnice v tom nebyl rozdíl, ale zde jsou obě možnosti stejně přirozené. Bez další specifikace nelze pokračovat v řešení. □

Komu ani předchozí argument nestačí, může uvažovat o „rovnoměrném" rozdělení na složitější křivce, třeba spirále (je jich mnoho druhů). Navíc mohou vyvstat i problémy s definicí délky křivky. Stejně tak bychom narazili na problémy, kdybychom chtěli definovat „rovnoměrné" rozdělení na různých plochách.

Ukázali jsme aspoň základní úlohy na geometrickou pravděpodobnost, které nelze řešit v Laplaceově modelu pravděpodobnosti. Vyžadují jiný aparát.

Cvičení 1.6.8: [Němičková 1990] Vybereme bod (X, Y) ze čtverce $\langle -1, 1\rangle \times \langle -1, 1\rangle$ s rovnoměrným rozdělením. Ukažte, že jevy A: $X > 0$, B: $Y > 0$, C: $X \cdot Y > 0$ jsou jsou závislé, ale *po dvou* nezávislé.

1.7 Kolmogorovův model pravděpodobnosti

Tento přístup odstraňuje většinu výhrad k Laplaceově definici pravděpodobnosti. Současně dovoluje popsat a řešit korektně i úlohy na geometrickou pravděpodobnost.

Základní pojmy

První změna je, že **elementárních jevů** (=prvků množiny Ω) může být *nekonečně mnoho.* Každý elementární jev odpovídá jedné možné kombinaci výsledků všech jevů, které lze v popisovaném náhodném pokusu pozorovat. Důsledkem je, že *nemůžeme předpokládat, že všechny elementární jevy jsou stejně pravděpodobné.*

Jevy jsou podmnožiny množiny Ω, ale *ne nutně všechny.* (Tato volba se ukáže nevyhnutelná.) Všechny jevy tvoří podmnožinu $\mathcal{A} \subseteq \exp \Omega$. Ta splňuje následující podmínky:

(A1) $\emptyset \in \mathcal{A}$.
(A2) $A \in \mathcal{A} \implies \overline{A} \in \mathcal{A}$.
(A3) $(\forall n \in \mathbb{N} : A_n \in \mathcal{A}) \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

V důsledku (A1) a (A2) dostáváme $\Omega = \overline{\emptyset} \in \mathcal{A}$. Podmínka (A3) říká, že disjunkce spočetně mnoha jevů je jevem. To samozřejmě platí i pro disjunkce konečně mnoha jevů; díky (A1) stačí doplnit posloupnost prázdnými množinami. (Volbu právě *spočetných* disjunkcí, nikoli všech nebo pouze konečných, zdůvodníme později.) Vlastnosti (A2), (A3) a de Morganův zákon zajišťují, že v $\mathcal{A}$ můžeme používat i spočetné průniky:

$$(\forall n \in \mathbb{N} : A_n \in \mathcal{A}) \implies \bigcap_{n\in\mathbb{N}} A_n = \overline{\bigcup_{n\in\mathbb{N}} \overline{A}} \in \mathcal{A}.$$

Definice 1.7.1: *Systém $\mathcal{A}$ podmnožin nějaké množiny Ω, který splňuje podmínky (A1)–(A3), se nazývá* ***σ-algebra.***

Poznámka 1.7.2: Písmeno σ v termínu σ-algebra je zkratkou znamenající, že tento pojem se vztahuje ke *spočetným* operacím (zde spočetné sjednocení).

Přirozený nápad $\mathcal{A} = \exp \Omega$ vede k nežádoucím paradoxům, jeden z nich uvedeme v příkladu 1.7.23.

Definice 1.7.3: ***Pravděpodobnost*** *(též* ***pravděpodobnostní míra)*** *je funkce P s hodnotami z $\langle 0, 1\rangle$, definovaná na σ-algebře a splňující podmínky*

(P1) $P(\Omega) = 1$,
(P2) $P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} P(A_n)$, *pokud jsou množiny (=jevy) A_n, $n \in \mathbb{N}$, po dvou disjunktní.*

Poznámka 1.7.4: Vlastnost (P2) se nazývá *spočetná aditivita* nebo *σ-aditivita.*

Definice 1.7.5: ***Pravděpodobnostní prostor*** *je trojice $(\Omega, \mathcal{A}, P)$, kde Ω je neprázdná množina, $\mathcal{A}$ je σ-algebra podmnožin množiny Ω a P: $\mathcal{A} \to \langle 0, 1\rangle$ je pravděpodobnost.*

Vlastnosti pravděpodobnosti

Porovnáme, které vlastností pravděpodobnosti z Laplaceova modelu zůstávají v platnosti a kde se Kolmogorovův model liší.

Tvrzení 1.7.6: *Každá pravděpodobnostní míra (dle definice 1.7.3) splňuje vlastnosti pravděpodobnosti z tvrzení 1.4.2.*

Důkaz: 1. $P(A) \in \langle 0, 1\rangle$ je součástí definice.

3. Jevy $A, \bar{A}$ jsou disjunktní, tedy

$$P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(\mathbf{1}) = 1\,,$$
$$P(\bar{A}) = 1 - P(A)\,.$$

2. $P(\mathbf{1}) = 1$ je součástí definice; podle bodu 3 je $P(\mathbf{0}) = 1 - P(\bar{\mathbf{1}}) = 0$.

4, 5. Je-li $A \subseteq B$, pak B je sjednocením disjunktních jevů $A, B \setminus A$,

$$P(A) + P(B \setminus A) = P(B)\,,$$
$$P(B) - P(A) = P(B \setminus A) \geq 0\,.$$

6. Aditivita pro dva jevy je speciální případ (P2).

7. Vztah $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ se dokáže stejně jako v tvrzení 1.4.2. □

V Laplaceově modelu pravděpodobnosti byl jev nemožný jediným jevem s nulovou pravděpodobností. V Kolmogorovově modelu tomu tak být nemusí; mohou existovat jevy s nulovou pravděpodobností, které přesto nejsou nemožné.

Příklad 1.7.7 (možný jev s nulovou pravděpodobností): Opakovaně házíme mincí a počítáme, kolikrát za sebou padne líc. Jakmile padne rub, pokus končí. Pravděpodobnost výsledku n je 2^{-n-1}. Pravděpodobnost, že výsledek je konečný, je

$$\sum_{n=0}^{\infty} 2^{-n-1} = 1\,,$$

takže je nulová pravděpodobnost, že by padal pořád líc. Přesto je to možné, protože nic nezaručuje, že někdy padne rub.

Příklad 1.7.8: Při analogovém měření veličiny je výsledek vyjádřen reálným číslem z nějakého intervalu. Možných výsledků je nespočetně mnoho, ale jen spočetně mnoho z nich může mít nenulovou pravděpodobnost (v typických případech žádný). Přesto všechny tyto reálné výsledky jsou možné a určitě některý z nich nastane.

Příklad 1.7.9: Jak se z pravděpodobností jevů A, B a jevů z nich složených pozná, že jsou neslučitelné?

Řešení: Není to možné! Jsou-li jevy A, B neslučitelné, pak $P(A \cap B) = 0$, ale obrácená implikace neplatí. (Jev s nulovou pravděpodobností nemusí být nemožný. V Laplaceově modelu pravděpodobnosti tento problém nebyl, zato však nebylo možno popsat některé náhodné pokusy.) Neslučitelnost jevů lze tedy poznat pouze z jejich množinové reprezentace (odpovídající množiny jsou disjunktní), nikoli z pravděpodobností jevů z nich složených. □

Pravděpodobnost zachovává limity monotónních posloupností množin.

Tvrzení 1.7.10: *Nechť $(A_n)_{n\in\mathbb{N}}$ je posloupnost jevů a P je pravděpodobnost.*

1. *Jestliže $A_1 \subseteq A_2 \subseteq \ldots$, pak*

$$P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} P(A_n)\,.$$

2. *Jestliže* $A_1 \supseteq A_2 \supseteq \ldots$, *pak*

$$P\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} P(A_n)\,.$$

Důkaz: 1. Zkonstruujeme posloupnost po dvou neslučitelných jevů následovně: $B_1 = A_1$, $B_{n+1} = A_{n+1} \setminus A_n$ $(n \in \mathbb{N})$. Pak

$$A_n = \bigcup_{j=1}^{n} B_j\,,$$

$$P(A_n) = P\Big(\bigcup_{j=1}^{n} B_j\Big) = \sum_{j=1}^{n} P(B_j)$$

pro všechna n, a tedy i v limitě pro $n \to \infty$,

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = P\Big(\bigcup_{j=1}^{\infty} B_j\Big) = \sum_{j=1}^{\infty} P(B_j) = \lim_{j\to\infty} \sum_{j=1}^{n} P(B_j) = \lim_{n\to\infty} P(A_n)\,.$$

2. Použijeme první výsledek pro doplňky:

$$P\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = P\Big(\overline{\bigcup_{n\in\mathbb{N}} \bar{A}}\Big) = 1 - P\Big(\bigcup_{n\in\mathbb{N}} \bar{A}\Big) = 1 - \lim_{n\to\infty} P(\bar{A}) = \lim_{n\to\infty} (1 - P(\bar{A})) = \lim_{n\to\infty} P(A_n)\,.$$

□

Laplaceův pravděpodobnostní model dostáváme jako speciální případ Kolmogorovova, kdy elementárních jevů je konečně mnoho a jsou všechny stejně pravděpodobné. Na rozdíl od Laplaceovy definice pravděpodobnosti, Kolmogorovův model nám nedává návod, jak pravděpodobnost počítat. Jen nám definuje podmínky, které pravděpodobnost musí splňovat. (S tím jsme se setkali již u rozšíření Laplaceova modelu.) Struktura jevů (σ-algebra) neurčuje jedinou pravděpodobnost, možností je více. Konkrétní pravděpodobnosti jednotlivých jevů je potřeba definovat zvlášť. Tím však získáváme prostor pro popis náhodných pokusů, které v Laplaceově modelu nebylo možno vyjádřit, jako v příkladech 1.2.3, 1.6.1, 1.6.2, 1.6.4.

Příklad 1.7.11: Kolik různých jevů lze popsat pomocí n jevů? (V nejobecnějším případě.) Kolik reálných čísel potřebujeme k popisu pravděpodobností všech těchto jevů, jestliže nerozlišujeme ekvivalentní jevy jako např. $\overline{A \cap B}$, $\bar{A} \cup \bar{B}$? (Srovnejte s příkladem 1.5.24, kde jsme předpokládali nezávislost.)

Řešení: Každý jev, složený z jevů $A_1, \ldots, A_n$, lze vyjádřit v úplné disjunktivní normální formě jako sjednocení jevů tvaru $B_1 \cap \ldots \cap B_n$, kde $B_j \in \{A_j, \bar{A}_j\}$, $j = 1, \ldots, n$. Množina $\mathcal{M}$ všech takových jevů tvoří úplný systém jevů a má 2^n prvků. Sjednocení libovolné podmnožiny množiny $\mathcal{M}$ odpovídá právě jednomu složenému jevu. Počet všech složených jevů je stejný jako počet všech podmnožin množiny o 2^n prvcích, tedy 2^{2^n}. Některé hodnoty pro představu uvádí tabulka. Pro určení jejich pravděpodobností stačí znát pravděpodobnosti $P(B_1 \cap \ldots \cap B_n)$ jevů z $\mathcal{M}$; ty jsou libovolné, až na to, že jejich součet je 1, proto potřebujeme $2^n - 1$ číselných parametrů.

počet výchozích jevů	n	1	2	3	4	5	6
počet složených jevů	2^{2^n}	4	16	256	65 536	4294 967 296	$\doteq 1.8 \cdot 10^{19}$
počet parametrů	$2^n - 1$	1	3	7	15	31	63

Nezávislost představuje značnou úsporu počtu parametrů při popisu pravděpodobnosti, jak ukazuje tabulka:

s předpokladem nezávislosti	n	4	6	10	16	32
bez předpokladu nezávislosti	$2^n - 1$	15	63	1023	65 535	$\doteq 4 \cdot 10^9$

Všimněte si, že rozdíl v počtu parametrů, $2^n - n - 1$, je počet rovnic, které je potřeba ověřit pro nezávislost výroků, viz příklad 1.5.23. □

Příklad 1.7.12: Pro příklad 1.2.1 (jeden hod kostkou) je přirozený následující pravděpodobnostní model: $\Omega = \{1, \ldots, 6\}$, $\mathcal{A} = \exp \Omega$, $P(A) = |A|/6$ pro všechna $A \in \mathcal{A}$.

Pro příklad 1.2.3 můžeme $\Omega, \mathcal{A}$ ponechat, pravděpodobnost se změní dle (1.2).

Příklad 1.7.13 (elementární jev nemusí být jev): Popišme jeden hod pravidelnou kostkou, kde však vidíme pouze střed horní stěny kostky, takže jediné, co můžeme rozlišit, je, zda padlo sudé či liché číslo. Množinou Ω všch elementárních jevů může nadále být všech 6 možných výsledků. Avšak všechny jevy pozorovatelné v tomto pokusu tvoří σ-algebru $\mathcal{B} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\} \subset \exp \Omega$. Pravděpodobnosti netriviálních pozorovatelných jevů jsou

$$P(\{1,3,5\}) = P(\{2,4,6\}) = \frac{1}{2}$$

(stejné jako v příkladu 1.2.1). Např. číslo 1 je elementární jev, ale ani toto číslo, ani jednobodová množina $\{1\}$ nejsou v $\mathcal{B}$, nejsou to jevy. Příčinou je, že za podmínek pokusu nedovedeme odlišit 1 od 3 a 5, a není tedy důvod, abychom se snažili tomuto výsledku přiřadit pravděpodobnost, když ani neumíme určit, kdy nastal.

Příklad 1.7.14: V předchozím příkladu můžeme zvolit úspornější pravděpodobnostní model $(\Omega', \mathcal{A}', P')$ se dvěma elementárními jevy, řekněme $\Omega' = \{L, S\}$ (lichý, resp. sudý výsledek), $\mathcal{A}' = \exp \Omega' = \{\emptyset, L, S, \Omega'\}$ a $P'(A) = |A|/2$.

Poznámka 1.7.15: Jak je vidět v příkladu 1.7.13, elementární jev $\omega \in \Omega$ nemusí být jevem (=prvkem σ-algebry $\mathcal{A}$), pokud jemu odpovídající jednobodová množina $\{\omega\}$ nepatří do $\mathcal{A}$. Je tedy potřeba chápat spojení *elementární jev* jako jediný (poněkud matoucí) termín, nikoli jako „jev, který je navíc elementární".

Otázka, který z možných pravděpodobnostních modelů je vhodný, zůstává v Kolmogorovově pravděpodobnostním modelu otevřená. Můžeme vycházet z dodatečných znalostí o náhodném pokusu a jeho fungování. Nejsou-li k dispozici, obvykle použijeme experimenty, při nichž sledujeme chování náhodného pokusu při mnoha opakováních. (K tomu nám bude sloužit statistika.) Zde se matematika dostává do podobné role jako jiné vědy, kdy navržený matematický model je testován dle skutečných výsledků a podle toho zamítnut, nebo dále používán.

Statistický odhad parametrů modelu použijeme i v některých případech, kdy máme dostatek informací pro exaktní stanovení parametrů, ale jejich využití by bylo pracné.

Příklad 1.7.16: Generování šifrovacího kódu typicky začíná vyhledáním velkého prvočísla mezi náhodně volenými čísly. Pro odhad rychlosti algoritmu i rizika uhodnutí kódu je třeba znát pravděpodobnost, že číslo v daném rozsahu je prvočíslo. Tato prvočísla nelze soudobou technikou spočítat. Můžeme však pokus mnohokrát opakovat a z toho statistickými postupy učinit závěr.

Navíc je poměrně obtížné testovat, že zvolené číslo je prvočíslo. Daleko rychlejší je série náhodných pokusů, které se snaží ukázat, že jde o číslo složené. Pokud opakované pokusy selžou, je vysoká pravděpodobnost, že se jedná o prvočíslo. Obvykle se spokojíme s malým nenulovým rizikem, že tomu tak není, např. $2^{-100} \doteq 7.889 \cdot 10^{-31}$ [Knuth 1997]. Pak je velmi pravděpodobné, že tato možnost pro daný program nikdy nenastane (a ještě pravděpodobnější, že si toho nikdo nevšimne).

Rovnoměrná rozdělení a geometrická pravděpodobnost

Ukážeme přístup, který dovolí řešit (některé) úlohy na geometrickou pravděpodobnost. Základem je rovnoměrné rozdělení na nějakém geometrickém útvaru; začneme od úsečky. Rovnoměrné rozdělení pravděpodobnosti na intervalu $\langle a, b\rangle$ popisuje pravděpodobnostní míra P taková, že každý interval $\langle c, d\rangle \subseteq \langle a, b\rangle$ má pravděpodobnost přímo úměrnou své délce, tj.

$$P(\langle c, d\rangle) = \frac{d-c}{b-a}. \tag{1.7}$$

To je korektní formulace původního záměru, že „všechny body z $\langle a, b\rangle$ jsou stejně pravděpodobné." (Pravděpodobnost každého bodu – přesněji jednobodové množiny – je ovšem nulová.) Můžeme P rozšířit na všechny intervaly $I \subseteq \mathbb{R}$ tak, že $P(I) = P(I \cap \langle a, b\rangle)$. Tedy σ-algebra, na které P definujeme, má obsahovat všechny intervaly z $\mathbb{R}$ a některé další množiny (např. sjednocení intervalů), aby se vyhovělo definici 1.7.1. Použijeme *nejmenší* takovou σ-algebru na $\mathbb{R}$, ale terpve až se ujistíme, že vůbec existuje.

Tvrzení 1.7.17: *Nechť Ω je neprázdná množina a $\mathcal{M} \subseteq \exp\Omega$. Pak existuje (jediná) nejmenší σ-algebra $\mathcal{A} \subseteq \exp\Omega$ obsahující $\mathcal{M}$.*

Důkaz: Nějaká σ-algebra obsahující $\mathcal{M}$ existuje, např. $\exp\Omega$. Snadno se lze přesvědčit, že průnik σ-algeber je opět σ-algebra. Vezmeme tedy za $\mathcal{A}$ průnik všech σ-algeber, které obsahují $\mathcal{M}$. □

Definice 1.7.18: *σ-algebra $\mathcal{A}$ z tvrzení 1.7.17 se nazývá σ-algebra* ***generovaná*** *množinou $\mathcal{M}$.* ***Borelova σ-algebra*** *je nejmenší σ-algebra podmnožin $\mathbb{R}$, která obsahuje všechny intervaly.*

Borelova σ-algebra obsahuje všechny intervaly otevřené, uzavřené i polouzavřené, dále jejich spočetná sjednocení a některé další množiny, ale je menší než $\exp\mathbb{R}$. Její prvky nazýváme **borelovské množiny**. Podobně je definována **Borelova σ-algebra** $\mathcal{B}(\mathbb{R}^n)$ podmnožin víceroměrného prostoru $\mathbb{R}^n$ jako σ-algebra generovaná všemi n-rozměrnými intervaly.

Pravděpodobnostní míry používané pro řešení úloh z geometrické pravděpodobnosti se zavádějí na Borelově σ-algebře v příslušném prostoru.

Příklad 1.7.19: * Nechť $\Omega \subseteq \mathbb{R}^2$ je omezený dvojrozměrný interval (tj. obdélník). Za $\mathcal{A} \subseteq \exp\Omega$ vezmeme σ-algebru generovanou všemi dvojrozměrnými podintervaly. Na ní lze definovat pravděpodobnost $P(A) = \varrho(A)/\varrho(\Omega)$, kde ϱ značí obsah plochy. Rovnoměrnost znamená, že pravděpodobnost každého intervalu (obdélníka rovnoběžného s osami) je úměrná jeho obsahu (součinu délek stran). Obsah plochy dalších množin, jako je např. kruh z příkladu 1.6.1, je dána obsahem obdélníků a σ-aditivitou (P2).

Toto je klíčový příklad pro geometrické pravděpodobnosti. Lze jej zobecnit na případy, kdy Ω má jiný tvar než dvojrozměrný interval (např. kruh), nebo na jiný počet dimenzí. Stejně jako rovnoměrné rozdělení pravděpodobnosti na reálném intervalu (=úsečce), považujeme rozdělení na křivce (např. kružnici) za rovnoměrné, jestliže pravděpodobnost výskytu na oblouku křivky je úměrná délce tohoto oblouku. Podobně lze vytvořit pravděpodobnostní model i pro příklad 1.6.4.

Poznámka 1.7.20: Pojem spojitého rovnoměrného rozdělení pravděpodobnosti na (borelovské) množině B ve skutečnosti není primární. Předpokládá, že na (borelovských) podmnožinách B už máme nějakou míru (ne nutně pravděpodobnostní; na celku nemusí být jednotková). Pravděpodobnost, že výsledek padne do množiny, je úměrná míře této podmnožiny. Požadujeme tedy, aby pravděpodobnostní míra byla přímo úměrná nějaké předem definované míře, jejíž volba nemusí být vždy zřejmá. Např. interval můžeme nelineárně zobrazit na stejný interval. Rozdělení rovnoměrné po této transformaci je jiné než rovnoměrné rozdělení před ní.

Tím se řeší Bertrandův paradox (příklad 1.6.6) i příklad 1.6.7; výsledky pokusu lze popsat ve vhodně zvoleném prostoru, ale zadání úlohy neurčuje, jaká na něm má být míra. Není ani žádná standardní volba, která by preferovala jednu míru oproti jiným. Zadání těchto úloh bylo neúplné.

Příklad 1.7.21: Nezavádí se rovnoměrné rozdělení na polopřímce. Přitom např. polopřímku $\langle 0, \infty)$ můžeme vzájemně jednoznačně (např. funkcí $u \mapsto 1/(u+1)$) zobrazit na omezený interval a na něm zavést rovnoměrné rozdělení. Tomu by odpovídalo nějaké rozdělení na původní polopřímce. Kdybychom však zobrazili polopřímku na stejný interval jinou funkcí, např. $u \mapsto \exp(-u)$, dostali bychom jiné rozdělení na polopřímce. Protože není žádné význačné zobrazení polopřímky na omezený interval, které bychom měli preferovat, nemůžeme vyčlenit jedno rozdělení na polopřímce, které bychom považovali za obdobu rovnoměrného rozdělení.

Kolmogorovova definice pravděpodobnosti se může zdát zbytečně komplikovaná. Pokusíme se naznačit, proč je právě taková.

Nabízí se myšlenka, že místo spočetné aditivity by v (P2) mohla stačit konečná aditivita; tu lze bez újmy na obecnosti redukovat na dva neslučitelné jevy:

$$A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)\,.$$

Pak bychom ale narazili na problém v příkladu 1.6.1; rovnoměrné rozdělení z příkladu 1.7.19 by nedávalo odpověď, jaká je pravděpodobnost, že vybereme bod z kruhu. Kruh lze vyjádřit jako sjednocení spočetně mnoha dvojrozměrných intervalů (i disjunktních), ale s konečně mnoha nevystačíme. Zde bychom ještě mohli problém vyřešit pomocí aproximace konečnými sjednoceními, ale v následujícím příkladu nikoli.

Příklad 1.7.22: Nechť $\Omega = \mathbb{N}$, $\mathcal{A} = \exp\Omega$. Pak lze definovat aditivní funkci $P\colon \mathcal{A} \to \langle 0,1\rangle$ takovou, že pro všechny konečné množiny $A \subseteq \mathbb{N}$ je $P(A) = 0$, $P(\bar{A}) = 1$. (Definice této funkce na zbývajících podmnožinách, které jsou nekonečné a mají nekonečný doplněk, není jednoduchá, ale je možná, viz např. [Sikorski 1969].) To znamená, že jakékoli číslo má nulovou pravděpodobnost (stejně jako každá konečná množina), ale přitom s jistotou nějaké číslo vybereme.

Další námitka je, proč si komplikujeme život pojmem σ-algebry a nepracujeme rovnou s množinou $\exp\Omega$ všech podmnožin množiny elementárních jevů. Následující příklad dokládá, že to není samoúčelné.

Příklad 1.7.23 (Banachův-Tarského paradox): Povrch koule (sféru) lze rozdělit na konečně mnoho disjunktních množin a ty do dvou skupin s následující vlastností: každá skupina množin po vhodných rotacích tvoří opět rozklad *celé* sféry na disjunktní množiny. To znamená, že z jedné sféry uděláme dvě stejně velké, aniž bychom měnili rozměry (a jakékoli metrické vlastnosti) použitých množin. To samozřejmě odporuje naší zkušenosti a chtěli bychom namítnout, že to není možné proto, že povrch dvou sfér je $2\times$ větší než povrch jedné sféry (a nenulový). Nicméně možné to je. (Důkaz je velmi náročný a není konstruktivní, ukazuje, že takové množiny existují, aniž by je přímo popsal.) Chyba naší námitky je v tom, že ne každé množině lze přiřadit obsah v obvyklém smyslu. Přesněji, pokud bychom zavedli (σ-aditivní) pravděpodobnostní míru na podmnožinách sféry tak, aby měla obvyklý význam obsahu plochy (děleného obsahem celé sféry), nepodařilo by se nám ji rozšířit na množiny, použité v tomto paradoxu; pro tyto množiny nelze definovat obsah v obvyklém významu. Následkem toho nelze například definovat a vyhodnotit pravděpodobnost, že bod sféry náhodně vybraný s rovnoměrným rozdělením padne do takovéto množiny.

Praktický závěr tohoto hluboce teoretického výsledku je, že nemůžeme definovat obsah *všech* podmnožin sféry. Abychom mohli používat pravděpodobnostní míru klíčovou pro geometrickou pravděpodobnost, musíme se omezit jen na *některé* podmnožiny, které tvoří σ-algebru. (Obvykle používáme Borelovu σ-algebru.) Proto musíme přijmout tento pojem jako nutný základ užitečné teorie, nikoli jako zbytečnou komplikaci.

Poznámka 1.7.24: Laplaceova definice pravděpodobnosti bývá v učebnicích nazývána **klasická definice pravděpodobnosti**. Nicméně v soudobých článcích bývá tento pojem používán pro Komogorovovu definici pravděpodobnosti, která se stává „novou klasikou" (na rozdíl od např. kvantové pravděpodobnosti a dalších zobecnění), což svědčí o jejím všeobecném přijetí. Proto se zde termínu „klasická pravděpodobnost" vyhýbáme.

Zavedli jsme pravděpodobnostní model, který navrhl Kolmogorov ve 30. letech 20. století. Tento přístup ukončil staletí hledání teorie, která jednotným způsobem pokrývá konečný Laplaceův model i geometrickou pravděpodobnost. Brzy se stal standardním nástrojem. Lze mu vytknout, že axiomatický přístup je nenázorný, říká nám pouze, jaké vlastnosti musí pravděpodobnost mít, nikoli, jakých hodnot má v konkrétním modelu nabývat. V tom nám ponechává

volnost (při volbě modelu konkrétního náhodného pokusu). Rovněž požadavak spočetné aditivity se jeví zvláštní. Nicméně jsme uvedli alespoň argumenty, proč aditivita pro konečné soubory jevů je příliš slabý požadavek a pro libovolné soubory jevů zase příliš silný. Nezbývá tedy než přijmout tento princip jako nejúspěšnější, který jiné přístupy zatlačil na okraj zájmu. Musíme se smířit i s tím, že pravděpodobnost (=pravděpodobnostní míra) nemusí být definována pro všechny množiny elementárních jevů, ale pouze pro vybrané jevy, tvořící σ-algebru. To není samoúčelný pojem, ale základ teorie pravděpodobnosti i míry a integrace. Vedlejším produktem je poznání, že některé úlohy na geometrickou pravděpodobnost žádné smysluplné řešení nemají.

Cvičení 1.7.25: Najděte pravděpodobnostní prostor, popisující hod pravidelnou kostkou, jestliže z horní stěny vidíme

1. pouze tečky uprostřed stran (např. 2 tečky u výsledku 6),
2. pouze rohové tečky (ty, které jsou u výsledku 4),
3. pouze jednu polovinu,
4. pouze jeden kvadrant (náhodně vybraný, všechny jsou stejně pravděpodobné).

Řešení: (Následující řešení nejsou jediná možná a s výjimkou případu 3 je lze zjednodušit.)

1. $\Omega = \{1, \ldots, 6\}$, $\mathcal{A} = \{\emptyset, \{6\}, \{1, \ldots, 5\}, \Omega\}$.
2. $\Omega = \{1, \ldots, 6\}$, $\mathcal{A} = \{\emptyset, \{1\}, \{2,3\}, \{4,5,6\}, \{1,2,3\}, \{1,4,5,6\}, \{2,3,4,5,6\}, \Omega\}$.
3. Jako v příkladu 1.2.1 (značky jsou symetrické, takže kterákoli polovina nám dává plnou informaci).

Při těchto popisech jsou pravděpodobnosti zúžením pravděpodobnosti P z příkladu 1.2.1 na $\mathcal{A}$, tj. $P(A) = |A|/6$.

4. Můžeme pozorovat následující konfigurace:
   a. žádná tečka,
   b. čtvrtina středové tečky,
   c. rohová tečka
   d. rohová tečka a čtvrtina středové tečky,
   e. rohová tečka a polovina tečky ve středu strany.

   $\Omega = \{a, b, c, d, e\}$, $\mathcal{A} = \exp\Omega$, $P(\{a\}) = 1/12$, $P(\{b\}) = P(\{c\}) = P(\{d\}) = 1/4$, $P(\{e\}) = 1/6$ (např. jev $\{c\}$ nastane vždy, když padne čtyřka, a v polovině případů, kdy padne dvojka, $P(\{c\}) = 1/6 + 1/12 = 1/4$). Z toho již vyplývají zbývající pravděpodobnosti. (Tento příklad je sice umělý, ale např. v počítačovém vidění je potřeba řešit složitější situace, kdy z objektu zájmu, např. obličeje, vidíme jen část.) □

Cvičení 1.7.26: Najděte pravděpodobnostní prostor, popisující hod pravidelnou kostkou, jestliže po hodu vidíme pouze jednu *boční* stěnu.

Řešení: Pravděpodobnostní popis může být stejný jako v příkladu 1.2.1; jediný rozdíl je, že jevy rozlišujeme podle té stěny kostky, kterou vidíme, nikoli podle té, která určuje výsledek pokusu.

Nabízí se též „úplnější" popis se 24 elementárními jevy, kterými jsou dvojice výsledků (horní stěna, boční stěna). Pozorovat však můžeme stále stejný počet jevů, např. můžeme pozorovat jev $\{(2,1), (3,1), (4,1), (5,1)\}$, ale již žádnou menší neprázdnou množinu. □

Příklad 1.7.27: * Alice a Bob se střídavě strefují míčem do koše, začíná Alice. Kdo se první strefí, vyhrává. Alice se strefí s pravděpodobností a, Bob s pravděpodobností b (nezávislou na jiných okolnostech, např. předchozím průběhu hry).

(a) Jaká je pravděpodobnost výsledků hry?
(b) Jak se pravděpodobnost změní, jestliže při každém hodu se může míč ztratit s pravděpodobností c? (V tom případě končí hra remízou.)

(c) Předpokládejme, že $a = b$. Navrhněte takové výše výher, aby hra byla spravedlivá (aby průměrná výhra obou hráčů byla stejná)

Řešení: (a) Po prvním hodu Alice vyhrává s pravděpodobností a, s pravděpodobností $1-a$ se pokračuje. Pravděpodobnost, že hra skončí 2. hodem (a výhrou Boba) je $(1-a)\,b$. Pravděpodobnost, že hra skončí 3. hodem (a výhrou Alice) je $(1-a)\,(1-b)\,a$ atd. Pravděpodobnosti výsledků bychom mohli dostat jako součty nekonečných geometrických řad s kvocientem $q_0 = (1-a)\,(1-b)$. Místo toho můžeme uvažovat následovně: Pokud po dvou hodech hra neskončila, což nastane s pravděpodobností q_0, je situace stejná jako na začátku. Je-li α pravděpodobnost výhry Alice ve hře, pak $\alpha = a + q_0\,\alpha$,

$$\alpha = \frac{a}{1-q_0} = \frac{a}{a+b-a\,b}\,.$$

Pravděpodobnost Bobovy výhry je

$$\beta = 1-\alpha = \frac{(1-a)\,b}{a+b-a\,b}\,.$$

To vše za předpokladu, že aspoň jedna z pravděpodobností a, b je nenulová.

(b) Zadání dává smysl pouze pro $c \le \min(1-a, 1-b)$. Po dvou pokusech se vracíme do výchozího stavu s pravděpodobností $q_c = (1-a-c)\,(1-b-c)$. Pravděpodobnost výhry Alice je analogicky

$$\alpha = \frac{a}{1-q_c} = \frac{a}{1-(1-a-c)\,(1-b-c)}\,.$$

Bob vyhraje s pravděpodobností

$$\beta = \frac{(1-a-c)\,b}{1-(1-a-c)\,(1-b-c)}\,.$$

Remíza nastává s pravděpodobností

$$\gamma = 1-\alpha-\beta = \frac{(2-a-c)\,c}{1-(1-a-c)\,(1-b-c)}\,.$$

(c) Je zřejmé, že při $a = b$ a při stejných výhrách by Alice byla zvýhodněna. Spravedlivé by bylo, kdyby Bobova výhra byla větší v poměru

$$\frac{\alpha}{\beta} = \frac{a}{b\,(1-c-a)} = \frac{1}{1-(a+c)}\,,$$

k tomu ale potřebujeme znát $a+c$. Bez této znalosti nemůžeme spravedlivá pravidla navrhnout, ledaže bychom vypláceli různé výhry v závislosti na délce hry. □

Cvičení 1.7.28: Foton vnikne do tenké vrstvy s rovnoběžnými stěnami.[1] Na jejím rozhraní může projít, nebo se odrazí a dojde k druhému rozhraní, kterým projde, nebo se odrazí atd. (a) Vypočtěte pravděpodobnosti, že foton projde do jednotlivých poloprostorů vymezených touto vrstvou. (b) Vyřešte modifikovanou úlohu, kdy foton může být při průchodu vrstvou též pohlcen.

Řešení: Jedná se o příklad 1.7.27 v jiné interpretaci. □

Poznámka 1.7.29: Adekvátním nástrojem pro efektivní popis podobných pokusů jsou *Markovovy řetězce*, viz např. [Hsu 1996, Papoulis 1990].

Cvičení 1.7.30: Dokažte tvrzení 1.4.5. (Návod: využijte distributivitu ve výrokovém počtu.)

[1] Takto se studují mj. vlastnosti vrstvy tiskové barvy na papíře, viz např. Hersch, R.D., Emmel, P., Collaud, F., Crété, F.: Spectral reflection and dot surface prediction models for color halftone prints. *Journal of Electronic Imaging* **14** (2005), No. 3, 33001–12, `http://diwww.epfl.ch/w3lsp/publications/colour/sradspmfchp.pdf`.

1.8 Jiné pohledy na pravděpodobnost

Statistická definice pravděpodobnosti

(Viz např. [Jaroš 1998].)

Provedeme velký počet náhodných pokusů a zjistíme *četnost* náhodného jevu A, tj. počet pokusů, při nichž jev A nastal. Vydělíme-li ji celkovým počtem pokusů, dostaneme *relativní četnost* jevu A, což je číslo z intervalu $\langle 0, 1\rangle$. Podle statistické definice je pravděpodobnost náhodného jevu A je takové číslo $P(A)$, k němuž se při rostoucím počtu realizací pokusu blíží relativní četnost jevu A.

Toto pojetí lze s výhodou použít tam, kde lze stanovit relativní četnosti na základě rozsáhlého empirického materiálu. Z matematického pohledu jako definice neobstojí, neboť postrádáme předpoklady zajišťující konvergenci. Na druhé straně toto pojetí připomíná praktický aspekt: pravděpodobnost nám dovoluje předpovídat budoucí chování na základě mnoha dosud provedených pozorování téhož pokusu. V Komogorovově definici pravděpodobnosti byla pravděpodobnost vnitřním parametrem systému, který lze vyčíst jen ve speciálních případech. Obvykle je nám tento parametr skryt a my pouze věříme, že pravděpodobnost existuje jako objektivní příčina výsledků, které pozorujeme. Jak dále uvidíme, z opakovaného pozorování pokusu můžeme pravděpodobnost pouze odhadovat. Náš odhad je zatížen chybou i nejistotou, zda je správný. Statistická definice pravděpodobnosti zdůrazňuje od začátku praktický aspekt jejího použití.

Regulérní sázkový koeficient

Příklad 1.8.1: Alice a Bob hrají následující hru, používající náhodný pokus se dvěma (nebo i více) výsledky: Alice zvolí kurs na možné výsledky, Bob jako bankéř jí vybere, na který výsledek si má při tomto kursu vsadit. Optimální strategie pro Alici odpovídá pravděpodobnosti jevu, při jakémkoli jiném kursu prodělává.

Řešením předchozí úlohy je tzv. **regulérní sázkový koeficient** (angl. *fair betting quotient*). Ten říká, kolik je člověk ochoten vsadit na to, že jev nastane. Toto pojetí zdůrazňuje, že pravděpodobnost je skrytý parametr, účastníkům pokusu typicky neznámý a pouze odhadovaný. Důležité je i praktické hledisko (kdo umí pravděpodobnost nejlépe odhadnout, vyhrává).

Subjektivní pravděpodobnost

Často v hovoru používáme slovo „pravděpodobně" ve významu, který bychom těžko přesně definovali. Řekneme-li: „zítřejší zkoušku pravděpodobně udělám" nebo „vlak pravděpodobně stihnu," pak tím ostatním sdělujeme informaci podstatnou pro jejich počínání (např. že s námi mohou o víkendu plánovat výlet). Situace má další podstatné rysy náhodného pokusu – výsledek závisí na mnoha okolnostech, které nejsou předem známy (např. zadání písemky a vylosovaná otázka u ústní zkoušky, zpoždění vlaku, v obou případech pak stav baterií v budíku), ale nakonec bude jednoznačně určen. Existuje tedy pro něj pravděpodobnostní model. Nicméně v tomto modelu neznáme pravděpodobnosti, jen na základě předchozích zkušeností soudíme, že pravděpodobnost úspěchu je „velká".

V těchto případech hovoříme o **subjektivní pravděpodobnosti**. Tak tomu bylo i v příkladu 1.4.1. Skutečné pravděpodobnosti zde neznáme, obvykle ani nemáme dostatečný materiál pro jejich seriózní odhad, neboť vypovídáme o jedinečných událostech, kdežto teorie pravděpodobnosti se zabývá náhodnými pokusy, které lze mnohanásobně opakovat.

Běžně však hovoříme o pravděpodobnosti i v situacích, které nevyhovují požadavkům, např.: „můj výkon u zítřejší zkoušky bude pravděpodobně chabý." Je problematické, jak definovat chabý výkon, aby po zkoušce bylo možno vždy jednoznačně rozhodnout, zda se předpověď naplnila, nebo ne.

Ukázali jsme některé z mnoha přístupů k pravděpodobnosti. I když nemohou konkurovat Kolmogorovovu modelu pravděpodobnosti, poskytují alternativní pohledy na příslušné pojmy. Orientují se více na to, k čemu chceme pravděpodobnostní model použít a jaký to pro nás má mít užitek. To neškodí připomenout.

2. Konstrukce pravděpodobností

V této kapitole uvedeme několik matematických konstrukcí, které pravděpodobnostní prostory umožňují. To nám dovolí popsat náhodné pokusy, které vznikly složením několika jednodušších náhodných pokusů podle různých pravidel.

2.1 Konvexní kombinace pravděpodobností

Příklad 2.1.1: V urně je $n_1 = 10$ správných hracích kostek, na nichž padají všechna čísla se stejnou pravděpodobností, a $n_2 = 5$ vadných hracích kostek, 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?

Až dosud jsme měli jeden pravděpodobnostní prostor a jedinou pravděpodobnost (tu „správnou", odpovídající skutečné situaci). To je samozřejmě velmi omezený pohled, protože si dovedeme představit jiné situace, kdy tytéž jevy budou mít jinou pravděpodobnost.

Mějme množinu Ω a na ní σ-algebru $\mathcal{A}$. Na ní lze definovat mnoho pravděpodobností, neboť za pravděpodobnost považujeme každou funkci $P\colon \mathcal{A} \to \langle 0,1\rangle$, která splňuje podmínky definice 1.7.3. Označme $\mathcal{P}(\mathcal{A})$ množinu všech pravděpodobností na $\mathcal{A}$; nazýváme ji **prostor pravděpodobností** a lze ji chápat jako podmnožinu lineárního prostoru všech reálných funkcí na $\mathcal{A}$. Vzhledem k normalizační podmínce $P(\mathbf{1}) = 1$ není prostor $\mathcal{P}(\mathcal{A})$ *uzavřený* na sčítání ani na násobení reálným číslem (známé operace s reálnými funkcemi), tj. když tyto operace s pravděpodobnostmi provedeme, může se stát, že jejich výsledek není pravděpodobnost (není z prostoru pravděpodobností, protože nesplňuje normalizační podmínku). Proto nemůžeme na prostoru pravděpodobností používat běžné operace z lineárního prostoru funkcí. Existuje však jedna povolená výjimka, a tou jsou *konvexní kombinace*, tj. lineární kombinace s nezápornými koeficienty, jejichž součet je 1. Nechť P_1, P_2 jsou pravděpodobnosti na $\mathcal{A}$ a nechť $w_1, w_2 \in \langle 0,1\rangle$, $w_1 + w_2 = 1$. Pak funkce

$$P = w_1 P_1 + w_2 P_2\,,$$

je pravděpodobnost na $\mathcal{A}$. (Stačí ověřit podmínky z definice 1.7.3.) V tom případě řekneme, že P je **konvexní kombinace pravděpodobností** P_1, P_2. Čísla w_1, w_2 nazýváme *váhy* nebo *koeficienty* konvexní kombinace.

Konvexní kombinaci pravděpodobností lze interpretovat následovně: Jsou 2 možné pravděpodobnostní modely na *stejné* σ-algebře, jeden popsaný pravděpodobností P_1, druhý pravděpodobností P_2. Provedeme náhodný pokus s pravděpodobnostmi výsledků w_1, w_2 a podle něj vybereme pravděpodobnostní model.

Řešení (příkladu 2.1.1): Označme P_1, resp. P_2, pravděpodobnost výsledků při hodu správnou, resp. vadnou kostkou,

$$P_1(j) = \frac{1}{6}\,, \qquad j = 1,\dots,6\,,$$
$$P_2(6) = \frac{1}{2}\,, \qquad P_2(j) = \frac{1}{10}\,, \qquad j = 1,\dots,5\,.$$

(Pravděpodobnosti ostatních jevů, např. že padne sudé číslo, jsou tímto jednoznačně určeny.) Pravděpodobnost, že vytáhneme správnou, resp. vadnou kostku, je

$$w_1 = \frac{n_1}{n_1 + n_2} = \frac{2}{3}\,, \qquad \text{resp. } w_2 = \frac{n_2}{n_1 + n_2} = \frac{1}{3}\,.$$

Pravděpodobnost výsledků pokusu je $P = w_1\,P_1 + w_2\,P_2$,

$$P\,(6) = \frac{w_1}{6} + \frac{w_2}{2} = \frac{5}{18}\,, \qquad P\,(j) = \frac{w_1}{6} + \frac{w_2}{10} = \frac{13}{90}\,, \qquad j = 1, \ldots, 5\,.$$

□

Konvexní kombinace lze např. indukcí zobecnit na libovolný konečný počet pravděpodobností,

$$P = \sum_{k=1}^{n} w_k\,P_k\,, \qquad \text{kde } w_k \in \langle 0, 1\rangle, \qquad \sum_{k=1}^{n} w_k = 1\,,$$

i na nekonečnou posloupnost pravděpodobností,

$$P = \sum_{k=1}^{\infty} w_k\,P_k\,, \qquad \text{kde } w_k \in \langle 0, 1\rangle, \qquad \sum_{k=1}^{\infty} w_k = 1\,.$$

(Předchozí případy dostaneme, zvolíme-li jen konečně mnoho koeficientů w_k nenulových.)

Poznámka 2.1.2: Obvyklá interpretace $\sum\limits_{k=1}^{\infty} w_k\,P_k$ jako limity částečných součtů

$$\sum_{k=1}^{\infty} w_k\,P_k = \lim_{n\to\infty} \sum_{k=1}^{n} w_k\,P_k$$

je možná v prostoru funkcí, nikoli však v prostoru pravděpodobností, neboť částečné součty této řady nejsou pravděpodobnosti, nenabývají jednotkové hodnoty na $\mathbf{1}$. Nicméně pro všechna $A \in \mathcal{A}$ konverguje řada

$$\sum_{k=1}^{\infty} w_k\,P_k\,(A) = \lim_{n\to\infty} \sum_{k=1}^{n} w_k\,P_k\,(A)\,,$$

takže funkce $P\colon A \mapsto \sum\limits_{k=1}^{\infty} w_k\,P_k\,(A)$ je korektně definována.

Ověříme definiční podmínky pravděpodobnosti z definice 1.7.3:

$$P(1) = \sum_{k=1}^{\infty} w_k\,P_k(1) = \sum_{k=1}^{\infty} w_k = 1\,,$$

$$P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{k=1}^{\infty} w_k\,P_k\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{k=1}^{\infty} w_k \sum_{n=1}^{\infty} P_k(A_n) = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} w_k\,P_k(A_n) = \sum_{n=1}^{\infty} P(A_n)\,,$$

pokud jsou jevy A_n, $n \in \mathbb{N}$, po dvou neslučitelné.

Prostor pravděpodobností $\mathcal{P}\,(\mathcal{A})$ je tedy uzavřený na tvoření konvexních kombinací (včetně spočetných), je konvexní podmnožinou příslušného prostoru reálných funkcí.

Poznámka 2.1.3: Konvexní kombinace nespočetně mnoha pravděpodobností by nepřinesly nic nového; má-li být součet vah konečný, může být jen spočetně mnoho z nich nenulových.

Zatímco dřívější přístupy se snažily popsat jedinou pravděpodobnost, „tu správnou" pro daný náhodný pokus, Kolmogorovův model připouští mnoho možných pravděpodobností. Na nich jsou definovány mj. konvexní kombinace, které mají navíc názorný význam v náhodných pokusech. Prostor pravděpodobností je konvexní.

Příklad 2.1.4: Předem je známo $n = 30$ zkouškových otázek; podle studenta je z nich $n_l = 5$ lehkých, $n_s = 15$ středních a $n_t = 10$ těžkých; odpovídající pravděpodobnosti výsledků P_l, P_s, P_t jsou v tabulce:

otázka	pravděpodobnost	1	2	3	4
lehká	P_l	0.8	0.2	0	0
střední	P_s	0.1	0.4	0.4	0.1
těžká	P_t	0	0	0.2	0.8

Jaká je pravděpodobnost hodnocení náhodně vybrané otázky? (Předpokládáme, že každá otázka může být vybrána se stejnou pravděpodobností.)

Řešení: Výsledná pravděpodobnost je $P = w_l P_l + w_s P_s + w_t P_t$, kde $w_l = n_l/n = 1/6$, $w_s = n_s/n = 1/2$, $w_t = n_t/n = 1/3$. Dostáváme

$$P(1) = \frac{1}{6}0.8 + \frac{1}{2}0.1 \doteq 0.183\,.$$
$$P(2) = \frac{1}{6}0.2 + \frac{1}{2}0.4 \doteq 0.233\,.$$
$$P(3) = \frac{1}{2}0.4 + \frac{1}{3}0.2 \doteq 0.267\,.$$
$$P(4) = \frac{1}{2}0.1 + \frac{1}{3}0.8 \doteq 0.317\,.$$

□

2.2 Konvergence posloupnosti pravděpodobností

V prostoru funkcí máme definovanou konvergenci; pokus o její použití v prostoru pravděpodobností v pozn. 2.1.2 selhal. Nicméně stejnou definici lze zavést:

Definice 2.2.1: *Nechť $(P_k)_{k\in\mathbb{N}}$ je posloupnost pravděpodobností na σ-algebře $\mathcal{A}$. Pokud existuje pravděpodobnost P na $\mathcal{A}$ taková, že*

$$\forall A \in \mathcal{A} : P(A) = \lim_{k\to\infty} P_k(A)\,,$$

pak ji nazýváme ***(bodová) limita*** *posloupnosti pravděpodobností $(P_k)_{k\in\mathbb{N}}$.*

Nelze vynechat předpoklad, že P je pravděpodobnost; může se stát, že limita v prostoru funkcí existuje, ale není to pravděpodobnost. (Nehrozí to pro konečně mnoho jevů.)

Příklad 2.2.2: Na σ-algebře $\mathcal{A} = \exp\mathbb{N}$ definujme pro každé $k \in \mathbb{N}$ pravděpodobnost P_k s rovnoměrným rozdělením na $\{1,\ldots,k\}$.

$$P_k(n) = \begin{cases} \frac{1}{k} & \text{pro } n \in \{1,\ldots,k\}\,, \\ 0 & \text{jinak.} \end{cases}$$

Pak neexistuje pravděpodobnost P, která by byla limitou posloupnosti $(P_k)_{k\in\mathbb{N}}$, neboť by musela splňovat

$$P(n) = \lim_{k\to\infty} P_k(n) = \lim_{k\to\infty} \frac{1}{k} = 0\,,$$
$$P(\mathbf{1}) = \lim_{k\to\infty} P_k(\mathbf{1}) = \lim_{k\to\infty} 1 = 1\,,$$

což je v rozporu se σ-aditivitou (P2).

Na prostoru pravděpodobností lze zavést více typů konvergence. Zde jsme se omezili na jeden. Prostor pravděpodobností není vzhledem němu uzavřený (v odpovídajícím prostoru reálných funkcí), tj. limita pravděpodobností *nemusí* být pravděpodobnost. Budeme se nicméně setkávat s případy, kdy lze nějakou pravděpodobnost vyjádřit jako limitu posloupnosti pravděpodobností, což nám umožní užitečné zjednodušující aproximace.

2.3 Součin pravděpodobnostních prostorů

Příklad 2.3.1: Hodíme hrací mincí a kostkou. Jaká je pravděpodobnost možných výsledků? Jaká je pravděpodobnost, že padne líc nebo liché číslo?

Řešení: Všechny možné výsledky označme Li, Ri, kde L značí líc, R rub a $i \in \{1,2,3,4,5,6\}$. Každý z nich má pravděpodobnost $1/12$. Pravděpodobnost, že padne líc nebo liché číslo, je

$$P(\{L1, L2, L3, L4, L5, L6, R1, R3, R5\}) = \frac{9}{12} = \frac{3}{4}.$$

□

To bylo snadné. Nicméně popíšeme tento postup obecně.

Původní dva náhodné pokusy jsou popsány pravděpodobnostními prostory $(\Omega_1, \mathcal{A}_1, P_1)$, $(\Omega_2, \mathcal{A}_2, P_2)$, kde např. $\Omega_1 = \{L, R\}$, $\Omega_2 = \{1,2,3,4,5,6\}$, $\mathcal{A}_j = \exp \Omega_j$, $P_j(A) = |A| / |\Omega_j|$ pro $j = 1, 2$. Ve složeném pokusu provedeme oba původní náhodné pokusy, prostor všech elementárních jevů bude $\Omega = \Omega_1 \times \Omega_2$, tj. množina všech dvojic, v nichž první složka je z Ω_1, druhá z Ω_2. Příslušná σ-algebra je zde $\mathcal{A} = \exp \Omega = \exp(\Omega_1 \times \Omega_2)$. Tak to bude vždy, pokud jsou množiny Ω_1, Ω_2 konečné a $\mathcal{A}_1 = \exp \Omega_1$, $\mathcal{A}_2 = \exp \Omega_2$. U nekonečných pravděpodobnostních modelů může nastat složitější situace.

Příklad 2.3.2: Nechť dva náhodné pokusy jsou měření analogových veličin, např. napětí a proudu. Pro libovolné dva reálné intervaly můžeme testovat jev, že první veličina padne do jednoho intervalu, druhá do druhého, tj. že dvojice výsledků padne do dvojrozměrného intervalu. Příslušná σ-algebra tedy obsahuje všechny dvojrozměrné intervaly, spočetná sjednocení dvojrozměrných intervalů atd. Zvolíme-li nejmenší takovou σ-algebru, je to σ-algebra *generovaná* všemi dvojrozměrnými intervaly (zde Borelova σ-algebra), nikoli však množina všech podmnožin dvojic reálných čísel.

Obecně σ-algebra $\mathcal{A}$ je *generovaná* všemi množinami tvaru $A_1 \cap A_2$, kde $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. Na $\mathcal{A}$ definujeme *součin pravděpodobností* P:

$$\forall A_1 \in \mathcal{A}_1 \; \forall A_2 \in \mathcal{A}_2 : P(A_1 \cap A_2) = P_1(A_1) \cdot P_2(A_2) . \tag{2.1}$$

Definovali jsme P jen na generujících prvcích, ale lze dokázat, že existuje jediná pravděpodobnost na $\mathcal{A}$ s touto vlastností.

Takto popíšeme situaci, kdy provedeme náhodný pokus popsaný pravděpodobnostním prostorem $(\Omega_1, \mathcal{A}_1, P_1)$ a také náhodný pokus popsaný pravděpodobnostním prostorem $(\Omega_2, \mathcal{A}_2, P_2)$. Celkový výsledek popíšeme dvojicí výsledků obou pokusů. Zásadně *musíme provést oba pokusy.*

Předchozí konstrukci lze zobecnit na libovolný konečný počet pravděpodobnostních prostorů:

Definice 2.3.3: *Nechť $(\Omega_1, \mathcal{A}_1, P_1), \ldots, (\Omega_k, \mathcal{A}_k, P_k)$ jsou pravděpodobnostní prostory.* ***Součin σ-algeber*** *je σ-algebra $\mathcal{A}$ podmnožin množiny $\Omega = \prod_{j=1}^{k} \Omega_j = \Omega_1 \times \cdots \times \Omega_k$ generovaná systémem množin $\left\{\bigcap_{j=1}^{k} A_j : A_j \in \mathcal{A}_j\right\}$. Na $\mathcal{A}$ je definován* ***součin pravděpodobností*** *$P_1, \ldots, P_k$, což je jediná pravděpodobnost P na $\mathcal{A}$ taková, že*

$$\forall A_1 \in \mathcal{A}_1 \; \ldots \; \forall A_k \in \mathcal{A}_k : P\Big(\bigcap_{j=1}^{k} A_j\Big) = \prod_{j=1}^{k} P_j(A_j) = P_1(A_1) \cdot \ldots \cdot P_k(A_k) .$$

Příklad 2.3.4: Příklad 2.1.4 lze modifikovat tak, že student musí odpovědět na všech 30 otázek, výsledkem je posloupnost 30 známek; vše popisuje součin třiceti σ-algeber a odpovídající součin pravděpodobností.

Lze zavést i součin nekonečně mnoha pravděpodobnostních prostorů.

Poznámka 2.3.5: Pravděpodobnost na součinu pravděpodobnostních prostorů byla vypočítána jako součin pravděpodobností, tomu odpovídá nezávislost příslušných jevů (viz kapitola 1.5), což je v tomto kontextu obvyklý předpoklad. Lze si však na stejné σ-algebře představit i jinou pravděpodobnost, např. kdyby v příkladu 2.3.1 byla kostka i mince zmagnetována. Tento případ však není předmětem pojmu součinu pravděpodobnostních prostorů.

Právě zavedený součin pravděpodobnostních prostorů nám dovoluje popsat jedním modelem dva nebo více náhodných pokusů, které jsou vždy *všechny provedeny* a „nesouvisí spolu" (jsou nezávislé). Jejich společný popis budeme potřebovat jako první krok dalších procedur.

2.4 Součet σ-algeber, směs pravděpodobností

Příklad 2.4.1: V urně je $n_1 = 10$ (správných) hracích kostek a $n_2 = 5$ mincí. Náhodně vybereme jeden předmět (všechny mají stejnou pravděpodobnost, že budou vybrány) a hodíme jím; jaká je pravděpodobnost možných výsledků?

Řešení: Množina všech možných výsledků je $\Omega = \{1,2,3,4,5,6,L,R\}$, kde L, resp. R, znamená líc, resp. rub, v případě, že jsme vybrali minci. Jelikož mince byla vybrána s pravděpodobností $\frac{n_2}{n_1+n_2} = \frac{1}{3}$ a kostka s pravděpodobností $\frac{n_1}{n_1+n_2} = \frac{2}{3}$, snadno stanovíme pravděpodobnosti

$$P(L) = P(R) = \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}.$$
$$P(j) = \frac{2}{3}\cdot\frac{1}{6} = \frac{1}{9}, \qquad j \in \{1,2,3,4,5,6\}.$$

□

Nyní popíšeme předchozí postup formálně. Nechť $(\Omega_1,\mathcal{A}_1,P_1),(\Omega_2,\mathcal{A}_2,P_2)$ jsou pravděpodobnostní prostory, $\Omega_1 \cap \Omega_2 = \emptyset$ a $w \in \langle 0,1\rangle$. Na $\Omega = \Omega_1 \cup \Omega_2$ definujeme σ-algebru $\mathcal{A} = \{A_1 \cup A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ zvanou *součet σ-algeber* $\mathcal{A}_1,\mathcal{A}_2$. Každý prvek $A \in \mathcal{A}$ lze *jednoznačně* vyjádřit ve tvaru $A = A_1 \cup A_2$, kde $A_1 = A \cap \Omega_1 \in \mathcal{A}_1$, $A_2 = A \cap \Omega_2 \in \mathcal{A}_2$. Vztahem

$$P(A) = P(A_1 \cup A_2) = w\,P_1(A_1) + (1-w)\,P_2(A_2)$$

definujeme pravděpodobnost $P = \mathrm{Mix}_w(P_1,P_2)$ zvanou *směs pravděpodobností P_1,P_2 s váhou (=koeficientem) w*.

Cvičení 2.4.2: Ověřte, že $\mathcal{A}$ je σ-algebra a P je pravděpodobnost.

Řešení: Ukážeme jen σ-aditivitu, ostatní podmínky jsou snadné. Nechť A_n, $n \in \mathbb{N}$, jsou po dvou disjunktní množiny z $\mathcal{A}$. S použitím rozkladů $A_{n1} = A_n \cap \Omega_1$, $A_{n2} = A_n \cap \Omega_2$, $n \in \mathbb{N}$, dostáváme

$$\begin{aligned}
P\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) &= P\Big(\bigcup_{n\in\mathbb{N}} (A_{n1}\cup A_{n2})\Big) = P\Big(\bigcup_{n\in\mathbb{N}} A_{n1} \cup \bigcup_{n\in\mathbb{N}} A_{n2}\Big) = \\
&= w\,P_1\Big(\bigcup_{n\in\mathbb{N}} A_{n1}\Big) + (1-w)\,P_2\Big(\bigcup_{n\in\mathbb{N}} A_{n2}\Big) = \\
&= w\sum_{n\in\mathbb{N}} P_1(A_{n1}) + (1-w)\sum_{n\in\mathbb{N}} P_2(A_{n2}) = \\
&= \sum_{n\in\mathbb{N}} \left(w\,P_1(A_{n1}) + (1-w)\,P_2(A_{n2})\right) = \sum_{n\in\mathbb{N}} P(A_n).
\end{aligned}$$

□

Poznámka 2.4.3: Součet σ-algeber $\mathcal{A}_1,\mathcal{A}_2$ se někdy nazývá *direktní součet*. Přitom jeho prvky lze chápat jako dvojice (A_1,A_2), tj. prvky *kartézského součinu* $\mathcal{A}_1 \times \mathcal{A}_2$, takže se nabízí i termín *součin*. Ovšem v definici 2.3.3 jsme již zavedli jiný součin, v algebře (nikoli lineární!) někdy nazývaný *tenzorový*. Na druhé straně je součet σ-algeber definován na sjednocení $\Omega_1 \cup \Omega_2$, proto se této konstrukci také říká *direktní sjednocení* [Sikorski 1969].

Směs pravděpodobností $\mathrm{Mix}_w(P_1, P_2)$ popisuje výsledek náhodného pokusu, při němž uděláme (navíc) nějaký *počáteční pokus*, který má pravděpodobnost úspěchu w. V případě úspěchu provedeme náhodný pokus popsaný pravděpodobnostním prostorem $(\Omega_1, \mathcal{A}_1, P_1)$, při neúspěchu v počátečním pokusu pokračujeme náhodným pokusem popsaným pravděpodobnostním prostorem $(\Omega_2, \mathcal{A}_2, P_2)$. Druhý z pokusů (který nebyl vybrán) nemá vliv na výsledek; obvykle jej ani *neprovedeme*, protože by nám to ubralo čas nebo způsobilo jinou újmu. To má mj. ten důsledek, že případná *závislost obou náhodných pokusů se nemůže projevit.*

Pojem směsi lze zobecnit na libovolný konečný počet pravděpodobností; váhy jsou libovolná nezáporná čísla, jejichž součet je 1.

Definice 2.4.4: *Nechť* $(\Omega_1, \mathcal{A}_1, P_1), \ldots, (\Omega_k, \mathcal{A}_k, P_k)$, *jsou pravděpodobnostní prostory a množiny* $\Omega_1, \ldots, \Omega_k$ *jsou po dvou disjunktní. Nechť* $w_1, \ldots, w_k \in \langle 0, 1\rangle$, $\sum_{j=1}^k w_j = 1$. *Na* $\Omega = \bigcup_{j=1}^k \Omega_j$ *definujeme* σ*-algebru* $\mathcal{A} = \left\{\bigcup_{j=1}^k A_j \mid A_j \in \mathcal{A}_j,\ j = 1, \ldots, k\right\}$ *zvanou* ***součet*** σ***-algeber*** $\mathcal{A}_1, \ldots, \mathcal{A}_k$. *Pak* ***směs pravděpodobností*** $P_1, \ldots, P_k$ ***s koeficienty*** *(=**váhami**)* $w_1, \ldots, w_k$ *je pravděpodobnost* P *na* $\mathcal{A}$ *definovaná vzorcem*

$$P\Big(\bigcup_{j=1}^k A_j\Big) = \sum_{j=1}^k w_j\, P_j(A_j)\,, \qquad A_j \in \mathcal{A}_j,\ j = 1, \ldots, k.$$

Značíme $P = \mathrm{Mix}_{(w_1, \ldots, w_k)}(P_1, \ldots, P_k)$.

Někdy pro stručnost zavádíme vektor vah $\boldsymbol{w} = (w_1, \ldots, w_k)$ a směs značíme $\mathrm{Mix}_{\boldsymbol{w}}(P_1, \ldots, P_k)$. Protože součet vah je 1, lze poslední váhu vynechat (je ostatními určena), což jsme využili v případě směsi dvou pravděpodobností v zápisu $\mathrm{Mix}_w(P_1, P_2)$ místo $\mathrm{Mix}_{(w, 1-w)}(P_1, P_2)$.

Poznámka 2.4.5: Pojem směsi pravděpodobností lze zobecnit i na posloupnosti pravděpodobností. Zobecnění na nespočetně mnoho pravděpodobností je sice možné, ale nepřináší nic nového, neboť pouze spočetně mnoho vah může být nenulových.

Poznámka 2.4.6: Pokud chceme vytvořit směs pravděpodobností a příslušné množiny elementárních jevů nejsou disjunktní, nahradíme je obdobnými (izomorfními) pravděpodobnostními prostory s disjunktními množinami elementárních jevů.

Poznámka 2.4.7: Také v příkladu 2.1.1 bychom mohli (a měli) použít směs pravděpodobností jako v příkladu 2.4.1. Došli bychom ke stejným závěrům. Zjednodušení na konvexní kombinace pravděpodobností je možné, pokud máme sice *různé pravděpodobnosti*, ale *na stejné* σ*-algebře.* Oba popisy lze navzájem převádět, proto se často nerozlišují a pojmem směsi se označuje i konvexní kombinace pravděpodobností. Zde to nečiníme zejména proto, že u náhodných veličin bude definován jiný pojem konvexní kombinace, který bude nutno striktně odlišovat od směsi.

Poznámka 2.4.8: I v příkladu 2.4.1 jsme mohli použít jedinou σ-algebru popisující jak hod mincí, tak i kostkou. Mohla to být σ-algebra $\exp\Omega$, kde $\Omega = \{1, 2, 3, 4, 5, 6, L, R\}$. Stačí dodefinovat, že u kostky mají výsledky L, R nulovou pravděpodobnost, stejně jako čísla u mince. U pravděpodobností není nutno důsledně rozlišovat směsi od konvexních kombinací (také se to mnohde nečiní), bude to však nezbytné u náhodných veličin, na což zde čtenáře připravujeme.

Poznámka 2.4.9: Čtenář purista by mohl požadovat, aby se v příkladu 2.4.1 rozlišovalo i to, kterou kostkou, resp. mincí se házelo. Tomu by odpovídal popis pomocí směsi 15 pravděpodobností, 10 pro kostky a 5 pro mince. Zde byly počty kostek a mincí použity jen k tomu, abychom modelovali náhodný pokus, který rozhodne, zda budeme házet kostkou nebo mincí. Tento počáteční pokus jsme mohli nahradit např. generátorem náhodných čísel, který rozhodne, zda máme hodit kostkou nebo mincí. Pak stačí jedna kostka, jedna mince a popis pomocí směsi dvou pravděpodobností.

Příklad 2.4.10: Příklad 2.1.4 popisuje směs 3 pravděpodobností $\mathrm{Mix}_{(1/6,1/2,1/3)}(P_l, P_s, P_t)$, ale též směs 30 pravděpodobností odpovídajících jednotlivým otázkám (z nichž mnohé lze popsat stejně). Pro směs je typické, že stačí provést jeden z náhodných pokusů, zde odpovědět na jednu otázku (náhodně vybranou na základě jiného náhodného pokusu). Rozdíl v náročnosti realizace náhodného pokusu je zřejmý.

Součet σ-algeber a na něm založený pojem směsi pravděpodobností nám dovoluje popsat jedním modelem dva nebo více náhodných pokusů, z nichž je vždy *proveden jen jeden* (vybraný dalším náhodným pokusem). Náhodné pokusy tohoto typu jsou v praxi časté a tyto modely se užívají mj. proto, že je lze popsat malým počtem parametrů.

Cvičení 2.4.11: Ukažte pojmy z definice směsi pravděpodobností na řešení příkladu 2.4.1.

2.5 Podmíněná pravděpodobnost

Příklad 2.5.1: Výše havarijního pojistného se stanovuje podle pravděpodobnosti, že vozidlo v pojistném období havaruje. Ta je odhadnuta na základě velkého počtu údajů. Jestliže víme, že majitel auta bydlí v Praze, riziko je vyšší a dále se zvyšuje, je-li začátečník nebo způsobil-li v nedávné minulosti nehodu. Také riziko odcizení je větší u vozidel, která parkují na ulici, nikoli v garáži. Tyto dodatečné informace upřesňují naši znalost pravděpodobnosti.

Příklad 2.5.2: Sázky na hokejový zápas může ovlivnit informace o tom, zda nastoupí některý z nejlepších hráčů.

Příklad 2.5.3: Šance na schválení zákona v parlamentu mohou být jiné, pokud víme, že některý poslanec se nebude moci zúčastnit jednání.

Dodatečná informace, že nějaký jev nastal, může změnit naše znalosti o pravděpodobnosti budoucích jevů. Dosud uvažovaný pravděpodobnostní popis systému přepokládal dvě fáze náhodného pokusu. Na začátku nemáme žádné jiné informace než pravděpodobnostní model. Na konci víme o všech jevech, zda nastaly, nebo ne.

Často však informace dostáváme postupně. O některých jevech už víme, zda nastaly, ale o některých to nevíme; pro ty nadále potřebujeme pravděpodobnostní popis, který vychází z původního a respektuje navíc nové poznatky.

Příklad 2.5.4: Dvě fotbalová družstva mohla mít před zápasem rovné šance na vítězství. Je-li však stav zápasu 5 minut před koncem 3 : 0, pravděpodobnosti výhry jsou podstatně jiné. To se v praxi projeví mj. tím, že sázková kancelář nedovolí uzavírat sázky po zahájení zápasu.

Příklad 2.5.5: Malé vlče by mohlo mít dojem, že skoro všechna zvěř v okolí je nemocná. Rodiče *téměř vždy uloví a přinesou nějaký nemocný kus.*

Jinou motivací může být, že některé informace jsou dostupné pouze některým. Jejich odhad pravděpodobností tím bude ovlivněn.

Příklad 2.5.6: Z balíčku 52 karet, mezi nimiž jsou 4 esa, dostali Alice a Bob po pěti kartách. *S jakou pravděpodobností* má Alice *a* es? *Jak* tuto pravděpodobnost odhadne Bob, má-li v ruce $b = 2$ esa?

Obě uvedené situace dovoluje popsat podmíněná pravděpodobnost.

Dostaneme-li dodatečnou informaci, že nastal jev B (neboli nenastal jev $\overline{B}$), změní se naše znalost o pravděpodobnostech jevů. Tomu odpovídá nový pravděpodobnostní model, v němž je pravděpodobnost jevu $\overline{B}$ nulová, což odráží naši znalost, že jev $\overline{B}$ nenastal. Tomu je potřeba přizpůsobit i pravděpodobnosti ostatních jevů v novém modelu.

Poznámka 2.5.7: Nový pravděpodobnostní model dostaneme jednoduše, přesto bychom měli vysvětlit, proč je právě takový a nemůže být jiný (z mnoha možných pravděpodobností). Tato otázka bývá většinou opomíjena, zde ji probereme podrobněji.

Původní model je vyjádřený pravděpodobnostní mírou P. Tu můžeme chápat jako konvexní kombinaci dvou pravděpodobností, P_B v případě, kdy jev B nastává, a $P_{\overline{B}}$ v případě, kdy jev B nenastává,

$$P = w\, P_B + (1 - w)\, P_{\overline{B}}\,.$$

(Jsou definovány na stejné σ-algebře.) V prvním případě je $P_B\left(\overline{B}\right) = 0$, tedy i $P_B\left(A \cap \overline{B}\right) = 0$ pro libovolný jev A. Podobně v druhém případě je $P_{\overline{B}}\left(B\right) = 0$, takže $P_{\overline{B}}\left(A \cap B\right) = 0$. S využitím (1.3) dostáváme

$$P\left(A \cap B\right) = w\, P_B\left(A \cap B\right) = w\, P_B\left(A\right)\,.$$

Speciálně pro $A = \mathbf{1}$:

$$P\left(B\right) = w\, P_B\left(\mathbf{1}\right) = w\,.$$

Dosazením $w = P\left(B\right)$ dostaneme

$$P_B(A) = \frac{P(A \cap B)}{P(B)}$$

i *vzorec pro úplnou pravděpodobnost* ve tvaru

$$P\left(A\right) = P\left(A \cap B\right) + P\left(A \cap \overline{B}\right) = P\left(B\right)\, P_B\left(A\right) + P\left(\overline{B}\right)\, P_{\overline{B}}\left(A\right)\,,$$

který ještě dále zobecníme. Pravděpodobnosti $P_B, P_{\overline{B}}$ nazýváme *podmíněné pravděpodobnosti*; vžilo se pro ně jiné značení.

Definice 2.5.8: *Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor a $B \in \mathcal{A}$, $P(B) > 0$. Pak* ***podmíněná pravděpodobnost jevu A za podmínky B*** *je číslo*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,. \tag{2.2}$$

Funkce

$$P(.|B)\colon \mathcal{A} \to \langle 0,1\rangle\,,\ A \mapsto \frac{P(A \cap B)}{P(B)}$$

se nazývá ***podmíněná pravděpodobnost za podmínky B*** *(nebo jen* ***podmíněná pravděpodobnost****, pokud nehrozí nedorozumění).*

Jevem s nulovou pravděpodobností nelze podmiňovat, protože bychom dělili nulou.

Úmluva. *Kdykoli pracujeme s podmíněnou pravděpodobností, předpokládáme, že jev, jímž je podmíněna, má nenulovou pravděpodobnost.*

Tvrzení 2.5.9 (vlastnosti podmíněné pravděpodobnosti):

1. $P(\mathbf{1}|B) = 1, \qquad P(\mathbf{0}|B) = 0.$
2. *Jsou-li jevy $A_1, A_2, \ldots$ jsou po dvou neslučitelné a $A = \bigcup_{n\in\mathbb{N}} A_n$, pak*

$$P\left(A|B\right) = \sum_{n\in\mathbb{N}} P(A_n|B)\,.$$

3. *Jevy A, B jsou* nezávislé, *právě když $P(A|B) = P(A)$.*

4. $B \subseteq A \implies P(A|B) = 1, \qquad P(A \cap B) = 0 \implies P(A|B) = 0.$

Důkaz těchto vlastností je jednoduchý a ponecháváme jej na čtenáři.

Poznámka 2.5.10: Vlastnost 2 je *σ-aditivita podmíněné pravděpodobnosti.* Spolu s vlastností 1 říká, že podmíněná pravděpodobnost je pravděpodobnost (v původním smyslu, pouze jsme z možných pravděpodobností vybrali jinou, srovnej s kapitolou 2). Značení $P(.|B)$ napovídá, že byla vytvořena z pravděpodobnosti $P(.)$ a jevu B, a to tak, že jsme položili pravděpodobnost jevu $\overline{B}$ rovnu nule (stejně jako všech jevů $C \subseteq \overline{B}$) a zbylé nenulové hodnoty vynásobili konstantou $1/P(B)$, aby pravděpodobnost jevu jistého byla opět 1.

Řešení (příkladu 2.5.6): Označme A_a jev, že Alice má a es a B_b jev, že Bob má b es. *(Později zavedeme univerzálnější značení pomocí náhodných veličin.)* Potřebujeme hypergeometrické rozdělení z příkladu 1.3.11. Bez další znalosti má Alice a es s pravděpodobností

$$P(A_a) = \frac{\binom{4}{a}\binom{52-4}{5-a}}{\binom{52}{5}}.$$

a	0	1	2	3	4
$P(A_a)$	0.659	0.299	0.040	0.00174	$1.847 \cdot 10^{-5}$

Bob odečte karty, které má v ruce, zbývá $52 - 5 = 47$ karet, mezi nimiž je $4 - b$ es, obecně

$$P(A_a|B_b) = \frac{\binom{4-b}{a}\binom{47-4+b}{5-a}}{\binom{47}{5}}.$$

Pro $b = 2$

a	0	1	2	3	4
$P(A_a\|B_2)$	0.796	0.194	$9.251 \cdot 10^{-3}$	0	0

□

Často nás zajímá podmiňování jevy, o nichž víme, že vždy nastane právě jeden z nich, tj. tvoří *úplný systém jevů.*

Věta 2.5.11 (věta o úplné pravděpodobnosti): *Nechť B_j, $j \in I$, je úplný systém jevů s nenulovou pravděpodobností. Pak pro každý jev A platí*

$$P(A) = \sum_{j \in I} P(B_j)\, P(A|B_j).$$

Důkaz: Existence sumy na pravé straně je zajištěna tím, že její členy jsou shora omezeny $P(B_j)$,

$$\sum_{j \in I} P(B_j)\, P(A|B_j) \leq \sum_{j \in I} P(B_j) = 1$$

(což mj. znamená, že množina I musí být spočetná). Použijeme σ-aditivitu pro jevy $B_j \cap A$, $j \in I$, které jsou po dvou neslučitelné:

$$P(A) = P\Big(\Big(\bigcup_{j \in I} B_j\Big) \cap A\Big) = P\Big(\bigcup_{j \in I} (B_j \cap A)\Big) = \sum_{j \in I} P(B_j \cap A) = \sum_{j \in I} P(B_j)\, P(A|B_j).$$

□

Věta o úplné pravděpodobnosti nám říká, že původní (nepodmíněná) pravděpodobnost $P(.)$ je konvexní kombinací podmíněných pravděpodobností $P(.|B_j)$; její koeficienty jsou $P(B_j)$, $j \in I$. Pokud víme, který z jevů B_j, $j \in I$, nastal, můžeme použít příslušnou podmíněnou pravděpodobnost. Dokud to nevíme a máme pro ně jen pravděpodobnostní popis, pravděpodobnost ostatních jevů počítáme jako nepodmíněnou. (Dostáváme jinou pravděpodobnost, protože vycházíme z jiné znalosti situace.)

Věta 2.5.12 (Bayesova věta): *Nechť B_i, $i \in I$, je úplný systém jevů s nenulovou pravděpodobností. Pak pro každý jev A splňující $P(A) \neq 0$ platí*

$$P(B_i|A) = \frac{P(B_i)\,P(A|B_i)}{\sum\limits_{j\in I} P(B_j)\,P(A|B_j)}\,.$$

Důkaz: S využitím věty o úplné pravděpodobnosti:

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)\,P(A|B_i)}{\sum\limits_{j\in I} P(B_j)\,P(A|B_j)}\,.$$

□

Význam Bayesovy nerovnosti spočívá v tom, že pravděpodobnosti $P(A|B_i)$ odhadneme z pokusů nebo z modelu, pomocí nich určíme pravděpodobnosti $P(B_i|A)$, které slouží k „optimálnímu" odhadu, který z jevů B_i nastal.

Pravděpodobnosti $P(B_i|A)$ nazýváme **aposteriorní pravděpodobnosti**, $P(B_i)$ **apriorní pravděpodobnosti**.

Problém 2.5.13: Ke stanovení aposteriorní pravděpodobnosti $P(B_i|A)$ potřebujeme znát i apriorní pravděpodobnost $P(B_i)$. Ta není vždy známa.

Příklad 2.5.14: * Na vstupu informačního kanálu mohou být znaky $1,\ldots,m$, výskyt znaku j označujeme jako jev B_j. Na výstupu mohou být znaky $1,\ldots,k$, výskyt znaku i označujeme jako jev A_i. (Obykle $k = m$, ale není to nutné.) Obvykle lze odhadnout podmíněné pravděpodobnosti $P(A_i|B_j)$ jevu, že znak j bude přijat jako i. Pokud známe apriorní pravděpodobnosti (vyslání znaku j) $P(B_j)$, můžeme pravděpodobnosti příjmu znaků vypočítat maticovým násobením:

$$\begin{bmatrix} P(A_1) & P(A_2) & \cdots & P(A_k) \end{bmatrix} =$$

$$= \begin{bmatrix} P(B_1) & P(B_2) & \cdots & P(B_k) \end{bmatrix} \cdot \begin{bmatrix} P(A_1|B_1) & P(A_2|B_1) & \cdots & P(A_k|B_1) \\ P(A_1|B_2) & P(A_2|B_2) & \cdots & P(A_k|B_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_1|B_m) & P(A_2|B_m) & \cdots & P(A_k|B_m) \end{bmatrix}.$$

Všechny matice v tomto vzorci mají jednotkové součty řádků (takové matice nazýváme **stochastické**). Pokud byl přijat znak i, pak pravděpodobnost, že byl vyslán znak B_j, je

$$P(B_j|A_i) = \frac{P(A_i|B_j)\;P(B_j)}{P(A_i)}\,.$$

Poznámka 2.5.15: Často se zavádí podmíněná pravděpodobnost tak, že z modelu odstraníme elementární jevy, které podmínce nevyhovují (nejsou z B). Tím přejdeme od původního pravděpodobnostního prostoru $(\Omega, \mathcal{A}, P)$ k pravděpodobnostnímu prostoru $(B, \mathcal{A} \cap \exp B, P(.|B))$, v němž B představuje množinu *všech* elementárních jevů. Podobně podmiňování jevem $\overline{B}$ vede na pravděpodobnostní prostor $\left(\overline{B}, \mathcal{A} \cap \exp \overline{B}, P\left(.|\overline{B}\right)\right)$. Původní pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ dostaneme pak jako *součet* pravděpodobnostních prostorů $(B, \mathcal{A} \cap \exp B, P(.|B))$ a $\left(\overline{B}, \mathcal{A} \cap \exp \overline{B}, P\left(.|\overline{B}\right)\right)$. Pravděpodobnost na něm je *směs* podmíněných pravděpodobností dle kapitoly 2.4. Vztah mezi oběma popisy je stejný jako mezi směsí a konvexní kombinací pravděpodobností dle pozn. 2.4.7.

Poznámka 2.5.16: Definice podmíněné pravděpodobnosti nedovoluje podmiňovat jevem s nulovou pravděpodobností. I takový pojem by však mohl mít smysl. Nechť $P(A) = 1/2$, $P(B) = 0$ (např. A je výsledek hodu mincí a B je jev, že při spojitém měření bylo výsledkem určité reálné

číslo, třeba $1/\pi$). Pak konjunkce $A \cap B$ má nulovou pravděpodobnost, ale rádi bychom sdělili, že je to jev „2× méně pravděpodobný" než B. To by vyjadřovala podmíněná pravděpodobnost $P(A|B) = 1/2$. Zde použitý přístup to nedovoluje, ale jiné modely to mohou umožnit.[1]

Podmíněná pravděpodobnost bývá někdy zaváděna jako základní pojem. I pro ni lze zavést axiomatický systém podobný tomu, kterým jsme definovali pravděpodobnost [Rényi 1966]. Jedním z motivů je, že v podstatě každý náhodný pokus předpokládá splnění určitých předpokladů a podmínek, které můžeme zahrnout do podmíněné pravděpodobnosti. Na „nepodmíněnou pravděpodobnost" lze navíc hledět jako na pravděpodobnost podmíněnou jevem jistým.

Podmíněná nezávislost

Náhodné jevy A, B jsou **podmíněně nezávislé** za podmínky C, jestliže

$$P(A \cap B|C) = P(A|C)\,P(B|C)\,.$$

Podobně definujeme podmíněnou nezávislost více jevů.

Podmíněná pravděpodobnost se může z matematického hlediska jevit jako nepotřebný pojem. Tak, jak je zde zavedena, je pouze jednou z mnoha možných pravděpodobností, vyznačuje se pouze podmínkami, za nichž jsme k ní došli. Nicméně je v praxi velmi důležitá pro rozhodování na základě neúplných údajů. Když o některých jevech víme, zda nastaly, a o jiných ještě ne, podmíněná pravděpodobnost dovoluje aktualizovat náš pravděpodobnostní model na podle těchto znalostí. Proto je východiskem pro rozhodování, a to lidské i automatické.

Příklad 2.5.17: * Test nemoci je u 1 % zdravých falešně pozitivní a u 10 % nemocných falešně negativní. Podíl nemocných v populaci je 0.001. Kolik procent ze všech lidí má pozitivní test? Jaká je pravděpodobnost, že pacient s pozitivním testem je nemocný?

Řešení: Označme jevy:
A: pozitivní test,
B: nemocný, $P(B) = 0.001$.
Z věty o úplné pravděpodobnosti

$$P(A) = P(B)\,P(A|B) + P(\bar{B})\,P(A|\bar{B}) = 0.001\ (1 - 0.1) + (1 - 0.001)\ 0.01 = 0.010\,89\,,$$

z Bayesovy věty

$$P(B|A) = \frac{P(B)\,P(A|B)}{P(A)} = \frac{0.001\ (1 - 0.1)}{0.010\,89} \doteq 0.0826\,.$$

□

Cvičení 2.5.18: Test nemoci je u 10 % zdravých falešně pozitivní a u 1 % nemocných falešně negativní. U náhodně vybraného vzorku byl pozitivní ve 20 %. Odhadněte výskyt nemoci v populaci. Jaká je pravděpodobnost, že pacient s pozitivním testem je nemocný?

Řešení: Označme jevy:
A: pozitivní test, $P(A) = 0.2$,
B: nemocný.
Z věty o úplné pravděpodobnosti

$$\begin{aligned} P(A) &= P(B)\,P(A|B) + P(\bar{B})\,P(A|\bar{B})\,,\\ 0.2 &= (1 - 0.01)\ P(B) + 0.1\ (1 - P(B)) = 0.89\,P(B) + 0.1\,,\\ P(B) &\doteq 0.112\,36\,, \end{aligned}$$

z Bayesovy věty

$$P(B|A) = \frac{P(B)\,P(A|B)}{P(A)} = \frac{0.112\,36\ (1 - 0.01)}{0.2} \doteq 0.556\,182\,.$$

□

[1] Viz např. Colleti, G., Scozzafava, R.: Probabilistic Logic in a Coherent Setting. Kluwer, Netherlands, 2002.

Cvičení 2.5.19: Mezi stoletými občany je 80 % žen. Mužů se rodí 52 %. Kolikrát měla žena narozená před 100 lety větší pravděpodobnost než muž narozený ve stejné době, že se dožije alespoň 100 let?

Řešení: Označme jevy:
Z: žena,
M: muž,
D: dožil(a) se 100 let.

$$\frac{P(Z|D)}{P(M|D)} = \frac{P(Z\cap D)}{P(M\cap D)} = 4\,, \qquad P(M) = 0.52\,, \qquad P(Z) = 0.48\,,$$

$$P(D|Z) = \frac{P(Z\cap D)}{P(Z)}\,, \qquad P(D|M) = \frac{P(M\cap D)}{P(M)}\,,$$

$$\frac{P(D|Z)}{P(D|M)} = \frac{P(Z\cap D)}{P(M\cap D)}\,\frac{P(M)}{P(Z)} = 4\,\frac{0.52}{0.48} \doteq 4.33\,.$$

(Nepotřebovali jsme neznámou pravděpodobnost, že se dožijí 100 let.) □

Cvičení 2.5.20: Diváci hlasováním vybírali ze tří možností, jimž dali následující počty hlasů:

možnost	A	B	C
četnost	20	30	50

Přitom každý s pravděpodobností $e = 10\,\%$ stiskl při hlasování nesprávné tlačítko (obě nesprávné možnosti jsou stejně pravděpodobné). Odhadněte rozdělení hlasů podle toho, jak diváci opravdu chtěli hlasovat.

Řešení: Tlačítka A, B, C byla stisknuta s pravděpodobnostmi 0.2, 0.3, 0.5. Pro pravděpodobnosti a, b, c *zamýšlené* volby A, B, C dostáváme soustavu lineárních rovnic

$$\begin{bmatrix}0.2 & 0.3 & 0.5\end{bmatrix} = \begin{bmatrix}a & b & c\end{bmatrix}\cdot\begin{bmatrix}0.9 & 0.05 & 0.05\\ 0.05 & 0.9 & 0.05\\ 0.05 & 0.05 & 0.9\end{bmatrix},$$

$$\begin{bmatrix}a & b & c\end{bmatrix} = \begin{bmatrix}0.2 & 0.3 & 0.5\end{bmatrix}\cdot\begin{bmatrix}0.9 & 0.05 & 0.05\\ 0.05 & 0.9 & 0.05\\ 0.05 & 0.05 & 0.9\end{bmatrix}^{-1} \doteq \begin{bmatrix}0.176\,47 & 0.294\,12 & 0.529\,41\end{bmatrix}.$$

□

Cvičení 2.5.21: Prezidentští kandidáti A, B dostali po řadě 49.9 % a 50.1 % odevzdaných hlasů. Někteří voliči se spletli; ti, kteří chtěli volit A, s pravděpodobností 0.002 zaškrtli B (který byl na hlasovacím lístku první), ti, kteří chtěli volit B, s pravděpodobností 0.001 zaškrtli A. Kterého z kandidátů si přálo více voličů (a kolik)?

Řešení: Pro pravděpodobnosti a, b zamýšlené volby A, B dostáváme soustavu lineárních rovnic

$$\begin{bmatrix}0.499 & 0.501\end{bmatrix} = \begin{bmatrix}a & b\end{bmatrix}\begin{bmatrix}0.998 & 0.002\\ 0.001 & 0.999\end{bmatrix},$$

$$\begin{bmatrix}a & b\end{bmatrix} = \begin{bmatrix}0.499 & 0.501\end{bmatrix}\begin{bmatrix}0.998 & 0.002\\ 0.001 & 0.999\end{bmatrix}^{-1} = \begin{bmatrix}0.499\,50 & 0.500\,5\end{bmatrix}.$$

Větší preference měl B. □

Cvičení 2.5.22: 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: žije
B: zažil ztroskotání.
Pak $P(A|B) = \frac{2}{3}$, $P(A|\bar{B}) = 1$, $P(B|A) = \frac{5}{60} = \frac{1}{12}$ (odhad). Z Bayesovy věty

$$P(B|A) = \frac{P(B)\,P(A|B)}{P(B)\,P(A|B) + P(\bar{B})\,P(A|\bar{B})}\,,$$

$$\frac{1}{12} = \frac{\frac{2}{3}P(B)}{\frac{2}{3}P(B) + (1 - P(B))}\,,$$

$$P(B) = \frac{3}{25} = 0.12\,.$$

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 1/3). □

Cvičení 2.5.23: V této hře je Bob bankéřem, kterému Alice zaplatí za účast ve hře. Bob umístí výhru 3 € pod jeden ze 3 kelímků, Alice hádá, pod který. Poté Bob otočí jiný kelímek a ukáže, že pod ním výhra není. Následně dá Alici ještě příležitost, aby změnila své rozhodnutí a vybrala si kterýkoli ze zbývajících 2 kelímků. Pokud je pod ním výhra, připadne Alici. Jaká je doporučená strategie obou hráčů a jaká je spravedlivá cena z účast ve hře?

Řešení: Jediné, co může Bob ovlivnit, je pravděpodobnost, s níž umístí výhru pod jednotlivé kelímky. Nejlépe pro něj bude, když tato pravděpodobnost bude pro všechny 3 kelímky stejná, tj. 1/3. Jinak by Alice mohla zvýšit své šance, kdyby se o této pravděpodobnosti dověděla (např. opakováním hry).

Alice nejprve vybírá ze 3 kelímků, optimální bude, když bude volit s rovnoměrnou pravděpodobností 1/3 pro každý kelímek (jinak by toho mohl Bob využít ke snížení jejích šancí). Pokud jeden z hráčů volí pravděpodobnost 1/3, na strategii druhého nezáleží.

Ve druhé části hry má Alice na vybranou. Může své rozhodnutí ponechat, změnit, nebo může použít i obecnější *smíšenou strategii*, kdy své rozhodnutí změní s pravděpodobností $c \in \langle 0, 1\rangle$. Označme jevy:
H: Alice napoprvé ukáže správný kelímek,
Z: Alice změní své rozhodnutí,
V: Alice vyhraje.
Pokud své rozhodnutí nezmění, její šance na výhru je $P(V|\bar{Z}) = P(H) = 1/3$. Pokud jej změní, je potřeba rozlišit 2 případy. Jestliže její první volba byla správná (což nastalo s pravděpodobností $P(H) = 1/3$), pak změnou rozhodnutí nic nevyhraje, $P(V|Z \cap H) = 0$. Pokud její první volba byla špatná (což nastalo s pravděpodobností $P(\bar{H}) = 2/3$), pak jí Bob otočením kelímku napověděl; výhra je pod zbývajícím kelímkem, tj. změna rozhodnutí vede k jisté výhře, $P(V|Z \cap \bar{H}) = 1$. Pro úplný soubor jevů $\bar{Z}, Z \cap H, Z \cap \bar{H}$ s pravděpodobnostmi po řadě $1 - c$, $c/3$, $2c/3$ (neboť jevy H, Z jsou nezávislé) dostáváme podle vzorce pro úplnou pravděpodobnost

$$P(V) = P(\bar{Z})\,P(V|\bar{Z}) + P(Z \cap H)\,P(V|Z \cap H) + P(Z \cap \bar{H})\,P(V|Z \cap \bar{H}) =$$
$$= (1 - c)\,\frac{1}{3} + \frac{c}{3}\,0 + \frac{2c}{3}\,1 = \frac{1}{3}\,c + \frac{1}{3}\,.$$

Tedy pro Alici je výhodné rozhodnutí vždy změnit, $c = 1$, šance na výhru tím zvýší na 2/3. Spravedlivá cena za hru je $\frac{2}{3}\,3 = 2\,€$. □

Cvičení 2.5.24: V domě vidíme $m = 6$ oken, která bývají v tuto denní dobu rozsvícena nezávisle s pravděpodobností $a_1 = 1/3$, pokud ovšem jde proud, což je splněno s pravděpodobností $b = 0.999$. Jestliže se v žádném okně nesvítí, jaká je pravděpodobnost, že v domě nejde proud? Jak se tato pravděpodobnost změní, pokud vidíme 20 oken?

Řešení: Označme jevy:
A: nesvítí žádné okno,
B: jde proud.
Pokud jde proud, pak podmíněná pravděpodobnost jevu A je

$$P(A|B) = (1-a_1)^m = \left(\tfrac{2}{3}\right)^6 \doteq 8.779 \cdot 10^{-2}\,.$$

Pokud nejde proud, pak $P(A|\bar{B}) = 1$. Ze vzorce pro úplnou pravděpodobnost

$$P(A) = P(B)\,P(A|B) + P(\bar{B})\,P(A|\bar{B}) = b\,(1-a_1)^m + (1-b) \doteq 8.870 \cdot 10^{-2}$$

a z Bayesova vzorce

$$P(\bar{B}|A) = \frac{P(\bar{B})\,P(A|\bar{B})}{P(A)} = \frac{1-b}{b\,(1-a_1)^m + (1-b)} \doteq 1.127 \cdot 10^{-2}\,.$$

Pro $m = 20$ dostáváme

$$P(A|B) \doteq 3.007 \cdot 10^{-4}\,,$$
$$P(A) \doteq 1.300 \cdot 10^{-3}\,,$$
$$P(\bar{B}|A) \doteq 0.769\,.$$

□

Cvičení 2.5.25: Na vstupu informačního kanálu mohou být znaky 0, 1, na výstupu jsou přečteny s nezávislou pravděpodobností chyby 0.1. Určete podmíněné pravděpodobnosti vstupu při známém výstupu, je-li apriorní pravděpodobnost jedničky (a) 0.4, (b) 0.1, (c) 0.05.

Řešení: Označme jevy:
B_0, B_1: vyslán znak 0, resp. 1,
A_0, A_1: přijat znak 0, resp. 1.

(a)

$$\begin{bmatrix} P(A_0) & P(A_1) \end{bmatrix} = \begin{bmatrix} P(B_0) & P(B_1) \end{bmatrix} \cdot \begin{bmatrix} P(A_0|B_0) & P(A_1|B_0) \\ P(A_0|B_1) & P(A_1|B_1) \end{bmatrix} =$$
$$= \begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.58 & 0.42 \end{bmatrix},$$

$$P(B_0|A_0) = \frac{P(A_0|B_0)\,P(B_0)}{P(A_0)} = \frac{0.9 \cdot 0.6}{0.58} \doteq 0.931\,03\,,$$
$$P(B_1|A_0) = \frac{P(A_0|B_1)\,P(B_1)}{P(A_0)} = \frac{0.1 \cdot 0.4}{0.58} \doteq 6.896\,6 \cdot 10^{-2}\,,$$
$$P(B_0|A_1) = \frac{P(A_1|B_0)\,P(B_0)}{P(A_1)} = \frac{0.1 \cdot 0.6}{0.42} \doteq 0.142\,86\,,$$
$$P(B_1|A_1) = \frac{P(A_1|B_1)\,P(B_1)}{P(A_1)} = \frac{0.9 \cdot 0.4}{0.42} \doteq 0.857\,14\,.$$

(b)

$$\begin{bmatrix} P(A_0) & P(A_1) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.82 & 0.18 \end{bmatrix},$$

$$\begin{aligned} P(B_0|A_0) &= \frac{P(A_0|B_0)\,P(B_0)}{P(A_0)} = \frac{0.9 \cdot 0.9}{0.82} = 0.9878, \\ P(B_1|A_0) &= \frac{P(A_0|B_1)\,P(B_1)}{P(A_0)} = \frac{0.1 \cdot 0.1}{0.82} = 1.2195 \cdot 10^{-2}, \\ P(B_0|A_1) &= \frac{P(A_1|B_0)\,P(B_0)}{P(A_1)} = \frac{0.1 \cdot 0.9}{0.18} = 0.5, \\ P(B_1|A_1) &= \frac{P(A_1|B_1)\,P(B_1)}{P(A_1)} = \frac{0.9 \cdot 0.1}{0.18} = 0.5. \end{aligned}$$

(c)

$$\begin{bmatrix} P(A_0) & P(A_1) \end{bmatrix} = \begin{bmatrix} 0.95 & 0.05 \end{bmatrix} \cdot \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.86 & 0.14 \end{bmatrix},$$

$$\begin{aligned} P(B_0|A_0) &= \frac{P(A_0|B_0)\,P(B_0)}{P(A_0)} = \frac{0.9 \cdot 0.95}{0.86} \doteq 0.99419, \\ P(B_1|A_0) &= \frac{P(A_0|B_1)\,P(B_1)}{P(A_0)} = \frac{0.1 \cdot 0.05}{0.86} \doteq 5.8140 \cdot 10^{-3}, \\ P(B_0|A_1) &= \frac{P(A_1|B_0)\,P(B_0)}{P(A_1)} = \frac{0.1 \cdot 0.95}{0.14} \doteq 0.67857, \\ P(B_1|A_1) &= \frac{P(A_1|B_1)\,P(B_1)}{P(A_1)} = \frac{0.9 \cdot 0.05}{0.14} \doteq 0.32143. \end{aligned}$$

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 2.5.26: Opisujeme data. Při čtení se z chybných položek ve vstupních datech odhalí a opraví 80 %. Při následném zápisu se do 1 % položek dostanou nové chyby. Při jakém výskytu chyb se opsáním jejich počet v průměru sníží?

Řešení: Apriorní pravděpodobnost chybné položky označme c, pravděpodobnost správné položky $1-c$. Změnu pravděpodobností čtením a zápisem kvantifikuje násobení maticemi podmíněných pravděpodobností (předpokládáme, že chybná položka se chybou při zápisu nemůže opravit):

$$\begin{bmatrix} c & 1-c \end{bmatrix} \cdot \begin{bmatrix} 0.2 & 0.8 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0.01 & 0.99 \end{bmatrix} = \begin{bmatrix} c & 1-c \end{bmatrix} \cdot \begin{bmatrix} 0.208 & 0.792 \\ 0.01 & 0.99 \end{bmatrix} =$$
$$= \begin{bmatrix} 0.198\,c + 0.01 & -0.198\,c + 0.99 \end{bmatrix}.$$

Pravděpodobnost (nikoli nutně četnost!) chybných položek se sníží, pokud $0.198\,c + 0.01 < c$, tj. přibližně $c > 0.0125 = 1.25\,\%$. □

Cvičení 2.5.27: Fotbalisté B, C, D promění penaltu s pravděpodobnostmi po řadě $b = 0.9$, $c = 0.8$, $d = 0.6$. Určili střelce losem (všichni měli stejnou pravděpodobnost). Jaká je pravděpodobnost proměnění penalty? Pokud víme, že se penalta nezdařila, jaká je pro jednotlivé hráče pravděpodobnost, že ji kopali?

Řešení: Pravděpodobnost úspěchu je $(b+c+d)/3 \doteq 0.767$, neúspěchu $q \doteq 1-0.767 = 0.233$. Podmíněné pravděpodobosti střelců neúspěšného pokusu vypočítáme z Bayesova vzorce:

$$\frac{(1-b)/3}{q} \doteq 0.143\,, \qquad \frac{(1-c)/3}{q} \doteq 0.286\,, \qquad \frac{(1-d)/3}{q} \doteq 0.572\,.$$

□

Cvičení 2.5.28: Mezi velkým množstvím hracích kostek je 1/2 falešných, na nichž padá šestka s pravděpodobností 1/3. Hodili jsme dvěma náhodně vybranými kostkami, padly dvě šestky. Jaká je pravděpodobnost, že obě kostky jsou falešné?

Řešení: Označme jevy:
F: falešná kostka,
S: padla šestka.
Je dáno: $P(S|F) = 1/3$, $P\left(S|\bar{F}\right) = 1/6$, $P(F) = 1/2$. Pravděpodobnost, že jedna kostka, na níž padla šestka, je falešná, je

$$P(F|S) = \frac{P(S|F)\ P(F)}{P(S|F)\ P(F) + P\left(S|\bar{F}\right)\ P\left(\bar{F}\right)} = \frac{P(S|F)}{P(S|F) - P\left(S|\bar{F}\right) + P\left(S|\bar{F}\right)/P\left(\bar{F}\right)} = \frac{2}{3}\,.$$

Pravděpodobnost, že nastanou dva nezávislé jevy s touto pravděpodobností, je

$$(P(F|S))^2 = \frac{4}{9}\,.$$

□

Cvičení 2.5.29: Opilí řidiči způsobili 10 % nehod. Při kontrolách bylo 1.5 % řidičů opilých. Porovnejte rizika havárie opilého a střízlivého řidiče.

Řešení: Označme jevy:
A = opilý,
H = havárie.
Je dáno: $P(A|H) = 0.1\,P(H)$, $P(A) = 0.015$. Pravděpodobnost havárie není uvedena, ale lze vypočítat, kolikrát větší je riziko havárie opilého řidiče než střízlivého:

$$P(H|A) = \frac{P(A|H)\ P(H)}{P(A)}\,,$$

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

$$\frac{P(H|A)}{P\left(H|\bar{A}\right)} = \frac{P(A|H)\ P\left(\bar{A}\right)}{P\left(\bar{A}|H\right)\ P(A)} = \frac{P(A|H)\ (1-P(A))}{(1-P(A|H))\ P(A)} = \frac{0.1\cdot 0.985}{0.9\cdot 0.015} \doteq 7.3\,.$$

□

Cvičení 2.5.30: [Hsu 1996] Dokažte implikace

$$P(A|B) > P(A) \implies P(B|A) > P(B)\,,$$
$$P(A) > P(B) \implies P(A|B) > P(B|A)\,.$$

3. Náhodné veličiny

Nyní zavedeme náhodné veličiny na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$, tj. budeme předpokládat Kolmogorovův model pravděpodobnosti. Čtenář dostatečně neobeznámený s tímto přístupem by však měl vystačit s tím, že Ω je množina elementárních jevů (=všech možných výsledků náhodného pokusu) a $\mathcal{A}$ je množina (přesněji σ-algebra) jevů, tj. *některých* podmnožin množiny Ω. Funkce $P\colon \mathcal{A} \to \langle 0,1\rangle$ je pravděpodobnost (=pravděpodobnostní míra) s vlastnostmi z kapitoly 1.4. Budeme především potřebovat speciální případ, kdy $\Omega = \mathbb{R}$ a $\mathcal{A}$ obsahuje všechna spočetná sjednocení intervalů (a tím i další množiny, společně nazývané *borelovské*).

Obrázky k pojmům probíraným v této kapitole jsou v dodatku B.

3.1 Definice náhodné veličiny

Příklad 3.1.1: Vojáci nově přijatí do armády přicházejí do skladu pro boty a oděv. Můžeme se pro každou velikost ptát, jak je pravděpodobné, že ji (náhodně vybraný) voják bude potřebovat. Adekvátnější popis bude, že velikosti (bot nebo oděvu) jsou funkce závislé na výběru vojáka, tedy náhodné veličiny, které mohou nabývat různých hodnot s různou pravděpodobností.[1]

Příklad 3.1.2: V řešení příkladu 2.5.6 na str. 57 jsme k popisu používali jev A_a, že Alice má a es, kde a probíhá množinu všech možných výsledků $\Omega = \{0, \ldots, 4\}$. Jevy A_a, $a \in \Omega$, tvoří úplný systém jevů a pravděpodobnosti $P(A_a)$, $a \in \Omega$, plně určují pravděpodobnost všech možných otázek na to, kolik má Alice es, např. „jaká je pravděpodobnost, že Alice má aspoň dvě esa." Nabízí se přirozená myšlenka reprezentovat tyto pravděpodobnosti jako funkci $p_A\colon \Omega \to \langle 0,1\rangle$, $p_A(a) = P(A_a)$. To bude obzvlášť výhodné, bude-li možných výsledků mnoho.

Příklad 3.1.3: *Ředitel podniku chtěl vědět, kolik procent obyvatel má nejvýše tak velký plat* jako on. Zadal statistický výzkum. Poté ho zajímalo, kolik procent obyvatel má nejvýše tak velký plat jako jeho náměstek. Zadal další výzkum atd. To je samozřejmě fiktivní situace, takový ředitel by ve funkci brzy skončil jednak proto, že se zabývá nepodstatnými otázkami, jednak proto, že to dělá neefektivně. Místo abychom u respondentů vybírali ze dvou možností, zda mají nejvýše takový plat jako zadaná mez, můžeme se (samozřejmě anonymně) ptát přímo na výši platu. Pak máme z jednoho průzkumu odpověď na všechny další podobné otázky. Abychom získaná data účelně reprezentovali, popíšeme pomocí nich *rozdělení* platů ve zkoumané populaci. K tomu potřebujeme chápat výši platu jako náhodnou veličinu. (V tomto případě náhodnost spočívá ve výběru respondentů pro průzkum.)

Většinou na pravděpodobnostním prostoru měříme veličiny s výsledkem, který lze vyjádřit reálným číslem. Ten může být při různých pokusech odlišný, proto *před provedením pokusu* máme jen pravděpodobnostní popis možných výsledků. Nástrojem k jeho vyjádření je pojem náhodné veličiny.

Definice 3.1.4: ***Náhodná veličina*** *na σ-algebře $\mathcal{A}$ podmnožin množiny Ω je zobrazení $X\colon \Omega \to \mathbb{R}$ takové, že pro všechny intervaly $I \subseteq \mathbb{R}$ množina $X^{-1}(I) = \{\omega \in \Omega \mid X(\omega) \in I\}$ patří do $\mathcal{A}$. Jestliže* po provedení náhodného pokusu *je jeho výsledek určen elementárním jevem $\omega \in \Omega$, nazýváme číslo $X(\omega) \in \mathbb{R}$* ***realizace náhodné veličiny*** *X.*

[1] Tyto náhodné veličiny byly poprvé studovány (statistickými metodami) a použity ve válce Severu proti Jihu.

Poznámka 3.1.5: Termín *náhodná veličina* je historický a zavádějící, z matematického hlediska je to *funkce*. Teprve *realizace náhodné veličiny* je *funkční hodnota*, tj. číslo dané výsledkem náhodného pokusu. Je to stejný rozdíl jako mezi funkcí $\exp\colon \mathbb{R} \to \mathbb{R}$ a funkční hodnotou $\exp(0)$, což je číslo 1.

Poznámka 3.1.6: Náhodná veličina je reálná funkce X na prostoru všech elementárních jevů, splňující navíc $X^{-1}(I) \in \mathcal{A}$ pro všechny intervaly I (a tedy i pro spočetná sjednocení intervalů a některé další množiny). Taková funkce se nazývá **měřitelná**; můžeme testovat (měřit) jev $[X \in I]$, jemuž odpovídá množina elementárních jevů $X^{-1}(I)$. V praxi s podmínkou měřitelnosti nebývají problémy; pokud by byla porušena, brzy bychom poznali, že nemůžeme postupovat dál, protože by nebyla definována pravděpodobnost jevu $[X \in I]$.

Příklad 3.1.7: Výsledkem jednoho hodu pravidelnou kostkou je náhodná veličina X, které v pravděpodobnostním modelu dle příkladu 1.2.1 odpovídá identická funkce na $\Omega = \{1, \ldots, 6\}$, $X(\omega) = \omega$. To je tím, že jsme elementární jevy přímo označili odpovídajícími výsledky.

Na tomto modelu však můžeme definovat i jiné náhodné veličiny. Např. náhodná veličina Y,

$$Y(\omega) = \begin{cases} 1 & \text{pro } \omega = 6, \\ 0 & \text{jinak,} \end{cases}$$

udává, zda padla šestka či nikoli (a to ve formě čísel 0, 1, ale mohli jsme zvolit i jakákoli jiná reálná čísla k reprezentaci výsledku). Podobně náhodná veličina Z,

$$Z(\omega) = \begin{cases} 1 & \text{pro } \omega \in \{2, 4, 6\}, \\ 0 & \text{jinak,} \end{cases}$$

udává, zda padlo sudé číslo.

Příklad 3.1.8: Pokud v předchozím příkladu pozměníme pravděpodobnostní model stejně jako v příkladu 1.7.13 (jsme schopni rozlišit jen sudé a liché výsledky), pak funkce X, Y přestávají být náhodnými veličinami, neboť nejsou měřitelné (nejsme schopni určit jejich hodnotu), např. $X^{-1}(\{6\}) = Y^{-1}(\{6\}) \notin \mathcal{B}$. Naproti tomu funkce Z je měřitelná i v tomto modelu, a je to tedy náhodná veličina.

Příklad 3.1.9: Pokud výsledkem hodu kostkou je číslo na horní stěně, ale my vidíme jen boční stěnu jako v příkladu 1.7.26, pak nemůžeme výsledek jednoznačně určit (i když víme, které číslo nepadlo) a nemůžeme pokus popsat náhodnou veličinou. Problém je opět v měřitelnosti.

Na náhodné veličiny přenášíme známé pojmy definované pro funkce. O náhodných veličinách X, Y řekneme, že $X = Y$, resp. $X \leq Y$, právě když jsou definovány na *stejné* σ-algebře a $X(\omega) = Y(\omega)$, resp. $X(\omega) \leq Y(\omega)$ pro všechna ω.

Náhodná veličina je zobecněním reálného čísla. Není to *číslo*, neboť závisí na proměnných (které určují náhodný výsledek a které někdy ani dobře neznáme). V matematickém popisu je to *funkce*, jejímž argumentem je elementární jev plně určující výsledek náhodného pokusu. Teprve provedením pokusu dostaneme číselnou hodnotu, *realizaci náhodné veličiny*.

3.2 Rozdělení pravděpodobnosti

Z defnice náhodné veličiny X je pro libovolný interval I definována pravděpodobnost jevu $\{\omega \in \Omega \mid X(\omega) \in I\}$, pro kterou zavedeme následující stručnější značení:

$$P[X \in I] = P(\{\omega \in \Omega \mid X(\omega) \in I\}).$$

Podobně budeme (v hranatých závorkách) používat i jiné podobné vztahy, např.

$$P[X \leq u] = P(\{\omega \in \Omega \mid X(\omega) \leq u\}) ,$$

a jejich konjunkce budeme pro přehlednější zápis značit čárkami, např.

$$P[X > u,\ X \leq v] = P(\{\omega \in \Omega \mid X(\omega) > u,\ X(\omega) \leq v\}) ,$$

přičemž poslední výraz můžeme zapisovat i zjednodušeně jako $P[u < X \leq v]$. Zápis $P[X \in I]$ přirozeně zobecníme z intervalů na sjednocení spočetně mnoha intervalů a na libovolnou borelovskou množinu $I \in \mathcal{B}(\mathbb{R})$ a zavedeme pro něj následující značení.

Definice 3.2.1: ***Rozdělení pravděpodobnosti*** *náhodné veličiny X (stručněji* ***rozdělení*** *náhodné veličiny, též* ***rozložení pravděpodobnosti*** *náhodné veličiny) je funkce $P_X : \mathcal{B}(\mathbb{R}) \to \langle 0, 1\rangle$ definovaná vztahem*

$$P_X(I) = P[X \in I] = P(\{\omega \in \Omega \mid X(\omega) \in I\}) .$$

Tvrzení 3.2.2: *Rozdělení pravděpodobnosti náhodné veličiny X je pravděpodobnostní míra na Borelově σ-algebře $\mathcal{B}(\mathbb{R})$.*

Poznámka 3.2.3: Zatímco původní pravděpodobnost P je definována na nějaké jiné σ-algebře (z pravděpodobnostního prostoru, na němž je definována náhodná veličina X), rozdělení pravděpodobnosti P_X je vždy pravděpodobnostní míra na Borelově σ-algebře $\mathcal{B}(\mathbb{R})$. To nám dovoluje sjednotit práci s náhodnými veličinami bez ohledu na to, na jakém prostoru jsou definovány. Můžeme např. počítat střední hodnotu a další kvantitativní charakteristiky náhodné veličiny. Tento pohled budeme nadále uplatňovat v maximální míře, takže čtenář může pokračovat i bez dobrého pochopení pojmů *σ-algebra* a *pravděpodobnostní prostor*; stačí, když se naučí pracovat s pravděpodobnostmi na Borelově σ-algebře, na kterou lze pohlížet jako na rozšíření množiny všech spočetných sjednocení intervalů. Význam všech těchto pojmů je však klíčový pro teoretický základ dále vykládaných postupů.

Opatrnost je potřeba, jakmile budeme mít více náhodných veličin. Pokud i pak chceme odhlédnout od původního pravděpodobnostního prostoru, budeme potřebovat jejich *sdružené rozdělení pravděpodobnosti*, které zavedeme v kapitole 5.

Poznámka 3.2.4: Zajímáme se víc o rozdělení než o pravděpodobnostní model, protože často ani nevíme, co vše náhodné výsledky způsobuje. I přesto například normální rozdělení dobře popisuje mnohé zdroje chyb při měření, aniž bychom věděli, čím jsou způsobeny.

Poznámka 3.2.5: Popis náhodné veličiny pomocí rozdělení pravděpodobnosti je užitečný pro sjednocení mnoha postupů, ale ztrácíme při něm informaci. Náhodná veličina má jednoznačně určené rozdělení pravděpodobnosti, ale z rozdělení pravděpodobnosti nelze zpětně určit náhodnou veličinu, a to ani se znalostí pravděpodobnostního prostoru, na němž je definována. Je třeba určit přiřazení elementárních jevů k hodnotám náhodné veličiny. Z rozdělení pravděpodobnosti nelze ani zcela určit obor hodnot náhodné veličiny. Je to nevyhnutelný důsledek rozdílu mezi jevem nemožným a jevem s nulovou pravděpodobností.

Příklad 3.2.6: Chceme-li uspořádat pokus s výsledky 0, 1, které mají oba pravděpodobnost 1/2, můžeme hodit mincí, hodit kostkou a rozlišit pouze sudé a liché výsledky (viz příklad 3.1.7) apod. Stejné rozdělení dostaneme, ať bude 0 znamenat sudé výsledky a 1 liché, nebo naopak.

Příklad 3.2.7: Má-li jedna náhodná veličina rovnoměrné rozdělení na uzavřeném intervalu $\langle 0, 1\rangle$ a druhá na otevřeném intervalu $(0, 1)$, mají obě stejná rozdělení, neboť jejich výsledky se liší s nulovou pravděpodobností (s níž u první může vyjít hodnota z $\{0, 1\}$, která je u druhé nemožná).

Příklad 3.2.8: V nekonečné ruletě je výsledkem náhodného pokusu bod na kružnici. (Předpokládáme rovnoměrné rozdělení; pravděpodobnost, že výsledek je na daném oblouku, je úměrná délce oblouku.) V případě, že Bob uhodne tento bod, získá milión, v opačném případě přijde

o jednotkový vklad do hry. Poté, co si Bob vsadí, je jeho zisk popsán náhodnou veličinou X, která nabývá hodnot 1 000 000 (v bodě, na který vsadil) a -1 (ve všech ostatních bodech kružnice). Protože pravděpodobnost výhry je nulová, je $P_X(\{-1\}) = P[X = -1] = 1$ a pro libovolný interval I

$$P_X(I) = \begin{cases} 1 & \text{pro } -1 \in I, \\ 0 & \text{pro } -1 \notin I. \end{cases}$$

Stejné rozdělení pravděpodobnosti bychom dostali, kdyby se vůbec nelosovalo a Bob vždy odevzdal vklad bez šance na výhru. Ztratili jsme informaci o tom, že jev $X = 1\,000\,000$ je možný, byť s nulovou pravděpodobností.

Tento problém nás bude provázet i nadále. Zjednodušení práce s rozdělením pravděpodobnosti místo s náhodnou veličinou nám to snad vynahradí.

Poznámka 3.2.9: K tomu, aby stačila znalost rozdělení pravděpodobnosti P_X na intervalech, se navíc potřebujeme omezit na tzv. *perfektní míry*; s jinými se v praxi nesetkáme.

Popsali jsme rozdělení náhodné veličiny pomocí pravděpodobnosti (=pravděpodobnostní míry) definované na Borelově σ-algebře (mj. na všech spočetných sjednoceních intervalů). To je univerzální přístup, z něhož odvodíme úspornější reprezentace. Popisem náhodné veličiny pomocí jejího rozdělení získáváme výhodu v tom, že můžeme abstrahovat od původního pravděpodobnostního prostoru a dále používat jednotné postupy na něm nezávislé. Na druhé straně tímto popisem ztrácíme informaci, zejména o tom, které jevy s nulovou pravděpodobností jsou možné; to už se z rozdělení pravděpodobnosti nedovíme. Naštěstí jevy s nulovou pravděpodobností nemají vliv na mnohé užitečné závěry.

3.3 Distribuční funkce

Úspornější reprezentaci rozdělení pravděpodobnosti dostaneme, pokud se omezíme na intervaly, a navíc zafixujeme jeden krajní bod, např. $-\infty$. Stačí nám znát rozdělení pravděpodobnosti na intervalech tvaru $(-\infty, u\rangle$ pro všechna $u \in \mathbb{R}$,

Definice 3.3.1: ***Distribuční funkce*** *náhodné veličiny X je funkce $F_X : \mathbb{R} \to \langle 0, 1\rangle$ definovaná vztahem*

$$F_X(u) = P_X\left((-\infty, u\rangle\right) = P[X \leq u].$$

Příklad 3.3.2: V příkladu 3.1.3 se ředitel ptal na jednotlivé hodnoty distribuční funkce.

Všechny intervaly lze vyjádřit pomocí intervalů tvaru $(-\infty, u\rangle$, neboť

$$\begin{aligned} (u, v\rangle &= (-\infty, v\rangle \setminus (-\infty, u\rangle, \\ (-\infty, u) &= \bigcup_{n\in\mathbb{N}} (-\infty, u - \frac{1}{n}\rangle, \\ (-\infty, \infty) &= \bigcup_{n\in\mathbb{N}} (-\infty, n\rangle, \\ (u, \infty) &= \mathbb{R} \setminus (-\infty, u\rangle, \\ \{u\} &= (-\infty, u\rangle \setminus (-\infty, u). \end{aligned} \tag{3.1}$$

Odtud už snadno dostaneme i otevřené a uzavřené intervaly. Díky tomu lze z distribuční funkce určit pro libovolný interval pravděpodobnost, že do něj náhodná veličina padne:

$$
\begin{aligned}
P_X((u,v\rangle) &= P[u < X \leq v] = F_X(v) - F_X(u)\,, \\
P_X((-\infty,u)) &= P[X < u] = \lim_{v \to u-} F_X(v) = F_X(u-)\,, \\
P_X((u,\infty)) &= P[X > u] = 1 - F_X(u)\,, \\
P_X(\{u\}) &= P[X = u] = F_X(u) - F_X(u-)\,.
\end{aligned} \tag{3.2}
$$

Definice 3.3.3: *Náhodná veličina X se nazývá* **omezená**, *jestliže existuje omezený interval I, pro který $P[X \in I] = 1$.*

Vlastnosti distribuční funkce shrnuje následující věta:

Věta 3.3.4: *Konjunkce následujících podmínek je nutná a postačující pro to, aby funkce $F\colon \mathbb{R} \to \langle 0,1\rangle$ byla distribuční funkcí nějaké náhodné veličiny:*

1. *F je neklesající.*
2. *F je zprava spojitá.*
3. $\lim_{u \to -\infty} F(u) = 0$, $\lim_{u \to \infty} F(u) = 1$.

Důkaz: Nejprve ověříme nutnost podmínek pro distribuční funkci F_X náhodné veličiny X:

1. Je-li $u \leq v$, pak

$$F_X(v) - F_X(u) = P[X \leq v] - P[X \leq u] = P[u < X \leq v] \geq 0\,.$$

2. Spojitost stačí ověřit pro posloupnost reálných čísel $(u_n)_{n\in\mathbb{N}}$ klesající k u. Podle tvrzení 1.7.10

$$\lim_{n\to\infty} F_X(u_n) = \lim_{n\to\infty} P_X((-\infty,u_n\rangle) = P_X\Big(\bigcap_{n\in\mathbb{N}} (-\infty,u_n\rangle\Big) = P_X((-\infty,u\rangle) = F_X(u)\,.$$

3. Nechť monotónní posloupnost reálných čísel $(u_n)_{n\in\mathbb{N}}$, resp. $(v_n)_{n\in\mathbb{N}}$, klesá do $-\infty$, resp. roste do $+\infty$. Podle tvrzení 1.7.10

$$\lim_{n\to\infty} F_X(u_n) = \lim_{n\to\infty} P_X((-\infty,u_n\rangle) = P_X\Big(\bigcap_{n\in\mathbb{N}} (-\infty,u_n\rangle\Big) = P_X(\emptyset) = 0\,,$$

$$\lim_{n\to\infty} F_X(v_n) = \lim_{n\to\infty} P_X((-\infty,v_n\rangle) = P_X\Big(\bigcup_{n\in\mathbb{N}} (-\infty,v_n\rangle\Big) = P_X(\mathbb{R}) = 1\,.$$

Abychom ověřili, že podmínky jsou postačující, musíme k funkci F, která je splňuje, najít pravděpodobnostní prostor $(\Omega,\mathcal{A},P)$ a na něm náhodnou veličinu X, pro kterou $F_X = F$. Stačí zvolit Borelovu σ-algebru, tj. $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ a na ní pravděpodobnost P určenou podmínkami $P((-\infty,u\rangle) = F(u)$ (těm, vzhledem ke konstrukci Borelovy σ-algebry, vyhovuje právě jedna pravděpodobnost, která je na intervalech dána vzorci (3.2)). Za náhodnou veličinu lze zvolit identickou funkci $X\colon u \mapsto u$. □

Poznámka 3.3.5: Rozdělení pravděpodobnosti P_X jsme mohli popsat hodnotami na intervalech tvaru $I = (-\infty,u)$, $u \in \mathbb{R}$, pravděpodobnostmi $P[X < u] = P_X((-\infty,u))$. Často se v literatuře (např. [Zvára, Štěpán 2002]) setkáme s takto zavedenou *zleva* spojitou distribuční funkcí. V bodech spojitosti distribuční funkce se obě definice shodují.

Předchozí věta nám mj. umožňuje definovat rozdělení pravděpodobnosti náhodné veličiny, aniž bychom specifikovali příslušný pravděpodobnostní prostor. To nám zjednoduší mnohé úvahy. Distribuční funkci budeme indexovat buď příslušnou náhodnou veličinou, nebo odpovídajícím typem rozdělení. Výjimku činíme u distribuční funkce normovaného normálního rozdělení (danou Gaussovým integrálem), pro kterou budeme používat i speciální značení Φ.

Příklad 3.3.6: * Náhodná veličina může být konstantní, takže nabývá na celém definičním oboru Ω stejnou hodnotu $r \in \mathbb{R}$. To znamená, že „není náhodná", neboť dává vždy stejný výsledek. Můžeme ji ztotožnit s reálným číslem r a také ji takto budeme značit. Její rozdělení je *Diracovo* a její distribuční funkce je posunutá Heavisideova funkce:

$$P_r(I) = \begin{cases} 0 & \text{pro } r \notin I, \\ 1 & \text{pro } r \in I, \end{cases} \qquad F_r(u) = \begin{cases} 0 & \text{pro } u < r, \\ 1 & \text{pro } u \geq r. \end{cases}$$

Tímto způsobem jsou všechna reálná čísla chápana jako speciální případy náhodných veličin, a tedy náhodné veličiny jsou zobecněním reálných čísel.

Tvrzení 3.3.7: *Distribuční funkce náhodné veličiny má spočetně mnoho bodů nespojitosti.*

Důkaz: V každém bodě nespojitosti v se distribuční funkce F_X mění skokem o kladnou hodnotu $F_X(v+) - F_X(v-)$. Součet všech těchto rozdílů je konečný, nejvýše

$$1 = \lim_{u \to \infty} F(u) - \lim_{u \to -\infty} F(u),$$

takže sčítanců musí být spočetně mnoho. □

Distribuční funkce je jedním ze základních popisů rozdělení náhodné veličiny. Později uvedeme jiné popisy, které jsou sice názornější, ale nejsou univerzálně použitelné.

Cvičení 3.3.8: [Zvára, Štěpán 2002, Příklad 6.5] Hodíme třemi kostkami; náhodná veličina X je počet dvojic kostek, na nichž padl stejný výsledek (nejvýše 3). Určete její rozdělení.

Řešení:

počet	pravděpodobnost obecně	pravděpodobnost číselně
0	$\frac{\binom{6}{3}\,3!}{6^3}$	$\frac{5}{9}$
1	$\frac{\binom{6}{2}\,3!}{6^3}$	$\frac{5}{12}$
2	0	0
3	$\frac{6}{6^3}$	$\frac{1}{36}$

Cvičení 3.3.9: Pokračujme v cvičení 1.5.26; jaké rozdělení má počet zásahů?

Řešení:

počet	pravděpodobnost obecně	pravd. číselně
0	$(1-a)(1-b)(1-c)$	0.016
1	$a(1-b)(1-c) + (1-a)b(1-c) + (1-a)(1-b)c$	0.212
2	$ab(1-c) + a(1-b)c + (1-a)bc$	0.628
3	abc	0.144

Cvičení 3.3.10: * Je nějaký vztah mezi nerovností náhodných veličin a nerovností jejich distribučních funkcí?

Řešení: Je-li $X \leq Y$, pak $F_X \geq F_Y$ (nerovnost má opačný směr!), obrácená implikace neplatí, neboť z rozdělení nelze rekonstruovat původní náhodné veličiny. *Čtenář, který má problémy s tímto výsledkem, by se měl vrátit na začátek kapitoly 3.*

3.4 Směs náhodných veličin

Příklad 3.4.1: V příkladu 2.1.1 jsme měli dva druhy hracích kostek; výsledky hodu můžeme popsat náhodnými veličinami s hodnotami $1,\dots,6$, řekněme U_1 pro správnou kostku, U_2 pro vadnou. Už jsme popsali rozdělení hodu náhodně vybranou kostkou. Nyní se ptáme, jaká operace s náhodnými veličinami (resp. jejich rozděleními) vede přímo k výsledku.

Příklad 3.4.2: V příkladu 2.1.4 je vybrána jedna otázka, na kterou má student odpovědět, a podle té je hodnocen. Hodnocení odpovědí na jednotlivé otázky lze přirozeně reprezentovat náhodnými veličinami. Z těchto náhodných veličin vytvoříme novou, která bude popisovat výsledek celého pokusu (zde zkoušky).

Ukáže se, že potřebujeme speciální operaci, *směs náhodných veličin*, k níž těžko hledáme analogii jinde. Budeme potřebovat konstrukci směsi pravděpodobností z kapitoly 2.4.

Definice 3.4.3: *Nechť $(\Omega_1,\mathcal{A}_1,P_1),\dots,(\Omega_k,\mathcal{A}_k,P_k)$, jsou pravděpodobnostní prostory a množiny $\Omega_1,\dots,\Omega_k$ jsou po dvou disjunktní. Nechť $U_1,\dots,U_k$, jsou náhodné veličiny na pravděpodobnostních prostorech $(\Omega_1,\mathcal{A}_1,P_1),\dots,(\Omega_k,\mathcal{A}_k,P_k)$ a $w_1,\dots,w_k\in\langle 0,1\rangle$, $\sum_{j=1}^k w_j=1$. Označme $\Omega=\bigcup_{j=1}^k\Omega_j$. Na součtu $\mathcal{A}$ σ-algeber $\mathcal{A}_1,\dots,\mathcal{A}_k$ definujeme* **směs náhodných veličin** *$U_1,\dots U_k$ s koeficienty (=váhami) $w_1,\dots,w_k$ jako náhodnou veličinu $X\colon\Omega\to\mathbb{R}$ vztahem*

$$\forall j\in\{1,\dots,k\}\ \ \forall\omega\in\Omega_j : X(\omega)=U_j(\omega)\ .$$

Značíme $X=\mathrm{Mix}_{(w_1,\dots,w_k)}(U_1,\dots,U_k)$.

Zúžení náhodné veličiny X dává původní náhodné veličiny,

$$X\restriction_{\Omega_j}=U_j\,,\qquad j=1,\dots,k\,.$$

Pro zápis používáme stejné konvence jako u směsí pravděpodobností. Je-li $\boldsymbol{w}=(w_1,\dots,w_k)$ vektor vah, můžeme psát $\mathrm{Mix}_{(w_1,\dots,w_k)}(U_1,\dots,U_k)=\mathrm{Mix}_{\boldsymbol{w}}(U_1,\dots,U_k)$. Poslední váhu někdy vynecháváme, je určena podmínkou $\sum_{j=1}^k w_j=1$. To nám pro $k=2$ dovoluje stručnější zápis $\mathrm{Mix}_{(w,1-w)}(U,V)=\mathrm{Mix}_w(U,V)$, kde w je číslo, nikoli vektor.

Ukážeme, že rozdělení směsi $X=\mathrm{Mix}_{(w_1,\dots,w_k)}(U_1,\dots,U_k)$ lze určit přímo z rozdělení náhodných veličin $U_1,\dots,U_k$, bez ohledu na jejich popis pomocí pravděpodobnostních prostorů. Označme směs pravděpodobností $P_{\mathrm{Mix}}=\mathrm{Mix}_{(w_1,\dots,w_k)}(P_1,\dots,P_k)$ (definovanou na $\mathcal{A}$),

$$P_{\mathrm{Mix}}\Big(\bigcup_{j=1}^k A_j\Big)=\sum_{j=1}^k w_j\,P_j(A_j)\ ,\qquad A_j\in\mathcal{A}_j,\ j=1,\dots,k.$$

Pro libovolný interval $I\subseteq\mathbb{R}$ (a také pro spočetné sjednocení intervalů a obecněji pro libovolnou borelovskou množinu) padne X do I právě tehdy, když náhodná veličina U_j, vybraná z $U_1,\dots,U_k$ počátečním pokusem, dá hodnotu z I, tedy

$$P_X(I)=P[X\in I]=P_{\mathrm{Mix}}\Big(\bigcup_{j=1}^k U_j^{-1}(I)\Big)=\sum_{j=1}^k w_j\,P_j\Big(U_j^{-1}(I)\Big)=\sum_{j=1}^k w_j\,P_{U_j}(I)\ .$$

($P_X,P_{U_1},\dots,P_{U_k}$ jsou pravděpodobnosti na Borelově σ-algebře.) *Rozdělení P_X směsi náhodných veličin $X=\mathrm{Mix}_{(w_1,\dots,w_k)}(U_1,\dots,U_k)$* je tedy *směs rozdělení $P_{U_1},\dots,P_{U_k}$*,

$$P_X=\sum_{j=1}^k w_j\,P_{U_j}\ .$$

Speciálně pro $I=(-\infty,u\rangle$ dostaneme obdobný vztah (konvexní kombinaci) pro distribuční funkce,

$$F_X = \sum_{j=1}^{k} w_j F_{U_j} .$$

(Neplést s *konvexní kombinací náhodných veličin*, což je jiný pojem!)

Poznámka 3.4.4: Směs náhodných veličin se plete s konvexní kombinací mnoha studentům a občas i autorům učebnic. Proto zde tomuto tématu věnujeme zvláštní pozornost. Z praktického hlediska je přitom rozdíl markantní, pokud uvážíme pracnost provedení jednoho vybraného pokusu (u směsi) a všech (u konvexní kombinace). Např. je rozdíl pro studenta i zkoušejícího, jestli se zkouší z jedné vybrané otázky, nebo ze všech (příklady 2.4.10 a 2.3.4).

Příklad 3.4.5: Směs reálných čísel $r_1, \ldots, r_k$ s váhami $w_1, \ldots, w_k$ je náhodná veličina $X = \mathrm{Mix}_{(w_1,\ldots,w_k)}(r_1, \ldots, r_k)$,

$$P_X(I) = P[X \in I] = \sum_{j:r_j \in I} w_j, \qquad F_X(u) = \sum_{j:r_j \le u} w_j .$$

Lze ji popsat též **pravděpodobnostní funkcí** $p_X : \mathbb{R} \to \langle 0, 1\rangle$,

$$p_X(u) = P_X(\{u\}) = P[X = u] = \begin{cases} w_j & \text{pro } u = r_j, \\ 0 & \text{jinak.} \end{cases}$$

Tyto vztahy lze zobecnit i na spočetně mnoho reálných čísel.

Příklad 3.4.6: Výsledky zkoušky z příkladu 2.4.10 lze přirozeně popsat směsí náhodných veličin odpovídajících výsledkům zkoušení jednotlivých otázek.

Na pořadí složek směsi nezáleží, pokud jim ponecháme příslušné váhy:

Tvrzení 3.4.7: *Pro libovolnou permutaci π indexů $1, \ldots, k$ má* $\mathrm{Mix}_{(w_{\pi(1)},\ldots,w_{\pi(k)})}(U_{\pi(1)}, \ldots, U_{\pi(k)})$ *stejné rozdělení jako* $\mathrm{Mix}_{(w_1,\ldots,w_k)}(U_1, \ldots, U_k)$.

Směs ze směsí lze opět popsat jednoduše, ukážeme to pro dvojice složek, zobecnění je zřejmé:

Tvrzení 3.4.8: *Je-li*

$$\begin{aligned} X &= \mathrm{Mix}_{(y_1,y_2)}(Y_1, Y_2) , \\ Y_1 &= \mathrm{Mix}_{(z_{11},z_{12})}(Z_{11}, Z_{12}) , \\ Y_2 &= \mathrm{Mix}_{(z_{21},z_{22})}(Z_{21}, Z_{22}) , \end{aligned}$$

pak

$$X = \mathrm{Mix}_{(y_1 \cdot z_{11}, y_1 \cdot z_{12}, y_2 \cdot z_{21}, y_2 \cdot z_{22})}(Z_{11}, Z_{12}, Z_{21}, Z_{22}) .$$

Poznámka 3.4.9: Pojem směsi náhodných veličin lze zobecnit i na posloupnosti náhodných veličin. Zobecnění na nespočetně mnoho náhodných veličin je sice možné, ale nepřináší nic nového, neboť pouze spočetně mnoho vah může být nenulových a ostatní neovlivní výsledek.

Poznámka 3.4.10: Pokud chceme vytvořit směs náhodných veličin a příslušné množiny elementárních jevů nejsou disjunktní, nahradíme je stejně rozdělenými náhodnými veličinami na obdobných (izomorfních) pravděpodobnostních prostorech s disjunktními množinami elementárních jevů.

Pojem směsi náhodných veličin je základní pro řadu pravděpodobnostních modelů, zejména pro počítačovou reprezentaci složitějších rozdělení. Jeho dobré pochopení je klíčové, často bývá mylně zaměňován s jinými operacemi.

3.5 Druhy náhodných veličin

Některé náhodné veličiny mohou nabývat jen konečně mnoha hodnot. (V krajním případě jedinou hodnotu, v tom případě je můžeme ztotožnit s příslušným reálným číslem.) Naopak u některých náhodných veličin je přirozené předpokládat, že jejich hodnotou může být libovolné reálné číslo z nějakého intervalu nenulové délky, tedy že mohou nabývat nespočetně mnoha hodnot. To klade zcela jiné nároky na popis.

Poznámka 3.5.1: Lze namítnout, že v praxi (např. při měření fyzikální veličiny) máme vždy omezenou informaci, která nedovoluje přesně určit reálné číslo, které je výsledkem. Jsme tedy schopni rozlišit jen konečně mnoho výsledků a matematický model, pracující s nespočetně mnoha hodnotami, se může zdát nepřiměřený. Ve skutečnosti bychom si tímto přístupem popis nezjednodušili, ale naopak zkomplikovali.

Především nebývá předem dána žádná největší možná přesnost, takže s ohledem na možná pozdější zpřesnění potřebujeme *v matematickém popisu* rozlišit i takové výsledky, které jsme dosud zvoleným postupem nerozlišili.

Náhodné veličiny s nespočetně mnoha hodnotami jsou abstrakcí, která dovoluje popsat některá velmi důležitá rozdělení, např. normální (Gaussovo), a teprve ta v případě potřeby diskretizovat. Snaha obejít tento krok a pracovat jen s diskretizovanými rozdělením vede paradoxně na mnohem složitější popis.

3.5.1 Diskrétní náhodné veličiny

Příklad 3.5.2: Hodíme kostkou. Náhodná veličina je číslo, které na kostce padne.

Příklad 3.5.3: Náhodná veličina je úroveň šedi jednoho pixelu obrázku, vyjádřená pomocí 8 bitů, tj. číslem z množiny $\{0, 1, \ldots, 255\}$.

Pro libovolnou náhodnou veličinu X definujeme množinu

$$O_X = \{u \in \mathbb{R} \mid P_X(\{u\}) > 0\} = \{u \in \mathbb{R} \mid P[X = u] > 0\}$$

(přitom X může nabývat ještě jiné hodnoty s nulovou pravděpodobností). Množina O_X je vždy spočetná, neboť

$$P_X(O_X) = \sum_{u \in O_X} P_X(\{u\}) \leq 1\,,$$

což by se pro nespočetnou množinu kladných čísel $P_X(\{u\})$ nemohlo stát.

Definice 3.5.4: *Náhodná veličin X se nazývá* ***diskrétní****, jestliže $P_X(O_X) = 1$. V tom případě definujeme* ***pravděpodobnostní funkci*** *(též* ***funkci pravděpodobnosti****) $p_X : \mathbb{R} \to \langle 0, 1\rangle$ vztahem*

$$p_X(u) = P_X(\{u\}) = P[X = u]\,.$$

Pravděpodobnostní funkce jednoznačně určuje rozdělení diskrétní náhodné veličiny, neboť pro libovolný interval I

$$P_X(I) = P[X \in I] = \sum_{u \in I} p_X(u)\,,$$

přičemž stačí sčítat přes $u \in I \cap O_X$ (ostatní členy jsou nulové). Příslušná suma je vždy absolutně konvergentní, neboť

$$\sum_{u \in O_X} p_X(u) = \sum_{u \in \mathbb{R}} p_X(u) = P[X \in \mathbb{R}] = P_X(\mathbb{R}) = 1\,.$$

Distribuční funkce F_X diskrétní náhodné veličiny X je po částech konstantní, mění hodnoty v bodech $u \in O_X$ skoky velikostí $p_X(u) = F_X(u+) - F_X(u-)$.

Tvrzení 3.5.5: *Pravděpodobnostní funkce směsi* $X = \mathrm{Mix}_w(U, V)$ *diskrétních náhodných veličin je*

$$p_X = w\, p_U + (1 - w)\, p_V\,.$$

Důkaz:

$$\begin{aligned} p_X(u) &= F_X(u) - F_X(u-) = \\ &= w\, F_U(u) + (1-w)\, F_V(u) - w\, F_U(u-) - (1-w)\, F_V(u-) = \\ &= w\, (F_U(u) - F_U(u-)) + (1-w)\, (F_V(u) - F_V(u-)) = w\, p_U(u) + (1-w)\, p_V(u). \end{aligned}$$

□

Poznámka 3.5.6: Každý náhodný pokus, který má pouze spočetně mnoho výsledků, lze považovat za realizaci diskrétní náhodné veličiny (jejíž hodnoty rozlišují výsledky pokusu).

Základní příklady diskrétních rozdělení jsou v kapitole B.1. Rozdělení hodu kostkou z příkladu 3.5.2 je na obr. B.3.

3.5.2 Spojité náhodné veličiny

Příklad 3.5.7: Hodíme kostkou. Náhodná veličina je vzdálenost, ve které se kostka zastaví. Realizací je nezáporné reálné číslo. Lze najít interval, v němž všechny výsledky jsou možné, proto všech možných výsledků je nespočetně mnoho.

Poznámka 3.5.8: Lze namítnout, že jsme schopni změřit jen konečně mnoho hodnot. Nicméně lze použít přesnější metodu, takže nelze *předem* stanovit konečnou množinu hodnot, do které musí výsledek patřit.

Příklad 3.5.9: Náhodná veličina je výsledek měření analogovým voltmetrem a je zaručeno, že měřené napětí je v odpovídajícím rozsahu. Realizací může být libovolné reálné číslo z rozsahu přístroje. Jelikož měření není přesné a je vždy zatíženo nějakou chybou, v typických případech je každá (přesná) hodnota pozorována s nulovou pravděpodobností (nicméně je možná).

Definice 3.5.10: *Náhodná veličin* X *se nazývá* ***(absolutně) spojitá****, jestliže její distribuční funkci lze vyjádřit ve tvaru*

$$F_X(u) = \int_{-\infty}^{u} f_X(v)\,\mathrm{d}v \tag{3.3}$$

pro nějakou nezápornou funkci $f_X : \mathbb{R} \to \langle 0, \infty)$*, nazývanou* ***hustota*** *náhodné veličiny* X*.*

Poznámka 3.5.11: Pro to, aby náhodná veličina byla spojitá, je nutné, aby její distribuční funkce byla spojitá. Není to však podmínka postačující. Existují i náhodné veličiny se spojitými distribučními funkcemi, které nelze vyjádřit ve tvaru (3.3). Takové náhodné veličiny se nazývají *spojité*, nikoli však *absolutně spojité*. Mají spíše teoretický než praktický význam a dále je neuvažujeme. Proto nadále budeme vynechávat přívlastek „*absolutně*". Budeme-li hovořit o *spojitých* náhodných veličinách, budeme mít na mysli *absolutně spojité*.

Poznámka 3.5.12: Integrál funkce se nezmění, pokud změníme její hodnoty na *množině nulové míry* (např. ve spočetně mnoha bodech). Totéž můžeme udělat s hustotou a dostaneme *jinou* hustotu téže náhodné veličiny; hustota není vztahem (3.3) definována jednoznačně. Dvě hustoty f_X, g_X téže náhodné veličiny splňují $\int_I (f_X(x) - g_X(x))\,\mathrm{d}x = 0$ pro všechny intervaly I. Přesněji bychom mohli zavést hustotu jako třídu *všech* funkcí, jejichž integrací dostaneme příslušnou distribuční funkci.

Lze volit

$$f_X(u) = \frac{\mathrm{d}F_X(u)}{\mathrm{d}u},$$

pokud derivace existuje; taková funkce je navíc určena jednoznačně. Tím bychom však vyloučili náhodné veličiny, jejichž distribuční funkce nemá derivaci v některých bodech, a nemohli bychom definovat hustotu mnoha důležitých spojitých rozdělení (rovnoměrné, exponenciální ...). Zde použitá definice hustoty to připouští.

Příklad 3.5.13: Nechť náhodná veličina X má spojité rovnoměrné rozdělení na $\langle 0, 1\rangle$. To znamená, že pro a, b splňující $0 \leq a \leq b \leq 1$ je $P_X(\langle a, b\rangle) = P[a \leq X \leq b] = b - a$. Její distribuční funkce je

$$F_X(u) = \begin{cases} 0 & \text{pro } u \leq 0, \\ u & \text{pro } 0 < u < 1, \\ 1 & \text{pro } u \geq 1. \end{cases}$$

Její hustota může být např.

$$f_X(u) = \begin{cases} 1 & \text{pro } 0 \leq u \leq 1, \\ 0 & \text{jinak.} \end{cases}$$

Nezáleží na tom, jak definujeme hodnoty v bodech 0, 1, a třeba i v jiných (spočetně mnoha) bodech, takže i funkce

$$g_X(u) = \begin{cases} 1 & \text{pro } 0 < u < 1, \\ 3.14 & \text{pro } u = 112, \\ 0 & \text{jinak,} \end{cases}$$

je hustotou náhodné veličiny X.

Pro libovolný interval $\langle a, b\rangle$ splňuje rozdělení *spojité* náhodné veličiny X vztah

$$P_X(\langle a, b\rangle) = P[X \in \langle a, b\rangle] = \int_a^b f_X(v)\,\mathrm{d}v = F_X(b) - F_X(a),$$

speciálně

$$P_X(\{u\}) = 0 \qquad \text{pro všechna } u \in \mathbb{R},$$

$$\int_{-\infty}^{\infty} f_X(v)\,\mathrm{d}v = P[X \in \mathbb{R}] = P_X(\mathbb{R}) = 1.$$

Při respektování tohoto vztahu může být hustota *větší než* 1 (na dostatečně malém intervalu).

Tvrzení 3.5.14: *Hustota směsi $X = \mathrm{Mix}_w(U, V)$ spojitých náhodných veličin splňuje vztah*

$$f_X = w\, f_U + (1 - w)\, f_V$$

až na množinu nulové míry.

Důkaz: Pokud mají distribuční funkce derivace, je uvedený vztah důsledkem linearity derivace. V obecném případě můžeme říci, že funkce

$$g = w\, f_U + (1 - w)\, f_V - f_X$$

splňuje rovnost

$$\int_{-\infty}^{u} g(v)\,\mathrm{d}v = \int_{-\infty}^{u} \left(w\, f_U(v) + (1-w)\, f_V(v) - f_X(v)\right)\,\mathrm{d}v =$$

$$= w \int_{-\infty}^{u} f_U(v)\,\mathrm{d}v + (1-w) \int_{-\infty}^{u} f_V(v)\,\mathrm{d}v - \int_{-\infty}^{u} f_X(v)\,\mathrm{d}v =$$

$$= w\, F_U(u) + (1-w)\, F_V(u) - F_X(u) = 0$$

pro všechna $u \in \mathbb{R}$, tedy g je nulová až na množinu nulové míry.

Základní příklady spojitých rozdělení jsou v kapitole B.2.

3.5.3 Náhodné veličiny se smíšeným rozdělením

Příklad 3.5.15: Náhodná veličina je výsledek měření analogovým voltmetrem, avšak není zaručeno, že měřené napětí je v odpovídajícím rozsahu, takže s nenulovou pravděpodobností dostáváme jeden z krajních výsledků.

Příklad 3.5.16: Náhodná veličina je množství dešťových srážek za 24 hodin. Pokud byly srážky nenulové, je přiměřené je chápat jako spojitou náhodnou veličinu. Nicméně s nenulovou pravděpodobností mohou být (přesně) nulové, což nelze postihnout hustotou pravděpodobnosti.

Příklad 3.5.17: Pojišťovna sleduje vyplacené náhrady z havarijního pojištění. Jejich výši lze modelovat spíše spojitým než diskrétním rozdělením. Nicméně část řidičů neměla žádnou pojistnou událost, tedy hodnota 0 má nenulovou pravděpodobnost.

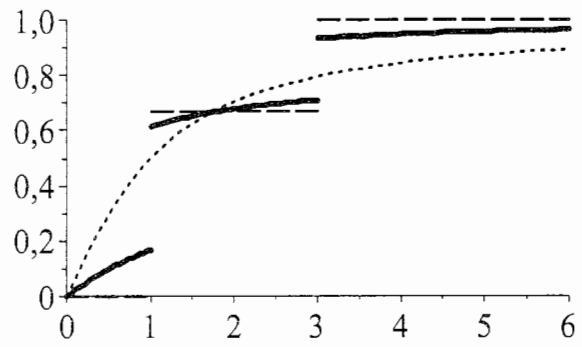
Předchozí příklady ukazují náhodné veličiny, které nejsou ani diskrétní, ani spojité. K jejich popisu nám nepostačí hustota (která nedovoluje, aby nějaká hodnota vycházela s nenulovou pravděpodobností) ani pravděpodobnostní funkce (tu lze definovat, ale nedovoluje popsat nenulovou pravděpodobnost na sjednocení nespočetně mnoha elementárních jevů, z nichž každý má nulovou pravděpodobnost). Taková náhodná veličina se nazývá **smíšená** (též náhodná veličina **se smíšeným rozdělením**). V širším smyslu lze i diskrétní a spojité náhodné veličiny považovat za speciální případy smíšených. Ačkoli mnohé učebnice se omezují pouze na diskrétní nebo spojité náhodné veličiny, z příkladů je vidět, že bychom se tím připravili o možnost popisu řady důležitých náhodných pokusů.

Poznámka 3.5.18: Smíšené náhodné veličiny dostaneme též při zobrazení spojitých náhodných veličin funkcemi, které jsou na nějakých intervalech konstantní. Tak lze interpretovat i příklad 3.5.15: původně spojitou náhodnou veličinu (napětí, které měříme) zobrazíme měřením do měřicího rozsahu, přičemž hodnoty mimo tento interval nahradíme nejbližšími povolenými, tj. krajními hodnotami stupnice voltmetru. Ještě pádnější argumenty najdeme v kapitole 5 o náhodných vektorech.

Pro smíšenou náhodnou veličinu X je typické, že množina O_X je neprázdná, ale $P_X(\mathbb{R}\setminus O_X) = P[X \notin O_X] \neq 0$. Na $\mathbb{R} \setminus O_X$ se X chová jako spojitá náhodná veličina, a proto ji popíšeme jako směs diskrétní a spojité náhodné veličiny.

Důležité je, že rozklad na diskrétní a spojitou složku je v podstatě jednoznačný.

Věta 3.5.19: *Každé rozdělení pravděpodobnosti náhodné veličiny je směsí diskrétního a spojitého rozdělení, tj. má stejné rozdělení jako náhodná veličina* $\mathrm{Mix}_w(U,V)$, *kde* U *je diskrétní a* V *spojitá náhodná veličina. Přitom* $w \in \langle 0,1\rangle$ *je určeno jednoznačně. Je-li* $w \neq 0$, *resp.* $w \neq 1$, *pak rozdělení pravděpodobnosti náhodné veličiny* U, *resp.* V, *je určeno jednoznačně.*



Obrázek 3.1. Distribuční funkce smíšeného rozdělení (tučně) jako směs diskrétního (čárkovaně) a spojitého (tečkovaně)

Důkaz: (Viz obr. 3.1.) Pro náhodnou veličinu X nejprve určíme spočetnou množinu

$$O_X = \{u \in \mathbb{R} : P_X(\{u\}) \neq 0\} .$$

Ta je oborem hodnot diskrétní složky U, jejíž váha musí být

$$w = P_X(O_X) .$$

Je-li $w = 0$, tj. $O_X = \emptyset$, je X spojitá a $X = V$ (diskrétní složka U je libovolná, ale má nulovou váhu). Pro $w \neq 0$ určíme pravděpodobnostní funkci p_U diskrétní složky pro všechna $u \in O_X$ ze vztahu

$$P_X(\{u\}) = w\, P_U(\{u\}) + (1-w) \underbrace{P_V(\{u\})}_{0} = w\, P_U(\{u\}) ,$$

$$p_U(u) = P_U(\{u\}) = \frac{P_X(\{u\})}{w}$$

a distribuční funkce je

$$F_U(u) = \frac{1}{w} \sum_{v:\, v \leq u} P_X(\{v\}) ,$$

kde stačí sčítat pro v ze spočetné množiny $O_X \cap (-\infty, u\rangle$.

Je-li $w = 1$, je X diskrétní a $X = U$ (spojitá složka V je libovolná, ale má nulovou váhu). Pro $w \neq 1$ určíme spojitou složku rozdělení $P_V(I)$ pro libovolný interval I ze vztahu

$$P_X(I) = w\, P_U(I) + (1-w)\, P_V(I) ,$$
$$P_V(I) = \frac{P_X(I) - w\, P_U(I)}{1-w} .$$

Náhodná veličina U byla zavedena tak, že $P_V(\{v\}) = 0$ pro všechna $v \in \mathbb{R}$, takže V má spojité rozdělení. Speciálně pro $I = (-\infty, u\rangle$ dostáváme distribuční funkci

$$F_V(u) = P_V((-\infty, u\rangle) = \frac{F_X(u) - w\, F_U(u)}{1-w} .$$

Pro $w \in (0,1)$ jsou rozdělení pravděpodobnosti náhodných veličin U, V určena jednoznačně. □

Diskrétní složku U lze ještě dále rozložit na směs Diracových rozdělení (konstantních náhodných veličin) dle příkladu 3.4.5.

Důsledek 3.5.20: *Každé rozdělení pravděpodobnosti náhodné veličiny je směsí spojitého rozdělení a Diracových rozdělení.*

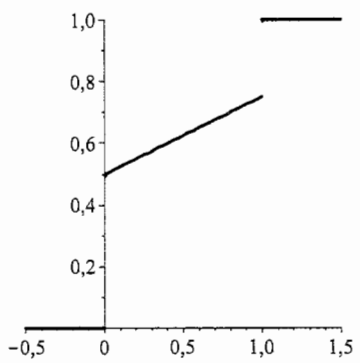
Poznámka 3.5.21: Dokázali jsme jednoznačnost rozdělení pravděpodobností, nikoli náhodných veličin, tento rozdíl je zmíněn v poznámce 3.2.5.

V dalším textu se omezíme jen na taková rozdělení pravděpodobnosti náhodných veličin, která lze vyjádřit jako směsi diskrétních a *absolutně* spojitých rozdělení.

Příklad 3.5.22: ** Náhodná veličina X má distribuční funkci (viz obr. 3.2)

$$F_X(u) = \begin{cases} 0 & \text{pro } u < 0, \\ \frac{u+2}{4} & \text{pro } 0 \le u < 1, \\ 1 & \text{pro } u \ge 1. \end{cases}$$

Najdeme rozdělení její diskrétní a spojité složky dle důkazu věty 3.5.19.



Obrázek 3.2. Distribuční funkce smíšené náhodné veličiny X

Diskrétní složka U bude nabývat hodnot, v nichž je F_X nespojitá, tj. $0, 1$. Váha bude

$$w = P_X(\{0,1\}) = P_X(\{0\}) + P_X(\{1\}) = F_X(0) - F_X(0-) + F_X(1) - F_X(1-) = 3/4\,.$$

Pravděpodobnostní funkce diskrétní složky má následující nenulové hodnoty:

$$p_U(0) = \frac{P_X(\{0\})}{w} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}\,, \qquad p_U(1) = \frac{P_X(\{1\})}{w} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}\,.$$

Nyní můžeme vyjádřit distribuční funkci spojité složky (viz obr. 3.3)

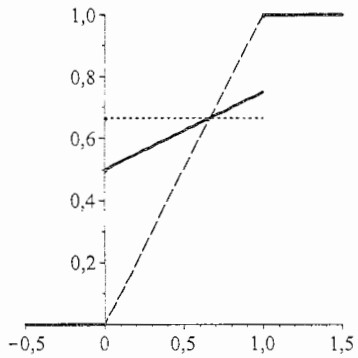
$$F_V(u) = \frac{F_X(u) - w\,F_U(u)}{1-w} = \begin{cases} 0 & \text{pro } u < 0, \\ u & \text{pro } 0 \le u < 1, \\ 1 & \text{pro } u \ge 1. \end{cases}$$

3.5.4 Obecnější náhodné veličiny

Komplexní náhodná veličina má za hodnoty (realizace) dvojice reálných čísel interpretované jako reálná a imaginární část. (Později ji popíšeme též jako *náhodný vektor* se dvěma složkami.) Zobecňujeme na ni pojmy z reálných náhodných veličin.

Někdy připouštíme i „náhodné veličiny", jejichž hodnoty jsou jiné než numerické. Podle míry zobecnění rozlišujeme následující náhodné veličiny (viz např. [Disman 2005, Novovičová 2002]):

- **kvantitativní (intervalové)**, s jejichž hodnotami můžeme provádět operace známé z lineárního prostoru reálných čísel (a po pokusu lze vyhodnotit, zda náhodná veličina padla do daného intervalu),



Obrázek 3.3. Distribuční funkce smíšeného rozdělení (tučně) rozložená na směs diskrétního (tečkovaně) a spojitého (čárkovaně)

- **pořadové** (**ordinální**), na jejichž hodnotách je definováno uspořádání, ale nikoli sčítání a násobení skalárem,
- **kvalitativní** (**nominální**), na jejichž hodnotách není definováno ani uspořádání.

Kvantitativní náhodné veličiny byly doposud jediné, kterými jsme se zabývali. I nadále, pokud budeme hovořit o náhodné veličině (bez další specifikace), budeme tím mínit kvantitativní náhodnou veličinu, jejíž hodnoty jsou reálná čísla.

Příkladem pořadové náhodné veličiny by mohla být známka u zkoušky: je jasné, že jednička je lepší než dvojka, ale není definováno, kolikrát. Stejně tak není zřejmé, zda je lepší jednička a trojka, nebo dvě dvojky. Zde ještě můžeme vypočítat aritmetický průměr, a pokud jej považujeme za smysluplný, můžeme známku chápat i jako intervalovou veličinu. Typičtějším příkladem pořadové proměnné je ukončené vzdělání, klasifikované hodnotami „žádné-základní-střední-vysokoškolské". Zde je jasné pořadí, ale nelze říci, kolikrát je „střední" větší než „základní", ani nemá smysl vypočítávat v této škále průměrné vzdělání skupiny obyvatel. Nelze tedy definovat střední hodnotu pořadové náhodné veličiny. (Kdybychom na tom trvali, mohli bychom zde použít počet roků školy, s tím už lze pracovat jako s kvantitativní veličinou [Disman 2005].)

Příklad 3.5.23: Pokud rozdělíme populaci lidí do skupin podle věku (dovršených let), např. 0–4, 5–9, ... , 75–79, 80 a více, pak obdržíme pořadovou náhodnou veličinu, protože nelze říci, kolikrát je skupina 5–9 starší než skupina 0–4. I samotný věk v letech má tuto vadu, nemluvě o skupině 80 a více, které neumíme přiřadit vhodnou reprezentativní hodnotu. Jestliže toto ignorujeme a pracujeme s náhodnou veličinou jako s kvantitativní, dopouštíme se nepřesnosti, která zkreslí výsledky. Například nemůžeme z takových údajů určit průměrný věk, pouze jej omezit zdola, ale nikoli shora.

Příklad 3.5.24: Příklady pořadových veličin: stupeň biometeorologické zátěže, stupeň utajení atd.

Příkladem kvalitativních náhodných veličin jsou ty, které nabývají konečně mnoha hodnot, jimž ponecháme jejich přirozené označení, např. „rub-líc", „kámen-nůžky-papír" apod. Mohli bychom všechny hodnoty očíslovat, ale není žádný důvod, proč bychom to měli udělat právě určitým způsobem (který by ovlivnil následné numerické výpočty).

Příklad 3.5.25: Politickým stranám v ČR jsou před volbami přidělena čísla bez dalšího významu. Nemá smysl spekulovat o významu toho, že jedna strana má větší číslo než jiná, nebo dokonce o průměru volených čísel.

Kvalitativní náhodné veličiny mohou být též např. náhodné množiny.

Příklad 3.5.26: Velitel nechá nastoupit vojáky do řady a vybere každého desátého. Pokud nastoupili v náhodném pořadí, lze to považovat za náhodný pokus, jehož výsledek bude přirozeně

popsán množinou vybraných vojáků. Očíslování všech možných výsledků je sice možné, ale neužitečné.

Příklad 3.5.27: Náhodný pokus z příkladu 1.7.26 (hrací kostka, z níž vidíme boční stěnu) můžeme popsat pomocí náhodné veličiny, jejímiž hodnotami budou *množiny* možných výsledků.

Ukázali jsme dva důležité speciální případy náhodných veličin – diskrétní, resp. spojité – a prostředky, které zjednodušují jejich popis (pravděpodobnostní funkci, resp. hustotu). Nepostihují však obecný případ; ten lze popsat pomocí směsi předchozích dvou. Naučíme-li se pracovat se směsí rozdělení, postihneme všechny prakticky důležité případy.

Kromě toho mějme na paměti, že ne všechny výsledky náhodných pokusů mají přirozený popis pomocí reálných čísel. Příležitostně se budeme věnovat i takovým případům.

3.6 Kvantilová funkce

Příklad 3.6.1: Jaký příjem má „horních deset tisíc", tj. 1/1000 nejbohatších ze země s 10^7 obyvateli?

Stejnou informaci jako distribuční funkce nám poskytne i *kvantilová funkce*. Zvolme pravděpodobnost $\alpha \in (0,1)$. Pokud je distribuční funkce F_X *spojitá a rostoucí*, je $u = F_X^{-1}(\alpha)$ jediné číslo, pro které $\alpha = F_X(u) = P[X \leq u]$. Pro obecné náhodné veličiny takové číslo existovat nemusí, nebo jich může být nekonečně mnoho. Vždy však existuje u takové, že

$$F_X(u-) = P[X < u] \leq \alpha \leq P[X \leq u] = F_X(u) = F_X(u+) \,.$$

Pokud je takových čísel u víc (tj. pokud je zde F_X konstantní), pak tvoří omezený interval a vezmeme z něj jeden bod. Mohl by to být jeden z krajních bodů, zde však volíme střed:

Definice 3.6.2: *Nechť X je náhodná veličina, $\alpha \in (0,1)$. Funkce $q_X : (0,1) \to \mathbb{R}$ daná vzorcem*

$$q_X(\alpha) = \frac{1}{2}\left(\sup\{v \in \mathbb{R} \mid P[X < v] \leq \alpha\} + \inf\{v \in \mathbb{R} \mid P[X \leq v] \geq \alpha\}\right)$$

je ***kvantilová funkce*** *náhodné veličiny X. Číslo $q_X(\alpha)$ se nazývá* ***α-kvantil*** *náhodné veličiny X (též jejího rozdělení).*

Příklad nespojité kvantilové funkce je na obr. B.4.

Kvantilovou funkci nedefinujeme v krajních bodech 0, 1. Důležité kvantily mají své názvy: $q_X(\frac{1}{2})$ je **medián**, $q_X(\frac{1}{3})$ **dolní tercil**, $q_X(\frac{2}{3})$ **horní tercil**, $q_X(\frac{1}{4})$ **dolní kvartil**, $q_X(\frac{3}{4})$ **horní kvartil**, ..., pro desetiny **decil** (první až devátý), pro setiny **centil** nebo **percentil** apod. Obvykle se zavádí značení pro kvantily význačných rozdělení, např. u_α pro α-kvantil normovaného normálního rozdělení; ten zde značíme $\Phi^{-1}(\alpha)$, neboť kvantilová funkce je v tomto případě inverzí distribuční funkce. Zvláštní značení ostatních kvantilů a kvantilových funkcí zde nezavádíme, avšak uvádíme je v dodatku B.2.

Věta 3.6.3: *Konjunkce následujících podmínek je nutná a postačující pro to, aby funkce $q : (0,1) \to \mathbb{R}$ byla kvantilovou funkcí nějaké náhodné veličiny:*

1. *q je neklesající,*
2. *$\forall \alpha \in (0,1) : q(\alpha) = \frac{1}{2}(q(\alpha-) + q(\alpha+))$.*

Tvrzení 3.6.4: *Kvantilová funkce má spočetně mnoho bodů nespojitosti.*

Důkaz: Každému bodu nespojitosti kvantilové funkce odpovídá interval nenulové délky, na němž je distribuční funkce konstantní. Tyto intervaly jsou disjunktní. Na $\mathbb{R}$ neexistuje nespočetně mnoho disjunktních intervalů nenulové délky. □

Operace s náhodnými veličinami můžeme s výhodou provádět ve vhodně zvolené reprezentaci. Např. směs lze dobře počítat z distribučních funkcí, ale pro kvantilové funkce nemá názorný význam. U jiných operací to bude naopak. Proto je důležité umět je navzájem převádět pomocí následujících vzorců:

$$q_X(\alpha) = \frac{1}{2}\left(\sup\{v \in \mathbb{R} \mid F_X(v-) \leq \alpha\} + \inf\{v \in \mathbb{R} \mid F_X(v) \geq \alpha\}\right),$$
$$F_X(u) = \inf\{\alpha \in (0,1) \mid q_X(\alpha) > u\} = \sup\{\alpha \in (0,1) \mid q_X(\alpha) \leq u\}.$$

Snazší k zapamatování však může být následující poučka:

Tvrzení 3.6.5: *Nechť X je náhodná veličina, $F_X : \mathbb{R} \to \langle 0,1\rangle$ její distribuční funkce a $q_X : (0,1) \to \mathbb{R}$ kvantilová funkce. Funkce F_X, q_X jsou navzájem inverzní tam, kde jsou spojité a rostoucí (tyto podmínky stačí ověřit pro jednu z nich). Svými hodnotami v bodech spojitosti jsou funkce F_X, q_X jednoznačně určeny.*

Poznámka 3.6.6: Kvantilovou funkci lze definovat i pro mnohé pořadové náhodné veličiny, pak ale nesmíme v bodech spojitosti použít aritmetický průměr limit zprava a zleva, nýbrž jednu z těchto limit (a ta musí existovat).

Kvantilová funkce poskytuje stejně univerzální popis rozdělení náhodné veličiny jako distribuční funkce. Každá z těchto reprezentací je výhodná pro některé operace s náhodnými veličinami a dává užitečnou odpověď na jinak formulované otázky.

Cvičení 3.6.7: Je nějaký vztah mezi nerovností náhodných veličin a nerovností jejich kvantilových funkcí?

Řešení: Je-li $X \leq Y$, pak $q_X \leq q_Y$, obrácená implikace neplatí. Na rozdíl od cvičení 3.3.10, zde mají nerovnosti stejný směr.

3.7 Intervalové odhady náhodných veličin

Příklad 3.7.1: Rozdělení počtu příchozích telefonních hovorů na zákaznickou linku považujeme za známé. Kolik zaměstnanců potřebujeme, aby riziko, že všechny linky budou současně obsazené, bylo menší než 5 %?

Často nás zajímá interval, v němž se náhodná veličina nachází (alespoň) s danou pravděpodobností. Pro dané $\alpha \in (0, 1/2)$ hledáme minimální interval I takový, že

$$P[X \in I] \geq 1 - \alpha.$$

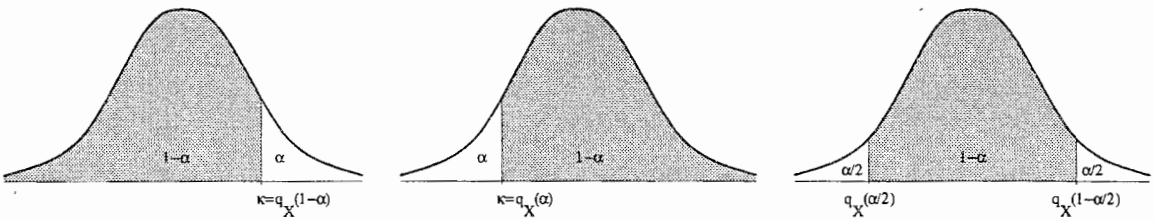
Výsledek je dán kvantily rozdělení. **Horní**, resp. **dolní jednostranný odhad** je

$$I = (-\infty, q_X(1-\alpha)), \text{ resp. } I = \langle q_X(\alpha), \infty).$$

Někdy potřebujeme **oboustranný odhad** (horní i dolní), ten však není jediný. Mohli bychom mezi nimi hledat např. interval, který je nejkratší, ale to by mohlo být velmi pracné. Nejčastěji se používá **symetrický oboustranný odhad**, pro který má překročení horní i dolní meze (pokud možno) stejnou pravděpodobnost $\alpha/2$,

$$I = \langle q_X(\alpha/2), q_X(1-\alpha/2)\rangle.$$

Symetrie intervalového odhadu neznamená, že by musel být symetrický kolem mediánu nebo jiné význačné hodnoty.



Dolní intervalový odhad | Horní intervalový odhad | Symetrický intervalový odhad

Poznámka 3.7.2: Uvedené vzorce lze bez problémů použít pro rozdělení, jejichž distribuční funkce je spojitá a rostoucí (alespoň v okolí potřebných kvantilů). Pak jsou hranice intervalového odhadu jednoznačně určeny a $P[X \in I] = 1 - \alpha$. Většinou budeme intervalové odhady používat jen pro rozdělení, která tuto podmínku splňují. V případě nespojitosti distribuční funkce se může skutečná pravděpodobnost lišit, ale nerovnost $P[X \in J] \geq 1 - \alpha$ lze splnit pro interval $J \supseteq I$, jehož meze se od mezí intervalu I liší libovolně málo. (Nemusí ani existovat nejmenší interval s touto vlastností.) Je-li distribuční funkce v příslušném místě konstantní, hodnota kvantilu nemusí být právě mezí nejmenšího vyhovujícího intervalu. Diskuse všech možných případů by byla zdlouhavá, ale rozhodnutí v konkrétních případech nečiní zvláštní obtíže.

Kvantilová funkce dává odpověď na to, v jakém intervalu se náhodná veličina nachází s danou pravděpodobností. Často o *přesnosti* náhodné veličiny, zejména statistického odhadu, můžeme podat pouze informaci tohoto druhu. Pokud existuje interval, do něhož náhodná veličina *zaručeně* patří, obvykle je tak velký, že nás nezajímá.

3.8 Diskretizace

Příklad 3.8.1: Při měření analogovým voltmetrem rozlišíme a zaznamenáme jen konečný počet číslic, které jsme schopni odečíst na stupnici. Pak je již jen konečně mnoho možných výsledků.

Spojitá náhodná veličina je důležitým teoretickým nástrojem, nicméně praktické aspekty nás nutí rozlišovat jen spočetně (lépe konečně) mnoho hodnot. Tomu odpovídá **diskretizace** náhodné veličiny.

Poznámka 3.8.2: Diskretizaci lze použít na libovolnou náhodnou veličinu. Má smysl i pro diskrétní náhodnou veličinu, neboť někdy je žádoucí snížit počet hodnot.

Typicky jedna diskrétní hodnota odpovídá intervalu hodnot původní náhodné veličiny. Reálnou přímku rozdělíme na spočetně mnoho disjunktních intervalů. (Stačí, když jimi pokryjeme obor hodnot náhodné veličiny.) Takto však dostaneme pouze *pořadovou* náhodnou veličinu (viz příklad 3.5.23). Chceme-li ji změnit na kvantitativní, musíme její hodnoty ztotožnit s čísly (obvykle z příslušných intervalů), což lze učinit více způsoby. Z hlediska přesnosti aproximace se mohou hodit středy intevalů. Problém však je, pokud chceme *neomezenou* náhodnou veličinu nahradit diskrétní náhodnou veličinou s *konečně mnoha* hodnotami. Nemáme jakou hodnotou smysluplně nahradit neomezený krajní interval.

Diskrétní aproximace náhodné veličiny

Volba hodnoty reprezentující interval bude tím méně podstatná, čím bude interval kratší. Zkracováním intervalů můžeme náhodnou veličinu aproximovat „libovolně přesně". Pro libovolné $\varepsilon > 0$ můžeme k diskretizaci použít funkci $h_\varepsilon \colon \mathbb{R} \to \{n\,\varepsilon \mid n \in \mathbb{Z}\}$,

$$h_\varepsilon(u) = \left\lfloor \frac{u}{\varepsilon} \right\rfloor \varepsilon,$$

která nabývá hodnotu $n\,\varepsilon$ na intervalu $\langle n\,\varepsilon, (n+1)\,\varepsilon)$. Aplikujeme-li ji na libovolnou náhodnou veličinu, výsledkem je diskrétní náhodná veličina s hodnotami z $\{n\,\varepsilon \mid n \in \mathbb{Z}\}$.

Tvrzení 3.8.3: *Pro každou náhodnou veličinu X a pro každé $\varepsilon > 0$ platí*

$$h_\varepsilon(X) \leq X \leq h_\varepsilon(X) + \varepsilon .$$

Je-li $X \geq 0$, je $h_\varepsilon(X) \geq 0$. Je-li X omezená, má $h_\varepsilon(X)$ konečně mnoho hodnot.

Tedy každou náhodnou veličinu X lze vyjádřit jako limitu diskrétních náhodných veličin $h_\varepsilon(X)$, $\varepsilon \in (0, \infty)$. Pro $\varepsilon \to 0$ konvergují kvantilové funkce $q_{h_\varepsilon(X)} \to q_X$. Pro distribuční funkce platí $F_{h_\varepsilon(X)}(u) \to F_X(u)$, je-li F_X v bodě u *spojitá*, jinde to být nemusí.

Spojitá aproximace diskrétního rozdělení

Poznámka 3.8.4: Zde budeme hovořit pouze o konvergenci *rozdělení*, nikoli *náhodných veličin*, neboť ty by musely být definovány na stejné σ-algebře. Přitom na *konečné* σ-algebře lze definovat pouze diskrétní náhodné veličiny.

Pro $n \to \infty$ konvergují např. normální rozdělení $N(0, 1/n)$ k Diracovu rozdělení v 0 v tom smyslu, že $q_{N(0,1/n)} \to 0$ a $F_{N(0,1/n)}$ konvergují k Heavisidově funkci v bodech její spojitosti, tj. mimo 0. Dané diskrétní rozdělení je směsí Diracových rozdělení v hodnotách u_k. Odpovídající směsi normálních rozdělení $N(u_k, 1/n)$ pak pro $n \to \infty$ konvergují k danému rozdělení v tom smyslu, že jejich distribuční, resp. kvantilové funkce konvergují k distribuční, resp. kvantilové funkci daného rozdělení *v jejích bodech spojitosti*.

V jistém smyslu lze každé rozdělení libovolně přesně aproximovat diskrétním nebo spojitým. Někdy odvodíme pojmy a tvrzení (např. definici střední hodnoty a její vlastnosti) pro jeden typ rozdělení a pomocí aproximace zobecníme na všechna.

Cvičení 3.8.5: * Navrhněte způsob diskretizace spojité náhodné veličiny tak, aby výsledných hodnot byl daný počet a všechny byly *stejně pravděpodobné*. (Návod: použijte kvantilovou funkci.) Nakolik lze tento postup aplikovat na diskrétní a smíšené náhodné veličiny?

Cvičení 3.8.6: Najděte posloupnosti spojitých rozdělení (jiných než normálních), které konvergují k Diracovu rozdělení.

3.9 Nezávislost náhodných veličin

Příklad 1.5.1 může posloužit i jako příklad na nezávislost náhodných veličin, které v něm lze přirozeně zavést.

Definice 3.9.1: *Náhodné veličiny X_1, X_2 jsou* ***nezávislé****, pokud pro všechny intervaly I_1, I_2 jsou jevy $X_1 \in I_1$, $X_2 \in I_2$ nezávislé, tj.*

$$P[X_1 \in I_1, X_2 \in I_2] = P[X_1 \in I_1] \cdot P[X_2 \in I_2] . \tag{3.4}$$

Při testování nezávislosti se stačí omezit na intervaly tvaru $(-\infty, u\rangle$:

Tvrzení 3.9.2: *Náhodné veličiny X_1, X_2 jsou nezávislé, právě když $P[X_1 \leq u_1, X_2 \leq u_2] = P[X_1 \leq u_1] \cdot P[X_2 \leq u_2]$ pro všechna $u_1, u_2 \in \mathbb{R}$.*

Důkaz: Ve vzorcích (3.1) bylo ukázáno, že všechny intervaly lze získat pomocí doplňků a spočetných disjunktních sjednocení z intervalů tvaru $(-\infty, u\rangle$, $u \in \mathbb{R}$. Stačí zkontrolovat, se při použitých operacích zachovává platnost vztahu (3.4), což ponecháváme čtenáři. □

Tvrzení 3.9.3: *Každé dvě náhodné veličiny, z nichž jedna je konstantní (tj. reálné číslo), jsou nezávislé.*

Důkaz: Je-li X konstantní náhodná veličina s hodnotou $r \in \mathbb{R}$, pak pro libovolný interval I je jev $X \in I$ (neboli $r \in I$) buď nemožný, nebo jistý. To dovoluje použít tvrzení 1.5.9. □

Definice 3.9.4: *Náhodné veličiny $X_1, \ldots, X_n$ jsou* ***nezávislé****, pokud pro libovolné intevaly $I_1, \ldots, I_n$ platí*

$$P[X_1 \in I_1, \ldots, X_n \in I_n] = \prod_{j=1}^{n} P[X_j \in I_j] \,. \tag{3.5}$$

Na rozdíl od definice nezávislosti více než dvou jevů zde není třeba požadovat nezávislost pro libovolnou podmnožinu náhodných veličin $X_1, \ldots, X_n$. Ta vyplývá z toho, že libovolnou náhodnou veličinu X_j lze „vynechat" tak, že zvolíme $I_j = \mathbb{R}$. Pak $P[X_j \in I_j] = 1$ a v součinu se tento činitel neprojeví. Opět stačí omezit se na intervaly tvaru $(-\infty, u)$:

Tvrzení 3.9.5: *Náhodné veličiny $X_1, \ldots, X_n$ jsou nezávislé, právě když*

$$P[X_1 \leq u_1, \ldots, X_n \leq u_n] = \prod_{j=1}^{n} P[X_j \leq u_j]$$

pro všechna $u_1, \ldots, u_n \in \mathbb{R}$.

Definice 3.9.6: *Nekonečná množina náhodných veličin se nazývá* ***nezávislá****, pokud každá její konečná podmnožina je nezávislá.*

Definice 3.9.7: *Náhodné veličiny $X_1, \ldots, X_n$ jsou* ***po dvou nezávislé****, pokud každé dvě (různé) z nich jsou nezávislé.*

Nezávislost jsme rozšířili z jevů na náhodné veličiny. Nezávislost *po dvou* je opět slabší podmínka.

Cvičení na nezávislost náhodných veličin jsou v kapitole 5.

3.10 Operace s náhodnými veličinami

Příklad 3.10.1: Známe-li rozdělení platů, známe rozdělení náhodné veličiny, kterou je plat náhodně vybraného zaměstnance. Z něj můžeme odvodit i rozdělení daně z příjmu, pokud pro zjednodušení přepokládáme, že závisí pouze na platu.

Příklad 3.10.2: Předpokládejme, že poloměr R dešťových kapek v mm má logaritmicko-normální rozdělení s parametry $\mu = 1$, $\sigma^2 = 1$. Jaké je rozdělení jejich objemu $V = \frac{4}{3}\pi R^3$?

Příklad 3.10.3: Lukostřelec míří do terče, chyba ve vodorovném, resp. svislém směru v cm má normální rozdělení $\mathrm{N}(0, 100)$, resp. $\mathrm{N}(5, 225)$. Máme určit rozdělení vzdálenosti zásahů od středu terče za předpokladu, že obě chyby jsou nezávislé.

V těchto příkladech je zřejmé, že máme dostatečnou informaci, která jednoznačně určuje odpověď, ale řešení nemusí být snadné. Uvedeme alespoň odpověď na nejjednodušší otázky tohoto typu.

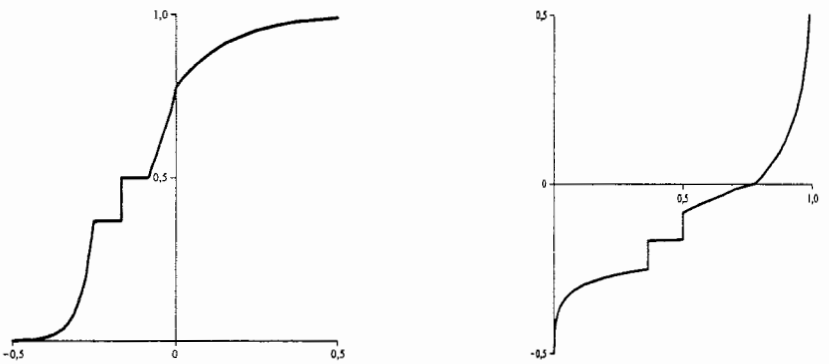
V této kapitole jsou X, Y náhodné veličiny, $I, J \subseteq \mathbb{R}$ intervaly (mohou to být i spočetná sjednocení intervalů a libovolné borelovské množiny), $u, v \in \mathbb{R}$. Operace budeme demonstrovat na náhodné veličině, jejíž rozdělení je na obr. 3.4.

Přičtení konstanty $r \in \mathbb{R}$ odpovídá vodorovné posunutí argumentu:

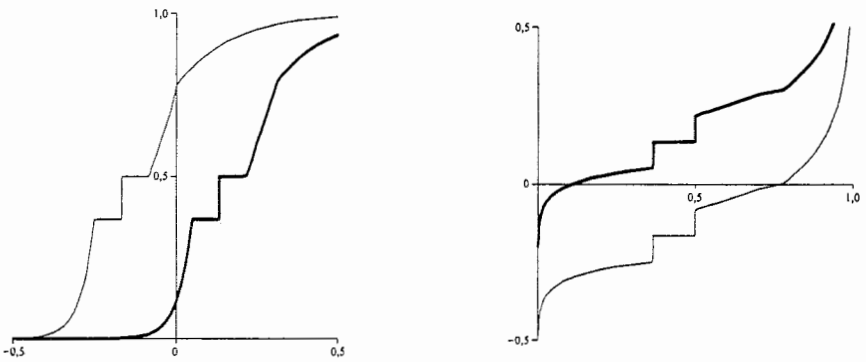
$$P_{X+r}(I + r) = P[X + r \in I + r] = P[X \in I] = P_X(I) \,,$$

kde $I + r = \{u + r \mid u \in I\}$. Po dosazení $J = I + r$

$$P_{X+r}(J) = P_X(J - r) \,.$$



Obrázek 3.4. Distribuční a kvantilová funkce náhodné veličiny X pro demonstraci operací



Obrázek 3.5. Distribuční a kvantilová funkce náhodné veličiny X (tence) a $X + 0.3$ (tučně)

Vzorce pro distribuční funkci dostaneme pro $I = (-\infty, u)$, $J = (-\infty, v)$, $v = u + r$:

$$F_{X+r}(u + r) = F_X(u), \qquad F_{X+r}(v) = F_X(v - r)\,.$$

Kvantilová funkce se zvýší o r:

$$q_{X+r}(\alpha) = q_X(\alpha) + r\,.$$

Vynásobení nenulovou konstantou $r \in \mathbb{R}\setminus\{0\}$ odpovídá podobnost ve směru vodorovné osy:

$$P_{r\,X}(r\,I) = P\,[r\,X \in r\,I] = P\,[X \in I] = P_X(I)\,,$$

kde $r\,I = \{r\,u \mid u \in I\}$. Po dosazení $J = r\,I$

$$P_{r\,X}(J) = P_X\left(\tfrac{1}{r}\,J\right)\,.$$

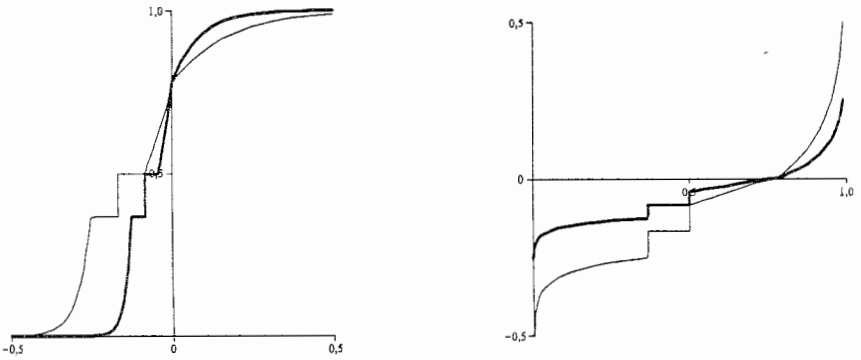
Pro distribuční funkci však musíme dát pozor na znaménko:

- Pro $r > 0$ postupujeme stejně:

$$F_{r\,X}(r\,u) = F_X(u)\,, \qquad F_{r\,X}(v) = F_X\left(\tfrac{v}{r}\right)\,,$$

pro kvantilovou funkci opět máme názornější vzorec

$$q_{r\,X}(\alpha) = r\,q_X(\alpha)\,.$$



Obrázek 3.6. Distribuční a kvantilová funkce náhodné veličiny X (tence) a $X/2$ (tučně)

- Pro $r < 0$ se projeví nesymetrie v definici distribuční funkce. Vyřešme napřed speciální případ násobení číslem -1:

$$F_{-X}(-u) = P\left[-X \leq -u\right] = P\left[X \geq u\right] = 1 - P\left[X < u\right] .$$

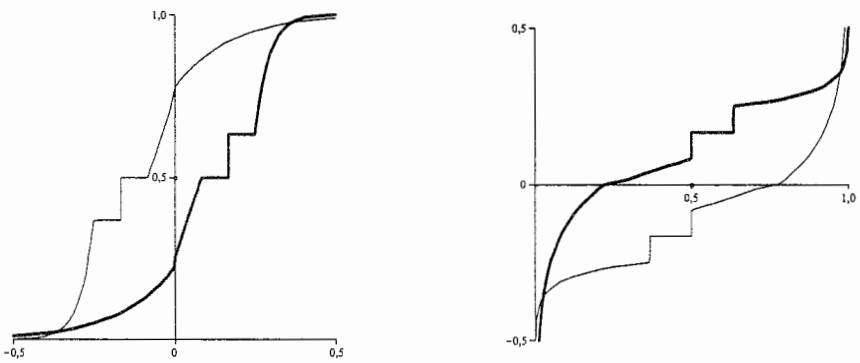
V bodech spojitosti distribuční funkce je $P[X = u] = 0$,

$$F_{-X}(-u) = 1 - P[X < u] = 1 - P[X \leq u] = 1 - F_X(u) ,$$
$$F_{-X}(v) = 1 - F_X(-v)$$

a graf F_{-X} je středově symetrický ke grafu F_X podle bodu $(0, 1/2)$; zbývá dodefinovat hodnoty *v bodech nespojitosti* jako limity zprava. Kvantilová funkce splňuje

$$q_{-X}(\alpha) = -q_X(1 - \alpha) ,$$

díky symetrii v její definici není potřeba provádět opravy v bodech nespojitosti.



Obrázek 3.7. Distribuční a kvantilová funkce náhodné veličiny X (tence) a $-X$ (tučně)

- Obecný případ $r < 0$ řešíme kombinací předchozích postupů jako postupné násobení -1 a $|r| > 0$.

Směs náhodných veličin jsme již probrali v kapitole 2.4. Na rozdíl od součtu je plně určena rozděleními argumentů a váhami směsi. Na směs lze funkci aplikovat „po složkách":

Tvrzení 3.10.4: *Obraz směsi náhodných veličin* $\mathrm{Mix}_w(U, V)$ *v měřitelné funkci* h *je směs obrazů,*

$$h(\mathrm{Mix}_w(U, V)) = \mathrm{Mix}_w(h(U), h(V)) . \tag{3.6}$$

Důkaz: Tvrzení vyplývá přímo z podstaty směsi náhodných veličin: Podle počátečního pokusu rozhodneme, která složka bude určovat výsledek, a je jedno, jestli jsme funkci aplikovali před tímto rozhodnutím, nebo až po něm. □

Zobrazení *spojitou rostoucí* funkcí h:

$$P_{h(X)}(h(I)) = P_X(I)\,, \qquad F_{h(X)}(h(u)) = F_X(u)\,, \qquad q_{h(X)}(\alpha) = h(q_X(\alpha))\,,$$

přičemž poslední vzorec platí jen *v bodech spojitosti* kvantilové funkce (čímž jsou zbývající hodnoty jednoznačně určeny).

Příklad 3.10.5: Náhodnou veličinu X s rovnoměrným spojitým rozdělením $\mathrm{R}(1/3, 1)$ zobrazíme funkcí $h(u) = u^2$. (Ta sice není monotónní, ale je ostře rostoucí na intervalu $\langle 1/3, 1\rangle$; mimo tento interval ji nepotřebujeme, takže bychom si ji mohli předefinovat tak, aby i tam byla rostoucí, např. $h(u) = u^2 \operatorname{sign}(u)$, a nezměnili bychom výsledek.)

Postup ukážeme na obrázku 3.8, kde je v každém ze čtyř kvadrantů graf jiné funkce. (Pro studenta elektrotechnické fakulty by to neměl být problém. Využili jsme toho, že u všech funkcí nás zajímají jen nezáporné hodnoty, jinak bychom museli grafy rozdělit a mít čtyři počátky souřadnic ve vrcholech obdélníka.) V prvním kvadrantu je graf funkce h, $v = h(u)$, ale lze jej chápat i jako graf inverzní funkce h^{-1}, $u = h^{-1}(v)$. Ve čtvrtém kvadrantu je graf distribuční funkce F_X,

$$\alpha = F_X(u) = \begin{cases} 0 & \text{pro } u < 1/3, \\ \frac{3u-1}{2} & \text{pro } 1/3 \le u < 1, \\ 1 & \text{pro } 1 \le u, \end{cases}$$

resp. kvantilové funkce q_X,

$$u = q_X(\alpha) = \frac{2\alpha + 1}{3}\,.$$

Ve druhém kvadrantu konstruujeme výsledek. Třetí kvadrant slouží jen k převedení hodnoty α ze svislé osy na vodorovnou (nebo naopak) pomocí identické funkce, zobrazené osou kvadrantu.

1. postup: K danému $v \in \mathbb{R}$ najdeme hodnotu distribuční funkce $F_{h(X)}(v)$ tak, že postupujeme po obdélníku ve směru hodinových ručiček: $u = h^{-1}(v) = \sqrt{v}$, $\alpha = F_X(u) = F_{h(X)}(h(u)) = F_{h(X)}(v)$. Např. $v = 9/16$, $u = 3/4$, $\alpha = 5/8$. Obecně

$$\alpha = F_{h(X)}(v) = \begin{cases} 0 & \text{pro } v < 1/9, \\ \frac{3\sqrt{v}-1}{2} & \text{pro } 1/9 \le v < 1, \\ 1 & \text{pro } 1 \le v. \end{cases}$$

2. postup: K danému $\alpha \in \langle 0, 1\rangle$ najdeme hodnotu kvantilové funkce $q_{h(X)}(\alpha)$ tak, že postupujeme po obdélníku proti směru hodinových ručiček: $u = q_X(\alpha) = (2\alpha + 1)/3$, $v = h(u) = h(q_X(\alpha)) = q_{h(X)}(\alpha)$ neboli $v = u^2 = (q_X(\alpha))^2$. Obecně

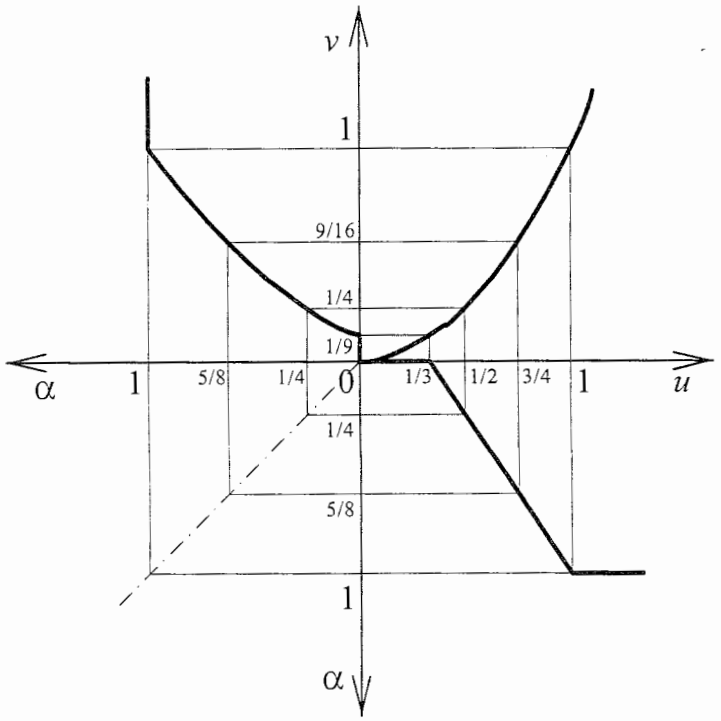
$$v = q_{h(X)}(\alpha) = \left(\frac{2\alpha + 1}{3}\right)^2\,.$$

Zobrazení *spojitou neklesající* funkcí h:

$$F_{h(X)}(u) = F_X(v)\,, \qquad \text{kde } v = \sup\{w \in \mathbb{R} \mid h(w) \le u\}.$$

(Zde stačila spojitost *zleva*.)

Zobrazení *spojitou nerostoucí* funkcí h: Zobrazíme neklesající funkcí $-h$ a vynásobíme -1.



Obrázek 3.8. Zobrazení náhodné veličiny spojitou rostoucí funkcí

Zobrazení funkcí h, která je ***po částech monotónní a spojitá*:** Obor hodnot náhodné veličiny X pokryjeme disjunktními intervaly I_j, na nichž je h monotónní a spojitá. Je-li počet takových intervalů k, vyjádříme X jako směs $X = \mathrm{Mix}_{\boldsymbol{w}}(U_1, \ldots, U_k)$ tak, že U_j nabývá hodnot z I_j. Podle tvrzení 3.10.4 je $h(X) = \mathrm{Mix}_{\boldsymbol{w}}(h(U_1), \ldots, h(U_k))$, přičemž jednotlivé složky této směsi spočítáme dle předchozích pravidel. Obdobně můžeme využít nekonečnou směs, je-li intervalů nekonečně mnoho (určitě je jich spočetně mnoho, protože jsou disjunktní a mají nenulovou délku).

Samozřejmě lze tento postup v mnoha konkrétních případech zjednodušit.

Příklad 3.10.6: Náhodná veličina X má rovnoměrné rozdělení na intervalu $\langle -1, 2\rangle$. Určete rozdělení náhodné veličiny (a) $|X|$, (b) $X + |X|$.

Řešení: (a) Funkce absolutní hodnota je klesající na $(-\infty, 0)$, rostoucí na $(0, +\infty)$. Náhodnou veličinu X lze vyjádřit jako směs, $X = Mix_w(U, V)$, kde náhodná veličina U nabývá pouze záporných hodnot (s rovnoměrným rozložením na $\langle -1, 0\rangle$), V nabývá pouze nezáporných hodnot (s rovnoměrným rozložením na $\langle 0, 2\rangle$) a $w = P[X < 0] = \frac{1}{3}$ (viz obr. 3.9),

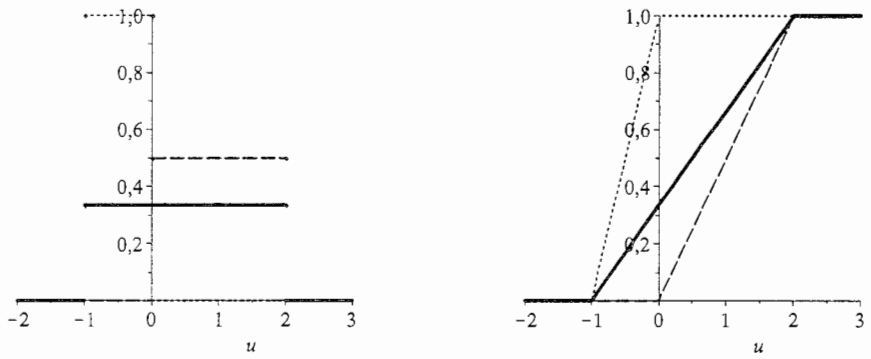
$$F_U(u) = \begin{cases} 0 & \text{pro } u < -1, \\ u+1 & \text{pro } -1 \le u < 0, \\ 1 & \text{pro } 0 \le u, \end{cases} \qquad F_V(u) = \begin{cases} 0 & \text{pro } u < 0, \\ \frac{u}{2} & \text{pro } 0 \le u < 2, \\ 1 & \text{pro } 2 \le u. \end{cases}$$

Po zobrazení absolutní hodnotou

$$F_{|V|}(u) = F_V(u)\,,$$

$$F_{|U|}(u) = F_{-U}(u) = 1 - F_U(-u) = \begin{cases} 0 & \text{pro } u < 0, \\ u & \text{pro } 0 \le u < 1, \\ 1 & \text{pro } 1 \le u, \end{cases}$$

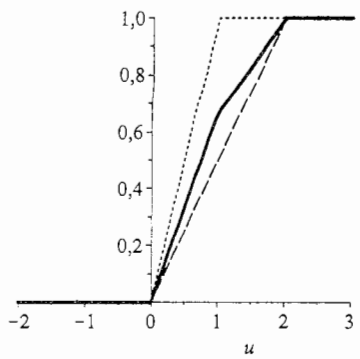
neboť F_U je všude spojitá.



Obrázek 3.9. Hustoty a distribuční funkce náhodných veličin U (tečkovaně), V (čárkovaně) a jejich směsi X

$$F_{|X|}(u) = w\,F_{|U|}(u) + (1-w)\,F_{|V|}(u) = \begin{cases} 0 & \text{pro} \quad u < 0, \\ \frac{2}{3}\,u & \text{pro} \quad 0 \le u < 1, \\ \frac{1}{3} + \frac{u}{3} & \text{pro} \quad 1 \le u < 2, \\ 1 & \text{pro} \quad 2 \le u. \end{cases}$$

(Viz obr. 3.10.) Uvedený postup se může zdát zdlouhavý, je však univerzálně použitelný i na složitější příklady.



Obrázek 3.10. Distribuční funkce náhodných veličin $|U|$ (tečkovaně), $|V|$ (čárkovaně) a jejich směsi $|X|$

(b) Jedná se o zobrazení funkcí

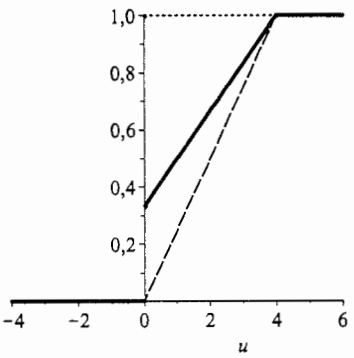
$$h(u) = \begin{cases} 0 & \text{pro} \quad u \le 0, \\ 2\,u & \text{pro} \quad u > 0, \end{cases}$$

konstantní na $(-\infty, 0)$, rostoucí na $(0, +\infty)$. Náhodnou veličinu X vyjádříme jako směs $X = Mix_w(U, V)$ stejně jako v případě (a). Po zobrazení funkcí h dostaneme

$$F_{h(V)}(u) = F_V\left(\frac{u}{2}\right) = \begin{cases} 0 & \text{pro} \quad u < 0, \\ \frac{u}{4} & \text{pro} \quad 0 \le u < 4, \\ 1 & \text{pro} \quad 4 \le u, \end{cases}$$

$$F_{h(U)}(u) = \begin{cases} 0 & \text{pro} \quad u < 0, \\ 1 & \text{pro} \quad u \ge 0, \end{cases}$$

neboť $h(U) = 0$ (tj. $h(U)$ nabývá hodnoty 0 s pravděpodobností 1).



Obrázek 3.11. Distribuční funkce náhodných veličin $h(U)$ (tečkovaně), $h(V)$ (čárkovaně) a jejich směsi $h(X)$

$$F_{h(X)}(u) = w\,F_{h(U)}(u) + (1-w)\,F_{h(V)}(u) = \begin{cases} 0 & \text{pro } u < 0, \\ \frac{1}{3} + \frac{1}{6}\,u & \text{pro } 0 \le u < 4, \\ 1 & \text{pro } 4 \le u \end{cases}$$

(viz obr. 3.11), což je distribuční funkce smíšeného rozdělení (nabývá hodnoty 0 s pravděpodobností 1/3). □

Příklad 3.10.7: Náhodnou veličinu X s rovnoměrným spojitým rozdělením $\mathrm{R}(-2,2)$,

$$F_X(u) = \begin{cases} 0 & \text{pro } u < -2, \\ \frac{u+2}{4} & \text{pro } -2 \le u < 2, \\ 1 & \text{pro } 2 \le u. \end{cases}$$

zobrazíme funkcí h z obr. 3.13

$$h(u) = \begin{cases} u & \text{pro } u \in \langle 0,1\rangle\,, \\ 0 & \text{pro } u < 0\,, \\ 1 & \text{pro } u > 1\,. \end{cases}$$

(Tento příklad ukazuje typickou situaci, kdy převodní charakteristika vykazuje saturaci. To může být např. operační zesilovač nebo měřicí přístroj z příkladu 3.5.15. Pro názornější zobrazení zde uvažujeme „zesilovač" se zesílením 1, ale lze si představit, že vstup se měří v jiných jednotkách.)

Funkce h je spojitá a neklesající.

1. postup: Funkce h je navíc prostá a rostoucí tam, kde nenabývá hodnot $0, 1$ (které budeme diskutovat zvlášť). Tedy pro $u \in (0,1)$ je

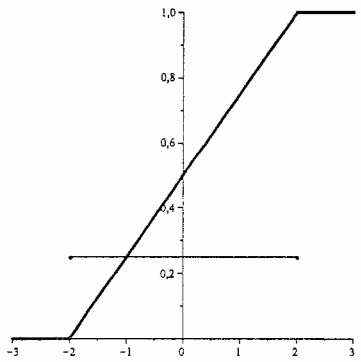
$$F_{h(X)}(h(u)) = F_X(u) = \frac{u+2}{4}\,,$$

zde navíc $h(u) = u$, takže máme přímo hodnoty distribuční funkce

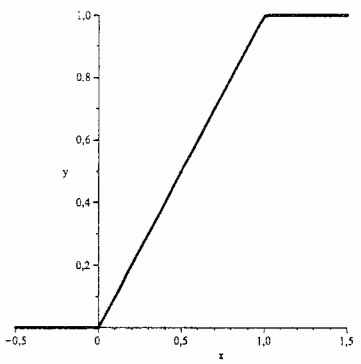
$$F_{h(X)}(u) = \frac{u+2}{4} \qquad \text{pro } u \in (0,1)\ .$$

Protože hodnoty záporné ani větší než 1 funkce h nenabývá, vychází distribuční funkce z obr. 3.2

$$F_{h(X)}(u) = \begin{cases} 0 & \text{pro } u < 0, \\ \frac{u+2}{4} & \text{pro } 0 \le u < 1, \\ 1 & \text{pro } u \ge 1. \end{cases}$$



Obrázek 3.12. Hustota (tence) a distribuční funkce (tučně) náhodné veličiny X s rovnoměrným rozdělením $\mathrm{R}(-2,2)$



Obrázek 3.13. Funkce h, kterou zobrazujeme

(Hodnoty v bodech $0, 1$ jsou dány spojitosti zprava.)

2. postup: Kvantilovou funkci

$$q_X(\alpha) = 4\alpha - 2$$

složíme s funkcí h,

$$q_{h(X)}(\alpha) = h(q_X(\alpha)) = \begin{cases} 0 & \text{pro } \alpha < 1/2, \\ 4\alpha - 2 & \text{pro } 1/2 \le \alpha \le 3/4, \\ 1 & \text{pro } \alpha > 3/4. \end{cases}$$

Předchozí příklady ukazují, že *zobrazením spojité náhodné veličiny funkcí, která není ryze monotónní, můžeme dostat náhodnou veličinu se smíšeným rozdělením.* I proto je nutno studovat smíšená rozdělení, neboť se s nimi setkáme zcela běžně. Pokusy vyhnout se jim pomocí aproximace spojitým (nebo naopak diskrétním) rozdělením nevedou na jednodušší a užitečnější model. Kvantilová funkce nám může práci usnadnit.

Příklad 3.10.8: * Náhodná veličina X má normované normální rozdělení $\mathrm{N}(0,1)$. Popište rozdělení náhodné veličiny $Y = X^2$.

Řešení: Pro $u \le 0$ je $F_Y(u) = 0$, $f_Y(u) = 0$. Pro $u > 0$ je

$$F_Y(u) = P[Y \le u] = P\left[-\sqrt{u} \le X \le \sqrt{u}\right] = \int_{-\sqrt{u}}^{\sqrt{u}} f_X(v)\,\mathrm{d}v =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{u}}^{\sqrt{u}} \exp\left(-v^2/2\right)\,\mathrm{d}v = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{u}} \exp\left(-v^2/2\right)\,\mathrm{d}v\,.$$

Po substituci $v = \sqrt{w}$, $\mathrm{d}v = \frac{1}{2\sqrt{w}}\,\mathrm{d}w$:

$$F_Y(u) = \frac{1}{\sqrt{2\pi}} \int_0^u \frac{\exp(-w/2)}{\sqrt{w}}\,\mathrm{d}w\,,$$
$$f_Y(u) = \frac{1}{\sqrt{2\pi}} \frac{\exp(-u/2)}{\sqrt{u}}\,.$$

(Distribuční funkce je transcendentní.) Odvodili jsme *rozdělení χ^2 s jedním stupněm volnosti* (viz dodatek B.2). Mohli jsme také použít rozklad X na směs dvou náhodných veličin, z nichž jedna nabývá nezáporných, druhá nekladných hodnot; v tomto speciálním případě mají jejich obrazy stejné rozdělení (jako Y). □

Součet náhodných veličin přináší ten problém, že jeho rozdělení není jednoznačně určeno rozdělením jeho argumentů. Z definice distribuční funkce vyplývá

$$F_{X+Y}(u) = P[X+Y \le u] = P\left(\{(x,y) \in \mathbb{R}^2 \mid x+y \le u\}\right)\,.$$

K určení výsledků potřebujeme znát *sdružené* rozdělení náhodných veličin X, Y (tj. *náhodného vektoru* (X,Y)). Pokud jsou náhodné veličiny X, Y *nezávislé*, pak jejich rozdělení jednoznačně určují i sdružené rozdělení a tím rozdělení jejich součtu. Není to však jednoduchá operace. Uvedeme ji ve speciálních případech.

Jsou-li X, Y *spojité nezávislé* náhodné veličiny, pak hustota jejich součtu je

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(v)\, f_Y(u-v)\,\mathrm{d}v = \int_{-\infty}^{\infty} f_Y(w)\, f_X(u-w)\,\mathrm{d}w = (f_X * f_Y)(u)\,, \quad (3.7)$$

což je **konvoluce** hustot argumentů. (Její základní vlastnosti jsou v dodatku C.)

Jsou-li X, Y *diskrétní nezávislé* náhodné veličiny s obory hodnot O_X, resp. O_Y, pak pravděpodobnostní funkce jejich součtu je

$$p_{X+Y}(u) = \sum_{v \in \mathbb{R}} p_X(v)\, p_Y(u-v) = \sum_{w \in \mathbb{R}} p_Y(w)\, p_X(u-w)\,,$$

což je vlastně diskrétní obdoba konvoluce. Ze sčítanců v sumách je jen spočetně mnoho nenulových, pro $v \in O_X \cap (u - O_Y)$ (kde $u - O_Y = \{u-w \mid w \in O_Y\}$) a $w \in O_Y \cap (u - O_X)$, proto nejsou problémy s konvergencí.

Jsou-li X, Y *nezávislé* náhodné veličiny, z nichž X je *spojitá* a Y *diskrétní* s oborem hodnot O_Y, pak jejich součet má spojité rozdělení. Můžeme je dostat např. tak, že Y vyjádříme jako směs Diracových rozdělení (jejichž přičtení odpovídá přičtení konstanty) a výsledek dostaneme jako směs součtů:

$$F_{X+Y}(u) = \sum_{w \in O_Y} p_Y(w)\, F_X(u-w)\,,$$
$$f_{X+Y}(u) = \sum_{w \in O_Y} p_Y(w)\, f_X(u-w)\,.$$

I přibližný výpočet konvoluce může být pracný, proto se nám bude hodit alternativní výpočet pomocí charakteristické funkce, který uvedeme v kapitole 4.7. Ten navíc současně řeší i obecný případ bez ohledu na typ rozdělení.

Speciálně výraz $a\,X + b\,Y$ se nazývá **konvexní kombinace náhodných veličin** X, Y, jestliže $a, b \in \langle 0, 1\rangle$, $a+b = 1$ (podobně pro konvexní kombinace libovolného spočetného počtu náhodných veličin).

Naučili jsme se základní početní operace s náhodnými veličinami. Ty dovolují stanovit rozdělení výsledku výpočtu ze znalosti rozdělení argumentů. Pokud zde některé základní operace chyběly, je to proto, že jsou pro náhodné veličiny definovány jen ve speciálních případech, nebo jsou příliš obtížné.

Cvičení 3.10.9: Demonstrujte rozdíl mezi směsí a konvexní kombinací náhodných veličin pomocí mísicí baterie na teplou a studenou vodu.

Řešení: Náhodné veličiny X, Y reprezentují teplotu přitékající teplé, resp. studené vody. Pokud pustíme $1/3$ teplé a $2/3$ studené vody, výsledná teplota je konvexní kombinace $1/3\,X + 2/3\,Y$. Směs $\mathrm{Mix}_{1/3}(X, Y)$ by popisovala pokus, kdy pustíme vodu z jednoho kohoutku náhodně zvoleného (přičemž teplovodní má pravděpodobnost $1/3$, že bude vybrán). Je zřejmé, že střední hodnota je v obou případech stejná, rozptyl nikoli.

Cvičení 3.10.10: Náhodná veličina X má rovnoměrné rozdělení na intervalu $\langle -4, 2\rangle$. Určete a znázorněte rozdělení náhodných veličin
(a) $X + 4$, (b) $-X/2$, (c) $|X|$.

Řešení: (a) rovnoměrné rozdělení na intervalu $\langle 0, 6\rangle$,
(b) rovnoměrné rozdělení na intervalu $\langle -1, 2\rangle$,
(c) $f_{|X|}(x) = \begin{cases} 1/3 & \text{pro } x \in \langle 0, 2\rangle, \\ 1/6 & \text{pro } x \in (2, 4\rangle, \\ 0 & \text{jinak.} \end{cases}$ □

Cvičení 3.10.11: Z jednotkového kruhu vybereme bod (předpokládáme rovnoměrné rozdělení). Náhodná veličina X je vzdálenost bodu od středu kruhu. Určete rozdělení náhodných veličin X a X^2.

Řešení: Pro $u \in \langle 0, 1\rangle$ jsou hustoty $f_X(u) = 2\,u$, $f_{X^2}(u) = 1$ (jinde nulové). To znamená, že X^2 má rovnoměrné rozdělení.

Cvičení 3.10.12: Náhodná veličina X má rovnoměrné diskrétní rozdělení s hodnotami $1, 2, 3, 4, 5, 6$. Zobrazte ji funkcí $h(u) = \cos \frac{1}{3}\pi\,u$. (Interpretace: náhodně vybereme vrchol pravidelného šestiúhelníka, funkční hodnota je jeho kartézská souřadnice.)

Řešení: Obrazem diskrétní náhodné veličiny bude opět diskrétní náhodná veličina $h(X)$ (protože počet hodnot se nezvýší). Její pravděpodobnostní funkce má následující nenulové hodnoty:

$$\begin{aligned} p_{h(X)}(1) &= P[h(X) = 1] = P[X = 6] = \tfrac{1}{6}, \\ p_{h(X)}\left(\tfrac{1}{2}\right) &= P\left[h(X) = \tfrac{1}{2}\right] = P[X = 1] + P[X = 5] = \tfrac{1}{3}, \\ p_{h(X)}\left(\tfrac{-1}{2}\right) &= P\left[h(X) = \tfrac{-1}{2}\right] = P[X = 2] + P[X = 4] = \tfrac{1}{3}, \\ p_{h(X)}(-1) &= P[h(X) = -1] = P[X = 3] = \tfrac{1}{6}. \end{aligned}$$

Cvičení 3.10.13: Náhodná veličina X má hustotu

$$f_X(u) = \begin{cases} c\,(u + 2) & \text{pro } u \in (-2, 1), \\ 0 & \text{jinak,} \end{cases}$$

kde $c \in \mathbb{R}$. Znázorněte a popište rozdělení náhodných veličin (a) $-2\,X$, (b) $X + 2$, (c) $\max(X, 0)$.

Cvičení 3.10.14: Náhodná veličina X má diskrétní rovnoměrné rozdělení s hodnotami $-1, 0, 1, 2$. Popište a znázorněte rozdělení náhodných veličin (a) $X - 1/2$, (b) $2\,X$, (c) $|X|$.

Cvičení 3.10.15: Náhodná veličina X má spojité rovnoměrné rozdělení na intervalu $\langle -1, 2\rangle$. Popište a znázorněte rozdělení náhodných veličin (a) $X + 3$, (b) $3\,X$, (c) $|X|$.

Cvičení 3.10.16: Náhodná veličina X má distribuční funkci

$$F_X(u) = \begin{cases} 1 - \mathrm{e}^{-x-2} & \text{pokud } x \geq -1, \\ 0 & \text{jinak.} \end{cases}$$

Určete a znázorněte rozdělení náhodných veličin (a) $X + 2$, (b) $X/2$, (c) $-2\,X$.

Cvičení 3.10.17: [Rogalewicz 2000] Najděte rozdělení, střední hodnotu a rozptyl náhodné veličiny $\min(X, 1)$, jestliže X má exponenciální rozdělení $\mathrm{Ex}(1)$.

4. Základní charakteristiky náhodných veličin

4.1 Střední hodnota

Velmi často se u náhodných veličin ptáme na jejich „průměrné" hodnoty. Ty se používají např. pro porovnání vlastností velkých souborů. Přesným matematickým pojmem pro tento účel je **střední hodnota** EX náhodné veličiny X. Kupodivu její zavedení není jednoduché a studentům dělá často problémy. Čtenář, kterému nevadí postup „definice–věta", může přeskočit k definici 4.1.3 a větě 4.1.4, kde se dozví, jak střední hodnotu počítat. Kdo chce pochopit další souvislosti, měl by projít i následující vysvětlení (volně upraveno dle [Rozanov 1971, Zvára, Štěpán 2002]). Na to se budeme odkazovat i u vlastností střední hodnoty.

Co nám říká speciální případ

Nejdříve uvažujme *nezápornou* náhodnou veličinu X s *konečně mnoha* hodnotami. Vzorec (1.1) posloužil pouze v Laplaceově modelu, kde všechny elementární jevy byly stejně pravděpodobné. I tam bychom ale efektivněji sčítali přes hodnoty než přes elementární jevy (hodnot náhodné veličiny určitě není více), tedy podle vzorce

$$EX = \sum_{u \in O_X} u \cdot p_X(u)\,, \tag{4.1}$$

kde sčítáme přes konečný obor hodnot O_X náhodné veličiny X. V Laplaceově modelu platí pro pravděpodobnosti

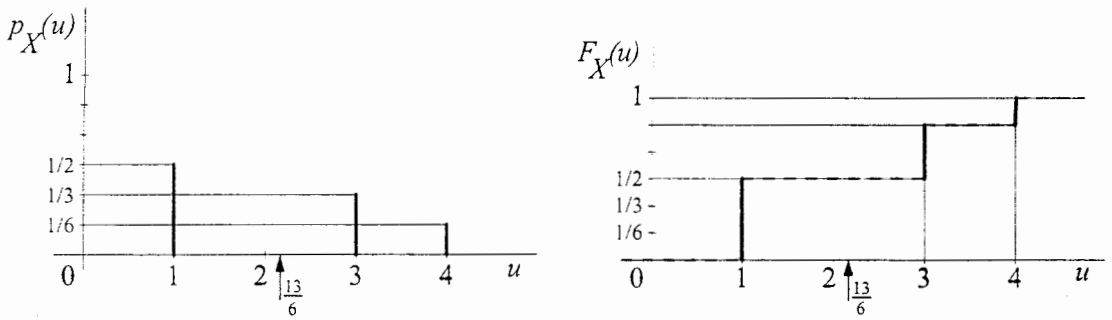
$$p_X(u) = \frac{\left|X^{-1}(u)\right|}{|\Omega|}\,,$$

kde $X^{-1}(u) = \{\omega \in \Omega \mid X(\omega) = u\}$. V rozšířeném Laplaceově modelu, kde jsou tyto pravděpodobnosti téměř libovolné, vzorec (4.1) zůstává v platnosti. Vyjdeme z něj i v Kolmogorovově modelu. *(To je obvyklý požadavek na zobecnění, kdy zachováme vyhovující původní pojmy a přidáme k nim něco nového, v tomto případě např. střední hodnotu náhodných veličin s nekonečně mnoha hodnotami.)*

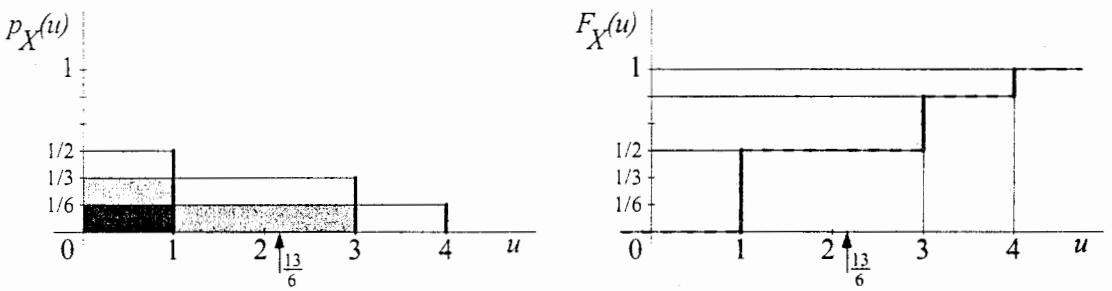
Podle vzorce (4.1) je střední hodnota váženým průměrem všech hodnot náhodné veličiny, přičemž váhy $p_X(u)$ jsou jejich pravděpodobnosti. Lze ji názorně interpretovat jako vodorovnou souřadnici těžiště úseček, tvořících čárový graf pravděpodobnostní funkce (viz obr. 4.1). Stejně tak ji lze vyčíst ze stejně dlouhých úseček odpovídajících skokům distribuční funkce. Ukážeme to na náhodné veličině X, která nabývá hodnot $1, 3, 4$ s pravděpodobnostmi po řadě $1/2, 1/3, 1/6$; její střední hodnota je

$$\frac{1}{2} \cdot 1 + \frac{1}{3} \cdot 3 + \frac{1}{6} \cdot 4 = \frac{13}{6}\,.$$

Příspěvek hodnoty u ke střední hodnotě je obsah obdélníka o stranách $u, p_X(u)$; tyto obdélníky lze opět zobrazit v grafu pravděpodobnostní, nebo lépe distribuční funkce (kde se nepřekrývají, viz obr. 4.2):



Obrázek 4.1. Stanovení střední hodnoty z pravděpodobnostní nebo distribuční funkce



Obrázek 4.2. Stanovení střední hodnoty z pravděpodobnostní nebo distribuční funkce

Vidíme, že střední hodnota má význam plochy nad grafem distribuční funkce, omezené shora jedničkou a zleva nulou (svislou osou). Ta je dána integrálem

$$\int_0^\infty (1 - F_X(u))\, \mathrm{d}u\,.$$

Je to současně plocha pod grafem *kvantilové* funkce q_X (která nabývá hodnotu $u \in O_X$ na intervalu délky $p_X(u)$, viz obr. 4.3), tedy integrál

$$\int_0^1 q_X(\alpha)\, \mathrm{d}\alpha\,. \tag{4.2}$$

Pro konstantní náhodnou veličinu s hodnotou $\mathrm{E}X$ by byla tato plocha stejně velká.

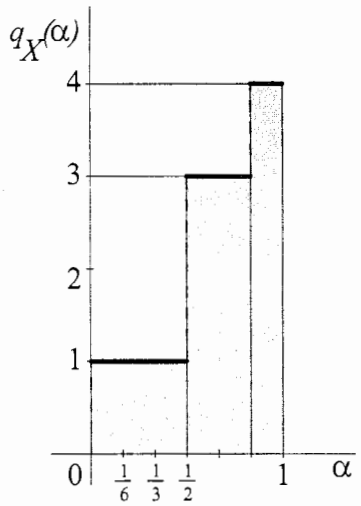
Kvantilová funkce závisí lineárně na *hodnotách* u náhodné veličiny, distribuční (i pravděpodobnostní) funkce závisí lineárně na *pravděpodobnostech* $p_X(u)$. Střední hodnota dle vzorce (4.1) je *bilineární*, tj. závisí lineárně jak na *hodnotách* u náhodné veličiny, tak na jejich *pravděpodobnostech* $p_X(u)$.To má následující klíčový důsledek:

Lemma 4.1.1: *Nechť X, U, V jsou nezáporné náhodné veličiny s konečně mnoha hodnotami, $r, s \in \mathbb{R}$. Jestliže platí pro kvantilové funkce $q_X = r\, q_U + s\, q_V$ nebo pro distribuční funkce $F_X = r\, F_U + s\, F_V$, pak střední hodnoty splňují rovnost $\mathrm{E}X = r\,\mathrm{E}U + s\,\mathrm{E}V$.*

Předchozí lemma se uplatní zejména pro směs $X = \mathrm{Mix}_w(U, V)$; pak $F_X = F_{\mathrm{Mix}_w(U,V)} = w\, F_U + (1 - w)\, F_V$ a

$$\mathrm{E}\,\mathrm{Mix}_w(U, V) = w\,\mathrm{E}U + (1 - w)\,\mathrm{E}V \tag{4.3}$$

(to vyplývá i přímo z tvrzení 3.5.5). Pro $X = r\,U + s$ máme $q_X = q_{r\,U+s} = r\, q_U + s$. Položíme $V = 1$ (konstantní náhodná veličina) a dostaneme



Obrázek 4.3. Střední hodnota z kvantilové funkce

$$\mathrm{E}\,(r\,U + s) = r\,\mathrm{E}U + s\,. \tag{4.4}$$

Bohužel stejný trik nelze použít pro součet náhodných veličin $U+V$, jeho rozdělení je složitější. Zde se musíme vrátit k podrobnějšímu popisu [Zvára, Štěpán 2002]:

$$\begin{aligned}
\mathrm{E}\,(U+V) &= \sum_{u\in O_U}\sum_{v\in O_V} (u+v)\;P\,[U=u, V=v] = \\
&= \sum_{u\in O_U}\sum_{v\in O_V} u\,P\,[U=u, V=v] + \sum_{u\in O_U}\sum_{v\in O_V} v\,P\,[U=u, V=v] = \\
&= \sum_{u\in O_U} u \sum_{v\in O_V} P\,[U=u, V=v] + \sum_{v\in O_V} v \sum_{u\in O_U} P\,[U=u, V=v] = \\
&= \sum_{u\in O_U} u\,P\,[U=u] + \sum_{u\in O_U} v\,P\,[V=v] = \mathrm{E}U + \mathrm{E}V\,.
\end{aligned}$$

Je-li $U \le V$, pak $\mathrm{E}U \le \mathrm{E}U + \mathrm{E}\,(V-U) = \mathrm{E}V$.

Střední hodnotu součinu náhodných veličin nelze obecně vypočítat z jejich středních hodnot, je to možné pouze pro *nezávislé* náhodné veličiny:

$$\begin{aligned}
\mathrm{E}\,(U\,V) &= \sum_{u\in O_U}\sum_{v\in O_V} u\,v\,P\,[U=u, V=v] = \sum_{u\in O_U}\sum_{v\in O_V} u\,v\,P\,[U=u]\cdot P\,[V=v] = \\
&= \Big(\sum_{u\in O_U} u\,P\,[U=u]\Big)\cdot\Big(\sum_{v\in O_V} v\,P\,[V=v]\Big) = \mathrm{E}U\cdot\mathrm{E}V\,.
\end{aligned}$$

Zobrazení nezápornou funkcí h se projeví jen drobnou změnou vzorce (4.1):

$$\mathrm{E}\,(h(U)) = \sum_{u\in O_U} h\,(u)\cdot p_U(u) = \int_0^1 h\,(q_X(\alpha))\;\mathrm{d}\alpha\,. \tag{4.5}$$

Zobecnění

Diskrétní náhodná veličina může nabývat nekonečně mnoha hodnot. I pak lze použít předchozí výsledky, pokud ovšem příslušné sumy konvergují, což nemusí být splněno:

Příklad 4.1.2 (neexistence střední hodnoty): Mějme diskrétní náhodnou veličinu Z s pravděpodobnostní funkcí

$$p_Z(k) = \frac{c}{k^2}\,, \qquad k\in\mathbb{N}\,,$$

kde $c = \frac{6}{\pi^2}$ na základě podmínky $\sum_{k=1}^{\infty} p_Z(k) = 1$ a vzorce (C.3). Vzorec pro střední hodnotu (4.1) dává známou divergentní sumu (C.2):

$$\sum_{k=1}^{\infty} k\, p_Z(k) = c \sum_{k=1}^{\infty} \frac{1}{k} = \infty\,.$$

I kdybychom připustili nekonečnou střední hodnotu, nemohli bychom ji použít pro další výpočty. Kromě toho jednoduchou úpravou získáme příklad náhodné veličiny, jejíž střední hodnotu nelze vůbec definovat. Např. pro diskrétní náhodnou veličinu W s pravděpodobnostní funkcí

$$p_W(k) = p_W(-k) = \frac{c}{2k^2}\,, \qquad k \in \mathbb{N}\,,$$

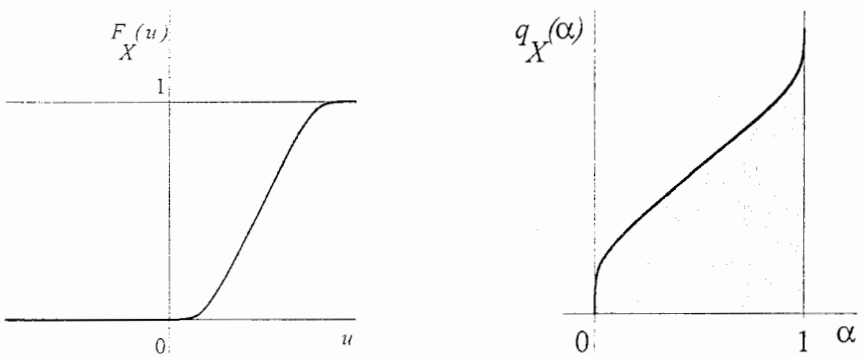
dostáváme sumu

$$\sum_{k=-\infty}^{\infty} k\, p_W(k) = \frac{c}{2} \sum_{k=-\infty}^{\infty} \frac{1}{k}\,,$$

která není definována, neboť není absolutně konvergentní a obsahuje členy s různými znaménky.

Podobně lze zkonstruovat příklady spojitých náhodných veličin, které nemají střední hodnotu.

Nezápornou náhodnou veličinu X lze podle tvrzení 3.8.3 vyjádřit jako limitu posloupnosti nezáporných diskrétních náhodných veličin. Jejich střední hodnoty konvergují k hodnotě dané integrálem (4.2), který lze opět interpretovat jako plochu nad grafem distribuční funkce, resp. pod grafem kvantilové funkce (viz obr. 4.4). Pokud je tato hodnota konečná, prohlásíme ji za střední hodnotu EX. *(Uvedený postup je korektní jen díky tomu, že limita středních hodnot nezávisí na volbě posloupnosti. Podrobné zdůvodnění zde neuvádíme, je však důsledkem toho, že konstrukce integrálu v matematické analýze kopíruje aproximaci z tvrzení 3.8.3. Přesné zavedení by vyžadovalo tzv.* Lebesgueův-Stieltjesův integrál.*)*

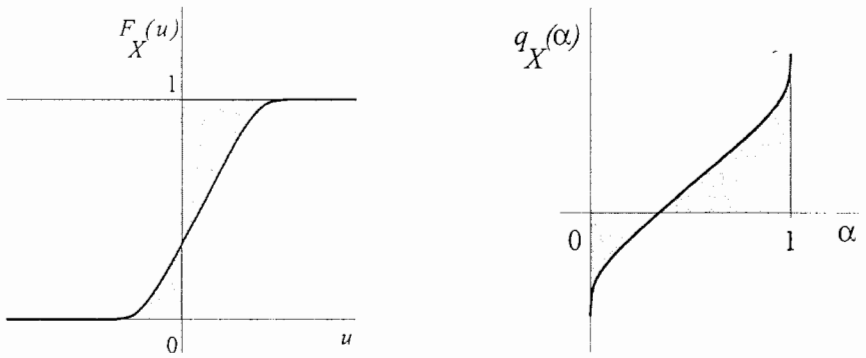


Obrázek 4.4. Střední hodnota jako plocha nad distribuční funkcí nebo pod kvantilovou funkcí nezáporné náhodné veličiny

Libovolnou náhodnou veličinu X lze rozložit na rozdíl dvou nezáporných náhodných veličin, $X = X^+ - X^-$. Pokud existují (konečné) střední hodnoty EX^+, EX^-, definujeme $EX = EX^+ - EX^-$. Opět dostáváme integrál (4.2) a lze jej interpretovat jako plochu pod křivkou, pouze je potřeba vzít plochu se záporným znaménkem tam, kde je kvantilová funkce záporná (viz obr. 4.5).

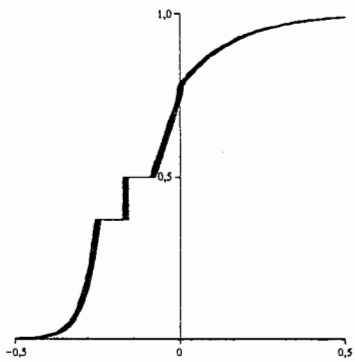
$$EX = \int_0^1 q_X(\alpha)\,\mathrm{d}\alpha = \int_0^{\infty} (1 - F_X(u))\,\mathrm{d}u - \int_{-\infty}^{0} F_X(u)\,\mathrm{d}u\,.$$

(Místo rozdílu jsme zde mohli použít také směs *nezáporné a nekladné náhodné veličiny a dospěli bychom přes jiné vzorce ke stejným výsledkům.)*



Obrázek 4.5. Střední hodnota jako plocha nad distribuční funkcí nebo pod kvantilovou funkcí obecné náhodné veličiny

Střední hodnota má i další názorný význam. Pro *spojitou* náhodnou veličinu je to vodorovná souřadnice těžiště plochy pod grafem hustoty. V obecném případě je střední hodnota vodorovnou souřadnicí těžiště grafu distribuční funkce, jsou-li jeho elementy váženy přírůstkem distribuční funkce (viz obr. 4.6).



Obrázek 4.6. Střední hodnota jako vodorovná souřadnice těžiště grafu distribuční funkce, váženého jejím přírůstkem

Dříve odvozené vlastnosti střední hodnoty se zachovávají limitou ve smyslu tvrzení 3.8.3 i přechodem od nezáporných náhodných veličin k obecným, takže zůstávají v platnosti a nebudeme je znovu dokazovat. V (4.5) navíc nemusí být funkce h nezáporná, pouze měřitelná, což není problém.

Někdy naopak odvodíme vlastnosti pro spojité rozdělení a rozšíříme na diskrétní rozdělení, které lze aproximovat spojitým dle kapitoly 3.8.

Obecná definice a vlastnosti

Definice 4.1.3: ***Střední hodnota*** *náhodné veličiny X je*

$$\mathrm{E}X = \int_0^1 q_X(\alpha)\,\mathrm{d}\alpha\,, \tag{4.6}$$

pokud tento integrál existuje a je absolutně konvergentní.

Alternativní značení: μ_X, μ.

Pro existenci integrálu je důležité, že kvantilová funkce má jen spočetně mnoho bodů nespojitosti (tvrzení 3.6.4); na výsledek nemá vliv, jak v nich byla definována. Zde uvedená definice střední hodnoty není obvyklá, běžně se zavádí zvlášť pro různé typy rozdělení:

Věta 4.1.4: *Střední hodnota* diskrétní *náhodné veličiny U s oborem hodnot O_U je*

$$\mathrm{E}U = \sum_{u \in O_U} u \cdot p_U(u)\,. \tag{4.7}$$

Střední hodnota spojité *náhodné veličiny V je*

$$\mathrm{E}V = \int_{-\infty}^{\infty} u \cdot f_V(u)\,\mathrm{d}u\,. \tag{4.8}$$

V obou předchozích případech požadujeme absolutní *konvergenci (sumy, resp. integrálu).*
Střední hodnota smíšené *náhodné veličiny X je*

$$\mathrm{E}X = w\,\mathrm{E}U + (1-w)\,\mathrm{E}V\,, \tag{4.9}$$

kde $X = \mathrm{Mix}_w(U,V)$, $w \in (0,1)$, U je diskrétní a V spojitá náhodná veličina (tento rozklad je dle věty 3.5.19 jednoznačný).

Důkaz: [Rozanov 1971] Nový je pouze vzorec (4.8). Použijeme stejnou diskretizaci jako v tvrzení 3.8.3, tj. na V aplikujeme funkci

$$h_\varepsilon(u) = \left\lfloor \frac{u}{\varepsilon} \right\rfloor \varepsilon\,,$$

kde $\varepsilon > 0$. Pak pro $n \in \mathbb{Z}$ je $h_\varepsilon(V) = n\,\varepsilon$, právě když $V \in \langle n\,\varepsilon, (n+1)\,\varepsilon)$, což nastává s pravděpodobností

$$P\left[h_\varepsilon(V) = n\,\varepsilon\right] = P\left[V \in \langle n\,\varepsilon, (n+1)\,\varepsilon)\right] = F_V\left((n+1)\,\varepsilon\right) - F_V\left(n\,\varepsilon\right) = \int_{n\,\varepsilon}^{(n+1)\,\varepsilon} f_V(u)\,\mathrm{d}u\,.$$

(Díky spojitosti F_V jsme nemuseli uvažovat její limity zleva.) Dosazením do vzorce pro střední hodnotu *diskrétní* náhodné veličiny $h_\varepsilon(V)$ dostaneme

$$\begin{aligned} Eh_\varepsilon(V) &= \sum_{n\in\mathbb{Z}} n\,\varepsilon\,P\left[h_\varepsilon(V) = n\,\varepsilon\right] = \sum_{n\in\mathbb{Z}} n\,\varepsilon \int_{n\,\varepsilon}^{(n+1)\,\varepsilon} f_V(u)\,\mathrm{d}u = \\ &= \sum_{n\in\mathbb{Z}} \int_{n\,\varepsilon}^{(n+1)\,\varepsilon} h_\varepsilon(u)\; f_V(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} h_\varepsilon(u)\; f_V(u)\,\mathrm{d}u\,. \end{aligned}$$

Pro $\varepsilon \to 0$ konverguje $h_\varepsilon(u) \to u$ a $\mathrm{E}h_\varepsilon(V) \to \mathrm{E}V$, čímž dostáváme (4.8). □

Úmluva: *Kdekoli nadále pracujeme se střední hodnotou, automaticky předpokládáme, že existuje a je konečná.*

Následující tvrzení byla odvozena pro nezáporné diskrétní náhodné veličiny, poté zobecněna přechodem k limitám a diskusí znamének.

Tvrzení 4.1.5: *Pro náhodnou veličinu X je*

$$\mathrm{E}X = \int_0^{\infty} \left(1 - F_X(u)\right)\,\mathrm{d}u - \int_{-\infty}^{0} F_X(u)\,\mathrm{d}u\,,$$

pokud jsou oba integrály konečné.

Pro existenci integrálu je důležité tvrzení 3.3.7, že distribuční funkce má jen spočetně mnoho bodů nespojitosti; na výsledek nemá vliv, jak v nich byla definována.

Tvrzení 4.1.6 (linearita střední hodnoty): *Jsou-li X, Y náhodné veličiny a $r, s \in \mathbb{R}$, pak*

$$\mathrm{E}\,(r\,X + s\,Y) = r\,\mathrm{E}X + s\,\mathrm{E}Y\,.$$

Důsledek 4.1.7:

$$\begin{aligned}
\mathrm{E}r &= r\,,\\
\mathrm{E}\,(X + r) &= \mathrm{E}X + r\,,\\
\mathrm{E}\,(r\,X) &= r\,\mathrm{E}X\,,\\
\mathrm{E}\,(X + Y) &= \mathrm{E}X + \mathrm{E}Y\,,\\
\mathrm{E}\,(X - Y) &= \mathrm{E}X - \mathrm{E}Y\,,\\
X \le Y &\implies \mathrm{E}X \le \mathrm{E}Y\,.
\end{aligned}$$

Tvrzení 4.1.8 (střední hodnota směsi): *Střední hodnota* směsi *náhodných veličin* $Z = \mathrm{Mix}_w(X, Y)$ *je*

$$\mathrm{E}Z = \mathrm{E}\,(\mathrm{Mix}_w(X, Y)) = w\,\mathrm{E}X + (1 - w)\,\mathrm{E}Y\,. \tag{4.10}$$

I směsi tedy v jistém smyslu „zachovávají" střední hodnotu. Vztah (4.10) *není* linearita střední hodnoty, neboť na levé straně je *směs*, nikoli *lineární kombinace* náhodných veličin.

Tvrzení 4.1.9 (střední hodnota součinu nezávislých náhodných veličin): *Jsou-li X, Y* nezávislé *náhodné veličiny, pak*

$$\mathrm{E}\,(X \cdot Y) = \mathrm{E}X \cdot \mathrm{E}Y\,.$$

Tvrzení 4.1.10: *Nechť X je náhodná veličina, $h\colon \mathbb{R} \to \mathbb{R}$ měřitelná funkce. Pak*

$$\mathrm{E}\,(h(X)) = \int_0^1 h\,(q_X(\alpha))\,\mathrm{d}\alpha\,.$$

Speciálně pro diskrétní *náhodnou veličinu*

$$\mathrm{E}\,(h(X)) = \sum_{u \in O_X} h\,(u) \cdot p_X(u)\,,$$

pro spojitou *náhodnou veličinu*

$$\mathrm{E}\,(h(X)) = \int_{-\infty}^{\infty} h\,(u) \cdot f_X(u)\,\mathrm{d}u\,.$$

Výhodou těchto vzorců je, že nepotřebujeme znát rozdělení náhodné veličiny $h(X)$, což by mohl být problém (h nemusí být monotónní). Nevadí ani to, že funkce spojité náhodné veličiny nemusí mít spojité rozdělení.

Interpretace

Poznámka 4.1.11: Když se hovoří o „průměrných" hodnotách, máme obvykle na mysli střední hodnotu náhodné veličiny. Nicméně není vždy zřejmé, *jaké* náhodné veličiny na *jakém* pravděpodobnostním prostoru, což může být i zneužito ke zkreslování informací.

Uvažme, co znamená např. „průměrná cena vajec". Může to být jedna z následujících veličin:

1. Průměrná cena, za kterou se vejce nabízejí v obchodě, tj. střední hodnota ceny v obchodě, náhodně vybraném ze všech obchodů se stejnou pravděpodobností.
2. Průměrná cena, za kterou spotřebitelé kupují vejce v obchodě, tj. střední hodnota ceny pro spotřebitele, náhodně vybraného ze všech spotřebitelů se stejnou pravděpodobností. (Ta bývá nižší, protože spotřebitelé častěji nakupují tam, kde jsou ceny nižší.)
3. Průměrná cena, za kterou se kupuje vejce v obchodě, tj. střední hodnota ceny vejce, náhodně vybraného ze všech prodaných vajec se stejnou pravděpodobností. (Ta bývá opět nižší, protože spotřebitelé nakupují více zboží tam, kde jsou ceny nižší.)

Všechny tři uvedené případy jsou přitom popsány jako střední hodnoty náhodných veličin, ale na různých pravděpodobnostních prostorech. Elementárním jevem je v prvním případě výběr obchodu, ve druhém spotřebitele, ve třetím vejce. Různé výsledky jsou zde způsobeny tím, že jde o různé náhodné pokusy a různé náhodné veličiny; ve všech případech je hledanou „průměrnou" cenou střední hodnota odpovídající náhodné veličiny. Mohli bychom také za elementární jev vzít vždy výběr vejce, ovšem s pravděpodobností úměrnou tomu, že právě toto vejce bude koupeno, tedy menší pro vejce v obchodech s vysokými cenami. Zde je potřeba připomenout, ze nemáme jen jedinou pravděpodobnost, ale máme jich celý prostor, viz kapitola 2.

Situace může být ještě komplikovanější, pokud se místo „průměrných" cen vajec budeme ptát třeba na „průměrnou" cenu benzínu. Ten se, na rozdíl od vajec, neprodává na kusy. Lze opět stejně popsat první dva případy (průměrnou cenu, za kterou se benzín nabízí u čerpadel a průměrnou cenu, kterou platí zákazník). Nicméně nelze k danému litru benzínu jednoznačně říci, kde, komu a za kolik byl prodán, neboť se prodává spojité množství. Přesto existuje i průměrná cena, za kterou se prodával litr benzínu (a získá se teoreticky snadno jako podíl tržeb a množství prodaného benzínu).

Pojem „průměr" nemusí mít bez dalšího upřesnění jednoznačný význam, což se může stát nástrojem zkreslování údajů. My zde vždy předpokládáme, že podmínky náhodného pokusu jsou stanoveny jeho matematickým popisem, takže další závěry jsou již jednoznačné.

Všechny předchozí situace však byly zjednodušené. Veškerá potřebná data byla (aspoň teoreticky) dostupná a náhodnost pokusu spočívala pouze ve výběru obchodu, zákazníka či zboží. Pokud však budeme hovořit o průměrné ceně v příštím roce, do pokusu zasáhnou další náhodné vlivy. Nicméně i pak si budeme muset vyjasnit, vzhledem k jaké pravděpodobnosti chceme počítat střední hodnotu.

Příklad 4.1.12 (nejvíce lidí čeká v nejdelších frontách): Předpokládáme, že počet lidí ve frontě má geometrické rozdělení s parametrem $q \in (0,1)$, tj. $p_X(k) = q^k(1-q)$. Jak dlouhá je „průměrná" fronta? V jak dlouhé frontě čeká „průměrný" člověk?

Řešení: Průměrná délka fronty je střední hodnota

$$\mathrm{E}X = \sum_{k=0}^{\infty} k\, p_X(k) = (1-q) \sum_{k=0}^{\infty} k\, q^k = \frac{q}{1-q}\,.$$

(Všimněte si, že tyto úvahy mají obecnou platnost bez ohledu na neznámý konečný počet front.) Počítali jsme průměr přes fronty. Ten odpovídá pokusu, kdy každá fronta má stejnou pravděpodobnost, že bude vybrána. Z pozice zákazníka ovšem uvažujeme rovnoměrné rozdělení přes lidi ve frontách, tedy pokus, kdy každý zákazník má stejnou pravděpodobnost, že bude tázán, v jak dlouhé frontě stojí. Pravděpodobnost $p_Z(k)$, že náhodně vybraný zákazník je ve frontě délky k, je úměrná $k\, p_X(k)$. z podmínky $\sum_{k=0}^{\infty} p_Z(k) = 1$ vychází

$$p_Z(k) = \frac{k\, p_X(k)}{\mathrm{E}X} = (1-q)^2\; k\, q^{k-1}\,.$$

Střední hodnota je

$$\mathrm{E}Z = \sum_{k=0}^{\infty} k\, p_Z(k) = (1-q)^2 \sum_{k=0}^{\infty} k^2\, q^{k-1} = \frac{1+q}{1-q} > \mathrm{E}X\,.$$

Tento výsledek nevypovídá přímo o tom, jak dlouhá je v průměru fronta ve chvíli, kdy do ní zákazník přichází, jen o tom, v jak dlouhé frontě stojí. □

Poznámka 4.1.13: Je třeba zdůraznit, že střední hodnota je obvykle skrytým parametrem rozdělení. Předpokládáme, že pro daný pokus objektivně existuje, ale máme o ní jen nepřímé informace z realizací, které nedovolují ji přímo měřit. (Více o tom bude v kapitole 9.2.)

Jestliže jsme z urny vytáhli (s opakováním) 10 bílých koulí a 90 černých, pak máme důvod se domnívat, že je v ní 10 % bílých koulí a 90 % černých (aniž bychom měli z čeho usuzovat na jejich počet). Je to však jen přibližný údaj, skutečnost může být jiná. Nemůžeme vyloučit ani to, že je v urně např. 90 % bílých koulí a 10 % černých. Dokonce musíme připustit i možnost, že je v ní např. 10 % bílých koulí, 10 % černých a 80 % červených (z nichž jsme náhodou dosud žádnou nevytáhli).

V tomto příkladu můžeme možná nahlédnout do urny a podívat se, co je v ní. Tím se dovíme pravděpodobnostní model, který generoval výsledky našeho pokusu, a tím i střední hodnoty náhodných veličin na něm definované. To je však netypický příklad. Obvykle takovou možnost nemáme.

Střední hodnota je základním prostředkem pro vyjádření „průměrných" údajů. Kupodivu není snadné ji zavést tak, aby se postihlo diskrétní a spojité rozdělení i obecný případ smíšeného rozdělení. Pokud neznáme Lebesgueův-Stieltjesův integrál, může v tom pomoci kvantilová funkce. Když už jsme tento pracný postup absolvovali, v maximální míře jej využijeme u budoucích pojmů, které často zavedeme pomocí střední hodnoty, aniž bychom znovu diskutovali každý typ rozdělení zvlášť.

Je třeba zdůraznit, že střední hodnota (na rozdíl od jejích odhadů) bývá skrytá, nepřístupná přímému pozorování, a dokonce nemusí vůbec existovat.

Cvičení 4.1.14: Určete střední hodnotu počtu zásahů ze cvičení 1.5.26, 3.3.9.

Řešení: $\mathrm{E}X = 0.016 \cdot 0 + 0.212 \cdot 1 + 0.628 \cdot 2 + 0.144 \cdot 3 = 1.9$. □

Cvičení 4.1.15: [Zvára, Štěpán 2002, Příklad 6.5] Určete střední hodnotu počtu dvojic kostek ze cvičení 3.3.8.

Řešení: $\mathrm{E}X = \frac{5}{9} \cdot 0 + \frac{5}{12} \cdot 1 + \frac{1}{36} \cdot 3 = \frac{1}{2}$. □

Cvičení 4.1.16: Určete střední hodnotu Poissonova rozdělení $\mathrm{Po}(w)$.

Řešení:

$$\sum_{k=0}^{\infty} \frac{k\,w^k}{k!}\,\mathrm{e}^{-w} = \mathrm{e}^{-w} \sum_{k=1}^{\infty} \frac{k\,w^k}{k!} = \mathrm{e}^{-w} \sum_{k=1}^{\infty} \frac{w^k}{(k-1)!} = w\,\underbrace{\mathrm{e}^{-w} \sum_{m=0}^{\infty} \frac{w^m}{m!}}_{1} = w$$

s využitím (C.6). □

Cvičení 4.1.17: Uveďte příklad náhodných veličin X, Y, pro něž $\mathrm{E}(X \cdot Y) \neq \mathrm{E}X \cdot \mathrm{E}Y$.

Cvičení 4.1.18: Zdůvodněte, proč $\mathrm{E}\mathrm{E}X = \mathrm{E}X$.

4.2 Medián

Střední hodnota jako charakteristika polohy má své jasné výhody, ale i nedostatky. Např. naprostá většina lidí má nadprůměrný počet nohou. Rovněž každé zveřejnění průměrných platů rozzlobí zhruba 2/3 zaměstnanců, jejichž plat je podprůměrný. Zde je to tím, že kladné odchylky mohou nabývat mnohem větších hodnot než záporné. (V případě počtu nohou naopak.) Medián

by v případě platů měl být takovou hodnotou, že polovina lidí má plat menší, polovina větší. (Musíme se ještě nějak vyrovnat s možností, že někteří mají plat právě takový.) V případě počtu nohou lze očekávat, že medián bude 2. Pokud do souboru přibude jedinec, značně se vymykající průměru, pak může citelně ovlivnit střední hodnotu, nikoli však medián. To odpovídá tomu, že přestěhováním boháče do chudé země se platové podmínky průměrného obyvatele nezmění. Vypovídací hodnota mediánu je tedy cenná, přesto bývá medián často opomíjen a bývá upřednostňována střední hodnota. Výpočetní důvody pro to uvedeme v kapitole 9.7. Zde uvedeme některé výhody a nevýhody mediánu ve srovnání se střední hodnotou; některé skutečnosti jsou jen opakováním obecnějších vlastností kvantilové funkce z kapitoly 3.6.

Na rozdíl od střední hodnoty medián vždy existuje (i když byl problém, jak jej jednoznačně definovat). Další výhodou mediánu jsou jeho vlastnosti vůči monotónním zobrazením.

Tvrzení 4.2.1: *Je-li X náhodná veličina a h monotónní funkce, pak*

$$q_{h(X)}(1/2) = h(q_X(1/2)),$$

(neboli medián funkční hodnoty je funkční hodnotou mediánu), pokud kvantilová funkce q_X je v bodě $1/2$ spojitá.

Kdyby např. daně závisely pouze na výši platu (monotónně), pak medián daní zaměstnance by byl daní vypočtenou z mediánu platů. O střední hodnotě to říci nelze a odhad střední hodnoty daní vyžaduje podrobnější znalosti o rozdělení.

Důsledek 4.2.2: *Je-li X náhodná veličina a $r, s \in \mathbb{R}$, pak*

$$q_{rX+s}(1/2) = r\, q_X(1/2) + s.$$

O mediánu součtu, součinu a směsi náhodných veličin nelze říci mnoho.

Tvrzení 4.2.3: *Jsou-li X, Y náhodné veličiny a $w \in \langle 0, 1\rangle$, pak medián směsi $\mathrm{Mix}_w(X, Y)$ splňuje nerovnosti*

$$\min(q_X(1/2), q_Y(1/2)) \leq q_{\mathrm{Mix}_w(X,Y)}(1/2) \leq \max(q_X(1/2), q_Y(1/2)).$$

Příklad 4.2.4: Předchozí tvrzení nelze v obecném případě zesílit. Vezmeme-li např. za X, Y konstantní náhodné veličiny $1, 0$, pak jejich směs má alternativní rozdělení s parametrem w. Medián tohoto rozdělení může být $0, 1/2$, nebo 1, v závislosti na tom, zda w je menší, rovno, nebo větší než $1/2$.

Medián může být užitečnou alternativou ke střední hodnotě, pokud chceme charakterizovat polohu, kolem níž jsou hodnoty náhodné veličiny rozloženy. Každý z těchto pojmů má své výhody a nevýhody.

4.3 Rozptyl

> „Když má soused k obědu celé kuře a já suším hubu, tak podle statistiky jsme měli každý půl kuřete."
>
> G.K. Chesterton

Známý spisovatel se v tomto případě hluboce mýlí. Zde probírané *charakteristiky polohy* (angl. *measures of central tendency*) – střední hodnota i medián – sice dávají v obou případech $1/2$, avšak známe další charakteristiky, zejména *charakteristiky variability* (angl. *measures of dispersion*). Podle těch se pozná, že nebyla všechna data stejná, dokonce, že nastal uvedený extrémní případ. Zde se budeme zabývat jednou z nich, a to *rozptylem.*

Poznámka 4.3.1: G.K. Chesterton se zde dopouští ještě vážnější chyby v nepochopení role statistiky, jak vysvětlíme v poznámce 8.1.6.

Definice 4.3.2: ***Rozptyl (disperze,*** *angl.* variance*)*** *náhodné veličiny X je*

$$\mathrm{D}X = \mathrm{E}\left((X - \mathrm{E}X)^2\right)$$

(pokud existuje).

Alternativní značení: σ_X^2, σ^2, $\operatorname{var} X$. Rozptyl je střední hodnota kvadrátu odchylky od střední hodnoty. Uvedený vzorec se odvolává na již zavedenou definici střední hodnoty, čímž se vyhýbá rozlišování diskrétních, spojitých a smíšených náhodných veličin. Často se používá ekvivalentní vzorec:

Tvrzení 4.3.3:

$$\mathrm{D}X = \mathrm{E}\left(X^2\right) - (\mathrm{E}X)^2 \tag{4.11}$$

Důkaz: $\mathrm{D}X = \mathrm{E}\left((X - \mathrm{E}X)^2\right) = \mathrm{E}\left(X^2 - 2\,X\,\mathrm{E}X + (\mathrm{E}X)^2\right) = \mathrm{E}\left(X^2\right) - 2\,\mathrm{E}X\,\mathrm{E}X + (\mathrm{E}X)^2 =$
$= \mathrm{E}\left(X^2\right) - (\mathrm{E}X)^2$. □

Z definice plyne, že rozptyl je vždy nezáporný, nulový je pouze pro náhodné veličiny s Diracovým rozdělením. Nelze jej definovat pro náhodné veličiny, které nemají střední hodnotu, ale i když ji mají, rozptyl nemusí existovat. Z tvrzení 4.1.10 vyplývají vzorce

$$\mathrm{D}X = \int_0^1 (q_X(\alpha) - \mathrm{E}X)^2 \,\mathrm{d}\alpha\,,$$

pro *diskrétní* náhodnou veličinu

$$\mathrm{D}X = \sum_{u\in\mathbb{R}} (u - \mathrm{E}X)^2 \cdot p_X(u) = \sum_{u\in O_X} (u - \mathrm{E}X)^2 \cdot p_X(u)\,,$$

pro *spojitou* náhodnou veličinu

$$\mathrm{D}X = \int_{-\infty}^{\infty} (u - \mathrm{E}X)^2 \cdot f_X(u)\,\mathrm{d}u\,.$$

Tvrzení 4.3.4 (vlastnosti rozptylu): *Je-li X náhodná veličina a $r \in \mathbb{R}$, pak*

$$\mathrm{D}(X + r) = \mathrm{D}X, \qquad \mathrm{D}(r\,X) = r^2\,\mathrm{D}X\,.$$

Důkaz: $\mathrm{D}(X + r) = \mathrm{E}\left((X + r - (\mathrm{E}X + r))^2\right) = \mathrm{E}\left((X - \mathrm{E}X)^2\right) = \mathrm{D}X,$
$\mathrm{D}(r\,X) = \mathrm{E}\left((r\,X - \mathrm{E}(r\,X))^2\right) = \mathrm{E}\left(r^2\,(X - \mathrm{E}X)^2\right) = r^2\,\mathrm{E}\left((X - \mathrm{E}X)^2\right) = r^2\,\mathrm{D}X.$ □

Tvrzení 4.3.5: *Jsou-li X, Y náhodné veličiny a $w \in \langle 0, 1\rangle$, pak*

$$\mathrm{D}(\mathrm{Mix}_w(X, Y)) = w\,\mathrm{D}X + (1 - w)\,\mathrm{D}Y + w\,(1 - w)\,(\mathrm{E}X - \mathrm{E}Y)^2\,.$$

Důkaz:

$$\begin{aligned}
\mathrm{D}(\mathrm{Mix}_w(X, Y)) &= \mathrm{E}\left(\mathrm{Mix}_w(X, Y)^2\right) - (\mathrm{E}(\mathrm{Mix}_w(X, Y)))^2 = \\
&= w\,\mathrm{E}\left(X^2\right) + (1 - w)\,\mathrm{E}\left(Y^2\right) - (w\,\mathrm{E}X + (1 - w)\,\mathrm{E}Y)^2 = \\
&= w\,\left(\mathrm{D}X + (\mathrm{E}X)^2\right) + (1 - w)\,\left(\mathrm{D}Y + (\mathrm{E}Y)^2\right) - \\
&\quad - \left(w^2\,(\mathrm{E}X)^2 + 2\,w\,(1 - w)\,\mathrm{E}X\,\mathrm{E}Y + (1 - w)^2\,(\mathrm{E}Y)^2\right) = \\
&= w\,\mathrm{D}X + (1 - w)\,\mathrm{D}Y + \\
&\quad + w\,(1 - w)\,(\mathrm{E}X)^2 - 2\,w\,(1 - w)\,\mathrm{E}X\,\mathrm{E}Y + w\,(1 - w)\,(\mathrm{E}Y)^2 = \\
&= w\,\mathrm{D}X + (1 - w)\,\mathrm{D}Y + w\,(1 - w)\,(\mathrm{E}X - \mathrm{E}Y)^2\,.
\end{aligned}$$

□

Tvrzení 4.3.6: *Jsou-li X, Y nezávislé náhodné veličiny, pak*

$$\mathrm{D}(X+Y) = \mathrm{D}(X-Y) = \mathrm{D}X + \mathrm{D}Y\,.$$

Důkaz:

$$\begin{aligned}\mathrm{D}(X+Y) &= \mathrm{E}\left((X+Y)^2\right) - (\mathrm{E}(X+Y))^2 = \\ &= \mathrm{E}(X^2) + 2\,\mathrm{E}(X\cdot Y) + \mathrm{E}(Y^2) - \left((\mathrm{E}X)^2 + 2\,\mathrm{E}X\cdot\mathrm{E}Y + (\mathrm{E}Y)^2\right) = \\ &= \mathrm{E}(X^2) - (\mathrm{E}X)^2 + \mathrm{E}(Y^2) - (\mathrm{E}Y)^2 = \mathrm{D}X + \mathrm{D}Y\,,\end{aligned}$$

kde poslední rovnost využívá (4.11). Nezávislost jsme potřebovali pro eliminaci výrazu $\mathrm{E}(X\cdot Y)$. □

Tvrzení 4.3.7: *Funkce $h(u) = \mathrm{E}\left((X-u)^2\right)$ nabývá svého minima, $\mathrm{D}X$, pro $u = \mathrm{E}X$.*

Důkaz:

$$\begin{aligned}h(u) &= \mathrm{E}\left((X-u)^2\right) = \mathrm{E}\left(((X-\mathrm{E}X) + (\mathrm{E}X-u))^2\right) = \\ &= \mathrm{E}\left((X-\mathrm{E}X)^2\right) + 2\,\mathrm{E}\left((X-\mathrm{E}X)\,(\mathrm{E}X-u)\right) + \mathrm{E}\left((\mathrm{E}X-u)^2\right) = \\ &= \mathrm{D}X + 2\,(\mathrm{E}X-u)\,\mathrm{E}(X-\mathrm{E}X) + (\mathrm{E}X-u)^2 = \mathrm{D}X + (\mathrm{E}X-u)^2\,.\end{aligned}$$

Vidíme, že $h(u)$ je kvadratickou funkcí u, která nabývá minima, právě když člen $(\mathrm{E}X-u)^2$ je nulový. □

Odmocnina z funkce h z předchozího tvrzení je **střední kvadratická odchylka** (angl. *mean square error*, MSE) náhodné veličiny X od u. Rovněž nabývá minima pro $u = \mathrm{E}X$.

Příklad 4.3.8: (Srovnej s příkladem 4.1.2.) Mějme diskrétní náhodnou veličinu U s pravděpodobnostní funkcí

$$p_U(k) = \frac{c}{k^3}\,, \qquad k \in \mathbb{N}\,,$$

kde kladnou konstantu c určíme z podmínky $\sum_{k=1}^{\infty} p_Z(k) = 1$ a vzorce (C.4). Dle (4.7) a (C.3) dostaneme střední hodnotu

$$\mathrm{E}U = \sum_{k=1}^{\infty} k\,p_U(k) = c\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{1}{6}\pi^2 c\,,$$

avšak rozptyl vychází dle (C.2)

$$\mathrm{D}U = \sum_{k=1}^{\infty}(k-\mathrm{E}U)^2\,p_U(k) = c\Big(\underbrace{\sum_{k=1}^{\infty}\frac{1}{k}}_{\infty} - 2\,\mathrm{E}U\sum_{k=1}^{\infty}\frac{1}{k^2} + (\mathrm{E}U)^2\sum_{k=1}^{\infty}\frac{1}{k^3}\Big) = \infty$$

Na stejném principu lze zkonstruovat příklady spojitých náhodných veličin, které nemají (konečný) rozptyl.

Úmluva. *Kdekoli nadále pracujeme se střední hodnotou, rozptylem a dalšími charakteristikami náhodných veličin, automaticky předpokládáme jejich existenci a konečnost. Pokud jimi ve vzorci dělíme, předpokládáme též nenulovost.* (Pokud tomu tak není, příslušný vzorec nelze uplatnit, tento předpoklad však již znovu neuvádíme.)

Rozptyl je jednou z nejčastěji užívaných charakteristik variability náhodné veličiny. Jeho vlastnosti budeme potřebovat pro výroky o přesnosti odhadů, zejména statistických.

4.4 Směrodatná odchylka

Definice 4.4.1: *Směrodatná odchylka* (*angl.* standard deviation) *náhodné veličiny* X *je*

$$\sigma_X = \sqrt{\mathrm{D}X} = \sqrt{\mathrm{E}\left((X - \mathrm{E}X)^2\right)}$$

(pokud existuje).

Je to vlastně střední kvadratická odchylka od střední hodnoty.

Poznámka 4.4.2: Na rozdíl od rozptylu má směrodatná odchylka stejný fyzikální rozměr jako původní náhodná veličina, což je důvodem, proč se užívá mnohem více. Je-li např. náhodnou veličinou napětí ve voltech (V), pak rozptyl má těžko interpretovatelný rozměr V^2, zatímco směrodatná odchylka se udává opět ve voltech. Nicméně matematická odvození vycházejí často jednodušší pro rozptyl.

Směrodatná odchylka je vždy nezáporná, nulová je pouze pro konstantní náhodné veličiny. Lze ji počítat i z kvantilové funkce:

$$\sigma_X = \sqrt{\int_0^1 (q_X(\alpha) - \mathrm{E}X)^2 \,\mathrm{d}\alpha}\,.$$

Tvrzení 4.4.3 (vlastnosti směrodatné odchylky): *Je-li* X *náhodná veličina a* $r \in \mathbb{R}$, *pak*

$$\sigma_{X+r} = \sigma_X\,, \qquad \sigma_{rX} = |r|\,\sigma_X\,.$$

Tvrzení 4.4.4: *Jsou-li* X, Y nezávislé *náhodné veličiny, pak*

$$\sigma_{X+Y} = \sqrt{\mathrm{D}X + \mathrm{D}Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}\,.$$

Směrodatná odchylka je asi nejpoužívanější charakteristikou variability náhodné veličiny. Oproti rozptylu má tu výhodu, že se měří ve stejných jednotkách jako původní náhodná veličina.

4.5 Normovaná náhodná veličina

Pro každou náhodnou veličinu X, která má nenulový rozptyl (a tedy má i střední hodnotu), definujeme jí příslušnou **normovanou** (též **standardizovanou**) **náhodnou veličinu** $\operatorname{norm} X$ vzorcem

$$\operatorname{norm} X = \frac{X - \mathrm{E}X}{\sigma_X}\,. \tag{4.12}$$

Ta má nulovou střední hodnotu a jednotkový rozptyl a splňuje

$$F_{\operatorname{norm} X}(u) = F_X(\sigma_X\, u + \mathrm{E}X)\,,$$
$$q_{\operatorname{norm} X} = \frac{q_X - \mathrm{E}X}{\sigma_X}\,.$$

Zpětná transformace je

$$X = \sigma_X \operatorname{norm} X + \mathrm{E}X\,, \tag{4.13}$$
$$F_X(u) = F_{\operatorname{norm} X}\left(\frac{u - \mathrm{E}X}{\sigma_X}\right)\,,$$
$$q_X = \sigma_X\, q_{\operatorname{norm} X} + \mathrm{E}X\,.$$

Pozor na rozdíl mezi *normovanou* náhodnou veličinou a náhodnou veličinou s *normálním* (=Gaussovým) *rozdělením* $\mathrm{N}(\mu, \sigma^2)$. *Normované normální rozdělení* $\mathrm{N}(0,1)$ je natolik důležité, že jsme mu vyhradili zvláštní symboly pro hustotu, distribuční a kvantilovou funkci: $\varphi = f_{\mathrm{N}(0,1)}$, $\Phi = F_{\mathrm{N}(0,1)}$, $\Phi^{-1} = q_{\mathrm{N}(0,1)}$. Najdeme je v každém souboru statistických programů nebo statistických tabulkách, zatímco ostatní normální rozdělení musíme často dopočítávat dle (4.13):

$$F_{\mathrm{N}(\mu,\sigma^2)}(u) = \Phi\left(\frac{u-\mu}{\sigma}\right),$$
$$f_{\mathrm{N}(\mu,\sigma^2)}(u) = \frac{1}{\sigma}\varphi\left(\frac{u-\mu}{\sigma}\right),$$
$$q_{\mathrm{N}(\mu,\sigma^2)} = \sigma\,\Phi^{-1} + \mu\,.$$

Normování náhodných veličin slouží mj. k tomu, abychom zmenšili počet tabulek nebo počítačových procedur nutných pro popis často používaných rozdělení.

4.6 Obecné momenty

Příklad 4.6.1 (kolik kapek je do litru?): Náhodná veličina R je poloměr kapek (dešťových, nebo z dávkovače léků, což je důležitější). Střední hodnota jejich objemu $V = \frac{4}{3}\pi R^3$ je $\mathrm{E}V = \frac{4}{3}\pi\,\mathrm{E}\left(R^3\right)$.

Mnohé předchozí pojmy lze dále zobecnit a doplnit tak popis náhodné veličiny o další charakteristiky.

Definice 4.6.2: *Nechť $k \in \mathbb{N}$. Pak k-tý* **obecný moment** *náhodné veličiny X je $\mathrm{E}\left(X^k\right)$ (pokud existuje).*

(Žádná speciální značení pro momenty zde nezavádíme. Závorky někdy vynecháváme: $\mathrm{E}X^k$. Alternativní značení: μ'_k nebo m_k.) Speciálně pro $k = 0$ bychom dostali $\mathrm{E}\left(X^0\right) = 1$, pro $k = 1$ střední hodnotu $\mathrm{E}X$, druhý obecný moment se vyskytl v (4.11).

Definice 4.6.3: *Nechť $k \in \mathbb{N}$. Pak k-tý* **centrální moment** *náhodné veličiny X je $\mathrm{E}((X - \mathrm{E}X)^k)$ (pokud existuje).*

(Žádná speciální značení pro centrální momenty zde nezavádíme. Alternativní značení: μ_k.) Speciálně pro $k = 0$ bychom dostali 1, pro $k = 1$ nulu, druhý centrální moment je rozptyl $\mathrm{D}X$.

Mezi druhým obecným a centrálním momentem platí vztah (4.11); obdobné vztahy lze odvodit i pro vyšší momenty. Dle tvrzení 4.1.10 lze momenty počítat též z kvantilové funkce:

$$\mathrm{E}\left(X^k\right) = \int_0^1 (q_X(\alpha))^k \,\mathrm{d}\alpha\,, \qquad \mathrm{E}\left((X - \mathrm{E}X)^k\right) = \int_0^1 (q_X(\alpha) - \mathrm{E}X)^k \,\mathrm{d}\alpha\,.$$

Jednotlivé momenty mají u spojitých rozdělení vztah k tvaru hustoty, zavádí se pomocí nich např. *koeficienty šikmosti* (angl. *skewness*), *špičatosti* (angl. *kurtosis*) atd.

Tvrzení 4.6.4: *Omezená náhodná veličina má všechny obecné i centrální momenty.*

Použití momentů vyšších řádů umožňuje doplnit popis rozdělení náhodné veličiny; střední hodnota a rozptyl o něm dávají jen velmi hrubou informaci. Nevýhodou je, že momenty vyšších řádů již nemají tak názorný význam.

Cvičení 4.6.5 (průměrná vzdálenost při hromadné obsluze): [Zvára, Štěpán 2002, 6.8] Obsluhujeme k pracovišť, umístěných v řadě v jednotkových vzdálenostech. Po obsluze jednoho přejdeme na další, které to vyžaduje, což může být kterékoli pracoviště (i to samé) se stejnou pravděpodobností. Jaká je průměrná délka cesty mezi obsluhovanými pracovišti? Jaký je rozptyl této náhodné veličiny?

Řešení: Za elementární jevy můžeme považovat dvojice pracovišť, mezi kterými přejdeme; každá dvojice má stejnou pravděpodobnost $1/k^2$. Tu máme násobit vzdáleností. Např. pro $k = 4$ bude střední hodnota $1/k^2\times$ součet čísel v následující tabulce:

$$\begin{array}{cccc} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{array}$$

Obecně sčítáním po diagonálách dostaneme dle (C.1)

$$\mathrm{E}X = \frac{2}{k^2}\sum_{j=1}^{k-1} j\,(k-j) = \frac{1}{3\,k}\,\left(k^2-1\right) = \frac{k}{3} - \frac{1}{3\,k}\,,$$

$$\mathrm{E}X^2 = \frac{2}{k^2}\sum_{j=1}^{k-1} j^2\,(k-j) = \frac{2}{k^2}\,\left(\frac{1}{12}\,k^4 - \frac{1}{12}\,k^2\right) = \frac{1}{6}\,\left(k^2-1\right)\,,$$

$$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{1}{6}\,\left(k^2-1\right) - \left(\frac{k}{3} - \frac{1}{3\,k}\right)^2 = \frac{1}{18}\,k^2 - \frac{1}{9\,k^2} + \frac{1}{18} =$$

$$= \frac{1}{18}\,\left(1 + \frac{2}{k^2}\right)\,(k+1)\,(k-1)\,.$$

□

Cvičení 4.6.6: Náhodné veličiny $X_k, k = 1, 2, 3$, jsou nezávislé a mají charakteristiky $\mathrm{E}X_k = k$, $\sigma_{X_k} = k$. Určete střední hodnotu a směrodatnou odchylku náhodné veličiny $Y = X_1 + 2X_2 - X_3$.

Řešení: $\mathrm{E}Y = \mathrm{E}X_1 + 2\,\mathrm{E}X_2 - \mathrm{E}X_3 = 2$,
$\sigma_Y = \sqrt{\mathrm{D}Y} = \sqrt{\mathrm{D}X_1 + 2^2\,\mathrm{D}X_2 + \mathrm{D}X_3} = \sqrt{1 + 2^2\cdot 4 + 9} = \sqrt{26} \doteq 5.099$. □

Cvičení 4.6.7: Náhodné veličiny X, Y mají charakteristiky $\mathrm{E}X = 1/3$, $\mathrm{D}X = 1/3$, $\mathrm{E}Y = 1/2$, $\mathrm{D}Y = 1/2$. Určete střední hodnotu a rozptyl náhodné veličiny $Z = 1/2 + X - Y$ za předpokladu, že X, Y jsou nezávislé. Co lze říci bez tohoto předpokladu?

Řešení: Platí $\mathrm{E}Z = 1/2 + \mathrm{E}X - \mathrm{E}Y = 1/3$ i bez předpokladu nezávislosti, $\mathrm{D}Z = \mathrm{D}X + \mathrm{D}Y = 5/6$ pouze za předpokladu nezávislosti; obecně $\sigma_Z \in \langle |\sigma_X - \sigma_Y|\,, \sigma_X + \sigma_Y\rangle$,

$$\mathrm{D}Z \in \langle (\sigma_X - \sigma_Y)^2\,, (\sigma_X + \sigma_Y)^2\rangle = \left\langle \left(\tfrac{1}{\sqrt{3}} - \tfrac{1}{\sqrt{2}}\right)^2, \left(\tfrac{1}{\sqrt{3}} + \tfrac{1}{\sqrt{2}}\right)^2 \right\rangle \doteq \langle 0.01683, 1.6498\rangle\,.$$

□

Cvičení 4.6.8: Nezávislé náhodné veličiny X, Y mají spojité rovnoměrné rozdělení na intervalu $\langle 0, 2\rangle$, resp. $\langle 0, 3\rangle$. Určete střední hodnotu a rozptyl veličiny $Y - 2\,X$.

Řešení: $\mathrm{E}\,(Y - 2\,X) = \mathrm{E}Y - 2\,\mathrm{E}X$, $\qquad \mathrm{D}\,(Y - 2\,X) = \mathrm{D}Y + 2^2\,\mathrm{D}X$.

□

veličina	X	Y	$Y - 2\,X$
střední hodnota	1	3/2	−1/2
rozptyl	1/3	3/4	25/12

Cvičení 4.6.9: Diskrétní náhodné veličiny X, Y jsou nezávislé, X má rovnoměrné rozdělení na množině $\{1, 2, 4\}$, Y nabývá hodnot z $\{0, 1\}$, $P\,[Y = 1] = 2/3$. Určete střední hodnotu a rozptyl veličiny $2\,Y - X$.

Řešení: $\mathrm{E}\,(2\,Y - X) = 2\,\mathrm{E}Y - \mathrm{E}X$, $\qquad \mathrm{D}\,(2\,Y - X) = 2^2\,\mathrm{D}Y + \mathrm{D}X$.

□

veličina	X	Y	$2\,Y - X$
střední hodnota	7/3	2/3	−1
rozptyl	14/9	2/9	22/9

Cvičení 4.6.10: Nezávislé náhodné veličiny X, Y mají střední hodnotu μ a rozptyl σ^2. Definujeme náhodné veličiny $U = 2X + 3$, $V = X + Y$, $W = X + \mathrm{E}Y$. Určete jejich střední hodnoty a rozptyly.

Cvičení 4.6.11: Náhodné veličiny X, Y mají stejné rozdělení se střední hodnotou μ a rozptylem σ^2. Určete střední hodnotu a rozptyl náhodných veličin (a) $X - Y$, (b) $X + Y$, (c) $2X$, (d) $-3Y$, (e) $\mathrm{Mix}_{1/3}(X, Y)$.

Cvičení 4.6.12: Náhodná veličina Z je konvexní kombinací dvou nezávislých náhodných veličin X, Y, $Z = wX + (1 - w)Y$, $w \in \langle 0, 1 \rangle$. Pro jakou hodnotu w je rozptyl náhodné veličiny Z minimální?

Cvičení 4.6.13: Náhodná veličina W je aritmetický průměr tří nezávislých náhodných veličin X, Y, Z, $W = (X + Y + Z)/3$, jejichž střední hodnoty a rozptyly jsou známy. Určete střední hodnotu a rozptyl náhodné veličiny W. Které výsledky platí i bez předpokladu nezávislosti?

4.7 Charakteristická funkce náhodné veličiny

(Dle [Rogalewicz 2000].)

Tato kapitola nabízí jiný pohled na reprezentaci rozdělení náhodné veličiny. Opírá se o náročnější aparát matematické analýzy, zejména Fourierovu transformaci. Proto je skriptum uzpůsobeno tak, aby čtenář neznalý příslušných partií mohl pokračovat i po vynechání této kapitoly. Bude však ochuzen o zajímavý pohled a bez výsledků zde uvedených nelze dokázat některé klíčové věty.

Budeme pracovat s *komplexními* náhodnými veličinami, nicméně budeme používat pro reálnou i imaginární část postupy obvyklé u reálných náhodných veličin.

Definice 4.7.1: ***Střední hodnota komplexní náhodné veličiny*** $X = U + \mathrm{i}V$ *je* $\mathrm{E}X = \mathrm{E}U + \mathrm{i}\,\mathrm{E}V$.

Definice 4.7.2: ***Charakteristická funkce náhodné veličiny*** X *je komplexní funkce* $\Psi_X : \mathbb{R} \to \mathbb{C}$ *definovaná vztahem*

$$\Psi_X(\omega) = \mathrm{E}\exp(\mathrm{i}\omega X) = \mathrm{E}\cos(\omega X) + \mathrm{i}\,\mathrm{E}\sin(\omega X)\,.$$

Alternativní značení: Ψ (pokud je zřejmé, o jakou náhodnou veličinu jde). V této kapitole výjimečně ω neznačí elementární jev, ale argument charakteristické funkce, aby se zdůraznila analogie s Fourierovou transformací. Fyzikální rozměr proměnné ω je převrácená hodnota rozměru náhodné veličiny X, např. je-li X čas, má ω rozměr frekvence.

Z tvrzení 4.1.10 dostáváme pro *diskrétní* náhodnou veličinu

$$\Psi_X(\omega) = \sum_{u \in \mathbb{R}} \exp(\mathrm{i}\omega u)\, p_X(u) \tag{4.14}$$

(kde sčítat stačí přes hodnoty, kterých X nabývá) a pro *spojitou* náhodnou veličinu

$$\Psi_X(\omega) = \int_{-\infty}^{\infty} \exp(\mathrm{i}\omega u)\, f_X(u)\, \mathrm{d}u\,, \tag{4.15}$$

což je vzorec pro Fourierovu transformaci, s jedinou výjimkou, že v exponentu není záporné znaménko, tj. ω bereme s opačným znaménkem. Bohužel se tento pojem historicky ustálil, takže pro charakteristickou funkci náhodné veličiny platí obdobné zákony jako pro Fourierovu transformaci, ale v některých vzorcích jsou opačná znaménka.

Charakteristickou funkci *smíšené* náhodné veličiny lze počítat z rozkladu na směs diskrétní a spojité náhodné veličiny (který je dle věty 3.5.19 jednoznačný) a následujícího důsledku tvrzení 3.10.4 a 4.1.8:

Tvrzení 4.7.3 (charakteristická funkce směsi): *Charakteristická funkce směsi náhodných veličin* $\mathrm{Mix}_w(U,V)$ *je*

$$\Psi_{\mathrm{Mix}_w(U,V)} = w\,\Psi_U + (1-w)\,\Psi_V\,.$$

Tvrzení 4.7.4 (vlastnosti charakteristické funkce): *Charakteristická funkce* Ψ_X *náhodné veličiny* X *je definována v každém bodě, je spojitá a má následující vlastnosti:*

1. $\Psi_X(0) = 1\,.$
2. $\forall\omega\in\mathbb{R} : |\Psi_X(\omega)| \le 1\,.$
3. $\forall a,b\in\mathbb{R} : \Psi_{a\,X+b}(\omega) = \exp(\mathrm{i}\,\omega\,b)\,\Psi_X(a\,\omega)\,.$

Důkaz: Ukážeme důkaz (kromě spojitosti) alespoň pro spojitou náhodnou veličinu. Definiční integrál je absolutně konvergentní, neboť

$$\int_{-\infty}^{\infty} |\exp(\mathrm{i}\,\omega\,u)\;f_X(u)|\;\mathrm{d}u = \int_{-\infty}^{\infty} f_X(u)\;\mathrm{d}u = 1\,.$$

$$\Psi_X(0) = \mathrm{E}\exp(0) = \mathrm{E}\,1 = 1\,.$$

$$|\Psi_X(\omega)| = \left|\int_{-\infty}^{\infty} \exp(\mathrm{i}\,\omega\,u)\;f_X(u)\;\mathrm{d}u\right| \le \int_{-\infty}^{\infty} |\exp(\mathrm{i}\,\omega\,u)\;f_X(u)|\;\mathrm{d}u = \int_{-\infty}^{\infty} f_X(u)\;\mathrm{d}u = 1\,.$$

$$\Psi_{a\,X+b}(\omega) = \mathrm{E}\exp(\mathrm{i}\,\omega\;(a\,X+b)) = \exp(\mathrm{i}\,\omega\,b)\;\mathrm{E}\exp(\mathrm{i}\,\omega\,a\,X) = \exp(\mathrm{i}\,\omega\,b)\;\Psi_X(a\,\omega)\;.$$

Vzorec (4.14) je pouze vyjádření charakteristické funkce směsi Diracových rozdělení.

Tvrzení 4.7.5: *Charakteristická funkce normálního rozdělení* $\mathrm{N}(\mu,\sigma^2)$ *je*

$$\Psi_{\mathrm{N}(\mu,\sigma^2)}(\omega) = \exp\left(\mathrm{i}\,\mu\,\omega - \frac{1}{2}\,\sigma^2\,\omega^2\right)\,, \tag{4.16}$$

speciálně

$$\Psi_{\mathrm{N}(0,1)}(\omega) = \exp\left(\frac{-\omega^2}{2}\right)\,. \tag{4.17}$$

Důkaz: Nejdříve uvažujme *normované* normální rozdělení $\mathrm{N}(0,1)$.

$$\begin{aligned}\Psi_{\mathrm{N}(0,1)}(\omega) &= \int_{-\infty}^{\infty} \exp(\mathrm{i}\,\omega\,u)\;\varphi(u)\;\mathrm{d}u = \int_{-\infty}^{\infty} \exp(\mathrm{i}\,\omega\,u)\;\frac{1}{\sqrt{2\,\pi}}\;\exp\left(\frac{-u^2}{2}\right)\;\mathrm{d}u = \\ &= \frac{1}{\sqrt{2\,\pi}}\int_{-\infty}^{\infty} \exp\left(\frac{-u^2+2\,\mathrm{i}\,\omega\,u}{2}\right)\;\mathrm{d}u = \\ &= \frac{1}{\sqrt{2\,\pi}}\;\exp\left(\frac{-\omega^2}{2}\right)\;\underbrace{\int_{-\infty}^{\infty} \exp\left(\frac{-(u-\mathrm{i}\,\omega)^2}{2}\right)\;\mathrm{d}u}_{\sqrt{2\,\pi}} = \exp\left(\frac{-\omega^2}{2}\right)\,,\end{aligned}$$

kde v poslední rovnosti jsme použili Gaussův integrál (C.10). Výsledek je (až na násobek konstantou) stejná (reálná) funkce jako původní hustota. (Odpovídající vzorec bychom našli v tabulkách Fourierovy transformace.) Obecný případ dostaneme použitím tvrzení 4.7.4. □

Věta o zpětné Fourierově transformaci má následující analogii:

Věta 4.7.6: *Nechť* X *je* spojitá *náhodná veličina, jejíž charakteristická funkce* Ψ_X *splňuje* $\int_{-\infty}^{\infty} |\Psi_X(\omega)|\;\mathrm{d}\omega < \infty$. *Pak funkce*

$$f_X(u) = \frac{1}{2\,\pi}\int_{-\infty}^{\infty} \exp(-\mathrm{i}\,\omega\,u)\;\Psi_X(\omega)\;\mathrm{d}\omega$$

je hustotou náhodné veličiny X.

Pokud v předchozí větě zintegrujeme hustotu na intervalu $\langle u_1, u_2\rangle$, dostaneme rozdíl hodnot distribuční funkce:

$$\begin{aligned} F_X(u_2) - F_X(u_1) &= \int_{u_1}^{u_2} f_X(u)\,\mathrm{d}u = \int_{u_1}^{u_2} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}\omega u)\,\Psi_X(\omega)\,\mathrm{d}\omega\,\mathrm{d}u = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{u_1}^{u_2} \exp(-\mathrm{i}\omega u)\,\mathrm{d}u \right) \Psi_X(\omega)\,\mathrm{d}\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(-\mathrm{i}\omega u_2) - \exp(-\mathrm{i}\omega u_1)}{-\mathrm{i}\omega} \Psi_X(\omega)\,\mathrm{d}\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(-\mathrm{i}\omega u_1) - \exp(-\mathrm{i}\omega u_2)}{\mathrm{i}\omega} \Psi_X(\omega)\,\mathrm{d}\omega, \end{aligned}$$

kde integrál je nutno chápat ve smyslu *Cauchyho hlavní hodnoty*, tj.

$$\int_{-\infty}^{\infty} \frac{\exp(-\mathrm{i}\omega u_1) - \exp(-\mathrm{i}\omega u_2)}{\mathrm{i}\omega} \Psi_X(\omega)\,\mathrm{d}\omega = \lim_{c\to\infty} \int_{-c}^{c} \frac{\exp(-\mathrm{i}\omega u_1) - \exp(-\mathrm{i}\omega u_2)}{\mathrm{i}\omega} \Psi_X(\omega)\,\mathrm{d}\omega.$$

Tento vztah lze zobecnit i na nespojité náhodné veličiny:

Věta 4.7.7: *Je-li distribuční funkce F_X náhodné veličiny X spojitá v bodech u_1, u_2, pak*

$$F_X(u_2) - F_X(u_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(-\mathrm{i}\omega u_1) - \exp(-\mathrm{i}\omega u_2)}{\mathrm{i}\omega} \Psi_X(\omega)\,\mathrm{d}\omega,$$

kde integrál je ve smyslu Cauchyho hlavní hodnoty.

Důsledek 4.7.8: *Charakteristická funkce náhodné veličiny jednoznačně určuje její rozdělení (distribuční funkci).*

Poznámka 4.7.9: Integrál z věty 4.7.7 by jinak než ve smyslu Cauchyho hlavní hodnoty nemusel být definován, neboť pro $\omega \to \infty$ čitatel $\exp(-\mathrm{i}\omega u_1) - \exp(-\mathrm{i}\omega u_2)$ nemusí konvergovat k nule, stejně jako $\Psi_X(\omega)$ (např. má-li rozdělení diskrétní složku) a jmenovatel $\mathrm{i}\omega$ se blíží nekonečnu, ale pouze lineárně, což nestačí k zajištění konvergence. Uvažovali jsme libovolný omezený interval, místo abychom – jako obvykle – zafixovali jeden konec intervalu, $u_1 = -\infty$, a dostali přímo distribuční funkci. Důvod je ten, že v komplexním oboru není výraz $\exp(\mathrm{i}\omega\infty)$ definován. Mohli jsme ovšem volit např. $u_1 = 0$, výsledný vzorec by již ale nebyl tak přehledný. Nicméně je vidět, že tato věta plně určuje distribuční funkci.

Věta 4.7.10: *Pro náhodné veličiny X a X_n, $n \in \mathbb{N}$, platí, že $\Psi_X = \lim_{n\to\infty} \Psi_{X_n}$, právě když $F_X(u) = \lim_{n\to\infty} F_{X_n}(u)$ v každém bodě u, v němž je F_X spojitá.*

Jedním z důležitých použití charakteristické funkce je výpočet momentů:

Věta 4.7.11: *Pro k-tý obecný moment náhodné veličiny X platí*

$$\mathrm{E}X^k = \frac{1}{\mathrm{i}^k} \Psi_X^{(k)}(0),$$

pokud tento moment existuje.

Důkaz: Pro *spojitou* náhodnou veličinu: k-krát zderivujeme integrand v (4.15) podle ω

$$\Psi_X^{(k)}(\omega) = \int_{-\infty}^{\infty} (\mathrm{i}u)^k \exp(\mathrm{i}\omega u)\, f_X(u)\,\mathrm{d}u = \mathrm{i}^k \int_{-\infty}^{\infty} u^k \exp(\mathrm{i}\omega u)\, f_X(u)\,\mathrm{d}u$$

a po dosazení $\omega = 0$

$$\Psi_X^{(k)}(0) = \mathrm{i}^k \int_{-\infty}^{\infty} u^k f_X(u)\,\mathrm{d}u = \mathrm{i}^k \mathrm{E}X^k.$$

Postup pro *diskrétní* náhodnou veličinu je obdobný a směsí se výsledek zachová.

Poznámka 4.7.12: První moment – střední hodnotu – umíme stanovit snadno jako součet středních hodnot. I druhý moment bychom vypočítali ze znalosti rozptylu-součtu nezávislých náhodných veličin. U vyšších momentů bychom se však bez charakteristické funkce těžko obešli.

Příklad 4.7.13: Máme najít první 4 obecné momenty normovaného normálního rozdělení.

Řešení: Nechť X je náhodná veličina s rozdělením $\mathrm{N}(0,1)$. Jelikož je normovaná, je $\mathrm{E}X = 0$, $\mathrm{E}X^2 = (\mathrm{E}X)^2 + \mathrm{D}X = 1$. Hustota f_X je sudá funkce, proto

$$\mathrm{E}X^3 = \int_{-\infty}^{\infty} u^3 \frac{1}{\sqrt{2\,\pi}} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u = 0\,,$$

neboť se jedná o integrál liché funkce. Čtvrtý moment je však nenulový a jeho výpočet z definičního vztahu vede po úpravách na Gaussův integrál. Místo toho můžeme použít větu 4.7.11 na (4.17):

$$\begin{aligned}
\Psi_X(\omega) &= \exp\left(\tfrac{-\omega^2}{2}\right)\,,\\
\Psi'_X(\omega) &= -\omega \exp\left(\tfrac{-\omega^2}{2}\right)\,,\\
\Psi''_X(\omega) &= \left(-1+\omega^2\right) \exp\left(\tfrac{-\omega^2}{2}\right)\,,\\
\Psi^{(3)}_X(\omega) &= \left(3\,\omega-\omega^3\right) \exp\left(\tfrac{-\omega^2}{2}\right)\,,\\
\Psi^{(4)}_X(\omega) &= \left(3-6\,\omega^2+\omega^4\right) \exp\left(\tfrac{-\omega^2}{2}\right)\,.
\end{aligned}$$

Ani jsme nemuseli všechny koeficienty počítat, neboť nás zajímá pouze hodnota pro $\omega = 0$:

$$\Psi^{(4)}_X(0) = 3\,.$$

Zbývá ji vydělit $\mathrm{i}^4 = 1$, tedy

$$\mathrm{E}X = 3\,.$$

I momenty vyšších řádů bychom tímto postupem dostali poměrně snadno, zatímco integrace by byla komplikovanější. □

Následující klíčový poznatek se opírá o známé vlastnosti Fourierovy transformace. Konvoluci funkcí odpovídá součin jejich Fourierových obrazů (a naopak, neboť zpětná transformace má analogické vlastnosti). To je užitečné, neboť součin funkcí se počítá snáze, a i s přímou a zpětnou Fourierovou transformací může být výpočetní složitost menší než u konvoluce.

Rozdělení součtu *nezávislých* náhodných veličin je konvoluce původních rozdělení. Té odpovídá součin charakteristických funkcí.

Věta 4.7.14: *Jsou-li* $X_1, \ldots, X_n$ nezávislé *náhodné veličiny, pak*

$$\Psi_{X_1+\cdots+X_n} = \Psi_{X_1} \cdot \ldots \cdot \Psi_{X_n}\,.$$

Důkaz:

$$\Psi_{X_1+\cdots+X_n}(\omega) = \mathrm{E}\exp\bigl(\mathrm{i}\,\omega \sum_k X_k\bigr) = \mathrm{E}\prod_k \exp(\mathrm{i}\,\omega\,X_k) = \prod_k \mathrm{E}\exp(\mathrm{i}\,\omega\,X_k) = \prod_k \Psi_{X_k}(\omega)\,.$$

Viděli jsme dříve, že rozdělení součtu nezávislých náhodných veličin může být složitější než původní rozdělení. Nyní máme nástroj pro jeho popis, protože součin charakteristických funkcí nebývá těžké vypočítat. *Násobení funkcí je mnohem jednodušší operace než sčítání nezávislých náhodných veličin.* Nejde o to, zda sčítáme nebo násobíme. Charakteristické funkce (reálného argumentu) se násobí „bod po bodu", tj. složitost výpočtu závisí lineárně na počtu hodnot. Náhodné

veličiny jsou sice také funkce, ovšem elementárních jevů, jejichž počet může být mnohem větší než počet hodnot náhodné veličiny (rozhodně ne menší). Pokud vyjdeme z rozdělení sčítanců, pak výsledek je dán (v podstatě) konvolucí, jejíž složitost závisí na počtu hodnot kvadraticky, a to může být příliš. Převod z distribuční funkce na charakterickou i zpět lze v příznivém případě provést se složitostí $n \log n$ (pomocí obdoby rychlé Fourierovy transformace, FFT). Lze tedy rychleji vypočítat rozdělení součtu nezávislých náhodných veličin převodem na charakterickou funkci, vynásobením obrazů a zpětným převodem, než pomocí konvoluce.

Věta 4.7.14 dovoluje také snadno dokázat některé poučky, uvedené v jiných kapitolách bez důkazu.

Tvrzení 4.7.15: *Součet dvou* nezávislých *náhodných veličin s normálním rozdělením má normální rozdělení.*

Důkaz: Využitím charakteristických funkcí se vyhneme počítání konvolucí. Můžeme k oběma náhodným veličinám, X a Y, přičíst vhodné konstanty a bez újmy na obecnosti předpokládat, že jejich střední hodnoty jsou nulové, rozptyly jsou σ_X a σ_Y. Pak

$$\Psi_{X+Y}(\omega) = \Psi_X(\omega) \cdot \Psi_Y(\omega) = \exp\left(-\frac{1}{2}\sigma_X^2\,\omega^2\right) \cdot \exp\left(-\frac{1}{2}\sigma_Y^2\,\omega^2\right) = \exp\left(-\frac{1}{2}\left(\sigma_X^2 + \sigma_Y^2\right)\omega^2\right),$$

což je charakteristická funkce normálního rozdělení $\mathrm{N}\left(0, \sigma_X^2 + \sigma_Y^2\right)$. □

Příklad 4.7.16: Najděte charakteristickou funkci binomického rozdělení.

Řešení: Mohli bychom postupovat dle vzorce (4.14). Jednodušší však bude využít toho, že binomické rozdělení $\mathrm{Bi}(m, q)$ dostaneme jako rozdělení součtu m nezávislých náhodných veličin s alternativním rozdělením. Vzorec (4.14) použijeme pouze pro alternativní rozdělení, jehož charakteristická funkce v ω je

$$(1 - q) + \exp(\mathrm{i}\omega)\, q\,.$$

Charakteristická funkce binomického rozdělení je součin m takovýchto funkcí, tedy

$$\Psi_{\mathrm{Bi}(m,q)}(\omega) = \left((1 - q) + \exp(\mathrm{i}\omega)\, q\right)^m\,.$$

□

Kombinace vět 4.7.11 a 4.7.14 dovoluje vypočítat libovolný moment součtu nezávislých náhodných veličin. V tomto směru je charakteristická funkce těžko nahraditelným nástrojem.

Charakteristická funkce poskytuje zcela jiný popis rozdělení náhodné veličiny. Integrální transformace (Fourierova, až na znaménko argumentu) vede na hodnoty, které charakterizují rozdělení (distribuční funkci) jako celek. Mají vztah i k momentům, které lze pomocí ní někdy výhodně počítat. Nenahraditelná je při studiu součtů nezávislých náhodných veličin, neboť charakteristická funkce výsledku je součinem charakteristických funkcí argumentů.

Cvičení 4.7.17: Najděte charakteristickou funkci součtu m hodů pravidelnou kostkou.

Cvičení 4.7.18: Najděte charakteristickou funkci součtu m nezávislých náhodných veličin, které mají rovnoměrné rozdělení na intervalu $\langle 0, 1\rangle$.

Cvičení 4.7.19: V příkladu 4.7.13 ověřte hodnoty prvního až třetího obecného momentu pomocí vzorce (4.17).

Cvičení 4.7.20: Odvoďte charakteristickou funkci exponenciálního a geometrického rozdělení.

Cvičení 4.7.21: Dokažte, že součet *nezávislých* exponenciálních rozdělení s parametry τ_j, $j = 1, \ldots, k$, má exponenciální rozdělení s parametrem $\sum_{j=1}^{k} \tau_j$.

Cvičení 4.7.22: Dokažte, že součet *nezávislých* Poissonových rozdělení s parametry w_j, $j = 1, \ldots, k$, má Poissonovo rozdělení s parametrem $\sum_{j=1}^{k} w_j$.

5. Náhodné vektory (vícerozměrné náhodné veličiny)

5.1 Motivace a definice

Máme-li více náhodných veličin, je často žádoucí popsat nejen každou zvlášť, ale i vztahy mezi nimi. Tak můžeme popsat vztah např. mezi výškou a váhou osob, mezi teplotou a životností součástek apod. K tomu nutně potřebujeme víc informací než jen popis jednotlivých náhodných veličin. Proto zavádíme vícerozměrné náhodné veličiny, kterým říkáme náhodné vektory. Jejich definice je jednoduchá, ale matematický popis nutně složitější. Kde to bude možné, budeme kopírovat jednorozměrný případ.

Definice 5.1.1: ***Náhodný vektor*** *o n složkách je n-tice náhodných veličin $\boldsymbol{X} = (X_1, \ldots, X_n)$, definovaných na stejném pravděpodobnostním prostoru.*

Ekvivalentně lze náhodný vektor chápat jako zobrazení $\boldsymbol{X} \colon \Omega \to \mathbb{R}^n$, které elementárním jevům přiřazuje n-rozměrné aritmetické vektory.

Poznámka 5.1.2: Komplexní náhodná veličina je náhodný vektor se dvěma složkami interpretovanými jako reálná a imaginární část.

Definice 5.1.3: ***(Sdružené) rozdělení pravděpodobnosti*** *náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_n)$, definovaného na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$, je pravděpodobnostní míra $P_{\boldsymbol{X}}$ na Borelově σ-algebře $\mathcal{B}(\mathbb{R}^n)$ taková, že*

$$P_{\boldsymbol{X}}(I) = P\left(\{\omega \in \Omega \mid (X_1(\omega), \ldots, X_n(\omega)) \in I\}\right)$$

pro všechna $I \in \mathcal{B}(\mathbb{R}^n)$.

Opět postačí znát rozdělení pravděpodobnosti $P_{\boldsymbol{X}}$ pro n-rozměrné intervaly $I = I_1 \times \ldots \times I_n$ (kde $I_1, \ldots, I_n$ jsou intervaly v $\mathbb{R}$), tj.

$$P_{\boldsymbol{X}}(I_1 \times \ldots \times I_n) = P[X_1 \in I_1, \ldots, X_n \in I_n].$$

Úspornější popis dostaneme pomocí intervalů tvaru $I_k = (-\infty, u_k\rangle$, $u_k \in \mathbb{R}$:

$$\begin{aligned} P_{\boldsymbol{X}}((-\infty, u_1\rangle \times \ldots \times (-\infty, u_n\rangle) &= P[X_1 \in (-\infty, u_1\rangle, \ldots, X_n \in (-\infty, u_n\rangle] = \\ &= P[X_1 \leq u_1, \ldots, X_n \leq u_n]. \end{aligned}$$

Definice 5.1.4: ***(Sdružená) distribuční funkce*** *náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_n)$ je funkce $F_{\boldsymbol{X}} \colon \mathbb{R}^n \to \langle 0, 1\rangle$ definovaná vzorcem*

$$F_{\boldsymbol{X}}(u_1, \ldots, u_n) = P[X_1 \leq u_1, \ldots, X_n \leq u_n].$$

V indexech vynecháváme závorky, např. $F_{X_1,\ldots,X_n} = F_{(X_1,\ldots,X_n)} = F_{\boldsymbol{X}}$.

Věta 5.1.5: *Distribuční funkce $F_{\boldsymbol{X}}$ náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_n)$ splňuje následující podmínky:*

1. *$F_{\boldsymbol{X}}$ je neklesající (ve všech proměnných),*
2. *$F_{\boldsymbol{X}}$ je zprava spojitá (ve všech proměnných),*

3. $\lim_{u_1 \to \infty} \cdots \lim_{u_n \to \infty} F_{\boldsymbol{X}}(u_1, \ldots, u_n) = 1,$
4. $\forall k \in \{1, \ldots, n\} : \lim_{u_k \to -\infty} F_{\boldsymbol{X}}(u_1, \ldots, u_n) = 0.$

Důkaz: První tři vztahy se dokazují analogicky jako v jednorozměrném případě ve větě 3.3.4, poslední dokážeme obecněji v tvrzení 5.1.7. □

Rozdělením pravděpodobnosti $P_{\boldsymbol{X}}$ náhodného vektoru $\boldsymbol{X} = (X_1, \ldots, X_n)$ jsou jednoznačně určena rozdělení pravděpodobnosti náhodných veličin $X_1, \ldots, X_n$, což jsou tzv. **marginální rozdělení**. Vztahy dokážeme v dvojrozměrném případě.

Tvrzení 5.1.6: *Pro náhodný vektor (X, Y) platí*

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) ,$$
$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y) .$$

Důkaz: Nechť $(y_n)_{n \in \mathbb{N}}$ je rostoucí posloupnost, $\lim_{n \to \infty} y_n = \infty$. Pak dle tvrzení 1.7.10

$$\begin{aligned} \lim_{n \to \infty} F_{X,Y}(x, y_n) &= \lim_{n \to \infty} P_{X,Y}((-\infty, x\rangle \times (-\infty, y_n\rangle)) = \\ &= P_{X,Y}\Big(\bigcup_{n \in \mathbb{N}} ((-\infty, x\rangle \times (-\infty, y_n\rangle))\Big) = \\ &= P_{X,Y}((-\infty, x\rangle \times \mathbb{R}) = P[X \leq x] = F_X(x) . \end{aligned}$$

V druhém vzorci jsou pouze vyměněny role X a Y. □

V n-rozměrném případě dostáváme složitější výrazy (důkaz je obdobný):

Tvrzení 5.1.7: *Pro náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ platí*

$$F_{X_1}(u_1) = \lim_{u_2 \to \infty} \cdots \lim_{u_n \to \infty} F_{\boldsymbol{X}}(u_1, \ldots, u_n)$$

a podobně pro další složky.

Část 4 věty 5.1.5 je limitním případem tvrzení 5.1.7 pro $u_1 \to -\infty$.

Náhodné vektory dovolují pomocí sdružené distribuční funkce korektněji formulovat nezávislost náhodných veličin, kterou jsme zavedli v kapitole 3.9. Ke znalosti sdruženého rozdělení náhodného vektoru nestačí znát marginální rozdělení, neboť ta neobsahují informace o vzájemné závislosti jednotlivých složek. Jsou-li složky náhodného vektoru *nezávislé*, lze sdruženou distribuční funkci vyjádřit jednodušeji pomocí marginálních rozdělení:

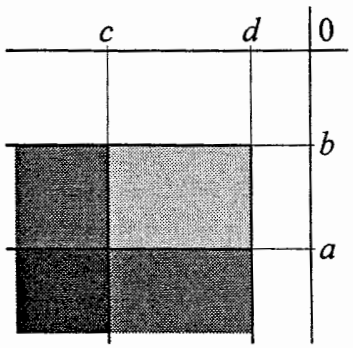
$$F_{\boldsymbol{X}}(u_1, \ldots, u_n) = \prod_{k=1}^{n} F_{X_k}(u_k) .$$

Poznámka 5.1.8: Součin pravděpodobností dle kapitoly 2.3 vede na sdružené rozdělení *nezávislých* náhodných veličin s danými marginálními rozděleními.

Na rozdíl od jednorozměrného případu, podmínky ve větě 5.1.5 jsou pouze nutné, nikoli postačující pro to, aby funkce $F_{\boldsymbol{X}} : \mathbb{R}^n \to \langle 0, 1\rangle$ byla distribuční funkcí náhodného vektoru. K tomu bychom museli splnit další podmínky, které nyní odvodíme pro dvojrozměrný případ.

Mějme dvojrozměrný náhodný vektor $\boldsymbol{X}$. Libovolný obdélník $(a, b\rangle \times (c, d\rangle$ lze vyjádřit pomocí polonekonečných intervalů (viz obr. 5.1), což umožňuje vyjádřit pravděpodobnost, že do něj padne $\boldsymbol{X}$:

$$\begin{aligned} P[\boldsymbol{X} \in (a, b\rangle \times (c, d\rangle] &= P[\boldsymbol{X} \in (-\infty, b\rangle \times (-\infty, d\rangle] - P[\boldsymbol{X} \in (-\infty, a\rangle \times (-\infty, d\rangle] - \\ &\quad - P[\boldsymbol{X} \in (-\infty, b\rangle \times (-\infty, c\rangle] + P[\boldsymbol{X} \in (-\infty, a\rangle \times (-\infty, c\rangle] = \\ &= F_{\boldsymbol{X}}(b, d) - F_{\boldsymbol{X}}(a, d) - F_{\boldsymbol{X}}(b, c) + F_{\boldsymbol{X}}(a, c) . \end{aligned} \tag{5.1}$$



Obrázek 5.1. Vyjádření obdélníka pomocí polonekonečných dvojrozměrných intervalů

Nezápornost všech těchto pravděpodobností klade na distribuční funkci další nutné podmínky

$$F_{\boldsymbol{X}}(b,d) - F_{\boldsymbol{X}}(a,d) - F_{\boldsymbol{X}}(b,c) + F_{\boldsymbol{X}}(a,c) \geq 0\,,$$

které musí být splněny pro všechna a, b, c, d, pro která $a < b$, $c < d$. Teprve přidáním těchto nerovností dostáváme i postačující podmínky.

Pro obecnou dimenzi n jsou tyto podmínky složitější, obsahují 2^n členů.

Diskrétní náhodné vektory

Náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ se nazývá **diskrétní**, jestliže všechny jeho složky jsou diskrétní náhodné veličiny. Lze jej popsat též **sdruženou pravděpodobnostní funkcí** $p_{\boldsymbol{X}}\colon \mathbb{R}^n \to \langle 0,1\rangle$

$$p_{\boldsymbol{X}}(u_1,\ldots,u_n) = P\,[X_1 = u_1, \ldots, X_n = u_n]\,,$$

která je nenulová jen ve spočetně mnoha bodech. Pomocí ní lze říci, že *diskrétní* náhodné veličiny $X_1, \ldots, X_n$ jsou *nezávislé*, právě když

$$p_{\boldsymbol{X}}(u_1,\ldots,u_n) = \prod_{j=1}^{n} p_{X_j}(u_j)\,,$$

neboli

$$P\,[X_1 = u_1, \ldots, X_n = u_n] = \prod_{j=1}^{n} P\,[X_j = u_j]$$

pro všechna $u_1, \ldots, u_n \in \mathbb{R}$.

Spojité náhodné vektory

Náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ se nazývá **spojitý**, jestliže všechny jeho složky jsou spojité náhodné veličiny. Lze jej popsat též **sdruženou hustotou pravděpodobnosti**, což je (každá) nezáporná funkce $f_{\boldsymbol{X}}\colon \mathbb{R}^n \to \langle 0,1\rangle$ taková, že

$$F_{\boldsymbol{X}}(u_1,\ldots,u_n) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_n} f_{\boldsymbol{X}}(v_1,\ldots,v_n)\,\mathrm{d}v_1 \ldots \mathrm{d}v_n$$

pro všechna $u_1, \ldots, u_n \in \mathbb{R}$. Pro intervaly $\langle a_j, b_j\rangle$ dostáváme

$$P[X_1 \in \langle a_1,b_1\rangle, \ldots, X_n \in \langle a_n,b_n\rangle] = P_{\boldsymbol{X}}(\langle a_1,b_1\rangle \times \ldots \times \langle a_n,b_n\rangle) = \\ = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\boldsymbol{X}}(v_1,\ldots,v_n)\,\mathrm{d}v_1 \ldots \mathrm{d}v_n\,.$$

Existuje-li parciální derivace

$$\frac{\partial^n F_{\boldsymbol{X}}(u_1,\ldots,u_n)}{\partial u_1 \ldots \partial u_n},$$

pak je hustotou náhodného vektoru $\boldsymbol{X}$.

Souvislost s marginálními rozděleními uvedeme pro dvojrozměrný náhodný vektor (X,Y), zobecnění je zřejmé.

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v\,,$$
$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v\,,$$
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v)\,\mathrm{d}v\,,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(u,y)\,\mathrm{d}u\,.$$

Nezávislost spojitých náhodných veličin nelze zavést tak jako pro diskrétní náhodné veličiny. Jev, že jedna z veličin nabývá danou hodnotu, má nulovou pravděpodobnost, a tudíž je nezávislý na hodnotách ostatních náhodných veličin. Např. jevy $X_1 = u_1$, $X_2 = u_2$ jsou pro *spojité* náhodné veličiny X_1, X_2 vždy nezávislé. Pomocí hustoty lze však říci, že *spojité* náhodné veličiny $X_1,\ldots,X_n$ jsou *nezávislé*, právě když

$$f_{\boldsymbol{X}}(u_1,\ldots,u_n) = \prod_{j=1}^{n} f_{X_j}(u_j)$$

pro všechna $u_1,\ldots,u_n \in \mathbb{R}$.

Tvrzení 5.1.9: *Jsou-li X,Y* spojité *náhodné veličiny se sdruženou hustotou $f_{X,Y}$, pak hustota jejich součtu je*

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(v,u-v)\,\mathrm{d}v\,.$$

Pro *nezávislé* spojité náhodné veličiny vede předchozí vzorec na konvoluci dle (3.7).

Kvantilovou funkci nelze definovat ani pro spojitý náhodný vektor, protože by jedné hodnotě pravděpodobnosti $\alpha \in \langle 0,1\rangle$ neodpovídal bod (kde má distribuční funkce hodnotu α), ale celá „vrstevnice" v grafu distribuční funkce.

Náhodné vektory se smíšeným rozdělením

Stejně jako u jednorozměrných náhodných veličin, mohou mít náhodné vektory obecně smíšené rozdělení, které lze získat směsí diskrétních a spojitých náhodných vektorů. Zde však může vzniknout ještě dalším přirozeným způsobem: náhodný vektor, jehož některé složky jsou diskrétní a některé spojité, má rovněž smíšené rozdělení. V tomto případě nemusí existovat rozklad na směs diskrétního a spojitého náhodného vektoru.

Příklad 5.1.10: Nechť náhodná veličina X má diskrétní rovnoměrné rozdělení s oborem hodnot $\{2,3\}$, náhodná veličina Y má spojité rovnoměrné rozdělení s oborem hodnot $\langle 0,1\rangle$. Obor hodnot náhodného vektoru (X,Y) je obecně podmnožinou množiny $M = \{2,3\}\times\langle 0,1\rangle$, což je sjednocení dvou úseček s krajními body $(2,0),(2,1)$, resp. $(3,0),(3,1)$. Jsou-li veličiny X,Y *nezávislé*, má (X,Y) na M rozdělení rovnoměrné v tom smyslu, že pravděpodobnost, že výsledek padne na danou úsečku, je úměrná její délce; distribuční funkce je

$$F_{X,Y}(x,y) = \begin{cases} 0 & \text{pro } x < 2 \text{ nebo } y < 0, \\ \frac{y}{2} & \text{pro } 2 \le x < 3,\ 0 \le y < 1, \\ y & \text{pro } x \ge 3,\ 0 \le y < 1, \\ \frac{1}{2} & \text{pro } 2 \le x < 3,\ y \ge 1, \\ 1 & \text{pro } x \ge 3,\ y \ge 1. \end{cases}$$

Bez předpokladu nezávislosti můžeme dostat i jiné distribuční funkce, např.

$$F_{X',Y'}(x,y) = \begin{cases} 0 & \text{pro } x < 2 \text{ nebo } y < 0, \\ \frac{y^2}{2} & \text{pro } 2 \le x < 3,\ 0 \le y < 1, \\ y & \text{pro } x \ge 3,\ 0 \le y < 1, \\ \frac{1}{2} & \text{pro } 2 \le x < 3,\ y \ge 1, \\ 1 & \text{pro } x \ge 3,\ y \ge 1, \end{cases}$$

kde náhodné veličiny X', Y' mají stejná rozdělení jako X, Y, ale nejsou nezávislé. Marginální distribuční funkce, získané jako limity pro $x \to \infty$, resp. $y \to \infty$, se nutně shodují s distribučními funkcemi původních náhodných veličin. Obor hodnot náhodného vektoru by mohl být i menší než M, např. sjednocení K dvou úseček s koncovými body $(2,0), (2,1/3)$, resp. $(3,1/3), (3,1)$; tomu vyhovuje jediná distribuční funkce se stejnými marginálními rozděleními,

$$F_{X'',Y''}(x,y) = \begin{cases} 0 & \text{pro } x < 2 \text{ nebo } y < 0, \\ \frac{3y}{2} & \text{pro } 2 \le x < 3,\ 0 \le y < \frac{1}{3}, \\ y & \text{pro } x \ge 3,\ 0 \le y < 1, \\ \frac{1}{2} & \text{pro } 2 \le x < 3,\ y \ge \frac{1}{3}, \\ 1 & \text{pro } x \ge 3,\ y \ge 1, \end{cases}$$

Žádný z uvedených náhodných vektorů nelze vyjádřit jako směs diskrétního a spojitého náhodného vektoru, neboť distribuční funkce se mění skokem v první proměnné a spojitě v druhé.

Rozdělení diskrétního náhodného vektoru je popsáno pravděpodobnostní mírou, která je soustředěna v bodech (odpovídajících výsledkům s nenulovou pravděpodobností). Pro spojitý náhodný vektor dostáváme hustotu, z níž se počítá plošný integrál (ve dvou dimenzích), obecně n-rozměrný integrál (v dimenzi n). V příkladu 5.1.10 bychom však odpovídající „hustotu" potřebovali integrovat po úsečce, jejíž dimenze je nenulová, ale menší než dimenze celého prostoru. I proto není možná redukce na spojitý nebo diskrétní případ. I přesto může použití směsí aspoň zjednodušit popis.

Příklad 5.1.11: Druhé rozdělení v příkladu 5.1.10 lze vyjádřit jako směs náhodných vektorů, jejichž první složky jsou konstantní $(X', Y') = \mathrm{Mix}_{1/2}((2,U),(3,V))$, kde U, V mají nabývají hodnot z intervalu $\langle 0,1\rangle$ s hustotami $f_U(u) = 2u$, $f_V(v) = 2(1-v)$.

Náhodné vektory potřebujeme všude tam, kde máme studovat více náhodných veličin i jejich souvislosti. Sdružené rozdělení opět dovoluje abstrahovat od původního pravděpodobnostního prostoru. Jeho popis je náročný (závisí na mnoha parametrech), proto využíváme nezávislost, kde je to možné. Ta dovoluje rozdělit jeden složitý popis na více jednodušších. Při popisu rozdělení náhodných vektorů je nenahraditelná distribuční funkce, neboť náhodné vektory se smíšeným rozdělením dostaneme snadno i tak, že některé složky jsou diskrétní a jiné spojité. Kvantilová funkce bohužel nemá ve vícedimenzionálním případě použitelnou obdobu.

Cvičení 5.1.12: [Spiegel et al. 2000] Náhodný vektor (X,Y) má sdruženou hustotu

$$f_{X,Y}(x,y) = \begin{cases} c\,x\,y & \text{pro } 0 \le x \le 2,\ 0 \le y \le x, \\ 0 & \text{jinak.} \end{cases}$$

Určete sdruženou distribuční funkci a marginální rozdělení. Zjistěte, zda X, Y jsou nezávislé. Vypočtěte pravděpodobnost $P[1 \le X \le 2,\ 1/2 \le Y \le 1]$.

Řešení: Konstanta c je dána podmínkou

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v$$

Hodnota sdružené distribuční funkce $F_{X,Y}(x,y)$ je nulová pro $x < 0$ nebo $y < 0$, jednotková pro $x \ge 2$, $y \ge 2$. Pro $0 \le x \le 2$, $0 \le y \le x$ dostáváme

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v = c \int_0^x u \left(\int_0^{\min(u,y)} v\,\mathrm{d}v \right) \mathrm{d}u = \\ &= c \int_0^y u \left(\int_0^u v\,\mathrm{d}v \right) \mathrm{d}u + c \int_y^x u \left(\int_0^y v\,\mathrm{d}v \right) \mathrm{d}u = \\ &= \frac{1}{2} c \int_0^y u^3\,\mathrm{d}u + \frac{1}{2} c \int_y^x u\,y^2\,\mathrm{d}u = \\ &= \frac{1}{2} c \left(\frac{1}{4} y^4 + \frac{1}{2} x^2 y^2 - \frac{1}{2} y^4 \right) = c \left(\frac{-1}{8} y^4 + \frac{1}{4} x^2 y^2 \right). \end{aligned}$$

(Lze zkontrolovat, že parciální derivací podle obou proměnných dostaneme zpět hustotu.) Ve zbývajících bodech jsou hodnoty sdružené distribuční funkce $F_{X,Y}(x,y)$ rovny hodnotám jedné z marginálních distribučních funkcí, které záhy odvodíme. Ze spojitosti

$$1 = \lim_{x\to\infty} \lim_{y\to\infty} F_{X,Y}(x,y) = F_{X,Y}(2,2) = 2\,c$$

vyplývá $c = 1/2$,

$$F_{X,Y}(x,y) = \frac{-1}{16} y^4 + \frac{1}{8} x^2 y^2 .$$

Již nyní je zřejmé, že sdruženou distribuční funkci nelze vyjádřit jako součin dvou funkcí, z nichž každá by závisela jen na jedné proměnné, takže náhodné proměnné X, Y jsou závislé. Marginální distribuční funkce F_X, F_Y jsou na intervalu $(-\infty, 0)$ nulové a na $\langle 2, \infty)$ jednotkové, na intervalu $\langle 0, 2)$ jsou dány limitami

$$\begin{aligned} F_X(x) &= \lim_{y\to\infty} F_{X,Y}(x,y) = F_{X,Y}(x,x) = \frac{1}{16} x^4 , \\ F_Y(y) &= \lim_{x\to\infty} F_{X,Y}(x,y) = F_{X,Y}(2,y) = \frac{-1}{16} y^4 + \frac{1}{2} y^2 . \end{aligned}$$

Hustoty marginálních rozdělení jsou

$$f_X(x) = F_X'(x) = \frac{1}{4} x^3 , \qquad f_Y(y) = F_Y'(y) = \frac{-1}{4} y^3 + y .$$

Mohli jsme je také určit přímo ze vzorců

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v)\,\mathrm{d}v , \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(u,y)\,\mathrm{d}u$$

a naopak distribuční funkce získat jejich integrací, ale nedošli bychom ke sdružené distribuční funkci. Snadno ověříme, že $F_X(x) \cdot F_Y(y) \ne F_{X,Y}(x,y)$, $f_X(x) \cdot f_Y(y) \ne f_{X,Y}(x,y)$, takže X, Y opravdu nejsou nezávislé. (Sdružená hustota je sice součin $\frac{1}{2} x\,y$, ale pouze na oboru, který není nezávislý na souřadnicích.)

Zbývající pravděpodobnost určíme integrací hustoty přes obdélník nebo ze vztahu (5.1):

$$P[1 \le X \le 2,\ 1/2 \le Y \le 1] = F_{X,Y}(2,1) - F_{X,Y}(1,1) - F_{X,Y}(2,1/2) + F_{X,Y}(1,1/2) = \frac{9}{32} .$$

Cvičení 5.1.13: V příkladu 5.1.10 existuje jediný náhodný vektor, který má obor hodnot $M \setminus K$ a stejná marginální rozdělení. Najděte jeho distribuční funkci.

5.2 Charakteristiky náhodného vektoru

Některé charakteristiky náhodného vektoru jsou přímými analogiemi jednodimenzionálního případu, některé mají smysl až ve větší dimenzi.

Definice 5.2.1: *Nechť* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *je náhodný vektor. Jeho* ***střední hodnota*** *je vektor středních hodnot jednotlivých složek,* $\mathrm{E}\boldsymbol{X} = (\mathrm{E}X_1, \ldots, \mathrm{E}X_n)$, ***rozptyl*** *je* $\mathrm{D}\boldsymbol{X} = (\mathrm{D}X_1, \ldots, \mathrm{D}X_n)$.

Připomeňme, že je-li U náhodná veličina, $a, b \in \mathbb{R}$, pak $aU + b$ má charakteristiky

$$\mathrm{E}(aU + b) = a\,\mathrm{E}U + b\,, \qquad \mathrm{D}(aU + b) = a^2\,\mathrm{D}U\,.$$

Na rozdíl od jednorozměrné náhodné veličiny, střední hodnota a rozptyl náhodného vektoru nedávají dostatečnou informaci pro výpočet rozptylu jeho lineárních funkcí. Proto zavádíme další charakteristiky. Např.

$$\mathrm{E}(X + Y) = \mathrm{E}X + \mathrm{E}Y\,,$$

ale

$$\begin{aligned}
\mathrm{D}(X+Y) &= \mathrm{E}\left((X+Y)^2\right) - (\mathrm{E}(X+Y))^2 = \\
&= \mathrm{E}\left(X^2 + Y^2 + 2XY\right) - (\mathrm{E}X + \mathrm{E}Y)^2 = \\
&= \mathrm{E}\left(X^2\right) + \mathrm{E}\left(Y^2\right) + 2\,\mathrm{E}(XY) - \left((\mathrm{E}X)^2 + (\mathrm{E}Y)^2 + 2\,\mathrm{E}X\,\mathrm{E}Y\right) = \\
&= \underbrace{\mathrm{E}\left(X^2\right) - (\mathrm{E}X)^2}_{\mathrm{D}X} + \underbrace{\mathrm{E}\left(Y^2\right) - (\mathrm{E}Y)^2}_{\mathrm{D}Y} + 2\,\underbrace{(\mathrm{E}(XY) - \mathrm{E}X\,\mathrm{E}Y)}_{\mathrm{cov}(X,Y)} = \\
&= \mathrm{D}X + \mathrm{D}Y + 2\,\mathrm{cov}(X,Y)\,,
\end{aligned}$$

kde veličina

$$\mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}X\,\mathrm{E}Y$$

se nazývá **kovariance** náhodných veličin X, Y. Ekvivalentně ji lze počítat dle vzorce

$$\mathrm{cov}(X,Y) = \mathrm{E}\left((X - \mathrm{E}X)\,(Y - \mathrm{E}Y)\right)\,,$$

neboť

$$\begin{aligned}
\mathrm{E}\left((X - \mathrm{E}X)\,(Y - \mathrm{E}Y)\right) &= \mathrm{E}(XY - X\,\mathrm{E}Y - Y\,\mathrm{E}X + \mathrm{E}X\,\mathrm{E}Y) = \\
&= \mathrm{E}(XY) - \mathrm{E}X\,\mathrm{E}Y - \underbrace{\mathrm{E}X\,\mathrm{E}Y + \mathrm{E}X\,\mathrm{E}Y}_{0}\,.
\end{aligned}$$

(První vzorec je vhodnější pro výpočet, druhý pro pochopení významu.)

Pro existenci kovariance je postačující existence rozptylů $\mathrm{D}X, \mathrm{D}Y$.

Tvrzení 5.2.2 (vlastnosti kovariance):

$$\begin{aligned}
\mathrm{cov}(X,X) &= \mathrm{D}X\,, \\
\mathrm{cov}(Y,X) &= \mathrm{cov}(X,Y)\,, \\
\mathrm{cov}(aX + b, cY + d) &= ac\,\mathrm{cov}(X,Y)\,, \qquad (a,b,c,d \in \mathbb{R})\,.
\end{aligned}$$

(Srovnejte s vlastnostmi rozptylu jako speciálního případu.) Speciálně $\mathrm{cov}(X,-X) = -\mathrm{D}X$. *Pro* nezávislé *náhodné veličiny* X, Y *je* $\mathrm{cov}(X,Y) = 0$.

Použitím kovariance pro *normované* náhodné veličiny vyjde **korelace** (též **koeficient korelace** nebo **korelační koeficient**)

$$\varrho(X,Y) = \operatorname{cov}(\operatorname{norm} X, \operatorname{norm} Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X\,\sigma_Y} = \mathrm{E}\,(\operatorname{norm} X \cdot \operatorname{norm} Y)\ .$$

(Předpokládáme, že směrodatné odchylky ve jmenovateli jsou *nenulové.*)

Tvrzení 5.2.3 (vlastnosti korelace):

$$\begin{aligned}
\varrho(X,Y) &\in \langle -1,1\rangle,\\
\varrho(X,X) &= 1,\\
\varrho(X,-X) &= -1,\\
\varrho(Y,X) &= \varrho(X,Y),\\
\varrho(a\,X+b, c\,Y+d) &= \operatorname{sign}(a\,c)\ \varrho(X,Y)\,, \qquad (a,b,c,d\in\mathbb{R},\quad a\neq 0\neq c)\ .
\end{aligned}$$

(Až na znaménko nezáleží na prosté lineární transformaci.) Důsledek:

$$\varrho(a\,X+b, X) = \operatorname{sign}(a)\ .$$

Podmínka $\varrho(X,Y)=0$ je nutná, ale nikoli postačující pro nezávislost náhodných veličin X, Y.

Definice 5.2.4: *Náhodné veličiny X, Y nazýváme* ***nekorelované****, jestliže $\varrho(X,Y)=0$.*

Definice 5.2.5: ***Kovarianční matice*** *náhodného vektoru $\boldsymbol{X} = (X_1,\dots,X_n)$ je matice*

$$\Sigma_{\boldsymbol{X}} = \begin{bmatrix} \operatorname{cov}(X_1,X_1) & \operatorname{cov}(X_1,X_2) & \cdots & \operatorname{cov}(X_1,X_n)\\ \operatorname{cov}(X_2,X_1) & \operatorname{cov}(X_2,X_2) & \cdots & \operatorname{cov}(X_2,X_n)\\ \vdots & \vdots & \ddots & \vdots\\ \operatorname{cov}(X_n,X_1) & \operatorname{cov}(X_n,X_2) & \cdots & \operatorname{cov}(X_n,X_n) \end{bmatrix} =$$

$$= \begin{bmatrix} \mathrm{D}X_1 & \operatorname{cov}(X_1,X_2) & \cdots & \operatorname{cov}(X_1,X_n)\\ \operatorname{cov}(X_2,X_1) & \mathrm{D}X_2 & \cdots & \operatorname{cov}(X_2,X_n)\\ \vdots & \vdots & \ddots & \vdots\\ \operatorname{cov}(X_n,X_1) & \operatorname{cov}(X_n,X_2) & \cdots & \mathrm{D}X_n \end{bmatrix}.$$

Kovarianční matice je symetrická pozitivně semidefinitní[1], na diagonále má rozptyly.

Definice 5.2.6: ***Korelační matice*** *náhodného vektoru $\boldsymbol{X} = (X_1,\dots,X_n)$ je matice*

$$\varrho_{\boldsymbol{X}} = \begin{bmatrix} 1 & \varrho(X_1,X_2) & \cdots & \varrho(X_1,X_n)\\ \varrho(X_2,X_1) & 1 & \cdots & \varrho(X_2,X_n)\\ \vdots & \vdots & \ddots & \vdots\\ \varrho(X_n,X_1) & \varrho(X_n,X_2) & \cdots & 1 \end{bmatrix}.$$

Korelační matice je symetrická pozitivně semidefinitní.

[1] Čtenář neznalý těchto pojmů z lineární algebry nemusí přestat číst, neboť je nebudeme v této učebnici potřebovat. Uvádíme je jen proto, že jsou to užitečné vlastnosti pro další práci s kovariančními maticemi.

Vícerozměrné normální rozdělení

Z rozdělení náhodných vektorů uvedeme jen to nejdůležitější, **vícerozměrné normální rozdělení** $\mathrm{N}(\mu, \Sigma)$. To popisuje speciální případ náhodného vektoru, jehož složky mají normální rozdělení a nemusí být nekorelované. Jeho hustota je

$$f_{\mathrm{N}(\mu,\Sigma)}(u) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\tfrac{1}{2}\,(u-\mu)^{\mathrm{T}}\,\Sigma^{-1}\,(u-\mu)\right),$$

kde $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$, $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ je vektor středních hodnot, $\Sigma \in \mathbb{R}^{n\times n}$ je symetrická pozitivně definitní kovarianční matice a Σ^{-1} je matice k ní inverzní. Marginální rozdělení j-té složky je $\mathrm{N}(\mu_j, \Sigma_{jj})$, kde Σ_{jj} je j-tý diagonální prvek matice Σ.

Charakteristiky náhodného vektoru kopírují jednodimenzionální případ, přibývají k nim však nové, vyjadřující míru závislosti náhodných veličin. K tomu slouží zejména korelace jako míra lineárního vztahu dvou náhodných veličin. Bez ní by se neobešla např. kompenzace chvění ruky u kamery nebo optická myš. Korelace nám však neříká nic o tom, zda se jedná o vztah příčiny a následku.

Cvičení 5.2.7: [Spiegel et al. 2000] Sdružené rozdělení náhodných veličin X, Y s hodnotami $0, 1, 2$ (přesněji sdruženou pravděpodobnostní funkci náhodného vektoru (X, Y)) udává tabulka:

X \ Y	0	1	2
0	1/18	1/9	1/6
1	1/9	1/18	1/9
2	1/6	1/6	1/18

Stanovte pravděpodobnost $P[1/2 \le X \le 3,\ Y \ge 1]$. Najděte marginální rozdělení a zjistěte, zda X, Y jsou nezávislé. Pokud ne, vypočtěte korelaci a popište rozdělení náhodného vektoru se stejnými marginálními rozděleními, jehož složky jsou nezávislé náhodné veličiny.

Řešení: $P[1/2 \le X \le 3,\ Y \ge 1] = 1/18 + 1/9 + 1/6 + 1/18 = 7/18$ (součet hodnot vyznačených v tabulce kurzívou). Pravděpodobnostní funkce marginálních rozdělení p_X, p_Y jsou dány sloupcovými, resp. řádkovými součty. Náhodné veličiny X, Y jsou závislé (jinak by hodnost matice v tabulce musela být 1, tj. každý řádek by musel být násobkem libovolného nenulového řádku, stejně tak pro sloupce). Nezávislé veličiny se stejnými rozděleními by určovaly sdruženou pravděpodobnostní funkci v druhé tabulce (každá hodnota je součin marginálních).

X \ Y	0	1	2	$\sum$
0	1/18	1/9	1/6	1/3
1	1/9	*1/18*	*1/9*	5/18
2	1/6	*1/6*	*1/18*	7/18
$\sum$	1/3	1/3	1/3	1

	0	1	2	$\sum$
0	1/9	1/9	1/9	1/3
1	5/54	5/54	5/54	5/18
2	7/54	7/54	7/54	7/18
$\sum$	1/3	1/3	1/3	1

Výpočet korelace:

$$\mathrm{E}(X\cdot Y)=\frac{1}{18}\cdot 1+\left(\frac{1}{9}+\frac{1}{6}\right)\cdot 2+\frac{1}{18}\cdot 4=\frac{5}{6},$$
$$\mathrm{E}X\cdot\mathrm{E}Y=\left(\frac{5}{18}\cdot 1+\frac{7}{18}\cdot 2\right)\cdot\frac{1}{3}=\frac{19}{18}\cdot\frac{1}{3}=\frac{19}{54},$$
$$\mathrm{cov}(X,Y)=\mathrm{E}(X\cdot Y)-\mathrm{E}X\cdot\mathrm{E}Y=\frac{5}{6}-\frac{19}{54}=\frac{13}{27}\doteq 0.481,$$
$$\sigma_X=\sqrt{\mathrm{E}(X^2)-(\mathrm{E}X)^2}=\sqrt{\left(\frac{5}{18}\cdot 1+\frac{7}{18}\cdot 2^2\right)-\left(\frac{19}{18}\right)^2}=\sqrt{\frac{233}{324}}\doteq 0.848,$$
$$\sigma_Y=\sqrt{\mathrm{E}(Y^2)-(\mathrm{E}Y)^2}=\sqrt{\frac{5}{3}-1}=\sqrt{\frac{2}{3}}\doteq 0.816,$$
$$\varrho(X,Y)=\frac{\mathrm{cov}(X,Y)}{\sigma_X\,\sigma_Y}\doteq\frac{0.481}{0.848\cdot 0.816}\doteq 0.695.$$

Cvičení 5.2.8: * Náhodné veličiny U_1, V_1 jsou nezávislé, rovněž náhodné veličiny U_2, V_2 jsou nezávislé. Všechny mají alternativní rozdělení s hodnotami 0, 1, s parametry 1/2 pro U_1, V_2, 3/4 pro U_2, V_1. Rozdělení náhodného vektoru (X, Y) je směsí s koeficientem 1/2 z rozdělení náhodných vektorů $(U_1, V_1), (U_2, V_2)$, tj. $X = \mathrm{Mix}_{1/2}(U_1, U_2)$, $Y = \mathrm{Mix}_{1/2}(V_1, V_2)$. Určete korelaci $\varrho(X, Y)$.

Řešení: Pravděpodobnostní funkce náhodných vektorů (včetně marginálních rozdělení) jsou v tabulkách:

V_1 / U_1	0	1	$\sum$
0	1/8	3/8	1/2
1	1/8	3/8	1/2
$\sum$	1/4	3/4	1

V_2 / U_2	0	1	$\sum$
0	1/8	1/8	1/4
1	3/8	3/8	3/4
$\sum$	1/2	1/2	1

Y / X	0	1	$\sum$
0	1/8	1/4	3/8
1	1/4	3/8	5/8
$\sum$	3/8	5/8	1

$$\mathrm{E}(X\cdot Y)=\frac{3}{8},$$
$$\mathrm{E}X\cdot\mathrm{E}Y=\left(\frac{5}{8}\right)^2=\frac{25}{64}.$$
$$\mathrm{cov}(X,Y)=\mathrm{E}(X\cdot Y)-\mathrm{E}X\cdot\mathrm{E}Y=\frac{3}{8}-\frac{25}{64}=\frac{-1}{64},$$
$$\sigma_X=\sigma_Y=\sqrt{\frac{3}{8}\frac{5}{8}}=\frac{\sqrt{15}}{8}\doteq 0.484.$$
$$\varrho(X,Y)=\frac{\mathrm{cov}(X,Y)}{\sigma_X\,\sigma_Y}=\frac{\frac{-1}{64}}{\frac{15}{64}}=\frac{-1}{15}\doteq -0.0667.$$

Na tomto příkladu je zajímavé, že směs dvou nezávislých rozdělení není nezávislá. Takto se dopustil poučné chyby student, kterému vyšla významná korelace mezi inteligencí a tělesnou výškou. To bylo překvapivé, neboť řada předchozích výzkumů nikdy žádnou takovou korelaci nezjistila. Ukázalo se, že udělal chybu ve výběru vzorku. Pro nedostatek dat použil dvě skupiny lidí, které nebyly typické. Jedna se vyznačovala velkou inteligencí, druhá výškou. Korelace mu vyšla na stejném principu jako zde.

Cvičení 5.2.9: [Rogalewicz 2000] Dvojrozměrný náhodný vektor má rovnoměrné rozdělení v jednotkovém kruhu se středem $(0, 0)$. Dokažte, že složky (=kartézské souřadnice) jsou závislé a nekorelované. Co se změní, pokud bude rovnoměrné rozdělení jen na průniku tohoto kruhu s prvním kvadrantem (tj. obě souřadnice budou nezáporné) nebo na kružnici?

5.3 Lineární prostor náhodných veličin

V této kapitole předpokládáme pevně zvolený pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$.

Označme $\mathcal{L}$ množinu všech náhodných veličin na $(\Omega, \mathcal{A}, P)$, tj. $\mathcal{A}$-měřitelných funkcí $\Omega \to \mathbb{R}$. Operacím sčítání náhodných veličin a jejich násobení reálným číslem odpovídají příslušné operace s funkcemi (prováděné na Ω bod po bodu). Stejně jako funkce, tvoří i náhodné veličiny z $\mathcal{L}$ lineární prostor.

Dále se omezíme na prostor $\mathcal{L}_2$ všech náhodných veličin z $\mathcal{L}$, které mají rozptyl; $\mathcal{L}_2$ je lineární podprostor prostoru $\mathcal{L}$. Na $\mathcal{L}_2$ lze definovat binární operaci $\bullet\colon \mathcal{L}_2 \times \mathcal{L}_2 \to \mathbb{R}$

$$X \bullet Y = \mathrm{E}(XY) .$$

Ta je *bilineární* (tj. lineární v obou argumentech) a komutativní.

Pokud ztotožníme náhodné veličiny, které se liší jen na množině nulové míry, pak $\bullet$ je *skalární součin. (Po ztotožnění náhodných veličin X, Y, pro které $P[X \neq Y] = 0$, považujeme za prvky prostoru třídy ekvivalence místo jednotlivých náhodných veličin.)*

Připomeňme, že skalární součin indukuje pojem ortogonality; vektory X, Y jsou **ortogonální**, právě když $X \bullet Y = 0$. Skalární součin $\bullet$ definuje také *normu*

$$||X|| = \sqrt{X \bullet X} = \sqrt{\mathrm{E}(X^2)}$$

a *metriku* (vzdálenost)

$$d(X,Y) = ||X - Y|| = \sqrt{\mathrm{E}\left((X-Y)^2\right)}.$$

(Bez předchozího ztotožnění by toto byla jen pseudometrika, mohla by vyjít nulová i pro $X \neq Y$.)

V $\mathcal{L}_2$ rozlišíme dva důležité podprostory:

- jednorozměrný prostor $\mathcal{R}$ všech konstantních náhodných veličin (které mají Diracovo rozdělení),
- prostor $\mathcal{N}$ všech náhodných veličin s nulovou střední hodnotou.

Prostor $\mathcal{N}$ je ortogonální doplněk podprostoru $\mathcal{R}$, tj.

$$\mathcal{N} = \{X \in \mathcal{L}_2 \mid (\forall Y \in \mathcal{R} : X \bullet Y = 0)\} .$$

Pomocí těchto podprostorů získají základní číselné charakteristiky náhodných veličin názorný význam. Střední hodnota $\mathrm{E}X$ je souřadnice vektoru X ve směru $\mathcal{R}$; pokud ztotožňujeme reálné číslo $\mathrm{E}X$ s příslušnou konstantní náhodnou veličinou, pak je to kolmý průmět vektoru X do podprostoru $\mathcal{R}$. Podobně $X - \mathrm{E}X$ je kolmý průmět X do podprostoru $\mathcal{N}$. Normovaná náhodná veličina

$$\mathrm{norm}\, X = \frac{X - \mathrm{E}X}{\sigma_X}$$

je jednotkový vektor, který má stejný směr jako kolmý průmět X do $\mathcal{N}$. Směrodatná odchylka $\sigma_X = ||X - \mathrm{E}X||$ je vzdálenost X od $\mathcal{R}$.

Z kolmosti vektorů $X - \mathrm{E}X \in \mathcal{N}$, $\mathrm{E}X \in \mathcal{R}$ a Pythagorovy věty plyne (4.11):

$$||X||^2 = ||X - \mathrm{E}X||^2 + ||\mathrm{E}X||^2 ,$$
$$\mathrm{E}(X^2) = \mathrm{D}X + (\mathrm{E}X)^2 .$$

Lineární podprostor náhodných veličin s nulovými středními hodnotami

Speciálně pro náhodné veličiny *s nulovými středními hodnotami* (tj. z podprostoru $\mathcal{N}$) vychází

$$\begin{aligned} \mathrm{D}X &= X \bullet X\,, \\ \sigma_X &= \|X\|\,, \\ \mathrm{cov}(X,Y) &= X \bullet Y\,, \\ \varrho(X,Y) &= \frac{\mathrm{cov}(X,Y)}{\sigma_X\,\sigma_Y} = \frac{X \bullet Y}{\|X\|\,\|Y\|}\,, \end{aligned}$$

takže kovariance je skalární součin a korelace $\varrho(X,Y)$ je kosinus úhlu vektorů $X, Y \in \mathcal{N}$.

Důsledek 5.3.1: *Náhodné veličiny* X, Y s nulovými středními hodnotami *jsou ortogonální, právě když jsou nekorelované.*

Pro obecné vektory z $\mathcal{L}_2$ není význam kovariance a korelace tak jednoduchý,

$$\mathrm{cov}(X,Y) = X \bullet Y - \mathrm{E}X\,\mathrm{E}Y = (X - \mathrm{E}X) \bullet (Y - \mathrm{E}Y)$$

je skalární součin *průmětů* X, Y do $\mathcal{N}$ a $\varrho(X,Y)$ je kosinus jejich úhlu.

Lineární prostor náhodných veličin (aspoň těch, které mají rozptyl a dovolují tudíž zavést přirozeně skalární součin) umožňuje geometrickou interpretaci důležitých charakteristik náhodných veličin. Zejména korelace náhodných veličin souvisí s úhlem, který svírají odpovídající vektory.

Cvičení 5.3.2: O náhodných veličinách X, Y víme následující údaje: $\mathrm{E}X = 10$, $\mathrm{E}Y = 15$, $\sigma_X = 2$, $\sigma_Y = 3$. Co lze říci o hodnotě $\mathrm{E}(X\,Y)$?

5.4 Lineární regrese

Příklad 5.4.1: Je dán náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ a náhodná veličina Y. *(Předpokládáme, že všechny tyto náhodné veličiny mají rozptyl, tj. jsou z* $\mathcal{L}_2$*).* Máme najít takové koeficienty $c_1, \ldots, c_n$, aby lineární kombinace $\sum_k c_k X_k$ byla co nejlepší aproximací náhodné veličiny Y ve smyslu minimalizace kritéria

$$\Big\|\sum_k c_k X_k - Y\Big\|\,.$$

Řešení: K vektoru Y hledáme nejbližší bod v lineárním podprostoru, který je lineárním obalem vektorů $X_1, \ldots, X_n$; řešením je kolmý průmět. Ten je charakterizován tím, že vektor $\sum_k c_k X_k - Y$ je kolmý na X_j, $j = 1, \ldots, n$,

$$\Big(\sum_k c_k X_k - Y\Big) \bullet X_j = 0\,,$$

$$\sum_k c_k\,(X_k \bullet X_j) = Y \bullet X_j\,. \tag{5.2}$$

To je soustava lineárních rovnic, jejímž řešením určíme koeficienty $c_1, \ldots, c_n$. □

Soustava (5.2) se nazývá **soustava normálních rovnic**. Speciálně pro náhodné veličiny *s nulovými středními hodnotami* má tvar

$$\sum_k c_k\,\mathrm{cov}\,(X_k, X_j) = \mathrm{cov}\,(Y, X_j)\,,$$

takže matice soustavy je kovarianční matice $\Sigma_{\boldsymbol{X}}$.

Dostali jsme návod, jak jednu náhodnou veličinu co nejlépe (ve smyslu *metody nejmenších čtverců*) aproximovat pomocí lineární kombinace daných náhodných veličin. V aplikacích se tato náhrada často používá, nebo se aspoň ptáme, jak je dobrá, tj. nakolik lze (neznámou) náhodnou veličinu předpovědět pomocí jiných (známých).

6. Limitní věty a zákony velkých čísel

6.1 Potřeba odhadu u náhodných pokusů

V příkladu 1.6.2 jsme navrhli postup, jak experimentálně odhadnout číslo π opakováním jednoduchého pokusu. Než bychom takový postup opravdu použili, stojí za to posoudit přesnost odhadu. Pro zjednodušení předvedeme obdobný postup u házení mincí, kde je rozdělení výsledků předem známo.

Příklad 6.1.1: **A.** Při $n = 1000$ hodech (správnou) mincí lze očekávat, že nám padne líc v přibližně $n/2 = 500$ případech. Počet líců je náhodná veličina X s binomickým rozdělením $\mathrm{Bi}(n, 1/2)$. Pravděpodobnost, že líc padne přesně 500×, je malá,

$$\binom{n}{\frac{n}{2}} \left(\frac{1}{2}\right)^n \doteq 0.0252\,.$$

Líc nemusí padnout vůbec, a to s pravděpodobností

$$\binom{n}{0} \left(\frac{1}{2}\right)^n = \frac{1}{2^n} \doteq 9.332\,6 \cdot 10^{-302}\,.$$

Lze očekávat, že s vysokou pravděpodobností bude počet líců v intervalu $\langle 400, 600\rangle$. Odhadněte tuto pravděpodobnost.

B. Jaká je pravděpodobnost, že při $N = 1\,000\,000$ hodech bude počet líců Y v intervalu $\langle 400\,000, 600\,000\rangle$?

Řešení (1. postup): **A.** Jelikož známe rozdělení, můžeme výsledek dostat „hrubou silou" jako součet pravděpodobností všech příznivých výsledků,

$$\left(\frac{1}{2}\right)^{1000} \sum_{k=400}^{600} \binom{1000}{k}\,.$$

Když jsem tento výraz počítal numericky poprvé, se standardním nastavením přesnosti, vyšlo 1. Víme však, že to je méně. Potvrdí to přesný výsledek ve tvaru zlomku *(lomítko najdete uprostřed)*: 1339 385 758 741 519 357 388 892 818 864 443 067 418 900 518 900 105 911 746 090 007 755 236 450 426 898 257 402 305 632 380 352 652 071 656 196 572 902 800 531 486 932 537 126 031 974 852 805 197 255 537 445 938 476 515 947 926 755 026 469 490 905 004 823 221 461 649 527 914 259 119 128 567 952 049 348 745 932 821 852 564 231 322 003 857 962 454 115 874 766 122 796 058 275 651 620 637 820 907 515/1339 385 758 982 834 151 185 531 311 325 002 263 201 756 014 631 917 009 304 687 985 462 938 813 906 170 153 116 497 973 519 619 822 659 493 341 146 941 433 531 483 931 607 115 392 554 498 072 196 837 321 850 491 820 971 853 028 873 177 634 325 632 796 392 734 744 272 769 130 809 372 947 742 658 424 845 944 895 692 993 259 632 864 321 399 559 710 817 770 957 553 728 956 578 048 354 650 708 508 672. Čitatel i jmenovatel se shodují v prvních 10 cifrách, podle toho nastavíme přesnost numerického výpočtu a dostaneme přibližně $0.999\,999\,999\,819 = 1 - 1.81 \cdot 10^{-10}$.

B. Úlohu B tímto způsobem řešit nebudeme. Raději předem odhadneme složitost. Při exaktním výpočtu by počet členů i počet násobení v nich vzrostly v poměru N/n, počet operací by se zvýšil v poměru $(N/n)^2 = 1\,000\,000$, ale vzrostla by navíc jejich pracnost s potřebou zpracování větších čísel. I paměťová náročnost by vzrostla. V podstatě totéž platí i pro numerický výpočet.

Ten by mohl kombinační čísla odhadovat pomocí přibližných vzorců se složitostí nižší než lineární; kdyby byla konstantní, mohl by výpočet trvat jen 1000× déle, ale byl by to jen přibližný odhad. Chyba součtu velkého počtu nepřesných sčítanců by jej mohla znehodnotit. Nelze ani vyloučit, že by hledaná pravděpodobnost, zřejmě blízká jedné, v důsledku numerických chyb vyšla větší než 1, což nemůže být správně. Pokud bychom chtěli eliminovat numerické chyby, mohli bychom sčítat jen malé sčítance odpovídající opačnému jevu a počítat výsledek ve tvaru

$$1-\left(\frac{1}{2}\right)^{1\,000\,000}\sum_{k=0}^{399\,999}\binom{1\,000\,000}{k}-\left(\frac{1}{2}\right)^{1\,000\,000}\sum_{k=600\,001}^{1\,000\,000}\binom{1\,000\,000}{k}=$$

$$=1-2\left(\frac{1}{2}\right)^{1\,000\,000}\sum_{k=0}^{399\,999}\binom{1\,000\,000}{k}.$$

Ani tento postup nelze doporučit, má více členů. □

V předchozím příkladu je známa střední hodnota $EX = n/2 = 500$ i rozptyl $DX = n/4 = 250$ (viz přehled základních rozdělení v kapitole B.1). Podobně $EY = 500\,000$, $DY = 250\,000$. Mohli bychom se pokusit odhadnout výslednou pravděpodobnost jen na základě těchto údajů, aniž bychom zacházeli do podrobností binomického rozdělení. V této kapitole vyložíme postupy, které to umožňují.

6.2 Čebyševova nerovnost

Intuitivně očekáváme, že náhodná veličiny nabývá „velkých" odchylek od své střední hodnoty s „malou" pravděpodobností. Přitom velikost odchylky je nutno měřit směrodatnou odchylkou. Kvantifikaci poskytuje následující věta:

Věta 6.2.1 (Čebyševova nerovnost): *Pro každou normovanou náhodnou veličinu* $\operatorname{norm} X$ *a pro všechna* $\delta > 0$ *platí nerovnost*

$$P\left[|\operatorname{norm} X| < \delta\right] \geq 1 - \frac{1}{\delta^2}\,.$$

Důkaz: Označme nezápornou náhodnou veličinu $Y = (\operatorname{norm} X)^2$ a odhadovanou pravděpodobnost

$$\beta = P\left[|\operatorname{norm} X| < \delta\right] = P[Y < \delta^2] = F_Y(\delta^2-)\,,$$

takže β-kvantil $q_Y(\beta) \geq \delta^2$. Podle (4.11) je

$$EY = E\left((\operatorname{norm} X)^2\right) = \underbrace{D\,(\operatorname{norm} X)}_{1} + \underbrace{(E\,(\operatorname{norm} X))^2}_{0} = 1\,.$$

Tuto střední hodnotu vyjádříme pomocí kvantilové funkce q_Y:

$$1 = \mathrm{E}Y = \int_0^1 q_Y(\alpha)\,\mathrm{d}\alpha = \int_0^\beta \underbrace{q_Y(\alpha)}_{\geq 0}\,\mathrm{d}\alpha + \int_\beta^1 \underbrace{q_Y(\alpha)}_{\geq \delta^2}\,\mathrm{d}\alpha \geq (1-\beta)\,\delta^2\,,$$

Odtud vyjádříme β:

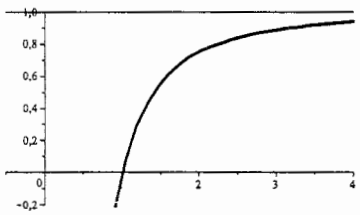
$$\beta \geq 1 - \frac{1}{\delta^2}\,.$$

Alternativní důkaz: Můžeme vyjádřit Y jako směs $Y = \operatorname{Mix}_\beta(L, U)$, kde L nabývá pouze hodnot z $\langle 0, \delta^2)$ (takže $\mathrm{E}L \geq 0$) a U nabývá pouze hodnot z $\langle \delta^2, \infty)$ (takže $\mathrm{E}U \geq \delta^2$). Pak

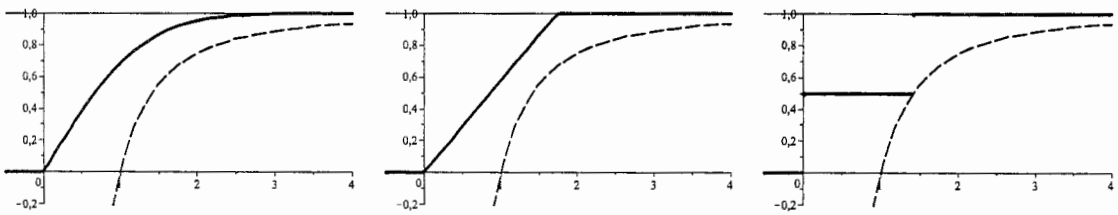
$$1 = \mathrm{E}Y = \beta \underbrace{\mathrm{E}L}_{\geq 0} + (1-\beta) \underbrace{\mathrm{E}U}_{\geq \delta^2} \geq (1-\beta)\,\delta^2 .$$

Rovnost nastává pro $U = \delta^2$, $L = 0$, tj. pro diskrétní rozdělení, které nabývá tří hodnot $\mathrm{E}X - \delta\,\sigma_X$, $\mathrm{E}X$, $\mathrm{E}X + \delta\,\sigma_X$ s pravděpodobnostmi po řadě $\frac{1-\beta}{2}$, β, $\frac{1-\beta}{2}$. □

Geometricky lze význam Čebyševovy nerovnosti interpretovat tak, že distribuční funkce $F_{|\mathrm{norm}\,X|}$ absolutní hodnoty z normované náhodné veličiny nemůže procházet pod grafem funkce $C(\delta) = 1 - \frac{1}{\delta^2}$, znázorněné v obr. 6.1. Pro $\delta \leq 1$ je $C(\delta) \leq 0$; v tom případě nám Čebyševova nerovnost neříká nic, co bychom bez ní nevěděli. Význam má jen pro $\delta > 1$.



Obrázek 6.1. Mez distribuční funkce daná Čebyševovou nerovností



Obrázek 6.2. Mez z Čebyševovy nerovnosti (čárkovaně) ve srovnání s distribučními funkcemi absolutních hodnot normalizovaných rozdělení: normálního, rovnoměrného a binomického Bi $(2, 1/2)$

Ve skutečnosti dává Čebyševova nerovnost velmi hrubý odhad, bývá splněna s velkou rezervou, viz srovnání na obr. 6.2. Přesto ji nelze bez dalších předpokladů nahradit lepším odhadem. Pro každé $\delta > 1$ existuje rozdělení, pro které je mez z Čebyševovy nerovnosti nejlepší možná. Toto rozdělení jsme odvodili v závěru důkazu Čebyševovy nerovnosti.

Čebyševovu nerovnost lze po úpravě použít i pro jakékoli náhodné veličiny, které lze normovat (dle vzorce $\mathrm{norm}\,X = \frac{X - \mathrm{E}X}{\sigma_X}$), tj. které *mají střední hodnotu a (nenulovou) směrodatnou odchylku.* Jiné požadavky na ně neklademe, zejména *nepotřebujeme znát rozdělení.*

Důsledek 6.2.2 (ekvivalentní tvary Čebyševovy nerovnosti): *Nechť X je náhodná veličina, která má nenulovou směrodatnou odchylku σ_X (tudíž má střední hodnotu $\mathrm{E}X$ a rozptyl $\mathrm{D}X = \sigma_X^2$). Pak pro všechna $\delta > 0$, $\varepsilon > 0$ platí*

$$P\left[\left|\frac{X-\mathrm{E}X}{\sigma_X}\right| < \delta\right] \geq 1 - \frac{1}{\delta^2}, \qquad P\left[\left|\frac{X-\mathrm{E}X}{\sigma_X}\right| \geq \delta\right] \leq \frac{1}{\delta^2},$$
$$P\left[|X - \mathrm{E}X| < \varepsilon\right] \geq 1 - \frac{\mathrm{D}X}{\varepsilon^2}, \qquad P\left[|X - \mathrm{E}X| \geq \varepsilon\right] \leq \frac{\mathrm{D}X}{\varepsilon^2}.$$

Důkaz: První dvě nerovnosti vznikly dosazením $\mathrm{norm}\,X = \frac{X-\mathrm{E}X}{\sigma_X}$ a přechodem k pravděpodobnosti negace. Poslední dvě nerovnosti dostaneme substitucí $\varepsilon = \delta\,\sigma_X$; zůstávají v platnosti i pro nulový rozptyl. □

Poznámka 6.2.3: Může se zdát překvapivé, že Čebyševovu nerovnost lze dokázat bez znalosti rozdělení. Nenajdeme rozdělení, které ji poruší. Podstata je v tom, že takové rozdělení by nutně mělo větší rozptyl, a ten byl v předpokladech omezen. Můžeme tedy Čebyševovu nerovnost chápat obráceně jako dolní odhad rozptylu .

Důsledek 6.2.4: *Nechť X je náhodná veličina, která má rozptyl $\mathrm{D}X$. Pak pro všechna $\varepsilon > 0$ platí*

$$\mathrm{D}X \geq \varepsilon^2 P\left[|X - \mathrm{E}X| \geq \varepsilon\right] ,$$

takže

$$\mathrm{D}X \geq \sup\left\{\varepsilon^2 P\left[|X - \mathrm{E}X| \geq \varepsilon\right] \mid \varepsilon > 0\right\} .$$

Řešení (příkladu 6.1.1, 2. postup): **A.** Dle Čebyševovy nerovnosti (důsledek 6.2.2) pro $\varepsilon = 101$:

$$P\left[|X - \mathrm{E}X| < \varepsilon\right] = P\left[|X - 500| < 101\right] = P\left[399 < X < 601\right] =$$
$$= P\left[400 \leq X \leq 600\right] \geq 1 - \frac{\mathrm{D}X}{\varepsilon^2} = 1 - \frac{250}{101^2} \doteq 0.975 .$$

Srovnáním s exaktním řešením vidíme, že jsme dostali pravdivý, ale velmi hrubý odhad.

B. Pro $\varepsilon = 100\,001$:

$$P\left[|X - \mathrm{E}X| < \varepsilon\right] = P\left[|X - 500\,000| < 100\,001\right] = P\left[400\,000 \leq X \leq 600\,000\right] \geq$$
$$\geq 1 - \frac{\mathrm{D}X}{\varepsilon^2} = 1 - \frac{250\,000}{100\,001^2} \doteq 0.999\,975 = 1 - 2.5 \cdot 10^{-5} .$$

Všimněte si, že pracnost odhadu byla v obou případech stejná, na rozdíl od 1. postupu. □

Poznámka 6.2.5: Čebyševova nerovnost v uvedených tvarech vypovídá jen o výskytu náhodné veličiny v intervalu, který je *symetrický kolem její střední hodnoty.* Kdybychom v příkladu 6.1.1 A chtěli odhadnout pravděpodobnost, že výsledek padne do intervalu $\langle 400, 700\rangle$, dovedli bychom ji pouze zdola omezit stejným výsledkem, jaký jsme odvodili pro interval $\langle 400, 600\rangle$. Zobecnění na intervaly, které nejsou symetrické kolem střední hodnoty, zde neuvádíme.

6.3 Centrální limitní věta

V Čebyševově nerovnosti jsme pracovali s téměř libovolným rozdělením (stačilo, že má rozptyl). V příkladu 6.1.1 však výsledné rozdělení vzniklo jako součet mnoha náhodných veličin se stejným rozdělením (v tomto případě jednotlivých hodů mincí, které mají alternativní rozdělení). Stejně je tomu v mnoha dalších náhodných pokusech, zejména při měřeních, která jsou zatížena chybami. Lze očekávat, že účinky *nezávislých* chyb se do jisté míry vyruší a umožní přesnější popis. Nyní zformulujeme základní výsledek tohoto druhu.

Věta 6.3.1 (centrální limitní věta): *Nechť $(U_j)_{j\in\mathbb{N}}$, je posloupnost* nezávislých *náhodných veličin se* stejným rozdělením *se střední hodnotou μ a směrodatnou odchylkou $\sigma \neq 0$. Částečné součty*

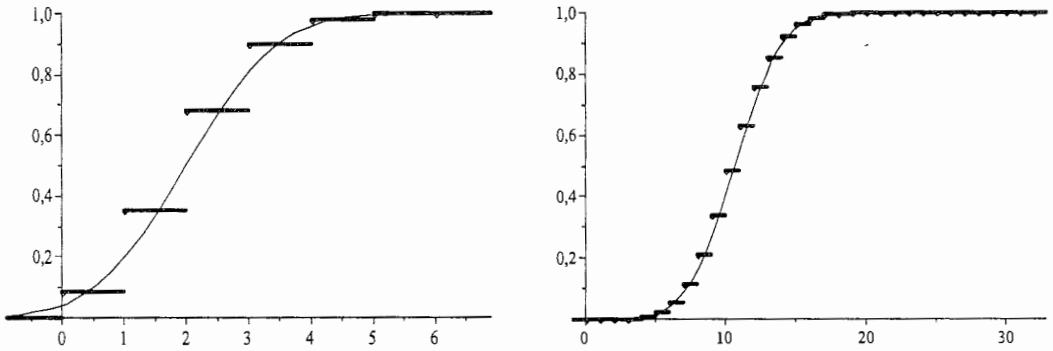
$$V_n = \sum_{j=1}^{n} U_j$$

mají charakteristiky $\mathrm{E}V_n = n\,\mu$, $\mathrm{D}V_n = n\,\sigma^2$, $\sigma_{V_n} = \sigma\sqrt{n}$. Rozdělení normovaných *náhodných veličin*

$$X_n = \operatorname{norm} V_n = \frac{V_n - n\,\mu}{\sigma\sqrt{n}}$$

konvergují k normovanému normálnímu rozdělení v následujícím smyslu:

$$\forall u \in \mathbb{R} : \lim_{n\to\infty} F_{X_n}(u) = \lim_{n\to\infty} F_{\operatorname{norm} V_n}(u) = \Phi(u) .$$



Obrázek 6.3. Aproximace distribučních funkcí binomických rozdělení Bi(6, 1/3) a Bi(32, 1/3) (tučně) normálními rozděleními (tence)

Toto je tzv. **Linderbergova-Lévyho věta**. Nejdříve byla odvozena pro případ výběru z alternativního rozdělení; tento speciální případ bývá uváděn jako **Moivreova-Laplaceova věta**.

Poznámka 6.3.2: Centrální limitní věta postihuje asymptotický tvar rozdělení vznikajících za specifických podmínek, viz obr. 6.3. Kdybychom vynechali normování náhodných veličin, směrodatná odchylka by rostla do nekonečna a limita rozdělení by nám nic užitečného neřekla.

Poznámka 6.3.3: Bohužel centrální limitní věta vypovídá jen o limitním chování rozdělení. V praxi vždy budeme pracovat s konečnou posloupností. Důsledkem centrální limitní věty je, že součet velkého počtu stejně rozdělených nezávislých náhodných veličin má *přibližně* normální rozdělení. Náhrada normálním rozdělením je pouze přibližná a přináší další chybu (kromě nevyhnutelné statistické chyby při realizaci náhodných veličin), která není v centrální limitní větě specifikována (nicméně existují její odhady, viz např. [Zvára, Štěpán 2002]).

Obvykle se tato chyba zanedbává pro „velký" počet náhodných veličin, avšak je problém, jaký počet je „velký". Chyba totiž záleží nejen na počtu náhodných veličin, ale i na jejich rozdělení, takže počet, který v jedné úloze může být vyhovující, může být v jiném kontextu zcela nedostatečný. Setkáme se s tím ve statistických úlohách. Přesto je centrální limitní věta velmi cenným nástrojem, protože dovoluje v některých případech složitá (někdy i neznámá) rozdělení aproximovat normálním, s nímž umíme pracovat.

Řešení (příkladu 6.1.1, 3. postup): **A.** Dle centrální limitní věty nahradíme binomické rozdělení normálním, neboť vzniklo součtem velkého počtu nezávislých hodů mincí (s alternativním rozdělením). Střední hodnotu a rozptyl už známe, zbývá náhodnou veličinu X znormovat a najít odpovídající hodnoty distribuční funkce normovaného normálního rozdělení:

$$\operatorname{norm} X = \frac{X - \frac{n}{2}}{\frac{1}{2}\sqrt{n}},$$

$$P\left[X \in \langle 400, 600\rangle\right] = P\left[\operatorname{norm} X \in \left\langle \frac{600 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}, \frac{400 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \right\rangle\right] \doteq$$

$$\doteq \Phi\left(\frac{600 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) - \Phi\left(\frac{400 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) \doteq \Phi(6.324) - \Phi(-6.324) \doteq$$

$$\doteq 0.999\,999\,999\,745 = 1 - 2.55 \cdot 10^{-10}.$$

(Tyto hodnoty mnohdy v tabulkách ani nenajdeme.) Získaný výsledek se liší od 1 zhruba o polovinu více než řešení z 1. postupu $1 - 1.81 \cdot 10^{-10}$. Vzhledem k tomu, že původní rozdělení bylo diskrétní, směli jsme použít i jiné meze intervalu než $\langle 400, 600\rangle$, např. $\langle 399.5, 600.5\rangle$; výsledek

je totiž beztak přibližný a jeho chyba zahrnuje i chybu náhrady diskrétního rozdělení spojitým. Pak bychom dostali odhad

$$P[X \in \langle 399.5, 600.5\rangle] \doteq \Phi\left(\frac{600.5 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) - \Phi\left(\frac{399.5 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) \doteq \Phi(6.356) - \Phi(-6.356) \doteq$$
$$\doteq 0.999\,999\,999\,793\,16 = 1 - 2.068 \cdot 10^{-10},$$

který je v dobré shodě s výsledkem 1. postupu, byl však získán s mnohem menším úsilím.

B.

$$\operatorname{norm} X = \frac{X - \frac{N}{2}}{\frac{1}{2}\sqrt{N}},$$

$$P[X \in \langle 400\,000, 600\,000\rangle] \doteq \Phi\left(\frac{600\,000 - \frac{N}{2}}{\frac{1}{2}\sqrt{N}}\right) - \Phi\left(\frac{400\,000 - \frac{N}{2}}{\frac{1}{2}\sqrt{N}}\right) \doteq$$
$$\doteq \Phi(200) - \Phi(-200) \doteq 1 - 2.57 \cdot 10^{-8689}.$$

Všimněte si, že pracnost odhadu byla v obou případech stejná, na rozdíl od 1. postupu. □

Poznámka 6.3.4: Centrální limitní větu lze použít i pro stanovení pravděpodobnosti, že náhodná veličina padne do *libovolného* intervalu, tedy i do takového, který *není symetrický kolem její střední hodnoty*. Čebyševova nerovnost o takových případech mnoho neříká.

Příklad 6.3.5: Vyřešíme modifikaci příkladu 6.1.1 A: odhadneme pravděpodobnost, že výsledek padne do intervalu $\langle 400, 700\rangle$:

$$P[X \in \langle 400, 700\rangle] \doteq \Phi\left(\frac{700 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) - \Phi\left(\frac{400 - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}\right) \doteq \Phi(12.649) - \Phi(-6.324) \doteq$$
$$\doteq 0.999\,999\,999\,915\,4 = 1 - 8.46 \cdot 10^{-11},$$

zatímco 1. postup dává

$$P[X \in \langle 400, 700\rangle] = \left(\frac{1}{2}\right)^{1000} \sum_{k=400}^{700} \binom{1000}{k} \doteq 0.999\,999\,999\,909\,9 = 1 - 9.01 \cdot 10^{-11}.$$

Uvedli jsme dva speciální postupy, jak vymezit interval, do něhož padnou hodnoty náhodné veličiny s danou velkou pravděpodobností. Oba jsou použitelné jen pro náhodné veličiny, které mají střední hodnotu a rozptyl. Čebyševovu nerovnost lze efektivně použít pouze pro intervaly symetrické podle střední hodnoty a vede na velmi hrubý odhad, který je však zaručený bez ohledu na typ rozdělení. Kdybychom znali rozdělení, např. normální, postupovali bychom dle kapitoly 3.7. Pokud rozdělení neznáme, ale víme, že náhodná veličina vznikla součtem velkého počtu *nezávislých* náhodných veličin se stejným rozdělením, pak můžeme použít centrální limitní větu a postupovat stejně jako u normálního rozdělení. Protože však tato věta popisuje pouze limitní případ a my máme konečný počet sčítanců, výsledek je zatížen dodatečnou chybou a není zaručený. Bývá to dobrý a užitečný odhad, nesmíme však zapomínat na podmínky jeho použitelnosti.

Cvičení 6.3.6: Rozhodněte, zda může existovat náhodná veličina X taková, že

$$P[\mathrm{E}X - 2\sigma_X \leq X \leq \mathrm{E}X + 2\sigma_X] < \frac{1}{2}.$$

Řešení: Nelze, z Čebyševovy nerovnosti vychází omezení

$$P\left[\mathrm{E}X - 2\sigma_X \leq X \leq \mathrm{E}X + 2\sigma_X\right] = P\left[|X - \mathrm{E}X| \leq 2\sigma_X\right] \geq 1 - \frac{1}{2^2} = \frac{3}{4}.$$

□

Cvičení 6.3.7: Pro $k = 1, \ldots, 10$ odhadněte pravděpodobnost, že se náhodná veličina liší od své střední hodnoty o méně než k-násobek své směrodatné odchylky (pokud ji má).

Řešení: Dle Čebyševovy nerovnosti (důsledek 6.2.2) pro $\varepsilon = k\,\sigma_X$,

$$P\left[|X - \mathrm{E}X| < k\,\sigma_X\right] = P\left[|X - \mathrm{E}X| < \varepsilon\right] \leq 1 - \frac{\mathrm{D}X}{\varepsilon^2} = 1 - \frac{1}{k^2},$$

dostáváme horní odhady dle tabulky:

k	1	2	3	4	5	6	7	8	9	10
horní odhad	1	0.75	0.889	0.9375	0.96	0.972	0.980	0.984	0.988	0.99

Porovnáním např. s distribuční funkcí normovaného normálního rozdělení (viz statistické tabulky) vidíme, že jsou to podstatně menší hodnoty. □

7. Informace, entropie a počítačové aspekty

7.1 Pojem informace, entropie diskrétního rozdělení

(Dle [Rényi 1966].)

Běžně pracujeme s pojmem informace, ale teprve rané období kybernetiky v polovině 20. století přispělo objevem, že informaci lze měřit. Tím lze kvantifikovat požadavky na přenos i uchovávání informace. Navíc se ukázalo, že souvisí i s fyzikálním pojmem entropie.

Příklad 7.1.1: Kolik binárních číslic potřebujeme k určení (a) dvanáctimístného čísla v dekadické soustavě, (b) data narození žijící osoby, (c) sekundy od začátku roku 1970, (d) obyvatele České republiky, (e) výsledku měření voltmetrem v rozsahu $\pm 2\,V$ s přesností 1 mV?

Vzhledem k tomu, že k binárních číslic dovoluje rozlišit 2^k možností, stačí k rozlišení n objektů $k = \lceil \log_2 n \rceil$ binárních číslic, kde $\lceil . \rceil$ značí zaokrouhlení na celá čísla nahoru. Proto Hartley[1] doporučil měřit informaci tak, že výběr jednoho z n objektů dává informaci $\log_2 n$. Volba základu logaritmu není podstatná, odpovídá jí vynásobení konstantou, tedy volba jednotky informace. Při použití dvojkového logaritmu je jednotkou jeden **bit** (angl. *bit*, zkratka z *binary digit*). Přirozené logaritmy se v tomto kontextu neosvědčily, neboť neodpovídají názornému náhodnému pokusu.

Řešení (příkladu 7.1.1): (a) K určení dekadické číslice stačí 4 bity, takže jistě stačí $4 \cdot 12 = 48$ bitů. Avšak 12 dekadických číslic dává 10^{12} možností, což je méně než $2^{40} = 1099\,511\,627\,776$, takže stačí 40 bitů. (b) K určení roku narození žijící osoby zatím postačuje 7 bitů, neboť není znám nikdo starší než $2^7 = 128$ let. Datum v roce bychom po číslicích kódovali $4 \cdot 4 = 16$ bity, ale stačí jich 9, neboť $2^9 = 512 \geq 366$. Celkem potřebujeme 16 bitů k určení roku i dne narození, a to až do věku přibližně $2^{16}/365.25 = 179.4$ let. Dekadických číslic by stačilo 5, místo toho se jich v rodném čísle používá 6, a ještě se tím nerozliší data odlišná o 100 let. (c) Na konci roku prvního vydání tohoto skripta, 2007, uplyne $60 \cdot 60 \cdot 24 \cdot (365 \cdot 38 + 10) = 1\,199\,232\,000$ s od začátku roku 1970 (s korekcí na přestupné roky, ale bez dalších korekcí, řádově v sekundách). Potřebujeme-li 31 bitů, vystačíme s nimi ještě dalších přibližně $2^{31}/(60 \cdot 60 \cdot 24 \cdot 365.25) - 38 = 30.05$ let. (d) Obyvatel České republiky je méně než $2^{24} = 16\,777\,216$, takže by postačovalo 24 bitů, což je např. 6 hexadecimálních číslic. Dekadických číslic by stačilo 7 až 8; místo toho se používá v rodném čísle 10 číslic, z nichž jedna je kontrolní. Zato je takový údaj srozumitelnější. (e) Výsledek je dán 4 dekadickými ciframi a znaménkem, ale je celkem 4001 možností, k jejichž kódování stačí 12 bitů. □

Hartleyho pojetí informace lze motivovat následovně:

Tvrzení 7.1.2 (Hartleyho vlastnosti informace): *[Rényi 1966] Funkce* $I\colon \mathbb{N} \to \mathbb{R}$,

$$I(n) = \log_2 n\,, \tag{7.1}$$

je jediná reálná funkce na $\mathbb{N}$ *následujících vlastností:*

1. $I(m\,n) = I(m) + I(n)$,
2. I *je neklesající,*
3. $I(2) = 1$.

[1] Hartley, R.V.: Transmission of Information. *Bell Syst. Techn. J.* **7** (1928), 535–563.

Poznámka 7.1.3: První Hartleyho vlastnost se vztahuje k následující situaci: $m\,n$ možností uspořádáme do m řádků a n sloupců. Máme-li určit jednu z nich, potřebujeme informaci $I(m)$ pro určení řádku a $I(n)$ pro určení sloupce. Tomu odpovídá informace $I(m\,n)$, určující jedinou možnost z $m\,n$. Tím, že jsme na pravé straně použili součet (a ne třeba součin), jsme vybrali pro měření informace logaritmické měřítko. Volba to byla v podstatě libovolná, ale prakticky užitečná. Někdy se tato vlastnost označuje jako *aditivita informace* [Rényi 1966], což však není vhodný termín, protože operace na levé straně není sčítání, ale násobení.

Hartleyho přístup se osvědčuje, když o možnostech nevíme nic a považujeme je všechny za stejně pravděpodobné (jako v Laplaceově modelu). Pak zůstanou jeho závěry zachovány i v následujícím zobecnění, které se však zaměří i na popis různě pravděpodobných možností.

Příklad 7.1.4: Bob si koupil 1000 losů, z nichž každý miliontý vyhrává. Bob nevidí výsledkovou listinu, Alice ano a zná i čísla Bobových losů. Jak dlouhou SMS zprávou mu může sdělit výsledek, pokud si předem dohodli způsob kódování?

Lze samozřejmě poslat řetězec 1000 binárních cifer, popřípadě vhodně zakódovaný povolenými znaky. Z toho se Bob doví, které losy vyhrávají. Pokud ho zajímá pouze počet vyhrávajících losů (které to jsou, si sdělí případně později), pak by bylo jen 1001 možností, které lze popsat pomocí 10 bitů.

Daleko spíše však Alice pošle krátkou zprávu, že Bob nic nevyhrál, neboť to se stane s pravděpodobností $(1 - 1/1000000)^{1000} \doteq 0.999\,000\,5$. Jestli vyhraje, jistě rád zaplatí delší podrobnou zprávu.

Rovněž výherní listina loterie obsahuje pouze čísla vyhrávajících losů, nikoli těch ostatních, čímž se podstatně zkrátí.

Podobně i noviny věnují více místa zprávám o vysokých výhrách, než většině soutěžících, kteří nevyhráli nic.

Předchozí příklad ukazuje nedostatek Hartleyho přístupu. Jestliže nejsou všechny výsledky stejně pravděpodobné, pak některé přinášejí málo informace, jiné více. Tomu můžeme přizpůsobit i kódování, které bude používat krátké zprávy pro nejpravděpodobnější výsledky a delší pro výsledky neočekávané, čímž se zkrátí *průměrná* délka zprávy (při mnohanásobném opakování pokusu). Při vhodném kódování může být střední délka zprávy kratší než dle (7.1). Obdržená informace závisí na výsledku, který sděluje, a stává se náhodnou veličinou. V tom případě bychom se měli ptát spíše na *průměrnou* informaci, kterou dostaneme, přesněji na *střední hodnotu* informace jako náhodné veličiny.

Příklad 7.1.5: Při digitálním přenosu televizního obrazu se využívá skutečnost, že jej lze do jisté míry předvídat a posílat jen odchylky, např. od předchozího obrazu. Tím se množství přenášené informace několikanásobně sníží. Ovšem ve chvíli, kdy se zcela změní obraz, potřebujeme poslat plné množství informace. Z hlediska kapacity přenosového kanálu je potřebné znát průměrné množství informace, či spíše horní odhad, který bude málokdy překročen.

V uvedených příkladech množství informace závisí na tom, jak neočekávaný (málo pravděpodobný) výsledek vyšel. Chceme měřit jak informaci jednoho výsledku, tak průměrnou informaci přes všechny možné výsledky (se zohledněním jejich pravděpodobnosti). Odpovídající vzorec lze odvodit i z Laplaceova pravděpodobnostního modelu.

Předpokládejme, že náhodný pokus má n možných stejně pravděpodobných výsledků (můžeme je ztotožnit s elementárními jevy, tvořícími množinu Ω). Ty rozlišujeme do k disjunktních tříd s četnostmi $n_1, \ldots, n_k$, $\sum_{j=1}^{k} n_j = n$. (Jinými slovy, máme úplný soubor jevů $A_1, \ldots, A_k$, přičemž A_j má pravděpodobnost $p_j = n_j/n$.) Pro plné určení každého z n výsledků potřebujeme informaci $I(n) = \log_2 n$. Pokud se dovíme, že výsledek patří do j-té třídy (nastal jev A_j), stačí nám pak k jeho určení informace $I(n_j) = \log_2 n_j$. To znamená, že příslušnost do j-té třídy nám poskytla informaci

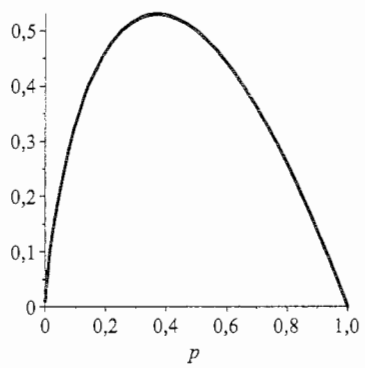
$$I(n) - I(n_j) = \log_2 n - \log_2 n_j = \log_2 \frac{n}{n_j} = -\log_2 p_j\,. \tag{7.2}$$

Tento vzorec vystihuje skutečnost, že málo pravděpodobné výsledky nám poskytly mnoho informace, pokud nastaly.

Průměrnou informaci, kterou nám přináší zařazení do třídy (před pokusem, kdy ještě nevíme nic o výsledku) dostaneme jako vážený průměr všech výsledků, přičemž váhami jsou příslušné pravděpodobnosti:

$$h(p_1, \ldots, p_k) = \frac{1}{n} \sum_{j=1}^{k} n_j \log_2 \frac{n}{n_j} = -\sum_{j=1}^{k} p_j \log_2 p_j \,. \tag{7.3}$$

Tento tzv. **Shannonův vzorec**[2] jsme odvodili pro racionální pravděpodobnosti (pro něž lze najít takové n, že $p_j = n_j/n$ pro nějaká celá čísla n_j), ale lze jej stejně použít i pro iracionální pravděpodobnosti. Funkce h se nazývá **Shannonova entropie** nebo pouze **entropie**. Závisí jen na pravděpodobnostech $p_1, \ldots, p_k$, nikoli na celkovém počtu elementárních jevů n. Vyjadřuje střední hodnotu informace, kterou dostaneme, pokud se dovíme, který z k disjunktních jevů nastal, jestliže jejich pravděpodobnosti byly $p_1, \ldots, p_k$. Příspěvek výsledku s pravděpodobností p je dán funkcí $\iota(p) = -p \log_2 p$. Ta je nezáporná a má nulové limity v 0 i v 1. Její hodnota v 0 dává neurčitý výraz, obvykle se zvlášť definuje $\iota(0) = 0$. (To jsme nepotřebovali, dokud jsme uvažovali pouze výsledky s nenulovou pravděpodobností.) Její graf je na obrázku 7.1. Důsledkem je, že entropie *diskrétního* náhodného pokusu je vždy nezáporná. Nulová je právě tehdy, když jedna z pravděpodobností je jednotková, ostatní nulové.



Obrázek 7.1. Příspěvek k průměrné informaci v závislosti na pravděpodobnosti

Poznámka 7.1.6: Toto pojetí informace odpovídá i psychologickým pozorováním. Při pokusech měly osoby co nejrychleji stisknout tlačítko odpovídající třídě ukázaného objektu. Přitom jednotlivé třídy byly ukazovány s různou pravděpodobností, takže některá tlačítka byla používána častěji. Průměrná doba reakce při různých podmínkách pokusu odpovídala Shannonově entropii [Jaglom 1964].

Zavedli jsme entropii, ač jsme chtěli měřit informaci. V matematice se tyto pojmy příliš nerozlišují. Obvykle se jako informace označuje míra množství neurčitosti nebo nejistoty o nějakém náhodném pokusu, odstraněné realizací tohoto pokusu a získáním výsledku [Apl. mat. 1978]. Při různých pravděpodobnostech výsledků se dovíme někdy více, někdy méně, takže takto chápaná informace je náhodná veličina. Entropie je její střední hodnota (v pravděpodobnostním popisu před pokusem, kdy ještě nevíme, jakou informaci dostaneme).

Naopak ve fyzice a v běžném životě jsou pojmy entropie a informace považovány za protichůdné. Entropie kvantifikuje neurčitost a neuspořádanost, informace je naopak prostředkem

[2] Shannon, E.: A mathematical theory of communication, *Bell Syst. Techn. J.* **27** (1948), 379–423, 623–653. Ke stejnému vzorci dospěl nezávisle i N. Wiener.

k jejímu omezení (což nebrání tomu, abychom velikost obou měřili stejně). V matematice neurčitost výsledku dává *možnost* vyjádřit mnoho informace, na rozdíl od systému, kde je vše předem určeno.[3]

Definičním oborem Shannonovy entropie h ze vzorce (7.3) je množina D_h všech k-tic nezáporných reálných čísel $p_1, \ldots, p_k$ takových, že $\sum_{j=1}^{k} p_j = 1$, přičemž k probíhá celou množinu přirozených čísel. Místo těchto pravděpodobností můžeme vyjít z rozdělení diskrétní náhodné veličiny X, která nabývá k hodnot s pravděpodobnostmi $p_1, \ldots, p_k$, a hovořit o **entropii diskrétní náhodné veličiny**, resp. jejího **rozdělení**; značí se obvykle $H(X)$. Entropie nezáleží na konkrétních (navzájem různých) hodnotách diskrétní náhodné veličiny, pouze na pravděpodobnostech rozlišitelných výsledků, proto platí následující výsledek:

Tvrzení 7.1.7: *Zobrazíme-li diskrétní náhodnou veličinu prostým zobrazením, její entropie se nezmění.*

Předchozí tvrzení platí i pro zobrazení, které je prosté na oboru hodnot příslušné náhodné veličiny.

Entropii lze motivovat i následovně:

Tvrzení 7.1.8 (vlastnosti Shannonovy entropie): *[Rényi 1966] Shannonova entropie h: $D_h \to \langle 0, \infty)$ daná vzorcem (7.3) je jediná reálná funkce na D_h následujících vlastností:*

1. *h nezávisí na permutaci argumentů,*
2. *funkce $p \mapsto h(p, 1-p)$ je spojitá,*
3. *$h(1/2, 1/2) = 1$,*
4. *$h(p_1, p_2, p_3, \ldots, p_n) = h(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)\, h\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$.*

První dvě Shannonovy vlastnosti jsou přirozené požadavky, třetí pouze určuje jednotku (bit). Čtvrtý vztah říká, že spojením dvou tříd (první a druhé) do jedné ztratíme entropii

$$(p_1 + p_2)\, h\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Dosazením ze vzorce (7.3) bychom dostali

$$\begin{aligned}(p_1 + p_2)\, h\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) &= \\ &= (p_1 + p_2)\left(\frac{-p_1}{p_1 + p_2} \log_2 \frac{p_1}{p_1 + p_2} + \frac{-p_2}{p_1 + p_2} \log_2 \frac{p_2}{p_1 + p_2}\right) = \\ &= -p_1 \log_2 p_1 - p_2 \log_2 p_2 + (p_1 + p_2) \log_2 (p_1 + p_2)\,,\end{aligned}$$

což odpovídá rozdílu ztráty entropie odpovídající první, resp. druhé třídě a zisku entropie odpovídající nové třídě, která vznikla spojením těchto dvou.

Pro n stejně pravděpodobných výsledků dostáváme ze vzorce (7.3) vzorec (7.1) jako speciální případ. V tomto případě je entropie největší:

Věta 7.1.9: *Pro daný počet výsledků $k \in \mathbb{N}$ je výraz $h(p_1, \ldots, p_k)$ maximální, právě když $p_j = 1/k$ pro všechna $j = 1, \ldots, k$.*

Důkaz: Pro $k = 2$ dostáváme

$$h(p_1, p_2) = h(p_1, 1 - p_1) = -p_1 \log_2 p_1 - (1 - p_1) \log_2 (1 - p_1)\,,$$

$$\frac{\partial}{\partial p_1} h(p_1, 1 - p_1) = \frac{\partial}{\partial p_1} \left(-p_1 \log_2 p_1 - (1 - p_1) \log_2 (1 - p_1)\right) = -\log_2 p_1 + \log_2 (1 - p_1)\,.$$

[3] Zde myslíme neurčitost *před* pokusem, nikoli neurčitost výsledku již *provedeného* pokusu, typickou pro princip neurčitosti v kvantové teorii.

Tento výraz je nulový a entropie nabývá maxima pro $p_1 = 1/2 = p_2$.

Pro $k > 2$ použijeme tvrzení 7.1.8. Předpokládejme, že dvě z hodnot $p_1, \ldots, p_k$ jsou různé. Díky první vlastnosti můžeme bez újmy na obecnosti předpokládat, že jsou to první dvě, tj. $p_1 \neq p_2$. Pak použijeme čtvrtou vlastnost:

$$h(p_1, p_2, p_3, \ldots, p_n) = h(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)\, h\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right),$$

kde $\frac{p_1}{p_1+p_2} \neq \frac{p_2}{p_1+p_2}$. Pokud bychom nahradili p_1, p_2 jejich aritmetickým průměrem $(p_1 + p_2)/2$, změnila by se entropie na hodnotu

$$\begin{aligned} h\left(\tfrac{p_1+p_2}{2}, \tfrac{p_1+p_2}{2}, p_3, \ldots, p_n\right) &= h(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)\, h\left(\tfrac{1}{2}, \tfrac{1}{2}\right) > \\ &> h(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)\, h\left(\tfrac{p_1}{p_1+p_2}, \tfrac{p_2}{p_1+p_2}\right) = \\ &= h(p_1, p_2, p_3, \ldots, p_n), \end{aligned}$$

kde nerovnost vyplývá z předchozího výsledku pro $k = 2$. Tedy pro $p_1 \neq p_2$ nenastává maximum. □

Shannonova entropie je největší, pokud jsou všechny výsledky stejně pravděpodobné, a nejmenší, pokud je výsledek pokusu předem jednoznačně určen. Lze ji tedy považovat i za míru neuspořádanosti systému, který je náhodným pokusem popsán.

Poznámka 7.1.10: Nejčastěji používanou mírou variability rozdělení náhodné veličiny je rozptyl. Ten lze počítat pouze u kvantitativních (intervalových) náhodných veličin. Naproti tomu entropii lze počítat i u kvalitativních (nominálních) náhodných veličin, a zde může posloužit také k vyjádření variability. Má však nezastupitelnou roli i u kvantitativních náhodných veličin.

Cvičení 7.1.11: [Rényi 1966] Ověřte, že funkce

$$h^*(p_1, p_2, \ldots, p_n) = -\log_2\left(p_1^2 + \cdots + p_n^2\right)$$

splňuje první tři podmínky z tvrzení 7.1.8, ale nesplňuje čtvrtou.

Jsou-li pravděpodobnosti výsledků různé a známé, můžeme to využít k úspornějšímu kódování. U úsporného kódu budou všechny použité znaky přibližně stejně pravděpodobné. Entropie určuje teoretickou dolní mez průměrné délky zprávy při nejúspornějším kódování. Tím nám ukazuje meze možné komprese dat. Častěji známe rozdělení jen přibližně, a proto nedosáhneme maximální účinnosti komprese.

Praktické měření entropie může být však obtížné, neboť předpokládá důkladnou znalost modelu, abychom věděli, jaké jsou pravděpodobnosti jednotlivých výsledků. To ukazuje následující příklad:

Příklad 7.1.12: Mimozemšťan by při pohledu na hlasování v Poslanecké sněmovně ČR měl dojem, že se jedná o chaotický proces s téměř maximální entropií, neboť kladné i záporné odpovědi se vyskytují zhruba stejně často bez jakéhokoli viditelného systému (jiné odpovědi pro jednoduchost neuvažujeme). Kdo však zná pozadí, ví, že často hlasují stejně celé poslanecké kluby, které navíc svá stanoviska předem oznámily, takže výsledek hlasování lze stručně shrnout „nic nečekaného se nestalo." Ve skutečnosti tedy výsledek nese velmi málo informace, pokud někdo nehlasuje jinak, než se všeobecně očekávalo. Abychom správně vyčíslili entropii tohoto náhodného pokusu, musíme odhadnout pravděpodobnost odpovědi pro každého poslance zvlášť, nikoli globálně za celý parlament.

Obvykle ale nebývá informace kódována nejúsporněji i z dalšího důvodu: vzrostla by citlivost k chybám. Kdyby každá posloupnost znaků měla význam, nebylo by možné odhalit chyby, natož je opravit. Proto se často zapisuje či přenáší více informace než je teoreticky nutné. Přebytečná

informace se nazývá **redundance**. Při jejím použití některé posloupnosti znaků nemají smysl; pokud se vyskytnou, víme, že nastala chyba. Redundance se vyvinula např. v přirozených jazycích. Když uvidíme v textu slovo „přelkep", asi se domyslíme, že to měl být „překlep". (Závisí ovšem na kontextu, neboť zde bylo schválně napsáno „přelkep", abychom demonstrovali překlep.) Díky redundanci můžeme rozumět hovoru i v hlučné místnosti, kde nezachytíme všechny hlásky a někdy přeslechneme i celé slovo. (Ale špatně se to daří v cizím jazyce, kde redundanci neumíme tak dobře využít.) Redundance většinou umožňuje porozumět češtině i bez diakritiky. Nicméně jsou případy, kdy ji potřebujeme, např. slovu „RADIM" můžeme různým doplněním diakritiky dát čtyři různé významy. V tomto případě je redundance malá. Naproti tomu v umělých kódech se obvykle dbá na to, aby byla dodržena určitá minimální vzdálenost mezi nejbližšími slovy, která dovoluje rozpoznání chyb a případně i jejich opravu. K tomu se obvykle používá tzv. *Hammingova vzdálenost*, což je počet znaků, v nichž se liší dva stejně dlouhé řetězce (slova). Nejjednodušším zabezpečením je tzv. *paritní bit*, tj. jeden bit binárního slova, vypočtený z ostatních bitů tak, aby počet jedniček ve slově byl vždy lichý (nebo vždy sudý, podle dohody). Pokud počet jedniček neodpovídá očekávání, víme, že jde o chybu, i když ji v tomto případě neumíme odstranit. Paritní bit nenese *při bezchybném přenosu* žádnou informaci (z ostatních bitů již víme jeho hodnotu), je nositelem redundance. Ovšem *při chybách přenosu* může přinášet informaci o tom, že chyba nastala.

Teorie informace dává užitečný nástroj i pro řešení některých kombinatorických problémů.

Příklad 7.1.13: Z n mincí je jedna těžší než ostatní. K jejímu určení máme rovnoramenné váhy. Z kolika mincí můžeme jednu vadnou určit pomocí 3 vážení?

Řešení: Každé vážení může mít 3 výsledky (rovnováha, těžší je pravá/levá miska). Nejvíce informace získáme (náhodný pokus bude mít největší entropii), budou-li všechny 3 výsledky stejně pravděpodobné. (Informaci z jednoho vážení můžeme vyčíslit jako $\log_2 3 \doteq 1.585$ bitů, ale snazší bude počítat případy, které je třeba rozlišit.) Posloupnost 3 vážení může dát celkem $3^3 = 27$ možných výsledků, můžeme tedy připustit maximálně $n = 27$ mincí, z nichž právě jedna je těžší. Při prvním vážení dáme na každou misku $27/3 = 9$ mincí, čímž zajistíme stejnou pravděpodobnost všech 3 výsledků. Při nerovnováze dále hledáme mezi mincemi na té misce, která byla těžší, při rovnováze mezi zbývajícími. V každém případě máme 2 vážení na určení jedné z 9 mincí, dále postupujeme analogicky. Při menším počátečním počtu než 27 postupujeme obdobně; i když není dělitelný třemi, stačí nám zajistit, že zbývajících k vážení nemusí rozlišit více než 3^k možností.

□

Příklad 7.1.14: Ze 13 mincí má jedna jinou hmotnost než ostatní; nevíme, zda je těžší nebo lehčí. Jak ji určíme pomocí 3 vážení na rovnoramenných vahách?

Řešení: Máme až 27 možných výsledků, pomocí nichž máme rozlišit ze 13 možností. Zdá se, že máme nadbytek informace, neboť se nestaráme o to, zda je vadná mince těžší nebo lehčí. Nicméně to má vliv na výsledek a ve skutečnosti to na konci budeme vědět, pokud se vadná mince objeví aspoň jednou na váze. (Nejvýše jedna mince může zůstat nevážená, abychom ji mohli odlišit od ostatních.) Pokud tedy přidáme i znaménko odchylky, máme celkem 26 možných výsledků, které považujeme za stejně pravděpodobné. Z nich rozlišíme alespoň 25 (z nichž jedna má $2\times$ větší pravděpodobnost než ostatní). Po této úvaze už nám mnoho nadbytečné informace nezbývá. Z cvičných důvodů rozdíl vyčíslíme, i když pro vlastní řešení úlohy stačí opět hlídat, že pomocí zbývajících k vážení můžeme rozlišit nejvýše 3^k možností.

Určitě nemůžeme získat více informace než $3 \log_2 3 \doteq 4.755$ bitů. Rozlišení 26 stejně pravděpodobných výsledků by vyžadovalo $\log_2 26 \doteq 4.700$ bitů. Pokud dva z nich sloučíme dohromady (jejich pravděpodobnost bude $1/13$), pak se entropie podle čtvrté vlastnosti z tvrzení 7.1.8 sníží o $1/13 \cdot 1$ bit, tj. na $4.700 - 1/13 \doteq 4.623$ bitů. Máme tedy rezervu pouze $4.755 - 4.623 = 0.132$ bitů.

Rezerva nám dovoluje použít i taková vážení, při nichž nejsou všechny 3 možné výsledky stejně pravděpodobné. To musíme připustit již při prvním vážení, neboť počet mincí není dělitelný

třemi. Tomuto ideálu se musíme aspoň přiblížit. Dáme-li na každou misku vah 5, 4, resp. 3 mince, docílíme entropii v bitech

$$h\left(\frac{3}{13},\frac{5}{13},\frac{5}{13}\right) = -\frac{3}{13}\log_2\frac{3}{13} - 2\cdot\frac{5}{13}\log_2\frac{5}{13} \doteq 1.549\,,$$
$$h\left(\frac{5}{13},\frac{4}{13},\frac{4}{13}\right) = -\frac{5}{13}\log_2\frac{5}{13} - 2\cdot\frac{4}{13}\log_2\frac{4}{13} \doteq 1.577\,,$$
$$h\left(\frac{7}{13},\frac{3}{13},\frac{3}{13}\right) = -\frac{7}{13}\log_2\frac{7}{13} - 2\cdot\frac{3}{13}\log_2\frac{3}{13} \doteq 1.457\,.$$

To vše by bylo dostatečně blízko teoretického maxima

$$h\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) = -\log_2\frac{1}{3} \doteq 1.585\,.$$

avšak my potřebujeme zajistit, aby i *minimální* získaná informace dle vzorce (7.2) byla dostatečná. Ta je největší v prvních dvou případech, kdy získáme informaci alespoň $-\log_2\frac{5}{13} \doteq 1.379$ bitů. Ani to však nestačí. To lze poznat i z počtu zbývajících možností. Pokud dáme na každou misku 5 mincí a rovnováha je porušena, pak buď je jedna z 5 mincí na jedné misce těžší, nebo jedna z 5 mincí na druhé misce lehčí. To je 10 případů, které nemůžeme rozlišit zbývajícími dvěma váženími.

Jediná zbývající možnost je dát při prvním vážení na každou misku 4 mince. Při nerovnováze zbývá 8 „podezřelých" mincí a pro každou z nich víme, jaké znaménko by mohla mít její odchylka hmotnosti. Rozlišíme je tak, že přibližně třetinu z nich dáme stranou, třetinu necháme na stejné misce a třetinu dáme na opačnou misku než při prvním vážení. Protože 8 není dělitelné třemi, rozdělíme podezřelé mince na skupiny po 3, 3 a 2 mincích. Případný rozdíl v počtu mincí na miskách dorovnáme pomocí zaručeně správných mincí, které se neúčastnily prvního vážení. Při třetím vážení už snadno určíme jednu ze tří mincí.

Při rovnováze v prvním vážení nám zbývá 5 podezřelých mincí, tedy 10 možností, z nichž však stačí rozlišit 9, pokud jedna mince nebude nikdy vážena. To dává naději na řešení. Entropie

$$h\left(\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{5}\right) = -8\cdot\frac{1}{10}\log_2\frac{1}{10} - \frac{1}{5}\log_2\frac{1}{5} \doteq 3.122$$

je pod teoretickým maximem informace ze zbývajících 2 vážení,

$$2\cdot h\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) = -2\cdot\log_2\frac{1}{3} \doteq 3.170\,.$$

Přesto bychom z 5 dosud nevážených mincí nemohli dvěma váženími určit vadnou, nebýt toho, že máme po prvním vážení k dispozici dostatek mincí, o nichž víme, že jsou správné. Do druhého vážení vezmeme přibližně 2/3 zbývajících 5 podezřelých mincí, konkrétně 3. Při nerovnováze dokončíme jako v předchozím případě. Při rovnováze zbývají 2 mince, z nichž jednu porovnáme s některou správnou a dostaneme odpověď.

Ač se jedná o kombinatorickou úlohu, měření informace zde dovoluje jednoznačné rozhodnutí, která strategie může vést (a vede) k cíli. Místo entropie dle vzorce (7.3) bychom však vystačili v každém kroku s kontrolou, zda není více než 3^k možností, které zbývá rozlišit pomocí k vážení. □

Jedním ze základních přínosů kybernetiky je, že informace se dá měřit a že její střední hodnota – entropie – se dá vyčíslit pomocí Shannonova vzorce. Tím dostáváme mj. kritérium a návod pro vhodné uspořádání náhodného pokusu, kódování, kompresi dat apod. Navíc má pojem entropie význam i např. ve statistické mechanice.

7.2 Entropie spojitého rozdělení

Pro spojité rozdělení nelze předchozí úvahy použít. I jediné reálné číslo (realizace spojité náhodné veličiny) obsahuje v tomto pojetí nekonečnou informaci. Z praktického hlediska naráží tento přístup na problémy. Proto se zavádí *jiný pojem entropie*, který je formálně podobný. **Entropie spojitého rozdělení**, nebo též **spojité náhodné veličiny** X s hustotou f_X je

$$H(X) = \int_{-\infty}^{\infty} \iota(f_X(u))\,\mathrm{d}u\,,$$

kde

$$\iota(u) = \begin{cases} 0 & \text{pro } u = 0, \\ -u \log_2 u & \text{jinak.} \end{cases} \tag{7.4}$$

Entropie spojité náhodné veličiny nemusí existovat, pokud příslušný integrál nekonverguje.

Že se jedná o úplně jiný pojem než v diskrétním případě, je patrné už z toho, že tato entropie může být záporná. Je to tím, že hustota může být větší než 1, a pak je její logaritmus kladný. Další rozdíl je v tom, že neplatí analogie tvrzení 7.1.7: když zobrazíme hodnoty náhodné veličiny prostou funkcí, entropie se může změnit.

Příklad 7.2.1: Nechť náhodná veličina X má spojité rovnoměrné rozdělení na intervalu $\langle a, b\rangle$. Její entropie je

$$H(X) = (b-a) \cdot \iota\left(\frac{1}{b-a}\right) = \log_2 (b-a)\,.$$

Je-li $r > 0$, pak náhodná veličina $r\,X$ má entropii

$$H(r\,X) = \log_2 (r\,(b-a)) = H(X) + \log_2 r\,.$$

Z toho je vidět, že *libovolné* reálné číslo může být entropií spojité náhodné veličiny, navíc s rovnoměrným rozdělením.

Cílem však je spíše porovnávání entropie různých rozdělení. Vychází např. následující zajímavé výsledky:

Věta 7.2.2: *[Apl. mat. 1978] Největší entropii má*

1. *z rozdělení na omezeném intervalu rovnoměrné rozdělení,*
2. *z rozdělení na intervalu $(0, \infty)$ s danou střední hodnotou exponenciální rozdělení,*
3. *z rozdělení na $\mathbb{R}$ s danou střední hodnotou a rozptylem normální rozdělení.*

Pojem entropie bývá dále studován v souvislosti s podmíněným rozdělením, kdy vyjadřuje, nakolik znalost jedné náhodné veličiny vypovídá o jiné. Tyto úvahy přesahují rámec našeho textu, čtenář se více dozví např. v [Rényi 1966].

Ještě je třeba se zmínit o jednom Shannonově klíčovém přínosu k teorii informace. Říká, že v praxi lze *veškerou* informaci měřit (v bitech, jak jsme to dělali u diskrétních náhodných veličin, resp. náhodných pokusů s konečně mnoha výsledky). Je to tím, že naše měření má vždy omezenou přesnost, nemůžeme ve skutečnosti rozlišit nespočetně mnoho výsledků pomocí reálných čísel, neboť nejsme schopni testovat rovnost reálných čísel. Tak např. televizní signál se nemění libovolně rychle (má omezené kmitočtové spektrum, už kvůli přenosovému kanálu). Jednotlivé hodnoty jasu (resp. barev) lze rovněž nahradit konečně mnoha hodnotami (diskretizovat), aniž bychom pozorovali rozdíl, tj. úbytek informace. Na těchto principech lze pak přenášet digitalizovaný signál v kvalitě, na níž příjemce nepozná žádnou ztrátu informace. Zbývá dodat, že tento obecný princip Shannon formuloval v době, kdy analogová televize byla v počátcích a digitální televize v daleké budoucnosti. Jde o princip mnohem obecnější platnosti. Díky němu se studium

entropie spojitých náhodných veličin stává pouze teoretickou pomůckou, zatímco entropie (informace) diskrétních náhodných veličin je běžným nástrojem práce s digitálními signály, který umožňuje mj. stanovit nároky na přenosové kanály a paměťová média.

Entropie (informace) spojitého rozdělení je zcela jiný pojem než u diskrétního rozdělení. Její definice je formálně podobná, a z toho vyplývají i obdobné vlastnosti. Nicméně je potřeba oba pojmy důsledně rozlišovat.

Cvičení 7.2.3: V příkladu 7.1.14 nám nic nebránilo dát na misky různý počet mincí. Jaká je pravděpodobnost, že váha bude v rovnováze?

Řešení: Bez další specifikace nemůžeme odpovědět. Nicméně může se to stát pouze tehdy, jestliže odchylka hmotnosti vadné mince je rovna „přesně" hmotnosti dobrých mincí, o než je na jedné misce více. Nebylo zadáno rozdělení odchylky hmotnosti vadné mince, ale pokud je spojité, je pravděpodobnost rovnováhy nulová. (Ovšem za předpokladu, že váha dovoluje rozlišit libovolně malé rozdíly.) Proto jsme s touto možností při řešení nepočítali. □

Cvičení 7.2.4: Kolik informace je potřeba k určení pořadí (permutace) k prvků? Řešte pro $k = 2, 3, \ldots, 10$.

Cvičení 7.2.5: Z miliónu obyvatel vybereme jeden tisíc (na pořadí nezáleží). Kolik informace je potřeba k určení provedeného výběru?

Cvičení 7.2.6: Náhodná veličina X má diskrétní rovnoměrné rozdělení s celočíselnými hodnotami $1, 2, \ldots, 100$. Jaká je entropie náhodných veličin (a) X, (b) $X/10$, (c) veličiny X zaokrohlené na desítky nahoru?

Cvičení 7.2.7: Video signál má rozlišení 640×480 pixelů, v každé ze tří barev se rozlišuje 256 úrovní. Frekvence snímků je 25 za sekundu. Jakou rychlostí (v bitech za sekundu) by se musel přenášet signál, pokud by nebyl žádným způsobem zkomprimován? Kolik informace by měl minutový záznam? Za jak dlouho by se záznamem zaplnilo na disku volné místo velikosti 10 GB? Pokud máme tento signál přenášet informačním kanálem, který nemá dostatečnou kapacitu, musíme slevit z požadavků a snížit množství přenášené informace. Čím se to může projevit u příjemce?

Cvičení 7.2.8: Alice a Bob hrají následující hru: Bob si myslí celé číslo od 1 do 1000. Alice smí klást pouze takové otázky, na které může Bob odpovědět „ano" nebo „ne". Má najít hledané číslo pomocí 10 otázek. Navrhněte vhodnou strategii. Kolik čísel je takto možno rozlišit k otázkami? Co se změní, jestliže se Bob smí v jedné odpovědi splést? Jak se úloha změní, jestliže Alice smí říci číslo a Bobovy (pravdivé) odpovědi mohou být „menší", „větší", „stejné" (tj. porovná dané číslo s hledaným)?

Cvičení 7.2.9: Vypočtěte entropii v příkladu 7.1.4. Navrhněte způsob kódování, který umožní zkrátit průměrnou délku zprávy o výsledku.

Cvičení 7.2.10: Sedmibitová slova zabezpečujeme osmým, paritním bitem. Chyba každého bitu nastává nezávisle s pravděpodobností $p = 10^{-3}$. Určete pravděpodobnost (a) výskytu chybného slova, (b) úspěšné detekce chyby, (c) chyby, která nebude detekována.

Cvičení 7.2.11: Dokažte, že diskretizací diskrétní náhodné veličiny se nezvýší její entropie. (Návod: použijte vzorec 4 z tvrzení 7.1.8.)

Cvičení 7.2.12: Dokažte, že funkce ι daná vzorcem (7.4) je konkávní. Určete její maximum a derivace v krajních bodech.

7.3 Reprezentace náhodných veličin v počítači

V praxi potřebujeme pracovat s mnohem širší škálou rozdělení než jsou ta, která jsou popsána v knihách. Proto potřebujeme nástroje na jejich reprezentaci.

Diskrétní náhodné veličiny

Nabývá-li náhodná veličina pouze *konečně mnoha* hodnot u_k, $k = 1, \ldots, n$, stačí k reprezentaci tyto hodnoty a jejich pravděpodobnosti $p_X(u_k) = P_X(\{u_k\}) = P[X = u_k]$. Těmito $2n$ čísly je pravděpodobnostní funkce plně popsána (až na nepřesnost zobrazení reálných čísel v počítači). Jsou-li navíc hodnoty např. $1, \ldots, n$, můžeme je popsat úsporněji než pomocí n čísel.

Pokud diskrétní náhodná veličina nabývá (*spočetně*) *nekonečně mnoha* hodnot, musíme je sdružit do konečně mnoha skupin; sdružujeme zejména ty, které jsou málo pravděpodobné. Pro každé $\varepsilon > 0$ lze vybrat konečně mnoho hodnot u_k, $k = 1, \ldots, n$, tak, že $P_X(\mathbb{R} \setminus \{u_1, \ldots, u_n\}) = P[X \notin \{u_1, \ldots, u_n\}] \leq \varepsilon$. Zbývá však problém, jakou hodnotu přiřadit zbývajícím (byť málo pravděpodobným) případům.

Příklad 7.3.1: Při studiu počtu dětí v rodinách se sice vyskytuje jen malý počet různých hodnot, ale není žádná pevná mez, která by nemohla být překročena. Mohli bychom s rezervou předpokládat, že vystačíme s hodnotami $0, 1, \ldots, 1000$, ale reprezentace pomocí 1001 pravděpodobností by nebyla úsporná, nehledě na to, že mnohé z nich jsou velmi blízké nule a nedovedeme je smysluplně odhadnout. Navíc poznáme metody (např. χ^2-test dobré shody), které nemůžeme použít, pokud některý výsledek je příliš málo pravděpodobný. Proto např. ponecháme třídy $0, 1, \ldots, 9$ a zbývající hodnoty sdružíme do třídy „10 a více". To však přináší nové problémy, protože poslední třídě neumíme přiřadit odpovídající hodnotu, víme pouze, že je větší než všechny ostatní. Tím se z kvalitativní náhodné veličiny stala pořadová.

Spojité náhodné veličiny

Hustotu můžeme přibližně popsat hodnotami $f(u_k)$ v „dostatečně mnoha" bodech u_k, $k = 1, \ldots, n$, ale jen za předpokladu, že je „dostatečně hladká". Zajímají nás z ní spíše integrály typu

$$F_X(u_{k+1}) - F_X(u_k) = \int_{u_k}^{u_{k+1}} f_X(v)\,\mathrm{d}v,$$

z nichž lze přibližně zkonstruovat distribuční funkci. Můžeme pro reprezentaci použít přímo hodnoty distribuční funkce $F_X(u_k)$. Tam, kde hustota nabývá velkou hodnotu, potřebujeme volit body blízko sebe.

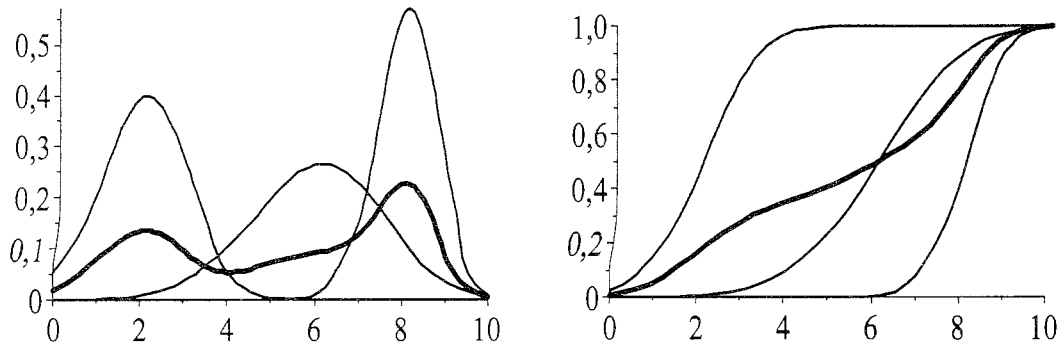
Můžeme volit body u_k, $k = 1, \ldots, n$, tak, aby přírůstky $F_X(u_{k+1}) - F_X(u_k)$ měly zvolenou velikost. Zvolíme tedy $\alpha_k \in (0, 1)$, $k = 1, \ldots, n$, a k nim najdeme čísla $u_k = q_X(\alpha_k)$. Tomu odpovídá přibližná reprezentace pomocí konečně mnoha hodnot kvantilové funkce. (Ve skutečnosti reprezentujeme body grafu funkce, a je jedno, zda ji chápeme jako distribuční, nebo kvantilovou, která je v tomto případě inverzní k distribuční.)

Paměťová náročnost je velká, závisí na jemnosti škály hodnot náhodné veličiny, resp. její distribuční funkce.

Často je rozdělení známého typu, pak stačí doplnit několik parametrů, aby bylo plně určeno. Mnohé obecnější případy se snažíme vyjádřit alespoň jako směsi náhodných veličin s rozděleními známého typu, abychom vystačili s konečně mnoha parametry. Velice často se používají směsi normálních rozdělení (viz obr. 7.2).

Smíšené náhodné veličiny

Rozdělení smíšené náhodné veličiny lze přibližně popsat pomocí konečně mnoha hodnot distribuční nebo kvantilové funkce jako u spojité náhodné veličiny. Tento popis je však pro diskrétní část zbytečně nepřesný. Můžeme použít rozklad na směs diskrétní a spojité náhodné veličiny a reprezentovat každou zvlášť dle předchozích bodů. Pokud diskrétní část nabývá nekonečně mnoha hodnot, pak použijeme rozklad na směs diskrétní a spojité náhodné veličiny s *konečně mnoha* hodnotami (těmi, které jsou nejčastější) a náhodné veličiny se smíšeným rozdělením, jejíž distribuční funkce má sice nekonečně mnoho bodů nespojitosti, ale mění se v nich jen málo.



Obrázek 7.2. Hustota a distribuční funkce směsi normálních rozdělení (složky tence, směs tučně)

Náhodné vektory

Reprezentace náhodných vektorů v počítači je obdobná jako u náhodných veličin, avšak s rostoucí dimenzí rychle roste paměťová náročnost. To by se nestalo, kdyby náhodné veličiny byly nezávislé; pak by stačilo znát marginální rozdělení. Proto velkou úsporu paměti může přinést i **podmíněná nezávislost**. Pokud najdeme úplný systém jevů, které zajišťují podmíněnou nezávislost dvou náhodných veličin, pak můžeme rozdělení náhodného vektoru popsat jako *směs* rozdělení nezávislých náhodných veličin (a tedy úsporněji).

Alespoň stručně jsme naznačili problémy, které v praxi přináší reprezentace rozdělení náhodných veličin pomocí konečného množství údajů. Vyhnuli bychom se jim, kdybychom používali pouze vybraná rozdělení charakterizovaná malým počtem parametrů. To by však bylo příliš omezující. Užitečným kompromisem je aproximace směsí známých rozdělení.

7.4 Generátory náhodných čísel

Různé aplikace generátorů náhodných čísel kladou velmi odlišné nároky na kvalitu a rychlost.

Příklad 7.4.1: Program (např. anitivirový) se má pravidelně aktualizovat. Kdyby to všechny instalace dělaly ve stejnou dobu, např. o půlnoci, přetížily by server. Proto by měly vygenerovat (v rámci omezení daných uživatelem) náhodný čas aktualizace, a to buď při instalaci jednou provždy, nebo při každém ukončení stanovit náhodně čas příští aktualizace. Podobný princip se využívá v komunikaci po síti Ethernet: pokud dva počítače začnou posílat data ve stejném čase, přenos přeruší a začnou po uplynutí náhodného času, který je s velkou pravděpodobností pro každý z nich jiný. Zde jsou nároky na rovnoměrnost rozdělení minimální, zato potřebujeme rychlý algoritmus.

Příklad 7.4.2: Kvalita náhodného generátoru je klíčová při šifrování. Obvykle generování kódu začíná nalezením nějakého velkého prvočísla. Tím se vybere jedna šifra z velkého množství, což ztěžuje rozluštění nepovolanou osobou. Pokud by však generátor náhodných (prvo-)čísel generoval málo možných kódů nebo s nerovnoměrným rozdělením (které lze alespoň zčásti předvídat), pak může luštitel zkusit generovat kódy stejným způsobem a má větší šanci šifru rozluštit. Proto musí být generátor náhodných čísel pro šifrování velmi kvalitní a nepredikovatelný. To znamená, že musí mít velkou entropii; ta zajišťuje malou pravděpodobnost, že šifru rozluští někdo nepovolaný. Rychlost zde není nejdůležitější.

Dříve sloužily k simulaci náhodných pokusů tabulky náhodných čísel. Ty byly prověřeny mnoha testy, zda nevykazují nějakou závislost. (Nejde jen o rozdělení čísel, ale i korelaci po sobě následujících čísel apod.) Náhodným pokusem se vybral počáteční prvek tabulky a pak se pokračovalo následujícími prvky.

Tabulky náhodných čísel by mohly být nahrány v počítači. Nicméně obvykle se generují dynamicky dle potřeby, a to buď *náhodně*, nebo *pseudonáhodně*.

V *generátoru náhodných čísel* nelze předpovědět výsledek, tj. za jiných podmínek a na jiném počítači dostaneme jinou posloupnost. K tomu je potřeba mít nějaký náhodný (nedeterministický) vstup zvenčí, neboť konstrukce i programové vybavení počítačů se naopak snaží o maximální determinismus. Existují speciální hardwarová řešení (např. šumové diody) produkující náhodný signál, je ovšem těžké kontrolovat jejich funkci. Pokud je nemáme k dispozici, pak náhodným vstupem může být např. čas volání procedury či jiný systémový parametr závislý na vnějších okolnostech. Samozřejmě si z času nevybereme rok; pak by mnoho opakování dalo tentýž výsledek. Použijeme čas obvykle celý, rozhodně jeho poslední číslice. Pak i malá změna času může způsobit velkou změnu výsledku. To je stejný princip jako v příkladu 1.2.5; budeme se s ním setkávat i nadále. Přímé použití takového údaje však přináší riziko, že některé hodnoty se budou vyskytovat častěji. Pokusy s takovým generátorem nejsou opakovatelné a je obtížné zajistit shodu s požadovaným rozdělením. Proto se náhodný vstup obvykle přepočítává vhodnou funkcí, která potlačí nedostatky jeho rozdělení.

Zde se uplatní princip z příkladů 1.5.11 a 1.5.12: případné nedostatky generátoru se podstatně zmenší, pokud vygenerujeme více výsledků a roztřídíme do vhodně zvolených tříd. V případě binárních výsledků bychom mohli testovat, zda počet jedniček vyšel sudý nebo lichý. Výsledné rozdělení by pak bylo velmi blízké rovnoměrnému.

Častěji se používá *generátor pseudonáhodných čísel*. Jeho výsledky vypadají na první pohled stejně, ale když opakujeme postup za stejných podmínek, dostáváme stále stejnou posloupnost čísel, podobně jako u tabulek náhodných čísel. Vypadají tedy jako náhodné a pozorovatel by neměl poznat rozdíl, ale ve skutečnosti jsou deterministické. Pro takový generátor se nabízí jednoduchá myšlenka, že provedeme např. nějaký výpočet s reálnými čísly a pak za výsledek vezmeme pouze lomenou část, tj. cifry za desetinnou čárkou. Pak výsledek znovu zobrazíme stejnou funkcí a vezmeme lomenou část. Jde opět o princip z příkladu 1.2.5. Tím rekurentně definujeme posloupnost; můžeme přitom použít i více než jednu poslední hodnotu. Nicméně laická implementace tohoto postupu může vést k vadným výsledům, např. že program uvázne na nějaké hodnotě a tu již neustále opakuje. Proto se používají důmyslnější postupy.[4] Jsou založeny na aritmetice nikoli reálné, nýbrž celočíselné s omezeným počtem hodnot. Díky tomu jsou výpočty přesné a nezávislé na hardwaru. Výsledná náhodná čísla z intervalu $\langle 0, 1\rangle$ jsou zlomky, jejichž jmenovatel je pevně zvolené velké číslo. Možné vzorce pro generování 32-bitových náhodných čísel jsou např.[5]

$$u_n = (1664525\, u_{n-1} + 1013904223) \bmod 2^{32}\,,$$
$$v_n = (279470273\, v_{n-1}) \bmod (2^{32} - 5)\,.$$

Pokud však požadujeme delší náhodný řetězec, potřebujeme jiný generátor, nestačí tento zopakovat. Lze zaručit, že postupně vyjdou všechny možné výsledky (kterých je konečný, ale velký počet) i posloupnosti několika výsledků. Generovaná posloupnost je nutně periodická (neboť její stav je určen konečnou informací), nicméně použitím složitějších rekurencí lze dosáhnout periody, která daleko převyšuje počet operací, které počítač za dobu své existence vykoná. Nevhodná volba konstant může generátor znehodnotit jako u generátoru RANDU, který se hojně používal zejména v 70. letech. Je založen na rekurzi

$$w_n = \left(2^{16} + 3\right)\, w_{n-1} \bmod 2^{31}\,,$$

začínající od nějakého lichého čísla. Bohužel se ukázalo, že když s ním generujeme body v trojrozměrném prostoru, tyto leží v 15 rovinách.[6] To zpochybňuje řadu publikovaných výsledků pomocí

[4] Viz např. L'Ecuyer, P.: Random number generation. Chapter 3 of S.G. Henderson, B.L. Nelson (eds.), *Simulation*, Elsevier Handbooks in Operations Research and Management Science, Elsevier Science, 2006, `http://www.iro.umontreal.ca/~lecuyer/papers.html`

[5] Viz [Press et al. 1986], `http://en.wikipedia.org/wiki/Linear_congruential_generator`

[6] Viz např. `http://en.wikipedia.org/wiki/RANDU`, [Knuth 1997], [Press et al. 1986].

něj testovaných. Také např. náhodný generátor ve starších verzích systému *Matlab* (od verze 5 po 7.3) vykazoval nežádoucí závislost výsledků.[7]

Aby generátor pseudonáhodných čísel nedával na výstupu vždy stejnou posloupnost, bývá různě inicializován (což odpovídá volbě počáteční položky z tabulky náhodných čísel). Při spuštění pak požaduje počáteční údaj (angl. *seed* nebo *key*), jehož délka (počet bitů) závisí na požadovaném počtu bitů na výstupu a požadované kvalitě náhodných čísel. (Umírněné jsou požadavky pro simulaci, extrémní pro kryptografické účely, kdy má mít zhruba tolik bitů, kolik potřebujeme na výstupu.) K tomu může být použit náhodný údaj (jako výše) nebo vstup od uživatele (často ještě zpracovaný jednodušším náhodným generátorem); druhá možnost má tu výhodu, že lze zvolit, zda příští pokus bude probíhat se zcela identickou posloupností náhodných čísel, nebo s nějakou jinou. Nevýhodou je možnost ovlivnění výsledku a jeho predikce; nemohli bychom takto např. losovat loterii.

Ať již používáme náhodný, nebo pseudonáhodný generátor, obvykle je jeho výstupem náhodná veličina W s přibližně rovnoměrným rozdělením na $\langle 0, 1)$ (obvykle diskrétním s velmi mnoha hodnotami v tomto intervalu). Jakékoli jiné rozdělení dostaneme aplikací jeho kvantilové funkce na náhodnou veličinu W, ve skutečnosti na každou její realizaci. Náhodná veličina $q_X(W)$ má stejné rozdělení jako náhodná veličina X. Např. pro náhodný generátor s normovaným normálním rozdělením můžeme použít kvantilovou funkci Φ^{-1} normovaného normálního rozdělení. Ta je ovšem transcendentní a lze ji pouze numericky aproximovat. Tento problém lze obejít následujícím trikem:[8]

Příklad 7.4.3 (Boxova-Mullerova transformace): * Uvedeme postup, jak místo jedné náhodné veličiny s normovaným normálním rozdělením vygenerovat rovnou dvě, X a Y, a to nezávislé. Ty můžeme chápat jako *kartézské* souřadnice náhodného bodu v rovině, konkrétně náhodného vektoru, který má dvojrozměrné normální rozdělení se střední hodnotou v počátku a jednotkovou korelační maticí. Toto rozdělení popíšeme pomocí *polárních* souřadnic, úhlu U a poloměru R (což jsou opět náhodné veličiny, ale s jinými rozděleními). Hustota v bodě $(x, y) \in \mathbb{R}^2$,

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) = \frac{1}{2\pi} \exp\left(-\frac{r^2}{2}\right),$$

závisí pouze na kvadrátu poloměru $r^2 = x^2 + y^2$, rozdělení je rotačně symetrické kolem počátku a úhel U má rovnoměrné rozdělení, např. na intervalu $\langle 0, 2\pi)$. Zbývá vygenerovat poloměr R s odpovídajícím rozdělením. Jeho kvadrát $R^2 = X^2 + Y^2$ má rozdělení $\chi^2(2)$ (podle definice tohoto rozdělení v kapitole B.2), což dle tvrzení B.2.2 je exponenciální rozdělení $\mathrm{Ex}(2)$ s kvantilovou funkcí

$$q_{R^2}(\alpha) = q_{\mathrm{Ex}(2)}(\alpha) = -2 \ln(1 - \alpha).$$

Tedy R má kvantilovou funkci

$$q_R(\alpha) = \sqrt{q_{R^2}(\alpha)} = \sqrt{-2 \ln(1 - \alpha)},$$

která se počítá snáze než kvantilová funkce normálního rozdělení.

Celý postup je následující: Vygenerujeme dvě nezávislé náhodné veličiny V, W s rovnoměrným rozdělením na intervalu $\langle 0, 1)$. Převedeme je na polární souřadnice s potřebným rozdělením,

$$U = 2\pi V,$$
$$R = q_R(W) = \sqrt{-2 \ln(1 - W)},$$

a převedeme do kartézských:

[7] Viz `http://www.cs.cas.cz/~savicky/papers/rand2006.pdf` nebo Savický, P., Robnik-Šikonja, M.: Learning random numbers: A Matlab anomaly. *Applied Artificial Intelligence*, přijato k tisku.

[8] O praktické užitečnosti svědčí, že tento algoritmus je popsán i v jedné z nejpoužívanějších učebnic počítačového vidění, Šonka, M., Hlaváč, V., Boyle, R.: *Image Processing, Analysis and Machine Vision*. 3rd ed., Thomson, 2007.

$$X = R\cos U\,,$$
$$Y = R\sin U\,.$$

Pak X, Y jsou nezávislé náhodné veličiny s normovaným normálním rozdělením. Místo $1 - W$ se používá přímo náhodná veličina s rovnoměrným rozdělením na intervalu $(0, 1\rangle$ (místo V rovněž). Tento postup se nazývá **Boxova-Mullerova transformace.**[9]

Všechna rozdělení *spojitých* náhodných veličin se liší pouze (nelineární) změnou měřítka. I diskrétní a smíšená rozdělení dostaneme stejným postupem. Dochází sice k odchylkám v bodech nespojitosti kvantilové funkce, ale těch je spočetně mnoho (tvrzení 3.6.4), takže se s nimi setkáváme s nulovou pravděpodobností.

Podobně implementace diskrétních náhodných veličin může posloužit k simulaci jakéhokoli náhodného jevu, který má spočetně mnoho možných výsledků. Jestliže (bez újmy na obecnosti) očíslujeme možné výsledky přirozenými čísly a výsledek k má mít pravděpodobnost p_k, pak vypočteme čísla

$$b_k = \sum_{j=1}^{k} p_j\,, \qquad k \in \mathbb{N}\,,$$

vygenerujeme realizaci x náhodné veličiny X s rovnoměrným rozdělením na $\langle 0, 1)$ a převedeme ji na výsledek k, kde k je jediné číslo, pro které platí $b_{k-1} \le x < b_k$. (Předpokládáme $b_0 = 0$.)

Počítačová realizace náhodných veličin je klíčová jak pro simulaci náhodných pokusů, tak pro zabezpečení dat. Generátor může být náhodný nebo pseudonáhodný. Nelze spoléhat na laické postupy a je třeba se držet vyzkoušených metod. Zejména pro kryptografii jsou požadavky velmi přísné. Přesto se i v profesionálních generátorech občas objeví chyby, kdy data vykazují nějakou nežádoucí závislost.

Cvičení 7.4.4: Navrhněte způsob, jak generovat (a) geometrické, (b) Poissonovo rozdělení s daným parametrem.

Cvičení 7.4.5: Navrhněte způsob, jak generovat rovnoměrné rozdělení na (a) kruhu, (b) kouli, (c) sféře (=povrchu koule), (d) trojúhelníku, (e) čtyřstěnu.

[9] Box, G.E.P., Muller, M.E.: A note on the generation of random normal deviates. *Annals Math. Statistics* **29** (1958), 610–611. Lze se vyhnout i výpočtu goniometrických funkcí za cenu opakovaného použití náhodného generátoru, viz `http://en.wikipedia.org/wiki/Box-Muller_transform`

Část II

Základy matematické statistiky

8. Základní pojmy statistiky

8.1 Na co je statistika

Dosud jsme předpokládali, že parametry pravděpodobnostního modelu jsou známy. To je málokdy splněno.

Příklad 8.1.1 (cena opatření): Při návrzích změn platů, daňových předpisů, sociálních dávek atd. by navrhovatel měl současně předložit odhad, kolik to bude stát (nebo kolik to přinese). Přitom se může opírat jen o přibližné odhady součtu velkého počtu položek. Problém je, jak shromáždit podklady pro kvalifikovaný odhad.

Příklad 8.1.2: Na Sportce normálně sázející prodělává, protože na výhry jde jen asi polovina sázek. Jelikož výhry jsou stanoveny podle počtu výherců, je výhodnější sázet *jinak než ostatní.* K tomu potřebujeme vědět, podle jakého modelu sázejí ostatní.

Příklad 8.1.3: U rulety se obě strany zajímají, zda padají všechna čísla se stejnou pravděpodobností, přesněji, jak velké jsou odchylky od rovnoměrného rozdělení. Jak to zjistit a jaké je riziko chybných závěrů?

I na to je statistika. Obecně je jejím tématem, jak z opakovaných pozorování náhodného pokusu stanovit odpovídající pravděpodobnostní model. Jde tedy opačným směrem než teorie pravděpodobnosti, která předpovídá, jaké výsledky lze očekávat u náhodného pokusu s daným pravděpodobnostním popisem. Ve statistice pozorujeme výsledky a z nich se snažíme usuzovat na pravděpodobnostní popis (tomuto postupu se říká **statistická indukce**, angl. *statistical inference*). Vnitřní charakteristiky modelu (např. střední hodnota) jsou nám obvykle skryty a dovídáme se o nich jen nepřímo, statistickými metodami.

Příklad 8.1.4: A. Z urny jsme vytáhli 60 černých a 40 bílých koulí. Z toho usuzujeme, že pravděpodobnost tažení černé koule je přibližně 60 %, tažení bílé koule přibližně 40 %. Vzhledem k náhodnému charakteru pokusu víme, že skutečné rozdělení koulí může být odlišné, i když pak by náš výsledek byl méně pravděpodobný. V tomto případě se můžeme přesvědčit otevřením urny a spočítáním všech koulí. Je možné, že v ní najdeme kromě černých a bílých třeba zelené koule, několik mincí a karet a další objekty, které nebyly dosud nikdy taženy.

B. Zkoumáme ryby v rybníce; z ulovených usuzujeme na výskyt jednotlivých druhů. I zde bychom mohli vypustit rybník a spočítat všechny ryby, ale možná nám to majitel nedovolí. Kdyby šlo o ryby v řece nebo v moři, neměli bychom ani tuto možnost.

C. Fyzik studuje elementární částice, které přilétly z vesmíru. O jejich skutečném zastoupení ve vesmíru může dělat jen statistické úsudky. Zachycení některých částic je tak vzácné, že je ani nemůže statisticky vyhodnotit.

Teorie pravděpodobnosti nám radí, jak se účelně chovat ve světě, který se chová náhodně. Přitom vychází z většinou nesplněného předpokladu, že víme, *jak* náhodně se chová, tj. že známe zákonitosti, byť pouze pravděpodobnostní. Naproti tomu statistika se snaží tyto zákonitosti odhalit a kvantifikovat na základě pozorování skutečných jevů. Proto teprve v kombinaci se statistikou dostáváme plnohodnotný prostředek, který nám radí, jak se chovat v reálném světě. Např. nám může poradit, jakou léčbu zvolit při daných příznacích nemoci. Přitom ovšem vychází z předpokladu, že náš případ má něco společného s mnoha dalšími, které byly v minulosti sledovány. („co

bylo dobré pro ostatní, bude dobré i pro mě"). Tato argumentace má své slabé stránky. Může se stát, že analogie selže a lék nám naopak uškodí. Teprve shromážděním dostatečného množství podobných případů pak někdo odhalí, že některé osoby jsou na lék alergické, a možná se je naučíme i rozpoznávat. Statistika není použitelná ke studiu *řídkých jevů*, které nastávají s tak malou pravděpodobností, že nemáme dostatek pozorování pro uplatnění statistických postupů. I proto je např. problém vědecky studovat kulové blesky, extrémy počasí apod.

Příklad 8.1.5: Statistika nám dovoluje vytipovat druhy zvířat ohrožené vyhynutím, ale nečekejme od ní odpověď na to, zda nějaký živočišný druh již vyhynul.

Další potíže statistického výzkumu (vyplývající např. z nedostatku času a dalších prostředků) dobře popisuje [Rogalewicz 2000, str. 99].

Předmětem statistiky je zkoumání *společných* vlastností *velkého počtu* obdobných jevů. Přitom nezkoumáme všechny, ale jen vybraný vzorek (kvůli ceně testů, jejich destruktivnosti apod.).

Poznámka 8.1.6: G.K. Chesterton se v citátu z úvodu kapitoly 4.3 zabývá případem, který *není předmětem statistiky*. Jestliže máme pouze dvě hodnoty, popíšeme je výčtem. Statistický popis se uplatní teprve tehdy, když máme velké množství dat a chceme stručně vyjádřit podstatné zákonitosti z nich vyplývající. Jedná se vlastně o extrémní formu *ztrátové komprese dat*.

Stejně jako v jiných vědách, základními postupy jsou *pozorování* a *pokus*. Ve statistickém pokusu se za předem stanovených podmínek snažíme posoudit platnost hypotézy. K jeho korektnímu provedení je potřeba splnit řadu požadavků.

Poznámka 8.1.7 (některé zásady statistického experimentu): Statistický experiment je nutno předem pečlivě připravit – rozhodnout, co se bude testovat, na jakém velkém souboru a podle jakých kritérií vybraném, jaké veličiny budeme měřit (nebo jak formulované otázky budeme klást). Je potřeba dopředu stanovit metody a kritéria pro vyhodnocení, aby se předešlo nežádoucím efektům s dodatečným vyhodnocením jinými postupy, třeba obvyklými, ale účelově zvolenými. Výběr souboru je velmi náročný a důležitý pro dosažení co nejobjektivnějších výsledků. Teprve pak lze přistoupit ke sběru dat. Na tom se často podílí více lidí. Ti mají mít jen takové informace, které jsou nutné pro jejich roli v experimentu. Požaduje se pokud možno tzv. *dvojitě slepý experiment*. Např. testujeme-li účinnost léku, pak jej podáváme jen části pacientů, ostatní dostávají neúčinnou náhražku (placebo). Přitom ani pacient, ani lékař nevědí, který z pacientů dostává lék, a nemohou tedy vědomě ovlivnit výsledek. Navíc musí být rozdělení do obou skupin provedeno natolik náhodně, aby se maximálně omezilo zkreslení jinými vlivy.

Pokud nemáme předem zformulovánu hypotézu, kterou chceme testovat, pak si musíme ponechat část vzorků jako tzv. *kontrolní skupinu*. Jestliže první výzkum vykázal např. nějakou novou závislost mezi dvěma náhodnými veličinami, pak bychom následně měli ověřit, zda i v kontrolní skupině (disjunktní s předchozí) tato závislost platí (je statisticky významná). Tím se vyvarujeme předčasného přijetí nedostatečně ověřených závěrů na základě výsledků, které mohly být statistickou chybou.

Často nelze zásady pokusu dodržet. Např. těžko můžeme zařídit, aby pacient nevěděl, zda byl nebo nebyl operován. Některé pokusy by trvaly nepřijatelně dlouho nebo měly katastrofální následky (např. experimentální posouzení vlivu lidské činnosti na globální oteplování). Pak se musíme spokojit s pozorováním. To bývá snazší, ale o to opatrněji se musí vyhodnotit. Pokud dodatečně zjistíme zajímavé souvislosti v datech, je potřeba se vyvarovat unáhlených úsudků, považovat to za hypotézu a hledat údaje z dalších podobných situací, než učiníme jakýkoli závěr.

Příklad 8.1.8: Po ztroskotání Titanicu se ukázalo, že příbuzní cestujících prokazatelně poslali předem asi 300 dopisů, v nichž vyjadřují předtuchu, že plavba špatně dopadne. U jiné podobně velké lodi, Normandie, bylo prokázáno zhruba 500 takových dopisů. Normandie ovšem doplula bez nehody.[1]

[1] Vinař, O.: Proč lidé rádi věří šarlatánům a neradi vědcům? Přednáška v Českém klubu skeptiků Sisyfos, 20.12.2006.

Lze říci, že statistika je dohodnutou metodologií vědecké práce, která přispívá k objektivnímu porovnání výsledků. Plní úlohu arbitra a zabraňuje neplodným debatám o tom, kdo má pravdu, pokud tato není zřejmá. Statistické postupy se mohou mýlit, ale jsou navrženy tak, abychom chybné závěry přijímali s velmi malou pravděpodobností.

Podle účelu se statistika dělí na *popisnou* (*deskriptivní*, angl. *descriptive statistics*), která se snaží účelně a názorně popsat skutečnost, a *induktivní* (angl. *inferential statistics*), která se pokouší činit závěry o platnosti hypotéz, souvislosti náhodných veličin apod.

Zde se zaměříme na následující témata (která nevyčerpávají všechny oblasti statistiky):

- Odhady charakteristik a parametrů pravděpodobnostního modelu
- Testování hypotéz

Tím se omezujeme na tzv. *konfirmační analýzu*, při níž sami formulujeme hypotézy a od statistiky očekáváme posouzení jejich platnosti. Za rámcem této knihy zůstává *explorační analýza*, která se naopak z dat snaží automaticky generovat hypotézy.

Poznámka 8.1.9 (minumum historie): Zkoumání rozsáhlých dat (hlavně o úrodě) se věnovali již staří Egypťané. Nicméně za počátky matematické statistiky je považováno použití metod vyrovnávacího počtu k co nejpřesnějšímu vyhodnocení většího počtu nepřesných měření, zejména pro účely astronomie. Tím se zabýval již G. Galilei (1564–1642). Metodu nejmenších čtverců objevili na počátku 19. století nezávisle hned tři vědci, A.M. Legendre (1752–1833), R. Adrain (1775–1843) a K.F. Gauss (1777–1855), který pomocí ní předpověděl pohyb planetky Ceres. Regresi zavedl a v genetice využil synovec Ch. Darwina F. Galton (1822–1911). W. Gossett (1876–1937) položil základy průmyslové statistiky. Jeho jméno není příliš známé; jako profesionální sládek a amatérský statistik publikoval své práce pod pseudonymem Student. Za zakladatele moderní statistiky bývá považován K. Pearson (1857–1936), kterému vděčíme za řadu běžně používaných testů a postupů. Přitom jeho aktivita spadá na přelom 19. a 20. století, tedy ještě před Kolmogorovův pravděpodobnostní model. R.A. Fisher (1890–1962) přispěl nejen dalšími testy a metodou maximální věrohodnosti, ale také zásadami plánování pokusů. Další důležitý impuls dalo rozvoji statistiky ve druhé polovině 20. století nasazení počítačů. Díky němu se staly schůdné mnohé dříve nepoužitelné postupy a zpracování rozsáhlých dat se stalo běžnou součástí práce v řadě oborů. Velkou inspiraci a využití poskytly statistice hlavně sociologie a pojišťovnictví.

Statistika poskytuje daleko víc než zpracování dat – nástroj pro zkoumání světa, pro hledání a ověřování závislostí, které nejsou zjevné. V tomto směru je nenahraditelným pomocníkem téměř všech vědeckých disciplín, a proto by každý vzdělaný člověk měl mít aspoň minimální znalosti o tom, co statistika poskytuje, co naopak nenabízí a jak interpretovat její výsledky. V běžném životě se setkáváme s překvapivě mnoha chybnými výklady statistických údajů.

8.2 Náhodný výběr

„Všechny lidi, bohužel, navštívit nemůžeme."

Václav Havel: Ztížená možnost soustředění

Známý dramatik má pravdu. Z této skutečnosti vychází statistika jako levnější nástroj pro získání informací o celku.

Příklad 8.2.1: Chceme odhadnout volební preference. Místo dotazování všech voličů se zeptáme vybraného vzorku několika tisíc respondentů a na základě něj uděláme úsudek. Není to tak přesné, ale je to levnější a rychlejší.

Příklad 8.2.2: Pro nastavení clony digitálního fotoaparátu chceme zjistit, jaké jsou úrovně jasu v jednotlivých pixelech. Nepoužijeme-li k tomu všechny, můžeme dostat přibližný odhad rychleji.

Příklad 8.2.3: Pro testování životnosti zářivek odebereme část produkce pro (destruktivní) testy a z výsledků usuzujeme na životnost těch zářivek, které jdou do prodeje.

Společným znakem uvedených příkladů je to, že máme velký soubor objektů, na nichž měříme určité veličiny (předpokládejme, že jsou pro každý objekt přesně dány a nejsou náhodné). Náhodnost je pouze ve výběru menšího vzorku, na základě něhož odhadneme vlastnosti celku. Rozlišujeme:

- **základní soubor** (nazývaný též **populace**), tj. množinu všech objektů, o nichž máme vypovídat,
- **výběrový soubor**, tj. množinu těch objektů, na nichž zkoumanou veličinu změříme.

Náhodný výběr jednoho prvku *základního souboru* (s rovnoměrným rozdělením) a stanovení odpovídající hodnoty veličiny je náhodný pokus, kterým je určeno rozdělení náhodné veličiny. Opakovaným nezávislým výběrem určíme *výběrový soubor*. Stanovením hodnot veličiny na něm dostaneme náhodný vektor, jehož složky jsou nezávislé náhodné veličiny se stejným rozdělením.

Definice 8.2.4: ***Náhodný výběr*** $\boldsymbol{X} = (X_1, \ldots, X_n)$ *je vektor náhodných veličin, které jsou* nezávislé *a mají* stejné rozdělení *(angl.* identically and indepently distributed, i.i.d.*). Provedením náhodného pokusu dostaneme* ***realizaci náhodného výběru,*** $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. *Číslo* $n \in \mathbb{N}$ *se nazývá* ***rozsah výběru.***

Nadále značíme náhodné veličiny a vektory velkými písmeny a jejich realizace odpovídajícími malými písmeny. Protože všechny distribuční funkce F_{X_k} jsou stejné, vynecháváme indexy a distribuční funkci kterékoli složky náhodného vektoru značíme F_X; podobně pro další charakteristiky náhodných veličin.

Poznámka 8.2.5: Výjimečně budeme používat i dvojrozměrný (vektorový) náhodný výběr, jehož složky jsou dvojice náhodných veličin, $((X_1, Y_1), \ldots, (X_n, Y_n))$, kde náhodné vektory (X_k, Y_k), $k = 1, \ldots, n$, mají stejné rozdělení a jsou nezávislé. Nicméně připouštíme závislost mezi X_k a Y_k (v rámci jednotlivých náhodných vektorů (X_k, Y_k)) nezávislou na $k = 1, \ldots, n$, která bude předmětem sledování. Kdybychom chtěli pracovat se dvěma náhodnými výběry $(X_1, \ldots, X_n)$, $(Y_1, \ldots, Y_n)$, ztratili bychom informaci o souvislostech, např. o jejich uspořádání, na kterém v případě jednorozměrného výběru nezáleží.

Poznámka 8.2.6: Podle poznámky 1.3.8 není podstatné, zda považujeme výběr za uspořádaný nebo neuspořádaný. Měli bychom však rozlišit, zda se jedná o *výběr s vracením*, nebo *výběr bez vracení*. Při výběru s vracením bychom připustili, že se nějaký objekt může dostat do výběrového souboru vícekrát. Protože na něm vždy dostaneme stejnou hodnotu měřeného parametru, porušili bychom tím nezávislost. Mnohdy ani nelze vybrat jeden objekt vícekrát, protože test může být destruktivní (např. životnost zářivek, pevnost mechanických součástek atd.). Také nepředpokládáme, že by agentura navštívila stejného respondenta dvakrát v rámci stejného výzkumu nebo opsala jím vyplněný dotazník. Proto obvykle vylučujeme vícenásobný výběr stejného prvku a používáme výběr bez vracení. Ani to není bez problémů. Výsledné rozdělení se může poněkud lišit od původního, neboť další objekty vybíráme ze souboru, v němž nejsou ty, které jsme již vybrali, viz příklad 1.5.14. Tento rozdíl se obvykle zanedbává, neboť

- rozsah základního souboru někdy není znám,
- pro velký rozsah základního souboru (ve srovnání s výběrovým souborem) není podstatné, zda jsme použili výběr s vracením nebo bez vracení (věta 1.3.9),
- výpočty se značně zjednoduší.

Klíčovou roli ve statistice má následující poněkud překvapivý výsledek, který zdůvodníme později: *Přesnost odhadu je dána velikostí výběrového souboru, nikoli populace.*

Příklad 8.2.7: [Disman 2005, Rogalewicz 2000] V prezidentských volbách v USA r. 1936 se předpovídalo vítězství kandidáta Landona rozdílem 14 %. Tento názor podporoval i týdeník Literary Digest na základě svého průzkumu vzorku dvou miliónů voličů. Když Franklin Delano Roosevelt zvítězil drtivou většinou, bylo to překvapením pro mnohé, nikoli však pro George Gallupa (1901–1984), který takový výsledek předpověděl na základě vzorku řádově tisíců respondentů.

V čem byl rozdíl? Literary Digest vyhledával voliče pomocí telefonních seznamů, adres držitelů řidičských průkazů, seznamů předplatitelů novin a časopisů atd. Do jeho vzorku se nedostaly některé skupiny, které pak ovlivnily výsledek voleb (nezaměstnaní, bezdomovci apod.). Naopak Gallup použil tzv. *kvótní výběr*, jehož složení odpovídá co nejlépe složení cílové populace v důležitých hlediscích (věk, příjem, místo bydliště, rodinná situace). Tato metodika se dnes používá standardně a Gallupův institut se zabývá statistickými výzkumy dodnes. Vzniká naopak otázka, jak je možné, že Literary Digest uspěl s předpověďmi v letech 1920, 1924, 1928 a 1932. Tehdy byla politická situace v USA stabilnější než po hospodářské krizi, a tak o výsledcích rozhodovala jiná než ekonomická hlediska.

Příklad 8.2.8: Vybereme-li pro průzkum 1000 z obyvatel Prahy, kterých je celkem přes milión, má každý občan pravděpodobnost o něco menší než 1/1000, že bude vybrán do výběrového souboru.

Nevadilo by, kdybychom vybrali pro průzkum 1000 z obyvatel Číny, kterých je celkem přes miliardu, takže každý občan má pravděpodobnost o něco menší než 1/1 000 000, že bude vybrán. Může se zdát, že takový odhad je příliš hrubý, bude však stejně přesný. *(Ve skutečnosti není snadné určit počet všech možných respondentů, viz pozn. 1.3.7.)*

Naopak by vadilo, kdybychom vybrali pro průzkum 1000 ze 2000 zaměstnanců podniku. Narazili bychom na problém, protože bychom už nemohli zanedbávat rozdíl mezi výběrem s vracením a bez vracení a nemohli bychom využívat zjednodušující předpoklady, za nichž jsou zde uvedené statistické metody odvozeny.

Slovo „*statistika*" označuje nejen vědní disciplínu, ale i následující pojem: **Statistika** je (každá) měřitelná reálná funkce G definovaná na náhodném výběru libovolného dostatečně velkého rozsahu.

Poznámka 8.2.9: Požadavkem na dostatečný rozsah výběru připouštíme důležité statistiky, jejichž definice nedává smysl pro velmi malé rozsahy. Nemůžeme např. odhadovat rozptyl z výběru o rozsahu 1.

Požadavek měřitelnosti znamená, že pro každé $u \in \mathbb{R}$ je definována pravděpodobnost

$$P[G(X_1, \ldots, X_n) \leq u] = F_{G(X_1,\ldots,X_n)}(u)\,.$$

To nás nijak neomezuje; neměřitelné funkce sice existují, ale nesetkáme se s nimi.

Typicky se statistiky používají k odhadům neznámých charakteristik či parametrů rozdělení. Při jejich výpočtu lze vycházet pouze z realizací náhodných veličin. Nemůžeme použít přímo charakteristiky či parametry rozdělení, neboť ty nebývají dostupné přímému pozorování, viz např. pozn. 4.1.13.

Příklad 8.2.10: Statistikou může být aritmetický průměr všech složek náhodného výběru. Může to být i jakýkoli jiný průměr (kvadratický, geometrický atd.), stejně jako maximum nebo minimum a jakékoli funkce předchozích, např. aritmetický průměr největšího a nejmenšího prvku. Tyto statistiky mohou např. sloužit k odhadu střední hodnoty.

Střední hodnota nemůže být statistikou, protože to je charakteristika rozdělení, která (možná!) objektivně existuje, ale nemůžeme ji z náhodného výběru určit, pouze o ní můžeme dělat úsudky zatížené pravděpodobnostní neurčitostí.

Statistika jako funkce náhodných veličin je rovněž *náhodná veličina*. Jako taková může mít střední hodnotu (obecně různou od střední hodnoty rozdělení, jež vygenerovalo náhodný výběr)

apod. *Realizace statistiky* (vypočtená stejným postupem z *realizace* náhodného výběru) je *reálné číslo*.

Zavedli jsme základní pojmy, které dovolují použít aparát teorie pravděpodobnosti k popisu statistického průzkumu. Východiskem je pojem náhodného výběru, který je abstraktně chápán jako vektor nezávislých stejně rozdělených náhodných veličin. Na něm jsou definovány funkce – statistiky – které se naučíme používat pro vyhodnocení výsledků.

9. Odhady charakteristik rozdělení

Častým cílem statistických průzkumů je odhad neznámých charakteristik (např. střední hodnoty či rozptylu) nebo jiných parametrů rozdělení, které vygenerovalo náhodný výběr. Často budeme hovořit pouze o odhadu parametrů, ale místo nich si lze představit i jakékoli charakteristiky rozdělení. Jako motivace mohou posloužit příklady ze začátku kapitoly 8.2.

9.1 Vlastnosti odhadů

Příklad 9.1.1: Pro odhad střední hodnoty můžeme použít

1. aritmetický průměr všech složek náhodného výběru,
2. aritmetický průměr všech složek náhodného výběru vynásobený číslem $\frac{n}{n+1}$, kde n je rozsah výběru,
3. aritmetický průměr největšího a nejmenšího prvku,
4. aritmetický průměr všech složek náhodného výběru kromě 10 % největších a 10 % nejmenších.

Všechny tyto statistiky jsou použitelné. První je asi nejpřirozenějším odhadem. Druhá se zdá být špatnou alternativou aritmetického průměru, ale podobné myšlenky mohou někdy mít opodstatnění. *(Navíc zde potřebujeme ukázat různá řešení, včetně nevhodných.)* Tušíme, že třetí možnost bude „horší", neboť využívá informaci jen o dvou vybraných složkách a ignoruje ostatní. Poslední možnost eliminuje vliv extrémních hodnot; proto se např. ještě na konci 20. století při známkování výkonů krasobruslařů používal podobný systém, kde se ignovorala nejvyšší a nejnižší známka a ze zbývajících se počítal průměr.

Abychom popsali výhody a nevýhody jednotlivých odhadů, zavedeme nejprve potřebné pojmy a následující značení:

- ϑ je neznámá skutečná hodnota parametru (reálné číslo),
- $\widehat{\Theta}_n$ je jeho odhad založený na náhodném výběru rozsahu n (statistika, náhodná veličina),
- $\widehat{\vartheta}_n$ je realizace odhadu $\widehat{\Theta}_n$ (reálné číslo).

Indexy vyznačující rozsah výběru často vynecháváme a značíme $\widehat{\Theta}, \widehat{\vartheta}$.

Poznámka 9.1.2: V angličtině se poslední dva pojmy rozlišují snáze: odhad se nazývá *estimator*, realizace odhadu *estimate*. Do češtiny se obě tato slova často překládají *odhad*.

Definice 9.1.3 (vlastnosti odhadů): *Odhad $\widehat{\Theta}_n$ parametru ϑ na základě náhodného výběru rozsahu n se nazývá*

- ***nestranný*** *(angl.* unbiased*), jestliže* $\mathrm{E}\widehat{\Theta}_n = \vartheta$ *(v opačném případě se nazývá* ***vychýlený****, angl.* biased*),*
- ***asymptoticky nestranný*** *(angl.* asymptotically unbiased*), jestliže* $\lim_{n\to\infty} \mathrm{E}\widehat{\Theta}_n = \vartheta$,
- ***nejlepší nestranný*** *(angl.* the best unbiased*), jestliže ze všech nestranných odhadů má nejmenší rozptyl,*
- ***konzistentní*** *(angl.* consistent*), jestliže pro všechna* $\varepsilon > 0$ *platí* $\lim_{n\to\infty} P\left[\left|\widehat{\Theta}_n - \vartheta\right| \geq \varepsilon\right] = 0$,

- *eficientní (též vydatný, angl.* efficient*), jestliže hodnota* $\mathrm{E}\left((\widehat{\Theta}_n - \vartheta)^2\right)$ *je malá,*
- ***robustní** (angl.* robust*), jestliže dává použitelné výsledky i tehdy, když jsou některá data vadná (zatížená šumem nebo chybami).*

Tvrzení 9.1.4 (postačující podmínka pro konzistenci): *Jestliže odhad* $\widehat{\Theta}_n$ *parametru* ϑ *na základě náhodného výběru rozsahu* n *splňuje podmínky* $\lim_{n\to\infty} \mathrm{E}\widehat{\Theta}_n = \vartheta$, $\lim_{n\to\infty} \sigma_{\widehat{\Theta}_n} = 0$, *pak je konzistentní.*

Uvedené podmínky konzistence jsou silnější, ale často používané, někdy se tak konzistence i definuje [Rogalewicz 2000].

Eficience odhadu je relativní vlastnost rozlišující odhady více či méně eficientní. Kritérium $\mathrm{E}\left((\widehat{\Theta}_n - \vartheta)^2\right) = \mathrm{MSE}\left(\widehat{\Theta}_n\right)$ se nazývá **střední kvadratická chyba** (angl. *mean square error*); je často užívaným *kvantitativním* měřítkem kvality odhadu. Pro *nestranné* odhady se zjednoduší na rozptyl. $\mathrm{E}\left((\widehat{\Theta}_n - \vartheta)^2\right) = \mathrm{D}\widehat{\Theta}_n$, zatímco v obecném případě postihuje i výchylku odhadu. S využitím (4.11) ji lze vyjádřit

$$\mathrm{MSE}\left(\widehat{\Theta}_n\right) = \mathrm{E}\left((\widehat{\Theta}_n - \vartheta)^2\right) = \mathrm{D}(\widehat{\Theta}_n - \vartheta) + \left(\mathrm{E}(\widehat{\Theta}_n - \vartheta)\right)^2 = \mathrm{D}\widehat{\Theta}_n + (\mathrm{E}\widehat{\Theta}_n - \vartheta)^2 .$$

Kdybychom i u *vychýlených* odhadů posuzovali pouze rozptyl, ten by byl minimální pro konstantní odhad (jakýkoli, nezávislý na datech!), což zřejmě není to, co chceme. Nejlepší nestranný odhad je ten, který je ze všech *nestranných* odhadů nejvíce eficientní. (Mohou však existovat eficientnější *vychýlené* odhady.)

Robustnost odhadu (angl. *robustness*) je rovněž relativní vlastnost, navíc ne zcela přesně definovaná. Připisuje se těm odhadům, které „i při zašuměných datech dávají správný výsledek". Obvykle se definuje jako procentuální podíl zašuměných dat, při kterém metoda ještě dává *správné* výsledky, což však nebývá vždy důsledně splněno. (Vliv šumu v datech nelze většinou zcela eliminovat.) Nicméně je to velmi praktická vlastnost. Její uplatnění vedlo k metodám, které se snaží rozpoznat tzv. **vychýlené hodnoty** (angl. *outliers*, tento termín se používá i v češtině), tj. hodnoty, které se od ostatních natolik liší, že je považujeme za chybné a pro další výpočet nepoužíváme. Výsledný odhad pak stanovujeme na základě těch dat, která přiměřeně odpovídají naší hypotéze (a pouze ji zpřesňují).

Příklad 9.1.5: Ukažme nové pojmy na odhadech z příkladu 9.1.1 (i když zatím bez přesných argumentů). První odhad (aritmetický průměr) je nestranný, druhý je jeho násobek číslem $\frac{n}{n+1} \neq 1$, tudíž je vychýlený, nicméně je asymptoticky nestranný, neboť $\frac{n}{n+1} \to 1$ pro $n \to \infty$. Oba jsou konzistentní. Třetí odhad (aritmetický průměr maxima a minima) bude zřejmě méně eficientní a také málo robustní, neboť jediná vychýlená hodnota (outlier) jej může zásadně ovlivnit a velký rozsah výběru to nenapraví. Poslední odhad je naopak pokusem o zvýšení robustnosti. Může však (stejně jako předchozí) být vychýlený, pokud rozdělení bude nesymetrické, tj. pokud budou mít náhodné veličiny $\widehat{\Theta}_n - \mathrm{E}\widehat{\Theta}_n$ a $-\left(\widehat{\Theta}_n - \mathrm{E}\widehat{\Theta}_n\right)$ (s nulovými středními hodnotami) různá rozdělení.

Formulovali jsme základní požadavky na odhady parametrů, pomocí nichž můžeme objektivně posoudit výhody a nevýhody různých metod odhadu. Nyní probereme nejdůležitější úlohy.

9.2 Výběrový průměr

Příklad 9.2.1: Plánujeme výrobu nového produktu, potřebujeme odhadnout, kolik by ho spotřebitelé koupili.

Asi nejčastějším dotazem na statistiky jsou „průměrné údaje" charakterizující základní soubor (počítané ovšem z výběrového souboru). Typicky se používají k odhadu střední hodnoty, která bývá skrytým parametrem rozdělení a nebývá přímo měřitelná.

Definice 9.2.2: *Výběrový průměr z náhodného výběru* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *je*

$$\overline{X} = \frac{1}{n} \sum_{j=1}^{n} X_j \,.$$

Alternativní značení: $\overline{\boldsymbol{X}}_n$ *(pokud potřebujeme zdůraznit rozsah výběru).* ***Realizaci výběrového průměru*** *značíme malým písmenem:*

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j \,.$$

Nezávisle na typu rozdělení lze snadno dokázat následující vlastnosti:

Tvrzení 9.2.3:

$$\mathrm{E}\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} \mathrm{E}X = \mathrm{E}X \,,$$

$$\mathrm{D}\overline{X}_n = \frac{1}{n^2} \sum_{j=1}^{n} \mathrm{D}X = \frac{1}{n} \mathrm{D}X \,,$$

$$\sigma_{\overline{X}_n} = \sqrt{\frac{1}{n} \mathrm{D}X} = \frac{1}{\sqrt{n}} \sigma_X \,,$$

kde $\mathrm{E}X = \mathrm{E}X_j$ *atd. (pokud příslušné charakteristiky existují), takže výběrový průměr je nestranným konzistentním odhadem střední hodnoty.*

Poznámka 9.2.4: Realizace výběrového průměru vždy existuje, a to i tehdy, když původní rozdělení střední hodnotu nemá. V takovém případě předchozí tvrzení nelze použít; nemůžeme odhadovat neexistující střední hodnotu.

Důkaz předchozího tvrzení je jen procvičením vlastností střední hodnoty a rozptylu. Konzistence vyplývá z toho, že výraz $\mathrm{D}\overline{X}_n = \frac{1}{n}\mathrm{D}X \to 0$ pro $n \to \infty$. Lidově se hovoří o „přesném součtu nepřesných čísel", což je chyba, neboť součet $\sum_{j=1}^{n} X_j$ má rozptyl $n\,\mathrm{D}X \to \infty$. *Relativní* chyba součtu *klesá, absolutní roste.* Bylo by tedy správnější hovořit o „přesném průměru nepřesných čísel". V kombinaci s Čebyševovou nerovností pro $\overline{X}_n$ můžeme vymezit i rychlost této konvergence:

Tvrzení 9.2.5: *Pro náhodný výběr* $(X_1, \ldots, X_n)$ *z rozdělení, které má rozptyl, platí*

$$P\left[\left|\overline{X}_n - \mathrm{E}X\right| \geq \varepsilon\right] \leq \frac{\mathrm{D}\overline{X}_n}{\varepsilon^2} = \frac{\mathrm{D}X}{n\,\varepsilon^2} \,. \tag{9.1}$$

Konvergence

$$P\left[\left|\overline{X}_n - \mathrm{E}X\right| \geq \varepsilon\right] \to 0$$

pro $n \to \infty$ vychází i za obecnějších předpokladů (X_j nemusí mít stejné rozdělení, ale musí mít stejnou střední hodnotu a rozptyl a musí být *nezávislé*), což je tzv. **slabý zákon velkých čísel** (angl. *weak law of large numbers*). Naproti tomu **silný zákon velkých čísel** (angl. *strong law of large numbers*) říká, že pro $n \to \infty$ konverguje výběrový průměr ke střední hodnotě *s pravděpodobností* 1 (což je nutno rozlišovat od jistého jevu).

Rozdělení výběrového průměru (stejně jako součtu náhodných veličin) může být podstatně složitější než původní, jen ve speciálních případech je jednoduchá odpověď. Pro normální rozdělení lze použít tvrzení 4.7.15:

Věta 9.2.6: *Výběrový průměr rozsahu* n *z* normálního *rozdělení* $\mathrm{N}(\mu, \sigma^2)$ *má normální rozdělení* $\mathrm{N}\left(\mu, \frac{1}{n}\sigma^2\right)$ *a je nejlepším nestranným odhadem střední hodnoty.*

Pro jiná rozdělení podobný výsledek platí alespoň asymptoticky jako důsledek centrální limitní věty (neboť rozdělení výběrového průměru se od rozdělení součtu liší jen lineární transformací a normováním se sjednotí):

Věta 9.2.7: *Nechť $(X_j)_{j\in\mathbb{N}}$ je posloupnost* nezávislých *náhodných veličin se* stejným rozdělením *se střední hodnotou* $\mathrm{E}X$ *a směrodatnou odchylkou* $\sigma_X \neq 0$. *Pak rozdělení* normovaných *náhodných veličin*

$$Y_n = \operatorname{norm} \overline{\boldsymbol{X}}_n = \frac{\sqrt{n}}{\sigma_X} \left(\overline{\boldsymbol{X}}_n - \mathrm{E}X\right)$$

konvergují k normovanému normálnímu rozdělení v následujícím smyslu:

$$\forall u \in \mathbb{R} : \lim_{n\to\infty} F_{Y_n}(u) = \lim_{n\to\infty} F_{\operatorname{norm}\overline{\boldsymbol{X}}_n}(u) = \Phi(u)\,.$$

Výběrový průměr je daleko nejpoužívanějším (byť ne jediným) odhadem střední hodnoty. Jeho rozdělení bývá složité, ale pro velký rozsah výběru jej můžeme přibližně nahradit normálním dle centrální limitní věty.

Cvičení 9.2.8: Sečetli jsme 10^6 čísel, která vznikla zaokrouhlením původních dat na 3 desetinná místa. Odhadněte směrodatnou odchylku součtu způsobenou zaokrouhlením.

Cvičení 9.2.9: Vysvětlete rozdíl mezi výběrovým průměrem a střední hodnotou. *Pokud s tím budete mít problém, zopakujte si kapitolu 4.1 a vraťte se na začátek kapitoly 9.2. Zvláštní pozornost věnujte pozn. 9.2.4.*

9.3 Výběrový rozptyl

Příklad 9.3.1: Následkem šumu měřicím přístrojem naměříme pokaždé jinou hodnotu. Chceme odhadnout velikost šumu.

Posoudíme několik možných odhadů rozptylu a jejich vlastnosti.

Definice 9.3.2: ***Výběrový rozptyl*** *náhodného výběru* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *je statistika*

$$S^2_{\boldsymbol{X}} = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{\boldsymbol{X}}_n)^2\,.$$

Alternativní značení: S^2. ***Realizaci výběrového rozptylu*** *značíme malým písmenem:*

$$s^2_{\boldsymbol{x}} = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}}_n)^2\,.$$

Tvrzení 9.3.3:

$$S^2_{\boldsymbol{X}} = \frac{1}{n-1} \sum_{j=1}^{n} X_j^2 - \frac{n}{n-1} \overline{\boldsymbol{X}}_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} X_j^2 - \frac{1}{n\,(n-1)} \Big(\sum_{j=1}^{n} X_j\Big)^2\,. \tag{9.2}$$

Důkaz:

$$\begin{aligned} S^2_{\boldsymbol{X}} &= \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{\boldsymbol{X}}_n)^2 = \\ &= \frac{1}{n-1} \sum_{j=1}^{n} X_j^2 - \frac{2\,\overline{\boldsymbol{X}}_n}{n-1} \underbrace{\sum_{j=1}^{n} X_j}_{n\,\overline{\boldsymbol{X}}_n} + \frac{n}{n-1} \overline{\boldsymbol{X}}_n^2 = \\ &= \frac{1}{n-1} \sum_{j=1}^{n} X_j^2 - \frac{n}{n-1} \overline{\boldsymbol{X}}_n^2\,. \end{aligned}$$

□

Tento vzorec se často používá pro praktický výpočet, zejména v kalkulátorech. Je totiž jednoprůchodový, stačí si průběžně střádat ve 3 registrech n, $\sum_{j=1}^{n} X_j$, $\sum_{j=1}^{n} X_j^2$ a není potřeba ukládat všechna data. Pokud je však střední hodnota ve srovnání se směrodatnou odchylkou velká, pak odečítáme podobně velká kladná čísla, čímž způsobíme velkou relativní numerickou chybu. Mohli bychom dokonce dostat záporný výsledek, který nemůže být správný. Pokud známe předem přibližně střední hodnotu, může pomoci následující vzorec:

Tvrzení 9.3.4: *Pro libovolné $c \in \mathbb{R}$ platí*

$$S_{\boldsymbol{X}}^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - c)^2 - \frac{1}{n\,(n-1)} \Big(\sum_{j=1}^{n} (X_j - c) \Big)^2 . \tag{9.3}$$

Důkaz: Jedná se o vzorec (9.2) pro výběrový rozptyl náhodného výběru $(X_1 - c, \ldots, X_n - c)$. Stejně jako rozptyl, ani výběrový rozptyl se přičtením konstanty nemění. □

Pokud máme dobrý odhad střední hodnoty, dosadíme jej za c ve vzorci (9.3) a odečítáme malé číslo

$$\frac{1}{n\,(n-1)} \Big(\sum_{j=1}^{n} (X_j - c) \Big)^2 .$$

Toto číslo by bylo velké, kdyby byl velký rozptyl, ale v tom případě by i výsledek byl velký a relativní chyba malá. Užitím vzorce (9.3) můžeme tedy snížit numerickou chybu. Nevýhodou je o něco větší složitost výpočtu, pokud ji nepřeneseme do předzpracování dat, která stačí zadávat v jiné stupnici. *(Přibližná znalost střední hodnoty nemůže být součástí obecného algoritmu pro výpočet výběrového rozptylu, nicméně je v praxi běžná. Obvykle se už podobná náhodná veličina někdy studovala a můžeme použít výsledky předchozích výzkumů. Např. o průměrném platu máme rámcovou představu, kterou nová data pouze upřesní.)*

Věta 9.3.5:

$$\mathrm{E} S_{\boldsymbol{X}}^2 = \mathrm{D} X .$$

Důkaz: Ze vzorce 9.2

$$\mathrm{E} S_{\boldsymbol{X}}^2 = \frac{n}{n-1} \mathrm{E} X^2 - \frac{n}{n-1} \mathrm{E} \overline{\boldsymbol{X}}_n^2 = \frac{n}{n-1} \Big(\mathrm{D} X + (\mathrm{E} X)^2 - \mathrm{D} \overline{\boldsymbol{X}}_n - \big(\mathrm{E} \overline{\boldsymbol{X}}_n\big)^2 \Big) =$$

$$= \frac{n}{n-1} \Big(\mathrm{D} X + (\mathrm{E} X)^2 - \frac{1}{n} \mathrm{D} X - (\mathrm{E} X)^2 \Big) = \mathrm{D} X .$$

□

Věta 9.3.6: *Výběrový rozptyl je nestranný konzistentní odhad rozptylu (pokud původní rozdělení má rozptyl a 4. centrální moment).*

Poznámka 9.3.7: Realizace výběrového rozptylu vždy existuje, a to i tehdy, když původní rozdělení rozptyl nemá. V takovém případě předchozí větu nelze použít.

Rozdělení výběrového rozptylu může být podstatně složitější. Odvodíme je speciálně pro normované normální rozdělení $\mathrm{N}(0,1)$ a rozsah výběru $n = 2$:

$$\overline{\boldsymbol{X}} = \frac{X_1 + X_2}{2} ,$$

$$X_1 - \overline{\boldsymbol{X}} = -(X_2 - \overline{\boldsymbol{X}}) = \frac{X_1 - X_2}{2} \text{ má rozdělení } \mathrm{N}\big(0, \tfrac{1}{2}\big) ,$$

$$S_{\boldsymbol{X}}^2 = (X_1 - \overline{\boldsymbol{X}})^2 + (X_2 - \overline{\boldsymbol{X}})^2 = 2 \left(\frac{X_1 - X_2}{2} \right)^2 = \left(\frac{X_1 - X_2}{\sqrt{2}} \right)^2 = U^2 ,$$

kde $U = \frac{X_1 - X_2}{\sqrt{2}}$ má rozdělení $\mathrm{N}(0,1)$. Rozdělení výběrového rozptylu $S_{\boldsymbol{X}}^2$ je v tomto případě rozdělení χ^2 s 1 stupněm volnosti. Obdobný výsledek platí obecněji:

Věta 9.3.8: *Pro výběrový rozptyl $S^2_{\boldsymbol{X}}$ výběru z* normálního *rozdělení* $\mathrm{N}(\mathrm{E}X, \mathrm{D}X)$ *má statistika* $\frac{(n-1)\,S^2_{\boldsymbol{X}}}{\mathrm{D}X}$ *rozdělení* $\chi^2(n-1)$ *(rozdělení* χ^2 *s* $n-1$ *stupni volnosti).*

Poznámka 9.3.9: Čitatel

$$(n-1)\,S^2_{\boldsymbol{X}} = \sum_{j=1}^{n} (X_j - \overline{\boldsymbol{X}}_n)^2$$

je součet kvadrátů odchylek od výběrového průměru.

Důsledek 9.3.10: *Rozptyl výběrového rozptylu výběru z* normálního *rozdělení* $\mathrm{N}(\mathrm{E}X, \mathrm{D}X)$ *je*

$$\mathrm{D}S^2_{\boldsymbol{X}} = \frac{2}{n-1}\,(\mathrm{D}X)^2\,.$$

Důkaz: Rozdělení $\chi^2(n-1)$ má rozptyl $2\,(n-1)$. Výběrový rozptyl dostaneme vynásobením statistiky $\frac{(n-1)\,S^2_{\boldsymbol{X}}}{\mathrm{D}X}$ konstantou $\frac{\mathrm{D}X}{n-1}$, ta se u rozptylu projeví v kvadrátu, proto

$$\mathrm{D}S^2_{\boldsymbol{X}} = 2\,(n-1)\left(\frac{\mathrm{D}X}{n-1}\right)^2 = \frac{2}{n-1}\,(\mathrm{D}X)^2\,.$$

□

Věta 9.3.11: *Pro náhodný výběr* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *z* normálního *rozdělení je* $\overline{\boldsymbol{X}}$ *nejlepší nestranný odhad střední hodnoty,* $S^2_{\boldsymbol{X}}$ *je nejlepší nestranný odhad rozptylu a statistiky* $\overline{\boldsymbol{X}}, S^2_{\boldsymbol{X}}$ *jsou konzistentní a* nezávislé.

Výběrový rozptyl je nejčastěji používaným odhadem rozptylu, ale nikoli jediným. Pokud známe typ rozdělení, můžeme odhadnout jeho parametry (některou z metod kapitoly 11) a z nich stanovit střední hodnotu i rozptyl. Kromě toho existuje i další často používaný odhad rozptylu, který je vychýlený, ale eficientnější:

$$\widehat{\Sigma}^2_{\boldsymbol{X}} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{\boldsymbol{X}}_n)^2 = \frac{n-1}{n}\,S^2_{\boldsymbol{X}}\,. \tag{9.4}$$

(Jako bychom zapomněli, že ve jmenovateli má být $n-1$, a použili n, stejně jako u výběrového průměru.) Jeho realizaci budeme značit $\widehat{\sigma}^2_X$. Je vidět, že pro velká n nebude rozdíl v obou odhadech podstatný.

Věta 9.3.12: *Odhad* $\widehat{\Sigma}^2_{\boldsymbol{X}}$ *dle (9.4) je vychýlený konzistentní odhad rozptylu.*

Důkaz:

$$\mathrm{E}\widehat{\Sigma}^2_{\boldsymbol{X}} = \frac{n-1}{n}\,\mathrm{D}X \to \mathrm{D}X\,,$$

$\widehat{\Sigma}^2_{\boldsymbol{X}}$ má rozptyl menší než $S^2_{\boldsymbol{X}}$, a to v poměru $\left(\frac{n-1}{n}\right)^2$. □

Eficienci obou odhadů rozptylu nemůžeme porovnat obecně; uděláme to aspoň pro *normální* rozdělení. Eficience odhadu $S^2_{\boldsymbol{X}}$ je

$$\mathrm{D}S^2_{\boldsymbol{X}} = \frac{2}{n-1}\,(\mathrm{D}X)^2\,.$$

Pro eficienci odhadu $\widehat{\Sigma}^2_{\boldsymbol{X}}$ dostáváme (jelikož $\mathrm{D}X$ je konstanta)

$$\mathrm{E}(\widehat{\Sigma}^2_{\boldsymbol{X}} - \mathrm{D}X)^2 = \mathrm{D}\left(\widehat{\Sigma}^2_{\boldsymbol{X}} - \mathrm{D}X\right) + \left(\mathrm{E}\left(\widehat{\Sigma}^2_{\boldsymbol{X}} - \mathrm{D}X\right)\right)^2 = \mathrm{D}\left(\widehat{\Sigma}^2_{\boldsymbol{X}}\right) + \left(\frac{1}{n}\,\mathrm{D}X\right)^2 =$$

$$= \left(\frac{n-1}{n}\right)^2 \frac{2}{n-1}\,(\mathrm{D}X)^2 + \frac{1}{n^2}\,(\mathrm{D}X)^2 = \frac{2n-1}{n^2}\,(\mathrm{D}X)^2\,.$$

Zbývá porovnat racionální funkce:

$$\frac{2n-1}{n^2} < \frac{2}{n} < \frac{2}{n-1}\,.$$

Z toho vyplývá, že odhad $\widehat{\Sigma}^2_{\boldsymbol{X}}$ je eficientnější než $S^2_{\boldsymbol{X}}$ (který je *nejlepší nestranný*). Máme tedy příklad toho, že eficience *nejlepšího nestranného* odhadu nemusí být nejlepší možná.

Příklad 9.3.13: Jaký odhad rozptylu tvaru $c_n\, S^2_{\boldsymbol{X}}$, kde c_n je konstanta závislá na n, je nejeficientnější, předpokládáme-li výběr z normálního rozdělení?

Řešení: $\frac{n-1}{n+1}\, S^2_{\boldsymbol{X}} = \frac{1}{n+1} \sum_{j=1}^{n} (X_j - \overline{\boldsymbol{X}}_n)^2$. Lze tedy jít ještě dále než ve vzorci (9.4). □

Poznámka 9.3.14: Vidíme, že mnohdy není snadné říci, „který odhad je nejlepší." Záleží na tom, jaké zvolíme kritérium optimality, ale i na tom, co víme o rozdělení, z něhož údaje pocházejí.

Seznámili jsme se s výběrovým rozptylem a jeho vlastnostmi. Jeho rozdělení pro výběr z *normálního* rozdělení popisujeme pomocí rozdělení χ^2. Výběrový rozptyl není jediným používaným odhadem rozptylu; jeho výhodou je nestrannost, nevýhodou může být nižší eficience.

9.4 Výběrová směrodatná odchylka

Příklad 9.4.1: Z předvolebního průzkumu jsou známy preference jednotlivých stran. Ptáme se, nakolik se skutečné výsledky voleb mohou lišit od těchto hodnot.

Druhým nejčastějším dotazem zadavatelů statistických výzkumů, hned po „průměrných hodnotách", je jejich „přesnost". Odpovíme jim odhadem směrodatné odchylky, neboť ta má názornější význam než rozptyl.

Definice 9.4.2: ***Výběrová směrodatná odchylka*** *náhodného výběru* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *je statistika*

$$S_{\boldsymbol{X}} = \sqrt{S^2_{\boldsymbol{X}}} = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{\boldsymbol{X}}_n)^2}\,.$$

Alternativní značení: S. ***Realizaci výběrové směrodatné odchylky*** *značíme malým písmenem:*

$$s_{\boldsymbol{x}} = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}}_n)^2}\,.$$

Věta 9.4.3: $\mathrm{E}S_{\boldsymbol{X}} \leq \sigma_X$.

Rovnost obecně nenastává:

Tvrzení 9.4.4: *Výběrová směrodatná odchylka je* vychýlený *odhad směrodatné odchylky.*

Důkaz:

$$\mathrm{D}X = \mathrm{E}S^2_{\boldsymbol{X}} = (\mathrm{E}S_{\boldsymbol{X}})^2 + \underbrace{\mathrm{D}S_{\boldsymbol{X}}}_{\geq 0} \geq (\mathrm{E}S_{\boldsymbol{X}})^2\,,$$

$$\sigma_X \geq \mathrm{E}S_{\boldsymbol{X}}\,.$$

Rovnost nastává pouze pro $\mathrm{D}S_{\boldsymbol{X}} = 0$. □

Vychýlenost výběrové směrodatné odchylky lze vysvětlit i následovně: Výběrový rozptyl je nestranný odhad rozptylu. Když z rozptylu i výběrového rozptylu vypočítáme odmocninu, aplikovali jsme nelineární funkci, a ta zobrazí kladné a záporné odchylky odlišně, čímž vznikne výchylka. *(Stejný princip se používá např. k demodulaci amplitudově modulovaného signálu pomocí nelinearity diody.)*

Věta 9.4.5: *Výběrová směrodatná odchylka je konzistentní odhad směrodatné odchylky (pokud původní rozdělení má rozptyl a 4. centrální moment).*

Výběrová směrodatná odchylka je hojně používaný odhad směrodatné odchylky, který však – na rozdíl od výběrového rozptylu – není nestranný. Výhodou je snazší interpretace, neboť výběrová směrodatná odchylka se měří ve stejných jednotkách jako původní náhodná veličina.

9.5 Výběrové obecné momenty

Příklad 9.5.1 (kolik kapek je do litru, pokračování): Náhodná veličina R je poloměr kapek. Těžko se určuje. Snáze změříme objem známého počtu kapek. Z něj dostaneme odhad třetího obecného momentu poloměru, viz příklad 4.6.1. Pro další použití potřebujeme vědět, jaké má tento odhad vlastnosti.

Zmíníme se o odhadech momentů náhodných veličin.

Definice 9.5.2: ***Výběrový k-tý obecný moment*** *náhodného výběru $\boldsymbol{X} = (X_1, \ldots, X_n)$ je statistika*

$$M_{X^k} = \frac{1}{n} \sum_{j=1}^{n} X_j^k .$$

Alternativní značení: M_k. ***Realizaci výběrového k-tého obecného momentu*** *značíme malým písmenem:*

$$m_{X^k} = \frac{1}{n} \sum_{j=1}^{n} x_j^k .$$

Výběrový k-tý obecný moment je nestranný odhad k-tého obecného momentu:

Věta 9.5.3: *Výběrový k-tý obecný moment je nestranný konzistentní odhad k-tého obecného momentu (pokud původní rozdělení má k-tý a $2\,k$-tý obecný moment).*

Důkaz:

$$\mathrm{E} M_{X^k} = \frac{1}{n} \sum_{j=1}^{n} \mathrm{E} X_j^k = \mathrm{E} X^k .$$

$$\mathrm{D} M_{X^k} = \frac{1}{n^2}\, n\, \mathrm{D} X^k = \frac{1}{n} \mathrm{D} X^k = \frac{1}{n} \left(\mathrm{E}(X^k)^2 - (\mathrm{E} X^k)^2 \right) = \frac{1}{n} \left(\mathrm{E} X^{2k} - (\mathrm{E} X^k)^2 \right) .$$

□

Poznámka 9.5.4: Realizace výběrových obecných momentů vždy existuje, a to i tehdy, když původní rozdělení tyto momenty nemá. V takovém případě předchozí větu nelze použít.

Výběrové obecné momenty slouží k odhadu obecných momentů. To využijeme k odhadu parametrů rozdělení metodou momentů v kapitole 11.1. Obdobně lze zavést a použít **výběrové centrální momenty**.

9.6 Histogram a empirické rozdělení

Příklad 9.6.1: Abychom správně nastavili clonu fotoaparátu, potřebujeme vyhodnotit rozdělení jasu. To bude mnohem snazší, jestliže jako vstup místo údajů z miliónů pixelů použijeme informaci o tom, kolikrát se vyskytla která z 256 úrovní šedi, které fotoaparát rozeznává; to nám k rozhodnutí stačí.

V (nenáhodném) vektoru $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ (získaném např. realizací náhodného výběru) nezáleží na pořadí složek (ale záleží na jejich opakování). Úsporněji je popsán množinou hodnot $H = \{x_1, \ldots, x_n\}$ (ta má nejvýše n prvků, obvykle méně) a jejich **četnostmi** n_u, $u \in H$. Tato data obvykle znázorňujeme **tabulkou četností** nebo grafem zvaným **histogram**.

Normováním dostaneme **relativní četnosti** $r_u = \frac{n_u}{n}$, $u \in H$. Jelikož $\sum_{u \in H} r_u = 1$, definují relativní četnosti pravděpodobnostní funkci $p_{\mathrm{Emp}(\boldsymbol{x})}(u) = r_u$ tzv. **empirického rozdělení** $\mathrm{Emp}(\boldsymbol{x})$. Je to diskrétní rozdělení s nejvýše n hodnotami charakterizující vektor $\boldsymbol{x}$. Indexem $\mathrm{Emp}(x)$ označujeme parametry jakékoli náhodné veličiny, která má empirické rozdělení $\mathrm{Emp}(x)$.

Věta 9.6.2 (vlastnosti empirického rozdělení):

$$\begin{aligned}
\mathrm{E}\,\mathrm{Emp}(\boldsymbol{x}) &= \sum_{u \in H} u\, r_u = \frac{1}{n} \sum_{u \in H} u\, n_u = \frac{1}{n} \sum_{j=1}^{n} x_j = \bar{\boldsymbol{x}}\,, \\
\mathrm{E}\,(\mathrm{Emp}(\boldsymbol{x}))^k &= \sum_{u \in H} u^k\, r_u = \frac{1}{n} \sum_{u \in H} u^k\, n_u = \frac{1}{n} \sum_{j=1}^{n} x_j^k\,, \\
\mathrm{D}\,\mathrm{Emp}(\boldsymbol{x}) &= \sum_{u \in H} (u - \mathrm{E}\,\mathrm{Emp}(\boldsymbol{x}))^2\, r_u = \frac{1}{n} \sum_{u \in H} (u - \bar{\boldsymbol{x}})^2\, n_u \\
&= \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}})^2 = \frac{n-1}{n}\, s_{\boldsymbol{x}}^2\,.
\end{aligned}$$

Rozptyl empirického rozdělení odpovídá odhadu $\widehat{\Sigma}_{\boldsymbol{X}}^2 = \frac{n-1}{n} S_{\boldsymbol{X}}^2$ dle vzorce (9.4), odlišnému od výběrového rozptylu $S_{\boldsymbol{X}}^2$.

Všimněme si, že např. suma $\sum_{j=1}^{n} x_j$ má n členů, kdežto $\sum_{u \in H} u\, n_u$ má $|H|$ členů (tolik, kolik je různých hodnot ve vektoru $\boldsymbol{x}$), což bývá podstatně méně.

Histogram je velmi často používán jako výchozí reprezentace dat pro další zpracování. Jeho použitím ušetříme nejen paměť, ale i čas pro výpočet.

9.7 Výběrový medián

Příklad 9.7.1: Je-li pravda, že „co Čech, to muzikant", pak by se to mělo projevit i na produkci hudebních nahrávek. Vezmeme-li za měřítko počet prodaných nosičů na obyvatele (včetně exportu), pak přestěhováním slavného umělce do jiné země se změní celková bilance, ač většina obyvatelstva zůstane stejná. Střední hodnota dobře nevypovídá o průměrných obyvatelích. (Podobně to dopadne, když se budeme ptát, kolik si obyvatelé státu v průměru vydělali tenisem nebo psaním knih pro mládež.)

Výběrový medián je medián empirického rozdělení, $q_{\mathrm{Emp}(\boldsymbol{x})}(\frac{1}{2})$. Poskytuje jinou informaci než výběrový průměr, mnohdy užitečnější (mj. *robustnější* – odolnější vůči vlivu vychýlených hodnot, *outliers*). Navíc víme, jak se změní monotónní funkcí, viz kapitola 4.2. Přesto se používá méně než výběrový průměr. Uvedeme některé důvody.

Výpočetní složitost střední hodnoty je lineární a paměťová náročnost je dána malou konstantou (kompletní data ani není nutno skladovat, stačí jediný průchod). Stanovení mediánu vyžaduje seřadit data podle velikosti; tento algoritmus má v nejlepším případě složitost přibližně $n \log_2 n$, kde n je rozsah souboru. Pro stanovení mediánu sice nepotřebujeme správně uspořádat nejmenší

a největší údaje, které nemají vliv, přesto je to procedura náročnější než výpočet střední hodnoty. Paměťová náročnost je úměrná rozsahu souboru, který potřebujeme mít stále celý k dispozici.

Ještě větším praktickým problémem je omezená možnost paralelizace. Pro výpočet střední hodnoty platu stačí získat z každého podniku 2 údaje: počet zaměstnanců a částku vyplacenou na mzdách. Pro výpočet mediánu potřebujeme kompletní tabulku platů jednotlivých zaměstnanců, kterou je nutno centrálně seřadit s údaji z ostatních míst. To vyžaduje zpracování rozsáhlých dat na jednom místě.

Problém může být i s pořízením dat, neboť např. platy jednotlivců jsou důvěrné údaje a lze předpokládat, že největší plat má ředitel, takže údaje přestávají být anonymní. Proto i oficiální výsledky statistického úřadu odhadují medián platu na základě neúplných údajů, protože se nepodařilo všechny shromáždit.

Naopak při testování hypotéz se může medián hodit. Můžeme jej totiž testovat např. znaménkovým testem (viz kapitola 12.5.1). Pak se respondentů ptáme, zda je jejich plat větší nebo menší než daná hodnota; lze předpokládat větší ochotu k odpovědi, než kdybychom se ptali přímo na výši platu.

Výhodou výběrového mediánu je i to, že se – na rozdíl od střední hodnoty – dá použít i pro pořadové náhodné veličiny. Ty mohou vzniknout i diskretizací kvantitativních náhodných veličin.

Příklad 9.7.2: Chceme vyhodnotit čas doručení zásilky nebo platby. Co máme udělat s těmi, které dosud nebyly doručeny? Možná ještě někdy přijdou a změní tím podstatně střední hodnotu dosažených časů. Naproti tomu medián můžeme vyhodnotit, jakmile dorazila první polovina.

Obdobný problém nastává, pokud hodnotíme věk, v němž lidé začali kouřit. Pak v lepším případě není ani medián definován.

Výběrový medián je robustnější alternativou k výběrovému průměru, přesto se používá méně, neboť:

- výpočetní náročnost je vyšší,
- paměťová náročnost je vyšší,
- možnosti paralelizace výpočtu jsou velmi omezené.

10. Intervalové odhady charakteristik rozdělení

10.1 Druhy intervalových odhadů

Obvykle první, na co se ptáme, je číselný odhad neznámé veličiny. Tento odhad je samozřejmě zatížen statistickou chybou. Proto následující přirozenou otázkou je, jak je odhad přesný, tj. v jakých mezích je správná hodnota.

Příklad 10.1.1: Pojišťovna stanovuje výši pojištění podle střední hodnoty vyplácených částek. Je pro ni však důležitý také jejich rozptyl a *horní odhad* pojistného plnění. Podle něj se rozhoduje, jak vysoké pojistky lze uzavřít s daným základním kapitálem, aby riziko bankrotu kvůli náhodné kumulaci pojistných událostí nepřesáhlo tolerovanou hodnotu.

Příklad 10.1.2: Na základě 100 pozitivních odpovědí mezi 1000 respondenty odhadneme volební preference určité politické strany na 10 %. Pokud bychom chtěli interval, v němž se skutečné preference *s jistotou* nacházejí, pak bychom mohli tvrdit pouze to, že stranu chce volit nejméně 100 a nejvýše $N-900$ lidí, kde N je rozsah základního souboru. Např. pro $N = 1\,000\,000$ bychom mohli říci, že preference jsou určitě v intervalu $\langle 0.0001, 0.9991\rangle$. To nám říká velmi málo, takže lepší je formulovat otázku, v jakém intervalu se hodnota nachází *s danou vysokou pravděpodobností*; přitom připouštíme riziko α, že tyto meze budou překročeny.

Dosud jsme pro skutečnou hodnotu parametru ϑ měli **bodový odhad**, což je náhodná veličina $\widehat{\Theta}$ taková, že $\widehat{\Theta} \doteq \vartheta$. Nyní místo toho hledáme **intervalový odhad**, tzv. **interval spolehlivosti** I (angl. *confidence interval*), což je minimální interval takový, že

$$P[\vartheta \in I] \geq 1 - \alpha\,,$$

kde $\alpha \in (0, 1/2)$ je *daná* pravděpodobnost, že meze intervalu I budou překročeny. Číslo $1-\alpha$ se nazývá **koeficient spolehlivosti.** Odpovědi vyplývají z kvantilů rozdělení náhodné veličiny $\widehat{\Theta}$. Na ni můžeme aplikovat výsledky z kapitoly 3.7, která říká, jaké meze intervalů překročí $\widehat{\Theta}$ s pravděpodobností α při *známém* rozdělení. Vyjde-li realizace $\widehat{\vartheta}$ mimo příslušný interval, bude porušen intervalový odhad, ovšem jen s pravděpodobností α. Nyní však budeme muset nerovnosti upravit, protože máme opačnou úlohu: $\widehat{\Theta}$ odhadneme realizací $\widehat{\vartheta}$ a ptáme se, při jakých hodnotách *neznámého* parametru ϑ by byla splněna příslušná nerovnost z intervalového odhadu náhodné veličiny $\widehat{\Theta}$. Pro ϑ můžeme dostat **horní**, resp. **dolní jednostranný odhad** nebo **symetrický oboustranný odhad**, pro který má překročení horní i dolní meze stejnou pravděpodobnost $\alpha/2$.

Intervalové odhady jsou žádané zejména kvůli hlediskům spolehlivosti. Potřebujeme k nim znát rozdělení příslušné statistiky, alespoň některé její kvantily. K tomu obvykle potřebujeme i předpoklady o původním rozdělení, z něhož byl proveden náhodný výběr (pokud nám nepomůže např. centrální limitní věta).

10.2 Intervalové odhady normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$

Řada metod je propracována pro normální rozdělení, zde uvedeme řešení nejčastějších úloh. Na závěr ukážeme, že tyto metody dovolují přibližně řešit i některé úlohy pro rozdělení, která nejsou normální.

10.2.1 Odhad střední hodnoty při známém rozptylu σ^2

Střední hodnotu μ odhadneme výběrovým průměrem $\overline{X}$; jeho rozdělení je $N\left(\mu, \sigma^2/n\right)$. Normovaná náhodná veličina norm $\overline{X} = \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu)$ má rozdělení $N(0,1)$;

$$1 - \alpha = P\left[\frac{\sqrt{n}}{\sigma}(\overline{X} - \mu) \leq \Phi^{-1}(1-\alpha)\right] = P\left[\mu \geq \overline{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right] =$$
$$= P\left[\mu \in \left\langle \overline{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha), \infty \right\rangle\right].$$

Dostáváme dolní intervalový odhad pro μ

$$\left\langle \overline{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha), \infty \right\rangle,$$

kde $\overline{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha) = \overline{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)$, ovšem $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$ pro $\alpha < 1/2$ nebývá v tabulkách. Obdobně dostaneme horní intervalový odhad

$$\left\langle -\infty, \overline{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha) \right\rangle.$$

Kombinací obou mezí po substituci $\alpha := \alpha/2$ dostaneme symetrický intervalový odhad

$$\left\langle \overline{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2), \overline{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha/2) \right\rangle.$$

Při výpočtu nahradíme výběrový průměr $\overline{X}$ jeho realizací $\bar{x}$.

Cvičení 10.2.1: Podle testů předpokládáme, že tachometr má směrodatnou odchylku $\sigma = 1.5$ *km/hod*. Jakou musí mít stálou chybu (tj. o kolik větší rychlost než skutečnou musí ukazovat), aby byla pravděpodobnost nejvýše $\alpha = 0.001$, že ukazuje menší rychlost než skutečnou?

Řešení: Za předpokladu normality rozdělení je příslušný kvantil $\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.999) \doteq 3.09$, takže tachometr má ukazovat alespoň o $\mu = \Phi^{-1}(1-\alpha)\,\sigma \doteq 3.09 \cdot 1.5 = 4.635$ *km/hod* více. Normální rozdělení $N\left(\mu, \sigma^2\right) = N\left(\sigma\,\Phi^{-1}(1-\alpha), \sigma^2\right)$ nabývá záporných hodnot s pravděpodobností α. Předpoklad normality je zde použitelný, avšak není jisté, zda dobře vystihuje pravděpodobnost velkých odchylek od průměru. □

10.2.2 Odhad střední hodnoty při neznámém rozptylu

Střední hodnotu μ odhadneme výběrovým průměrem $\overline{X}$; jeho rozdělení je $N\left(\mu, \sigma^2/n\right)$. Rozptyl σ^2 odhadneme výběrovým rozptylem S_X^2; jeho násobek $\frac{(n-1)S_X^2}{\sigma^2}$ má rozdělení $\chi^2(n-1)$.

Testujeme analogicky náhodnou veličinu

$$T = \frac{\sqrt{n}}{S_X}(\overline{X} - \mu),$$

jejíž rozdělení však není normální, ačkoli $\overline{X}, S_X$ jsou nezávislé. Zde potřebujeme Studentovo t-rozdělení $t(\eta)$ s η stupni volnosti, což je rozdělení náhodné veličiny

$$\frac{U}{\sqrt{\frac{V}{\eta}}},$$

kde U má rozdělení $N(0,1)$, V má rozdělení $\chi^2(\eta)$ a U, V jsou nezávislé.

V našem případě dosadíme

$$U = \frac{\sqrt{n}}{\sigma}\,(\overline{X} - \mu) \text{ má } \mathrm{N}(0,1)\,,$$
$$V = \frac{(n-1)\,S_X^2}{\sigma^2} \text{ má } \chi^2(n-1),$$
$$\eta = n - 1\,,$$
$$\frac{U}{\sqrt{\frac{V}{\eta}}} = \frac{\frac{\sqrt{n}}{\sigma}\,(\overline{X} - \mu)}{\sqrt{\frac{S_X^2}{\sigma^2}}} = \frac{\sqrt{n}}{S_X}\,(\overline{X} - \mu) = T \text{ má } \mathrm{t}(n-1).$$

Z toho vyplývají intervalové odhady pro μ:

$$\left(-\infty, \overline{X} + \frac{S_X}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(1-\alpha)\right\rangle,$$
$$\left\langle \overline{X} - \frac{S_X}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(1-\alpha), \infty\right),$$
$$\left\langle \overline{X} - \frac{S_X}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(1-\alpha/2), \overline{X} + \frac{S_X}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(1-\alpha/2)\right\rangle.$$

Při výpočtu nahradíme výběrový průměr $\overline{X}$ jeho realizací $\bar{x}$ a výběrovou směrodatnou odchylku S_X její realizací s_x.

Cvičení 10.2.2: Pro dopravu z kolejí Strahov do budovy ČVUT na Karlově náměstí byly z $n = 18$ měření pro cestu autobusem, resp. tramvají, získány následující odhady parametrů (v minutách)[1]:

$\overline{X} = 18.83$, $s_a = 2.706$, resp. $\overline{Y} = 20.90$, $s_t = 1.348$.

Kolik času je potřeba u těchto dopravních prostředků na cestu, abychom s pravděpodobností 95 % přišli včas?

Řešení: Potřebujeme 95 %-ní horní intervalový odhad. Za předpokladu normality původního rozdělení je intervalový odhad dán kvantilem Studentova t-rozdělení, konkrétně

$$(-\infty, \overline{X} + s_a\, q_{\mathrm{t}(n-1)}(0.95)\rangle \doteq (-\infty, 23.537\rangle\,, \quad (-\infty, \overline{Y} + s_t\, q_{\mathrm{t}(n-1)}(0.95)\rangle \doteq (-\infty, 23.245\rangle\,.$$

Z toho vychází, že tramvaj je časově výhodnější, přestože střední hodnota času je pro ni vyšší. Předpoklad normality však není vhodný. Připouští i velmi malé nebo záporné hodnoty času, což nelze, vhodnější by bylo logaritmicko-normální rozdělení. Kromě toho velké kladné odchylky mohou nastat z důvodů špatného počasí či dopravního kolapsu mnohem častěji, než předpokládá normální rozdělení, a navíc se některé příčiny projevují jinak u tramvají (výpadky proudu) a jinak u autobusů (špatná sjízdnost silnic). □

Cvičení 10.2.3: Náhodná veličina má alternativní rozdělení s pravděpodobností $p = 1/4$. Při jakém rozsahu výběru n vyjde výběrový průměr v intervalu $\langle p - \varepsilon, p + \varepsilon\rangle$ s pravděpodobností $1 - \alpha$, kde $\alpha = 5\,\%$ a (a) $\varepsilon = 0.05$, (b) $\varepsilon = 0.001$, (c) $\varepsilon = 10^{-6}$?

Řešení: Alternativní rozdělení má střední hodnotu p a rozptyl $p\,(1-p)$; výběrový průměr z výběru o rozsahu n má střední hodnotu p a rozptyl $p\,(1-p)\,/n$. Podle centrální limitní věty můžeme jeho rozdělení přibližně nahradit normálním, tedy $\mathrm{N}\,(p, p\,(1-p)\,/n)$ (pokud n bude dostatečně velké). Oboustranný intervalový odhad $\langle p-\varepsilon, p+\varepsilon\rangle$ dostaneme s koeficientem spolehlivosti $1-\alpha$, pokud bude

$$\frac{\varepsilon}{\sqrt{p\,(1-p)\,/n}} = \Phi^{-1}\,(1-\alpha/2) = \Phi^{-1}\,(0.975) \doteq 1.96\,.$$

[1] Tomáš Valenta: *Je rychlejší dostat se do školy (budovy ČVUT na Karlově náměstí) ze strahovských kolejí pomocí autobusu, nebo tramvaje?* Zápočtová práce z Matematiky pro výpočetní techniku, ČVUT, 2006, `http://cmp.felk.cvut.cz/~navara/MVT/Ze_Strahova.pdf`

Dostáváme

$$n \geq p\,(1-p)\left(\frac{\Phi^{-1}(1-\alpha/2)}{\varepsilon}\right)^2 \doteq p\,(1-p)\left(\frac{1.96}{\varepsilon}\right)^2 \doteq \frac{0.720\,3}{\varepsilon^2}\,,$$

(a) $n \geq 289$, (b) $n \geq 7.203 \cdot 10^5$, (c) $n \geq 7.203 \cdot 10^{11}$. Použití centrální limitní věty pro takto velké počty se zdá být oprávněné. □

Cvičení 10.2.4: Máme generátor náhodných čísel z intervalu $\langle 0, 1\rangle$ s rovnoměrným rozdělením. Odhadněte (co nejmenší) mez, kterou s pravděpodobností $\alpha = 0.05$ nepřevýší aritmetický průměr z $n = 10000$ takových čísel.

Řešení: Původní náhodná veličina Z má charakteristiky

$$\mathrm{E}Z = \frac{1}{2}\,, \qquad \mathrm{D}Z = \frac{1}{12}\,, \qquad \sigma_Z = \sqrt{\frac{1}{12}} \doteq 0.288\,68\,.$$

Výběrový průměr $\overline{Z}$ má charakteristiky

$$\mathrm{E}\overline{Z} = \frac{1}{2}\,, \qquad \mathrm{D}\overline{Z} = \frac{1}{12\,n} \doteq 8.333\,3 \cdot 10^{-6}\,, \qquad \sigma_{\overline{Z}} = \sqrt{\frac{1}{12\,n}} = \frac{1}{600}\sqrt{3} \doteq 2.886\,8 \cdot 10^{-3}\,.$$

Pomocí centrální limitní věty a kvantilu $\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95) \doteq 1.64485$ dostáváme horní intervalový odhad $\mathrm{E}\overline{Z} + \Phi^{-1}(1-\alpha)\,\sigma_{\overline{Z}} \doteq \mathrm{E}\overline{Z} + 1.64485\,\sigma_{\overline{Z}} \doteq 0.504\,75$. □

Cvičení 10.2.5: V letadle je $j = 216$ míst pro cestující. V průměru $q = 5\,\%$ cestujících se k odletu nedostaví. Kolik letenek lze prodat, aniž by riziko, že se do letadla nevejdou, nebylo větší než $\alpha = 2\,\%$? Posuďte použité předpoklady.

Řešení: Z m cestujících přijde počet, daný náhodnou veličinou X s binomickým rozdělením $\mathrm{Bi}(m, 1-q)$. Potřebujeme splnit nerovnost $q_X(1-\alpha) \leq j$. Exaktní výpočet by vyžadoval určit m zkusmo. Místo toho použijeme centrální limitní větu a s rozdělením náhodné veličiny X budeme pracovat, jako by bylo normální, s parametry $\mathrm{E}X = m\,(1-q)$, $\mathrm{D}X = m\,q\,(1-q)$. Znormováním X dostaneme

$$\frac{j - \mathrm{E}X}{\sigma_X} \geq \Phi^{-1}(1-\alpha)\,,$$

$$j \geq \mathrm{E}X + \sigma_X\,\Phi^{-1}(1-\alpha) = m\,(1-q) + \sqrt{m\,q\,(1-q)}\,\Phi^{-1}(1-\alpha)\,.$$

Z nerovnice

$$216 \geq 0.95\,m + \sqrt{0.05\,0.95\,m}\,\Phi^{-1}(0.98) \doteq 0.95\,m + 2.05375\,\sqrt{0.05\,0.95\,m}$$

pro kladná m dostáváme $m \leq 220.374$, tedy celočíselná řešení $m \leq 220$. Použili jsme předpoklady centrální limitní věty, tedy nezávislost alternativních náhodných veličin, vytvářejících binomické rozdělení, a jejich dostatečně velký počet. Obojí lze zpochybnit. Zrušení letu dvou cestujících nemusí být nezavislé, pokud např. chtěli letět na společnou dovolenou. Počet $m = 220$ se zdá být dostatečný na to, abychom směli použít přibližné řešení pomocí centrální limitní věty, avšak z něj by v průměru zrušilo let jen $m\,(1-q) = 11$ cestujících. Je tedy žádoucí výsledek ověřit pomocí binomického rozdělení. Pro $m = 220$ vyjde riziko

$$\sum_{k=217}^{m} \binom{m}{k} (1-q)^k\, q^{m-k} \doteq 4.2 \cdot 10^{-3}\,,$$

zadání vyhovuje i $m = 221$ s rizikem $1.2\,\%$, nikoli však $m = 222$ s rizikem $3.2\,\%$. □

Poznámka 10.2.6: Čísla v předchozím příkladu jsou fiktivní, ale úloha skutečná. Tzv. *overbooking* se skutečně používá, neboť pro letecké společnosti je výhodnější zaplatit občas dodatečné náklady cestujícím, kteří se nevešli na palubu, než soustavně nevyužívat kapacitu letadla. Zde by se např. prodalo 5 letenek navíc a jen zhruba v jednom letu z 80 by si toho někdo všiml. Vzhledem k pochybám o nezávislosti je však potřebná podrobnější analýza problému.

Cvičení 10.2.7: Ve vzorku je 10^6 atomů radioaktivního izotopu. Určete symetrický 99 %-ní intervalový odhad počtu atomů, které se rozpadnou za dvojnásobek poločasu rozpadu.

Řešení: Odhadujeme náhodnou veličinu X s rozdělením $\mathrm{Bi}(n,p)$, $n = 10^6$, $p = 1-(1/2)^2 = 3/4$ (=pravděpodobnost, že se atom v daném čase rozpadne), $\mathrm{E}X = np = 750\,000$, $\mathrm{D}X = np(1-p) = 187\,500$, $\sigma_X = \sqrt{np(1-p)} = 250\sqrt{3} \doteq 433$. Při aproximaci normálním rozdělením vyjdou meze $\mathrm{E}X \pm \sigma_X \Phi^{-1}(0.995) \doteq 750\,000 \pm 433 \cdot 2.576 \doteq 750\,000 \pm 1\,115$, interval přibližně $\langle 748\,885, 751\,115\rangle$. □

Cvičení 10.2.8: Generátor náhodných čísel realizuje rovnoměrné rozdělení na intervalu $\langle 0,1\rangle$. Určete interval se středem $1/2$, do něhož padne výběrový průměr z 1000 náhodných čísel s pravděpodobností 99 %.

10.2.3 Odhad rozptylu a směrodatné odchylky

Rozptyl σ^2 odhadneme výběrovým rozptylem $S^2_{\boldsymbol{X}}$; jeho násobek $\frac{(n-1)S^2_{\boldsymbol{X}}}{\sigma^2}$ má rozdělení $\chi^2(n-1)$.

$$
\begin{aligned}
1-\alpha &= P\left[\frac{(n-1)\,S^2_{\boldsymbol{X}}}{\sigma^2} \le q_{\chi^2(n-1)}(1-\alpha)\right] = \\
&= P\left[\frac{(n-1)\,S^2_{\boldsymbol{X}}}{q_{\chi^2(n-1)}(1-\alpha)} \le \sigma^2\right] = \\
&= P\left[\sigma^2 \in \left\langle \frac{(n-1)\,S^2_{\boldsymbol{X}}}{q_{\chi^2(n-1)}(1-\alpha)}, \infty \right)\right] .
\end{aligned}
$$

Dostali jsme *dolní* odhad

$$\left\langle \frac{(n-1)\,S^2_{\boldsymbol{X}}}{q_{\chi^2(n-1)}(1-\alpha)}, \infty \right) .$$

Obdobně dostaneme i další intervalové odhady,

$$\left(-\infty, \frac{(n-1)\,S^2_{\boldsymbol{X}}}{q_{\chi^2(n-1)}(\alpha)} \right\rangle ,$$

$$\left\langle \frac{(n-1)\,S^2_{\boldsymbol{X}}}{q_{\chi^2(n-1)}\,(1-\alpha/2)}, \frac{(n-1)\,S^2_{\boldsymbol{X}}}{q_{\chi^2(n-1)}\,(\alpha/2)} \right\rangle .$$

Odmocněním mezí dostaneme *intervalové odhady směrodatné odchylky.*
Při výpočtu nahradíme výběrový rozptyl $S^2_{\boldsymbol{X}}$ jeho realizací $s^2_{\boldsymbol{x}}$.

Cvičení 10.2.9: Náhodná veličina X má normované normální rozdělení $\mathrm{N}(0,1)$. Určete 90 %-ní symetrické intervalové odhady náhodných veličin (a) $2X+4$,(b) $|X|$, (c) X^2.

Řešení: 90 %-ní symetrický intervalový odhad náhodné veličiny $f(X)$ je $\langle q_{f(X)}(0.05), q_{f(X)}(0.95)\rangle$.
(a) Lineární transformací $x \mapsto 2x+4$ dostaneme odhad

$$
\begin{aligned}
\langle q_{2X+4}(0.05), q_{2X+4}(0.95)\rangle &= \langle 2\Phi^{-1}(0.05)+4,\ 2\Phi^{-1}(0.95)+4\rangle = \\
&= \langle -2\Phi^{-1}(0.95)+4,\ 2\Phi^{-1}(0.95)+4\rangle = \\
&\doteq \langle -2\cdot 1.64+4,\ 2\cdot 1.64+4\rangle = \langle 0.72,\ 7.28\rangle .
\end{aligned}
$$

(b) $q_{|X|}(0.05)$ je takové číslo c, že

$$\begin{aligned} 0.05 &= P[|X| \le c] = P[-c \le X \le c] = \Phi(c) - \Phi(-c) = \\ &= 2\,(\Phi(c) - \Phi(0)) = 2\Phi(c) - 1\,, \\ \Phi(c) &= \frac{1.05}{2} = 0.525\,, \\ q_{|X|}(0.05) &= c = \Phi^{-1}(0.525) \doteq 0.0627\,. \end{aligned}$$

Podobně

$$q_{|X|}(0.95) = \Phi^{-1}(0.975) \doteq 1.96\,.$$

(c) Jelikož $X^2 = |X|^2$, stačí výsledek (b) zobrazit funkcí $x \mapsto x^2$, která je na příslušném oboru (kladných reálných čísel) rostoucí. Jedná se též o rozdělení χ^2 s 1 stupněm volnosti.

$$\begin{aligned} q_{X^2}(0.05) &= \left(q_{|X|}(0.05)\right)^2 = q_{\chi^2(1)}(0.05) \doteq 0.0627^2 \doteq 3.9 \cdot 10^{-3}\,, \\ q_{X^2}(0.95) &= \left(q_{|X|}(0.95)\right)^2 = q_{\chi^2(1)}(0.95) \doteq 1.96^2 \doteq 3.84\,. \end{aligned}$$

□

10.3 Intervalové odhady rozdělení, která nejsou normální

Náhodnou veličinu X se spojitým rozdělením lze převést na náhodnou veličinu $h(X)$ s normovaným normálním rozdělením; h je neklesající nelineární transformace

$$h(t) = \Phi^{-1}(F_X(t))\,.$$

(Mezivýsledek $F_X(X)$ je náhodná veličina s rovnoměrným rozdělením na $\langle 0, 1\rangle$.)

Použijeme intervalový odhad pro normální rozdělení a transformujeme jej zpět podle vzorce

$$h^{-1}(u) = q_X(\Phi(u))\,.$$

Intervalové odhady z jiných než spojitých rozdělení bývají komplikovanější. Někdy je lze přibližně nahradit normálním rozdělením pomocí centrální limitní věty; pak opět použijeme alespoň přibližný postup založený na normálním rozdělení.

Intervalové odhady dávají odpověď na otázky o přesnosti statistických odhadů. Jejich formulace (připouštějící malé riziko překročení mezí) není pro laickou veřejnost příliš názorná ani srozumitelná. O to víc je potřeba, aby alespoň část populace uměla tyto údaje správně interpretovat. Jednodušší a užitečná náhrada za tento nástroj totiž neexistuje.

11. Odhady parametrů rozdělení

Často nechceme odhadnout jen střední hodnotu či rozptyl (resp. směrodatnou odchylku), ale i jiné parametry rozdělení.

Příklad 11.1.1: Ve hře „kámen-nůžky-papír" chceme odhadnout strategii soupeře a tím zvýšit své šance na výhru. I kdybychom možné výsledky označili číselně, střední hodnota ani rozptyl nejsou ty parametry, které by nám byly užitečné. Místo toho nás mohou zajímat pravděpodobnosti jednotlivých výsledků, případně jejich posloupností.

Rozdělení náhodné veličiny X závisí na vektoru parametrů $\vartheta = (\vartheta_1, \ldots, \vartheta_k) \in \Pi$, kde $\Pi \subseteq \mathbb{R}^k$ je *parametrický prostor*, tj. množina všech přípustných hodnot parametrů; pravděpodobnostní funkci značíme $p_X(t; \vartheta) = p_X(t; \vartheta_1, \ldots, \vartheta_k)$ atd. Hledáme odhad $\widehat{\boldsymbol{\Theta}} = (\widehat{\Theta}_1, \ldots, \widehat{\Theta}_k)$, resp. realizaci odhadu $\widehat{\vartheta} = (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_k)$ pomocí realizace $\boldsymbol{x} = (x_1, \ldots, x_n)$.

Příklady k této kapitole jsou soustředěny na jejím konci, až po vyložení metod, čímž se umožní jejich srovnání.

11.1 Metoda momentů

Poznámka 11.1.2: Motivace (nepovinná pro ty, kdo neznají charakteristickou funkci náhodné veličiny): Charakteristická funkce jednoznačně určuje rozdělení náhodné veličiny. Z charakteristické funkce lze určit všechny existující obecné momenty, mnohdy však i naopak. Pokud má charakteristická funkce rozvoj v Taylorovu řadu (což nemusí být, protože řada nemusí konvergovat), pak obecné momenty jednoznačně určují koeficienty řady, a tím i charakteristickou funkci. (Až na násobek konstantou závislou na řádu jsou obecné momenty přímo koeficienty Taylorova rozvoje charakteristické funkce v 0, jak vyplývá z věty 4.7.11.)

Znalost všech momentů by vyžadovala nekonečnou informaci, kterou nemáme k dispozici. Proto na základě konečné informace (realizace náhodného výběru) hledáme takový model, který by se shodoval s pozorováním v několika prvních momentech. Požadujeme tedy shodu empirického rozdělení s teoretickým v momentech nejnižších řádů.

Pro $i = 1, 2, \ldots$ je i-tý obecný moment funkcí ϑ,

$$\mathrm{E}X^i(\vartheta) = \mathrm{E}X^i(\vartheta_1, \ldots, \vartheta_k)$$

(závislost na parametrech lze stanovit dle *teoretického* pravděpodobnostního modelu). Na základě *realizace* odhadneme i-tý obecný moment pomocí *výběrového* i-tého obecného momentu

$$m_{X^i} = \frac{1}{n} \sum_{j=1}^{n} x_j^i .$$

Hledáme takovou realizaci odhadu $\widehat{\vartheta} = (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_k)$, při níž nastává rovnost příslušných momentů a jejich odhadů:

$$\mathrm{E}X^i(\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_k) = m_{X^i} = \frac{1}{n} \sum_{j=1}^{n} x_j^i, \qquad i = 1, 2, \ldots .$$

K jednoznačnému určení k proměnných obvykle potřebujeme (prvních) k rovnic pro $i = 1, 2, \ldots, k$.

Použitelnost metody momentů

Můžeme narazit na řadu problémů typických pro řešení soustavy (nelineárních) rovnic:

1. Řešení neexistuje. Může se dokonce stát, že některá z rovnic sama o sobě nemá řešení. Jelikož není nijak předepsán počet rovnic (porovnávaných momentů), můžeme z nich nějakou ubrat.
2. Je nekonečně mnoho řešení. To se typicky stává proto, že některá z rovnic řešení vůbec neomezuje; např. proměnné se v ní vůbec nevyskytují, protože příslušný moment nezávisí na hledaných parametrech. Můžeme přidat další rovnice pro větší i.
3. Je více než jedno řešení. To se typicky stává např. u soustavy kvadratických rovnic. Bohužel, pokud všechna řešení odpovídají podmínkám na ně kladeným, nemáme kritérium, podle kterého bychom rozhodli, které řešení je lepší.
4. Je jediné řešení, ale je obtížné je nalézt. To je velmi častý případ, že u složitějších soustav rovnic symbolické řešení není možné a numerické může být komplikované.
5. Soustava je špatně podmíněná. I malé nepřesnosti ve vstupních hodnotách způsobují velké chyby výsledných odhadů (typicky pro velký počet parametrů).
6. Našli jsme jediné řešení, které však *nesplňuje předpoklady*, tedy $\widehat{\vartheta} \notin \Pi$. Obvykle parametry nemohou být libovolná čísla, abychom v modelu nedostali např. záporné pravděpodobnosti. *Vždy je třeba kontrolovat řešení*, zda odpovídá zadání. Mohli bychom se sice spokojit i s přibližným řešením, které předpoklady splňuje, avšak je otázka, které je „nejbližší": postrádáme kritérium pro porovnání chyb v jednotlivých momentech, jelikož mohou mít různý fyzikální rozměr.

Kromě toho lze metodě momentů vytknout, že všem rovnicím je přikládána stejná důležitost. Např. rovnost střední hodnoty a jejího odhadu nám připadá přirozená (její porušení by se při použití modelu brzy projevilo). Těžko však zdůvodnit, proč by stejně důležitá byla rovnost desátého obecného momentu, jehož vliv stěží dovedeme interpretovat. I z řady výše uvedených důvodů *metodu momentů nelze doporučit, pokud je hledaných parametrů mnoho.*

Metodu momentů nelze použít pro nenumerická data (pokud je nelze smysluplně očíslovat). Výjimku povolíme v pozn. 11.3.6.

Výhodou metody momentů je, že ji lze beze změn použít pro diskrétní, spojité i *smíšené* rozdělení.

11.2 Metoda maximální věrohodnosti

Příklad 11.2.1: V příkladu 3.5.23 jsme řešili úlohu, jak odhadnout průměrný věk populace *(=střední hodnotu věku), máme-li data pouze roztříděná do intervalů, tedy pořadovou náhodnou veličinu.* Můžeme navrhnout vhodný typ rozdělení s několika parametry, a ty odhadneme ze zjištěných dat. Metodu momentů nelze použít, protože nedovedeme odhadnout ani první moment (=střední hodnotu). Přesto data dávají podklady použitelné pro stanovení neznámých parametrů.

Myšlenkou metody maximální věrohodnosti je, že hledáme takové hodnoty parametrů, které by nejlépe vysvětlovaly realizaci náhodného výběru, tj. při kterých by pozorované výsledky byly „nejméně nepravděpodobné". Ve skutečnosti se jedná o dvě metody, které pracují se zcela jinými pojmy v případech diskrétního a spojitého rozdělení.

11.2.1 Metoda maximální věrohodnosti pro diskrétní rozdělení

Definice 11.2.2: *Nechť $\boldsymbol{x} = (x_1, \ldots, x_n)$ je realizace náhodného výběru z diskrétního rozdělení s pravděpodobnostní funkcí $p_X(.; \vartheta)$ závislou na vektoru parametrů $\vartheta \in \Pi$. Pak definujeme* věrohodnost realizace diskrétního rozdělení *(zkráceně* věrohodnost, *angl.* likelihood*)* $L\colon \Pi \to \langle 0, 1\rangle$ *vztahem*

$$L(\vartheta) = p_{\boldsymbol{X}}(\boldsymbol{x};\vartheta) = \prod_{j=1}^{n} p_X(x_j;\vartheta)\,.$$

Poznámka 11.2.3: Je třeba rozlišovat mezi věrohodností a pravděpodobností. Obě nabývají hodnot z intervalu $\langle 0,1\rangle$, neboť věrohodnost realizace diskrétního rozdělení je pravděpodobnost této realizace (využíváme zde předpoklad nezávislosti složek náhodného výběru),

$$L(\vartheta) = p_{\boldsymbol{X}}(\boldsymbol{x};\vartheta) = P\left[X_1 = x_1,\ldots,X_n = x_n;\vartheta\right] = \prod_{j=1}^{n} P\left[X_j = x_j;\vartheta\right] = \prod_{j=1}^{n} p_X(x_j;\vartheta)\,.$$

Mají však odlišný definiční obor. Pravděpodobnost je definována pro jevy z nějaké σ-algebry, kdežto věrohodnost je definována na prostoru Π všech možných hodnot parametrů rozdělení. (Obvykle $\Pi \subseteq \mathbb{R}^k$.) Pravděpodobnost je určena pro předvídání výsledků *budoucího* náhodného pokusu při *známém* pravděpodobnostním modelu. Naopak věrohodnost vyhodnocuje sérii pokusů již *realizovaných* a dovoluje jim přizpůsobit *neznámé* parametry pravděpodobnostního modelu, který k nim vedl.

Metoda maximální věrohodnosti považuje za správný odhad parametrů takové hodnoty $\widehat{\vartheta} = (\widehat{\vartheta}_1,\ldots,\widehat{\vartheta}_k)$, které maximalizují věrohodnost. Stejný výsledek dostaneme, jestliže místo věrohodnosti maximalizujeme její logaritmus (angl. *log-likelihood*),

$$\ell(\vartheta) = \ln L(\vartheta) = \sum_{j=1}^{n} \ln p_X(x_j;\vartheta)\,.$$

(Nutno vyloučit případ $p_X(x_j;\vartheta) = 0$, který však nevede na maximum.)

Obvykle se hodnoty v realizaci náhodného výběru opakují, proto uvedeme ještě vzorce vycházející z četností, které dávají stejný výsledek typicky s menší složitostí. Označme $u_1,\ldots,u_k$ ($k \le n$) všechny hodnoty, které se vyskytují v realizaci $\boldsymbol{x} = (x_1,\ldots,x_n)$, n_i četnost a $r_i = n_i/n$ relativní četnost hodnoty u_i v realizaci $\boldsymbol{x}$. (Jelikož uvažujeme jen hodnoty, které se vyskytly, jsou všechny tyto četnosti nenulové.) Pak

$$L(\vartheta) = \prod_{j=1}^{n} p_X(x_j;\vartheta) = \prod_{i=1}^{k} \left(p_X(u_i;\vartheta)\right)^{n_i}\,,$$

$$\ell(\vartheta) = \sum_{j=1}^{n} \ln p_X(x_j;\vartheta) = \sum_{i=1}^{k} n_i \ln p_X(u_i;\vartheta)\,.$$

Odhad na základě maxima věrohodnosti odpovídá bayesovskému odhadu ve speciálním případě, kdy všechny hodnoty parametrů mají stejnou apriorní pravděpodobnost (nebo hustotu pravděpodobnosti). Používá se, pokud apriorní pravděpodobnosti parametrů neznáme. To však může vést k neuspokojivým závěrům:

Příklad 11.2.4: Přítel neodpovídá na dotazy poslané elektronickou poštou. Důvodem může být cestování, nemoc, nebo smrt. Apriorní pravděpodobnosti těchto příčin neznáme, proto rozhodneme podle maximální věrohodnosti, tj. podle podmíněné pravděpodobnosti, že nepíše, je-li příslušný důvod pravdivý. Dojdeme k závěru, že přítel je mrtev, neboť v tom případě podmíněná pravděpodobnost, že nemůže odpovídat, je 1.

11.2.2 Metoda maximální věrohodnosti pro spojité rozdělení

Definici 11.2.2 nelze použít pro spojité rozdělení, neboť každá jeho realizace má nulovou pravděpodobnost. Používá se obdobný vzorec, kde místo pravděpodobnostní funkce použijeme hustotu pravděpodobnosti, což ale vede na zcela *jiný pojem* (byť stejně pojmenovaný; zde je rozlišíme aspoň značením):

Definice 11.2.5: *Nechť* $\boldsymbol{x} = (x_1, \ldots, x_n)$ *je realizace náhodného výběru ze spojitého rozdělení se* spojitou *hustotou* $f_X(.;\vartheta)$ *závislou na vektoru parametrů* $\vartheta \in \Pi$. *Pak definujeme* věrohodnost realizace spojitého rozdělení *(zkráceně* věrohodnost*)* $\Lambda\colon \Pi \to \langle 0, \infty)$ *vztahem*

$$\Lambda(\vartheta) = f_{\boldsymbol{X}}(\boldsymbol{x};\vartheta) = \prod_{j=1}^{n} f_X(x_j;\vartheta)\,.$$

(Nutno vyloučit případ $f_X(x_j;\vartheta) = 0$, *který však nevede na maximum.)*

Poznámka 11.2.6: Na rozdíl od diskrétního rozdělení může věrohodnost realizace spojitého rozdělení nabývat libovolně velkých hodnot, stejně jako hustota. Rozdíl mezi věrohodností a pravděpodobností je zde nápadnější, i když základní je odlišnost definičních oborů dle poznámky 11.2.3.

Podle **metody maximální věrohodnosti** opět považujeme za správný odhad parametrů takové hodnoty $\widehat{\vartheta} = (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_k)$, které maximalizují věrohodnost. Místo ní často maximalizujeme její logaritmus,

$$\lambda(\vartheta) = \ln \Lambda(\vartheta) = \sum_{j=1}^{n} \ln f_X(x_j;\vartheta)\,.$$

Poznámka 11.2.7: Bez předpokladu spojitosti hustoty by definice 11.2.5 nebyla korektní. Hustota rozdělení je určena svým integrálem (distribuční funkcí), a je tedy definována až na množinu nulové míry (viz pozn. 3.5.12). Změníme-li hodnoty hustoty např. v konečně mnoha bodech, považujeme ji stále za hustotu téhož rozdělení. Těmito body mohou být i všechny hodnoty realizace náhodného výběru, z nichž vycházíme při výpočtu věrohodnosti. Striktně vzato, nemáme právo hovořit o hodnotách hustoty v jednotlivých bodech, neboť nejsou definovány; korektní jsou pouze integrální charakteristiky hustoty.

Pokud požadujeme navíc *spojitou* hustotu, pak je již jednoznačná a postup je korektní. Tímto způsobem se také metoda maximální věrohodnosti obvykle používá, ale v definici bývá požadavek spojitosti hustoty často opomíjen. Někteří autoři definují hustotu pouze jako derivaci distribuční funkce. Pak je jednoznačně určena a zavedení věrohodnosti je v pořádku, ale připravujeme se tím o možnost zavedení hustoty u řady běžně používaných rozdělení. Proto je zde volena obecnější definice hustoty, avšak při zavedení věrohodnosti je přidán předpoklad spojitosti.

Nepříjemným důsledkem je, že není definována věrohodnost pro taková rozdělení, která nemají žádnou spojitou hustotu. Např. exponenciální rozdělení má hustotu nespojitou v nule. Pokud by se v realizaci vyskytla nulová hodnota, neměli bychom vůbec žádný korektní způsob, jak určit věrohodnost. To je neodstranitelná vada pojmu věrohodnosti pro spojité rozdělení. Problém by nebyl, kdyby nulová hodnota nebyla možná. To je často zajištěno, a pak by postup metodou maximální věrohodnosti byl použitelný. Bohužel popis rozdělení náhodné veličiny, který nerozlišuje mezi jevem nemožným a jevem s nulovou pravděpodobností, nedovoluje vyjádřit, zda je nulová hodnota možná, a zavést věrohodnost pouze na základě popisu pravděpodobnostního rozdělení, takže definice 11.2.5 to nepřipouští.

11.2.3 Metoda maximální věrohodnosti pro smíšené rozdělení

Věrohodnost smíšeného rozdělení není definována. Žádná z předchozích definic nedovoluje rozšíření na druhý případ. Přesto se metoda maximální věrohodnosti používá i jako součást odhadů parametrů smíšeného rozdělení. Postup je založen na následující úvaze: Nutným předpokladem je, že známe množinu D všech hodnot, jichž nabývá (diskrétní složka) s nenulovou pravděpodobností. Hodnoty z D mohou být též hodnotami spojité složky, avšak ta je nabývá s nulovou pravděpodobností, což neovlivní výsledek. Předpokládáme tedy, že v realizaci každá hodnota z D pocházela z diskrétní složky směsi. Pravděpodobnost množiny D při empirickém rozdělení je

$$P\left[\mathrm{Emp}\,(\boldsymbol{x}) \in D\right] = \frac{n_D}{n}\,,$$

kde n je rozsah výběru a n_D je počet složek realizace $(x_1, \ldots, x_n)$, které patří do D. Tuto pravděpodobnost lze považovat za odhad váhy diskrétní složky hledaného rozdělení. Parametry diskrétní složky směsi lze pak stanovit metodou maximální věrohodnosti pro realizaci *diskrétního* rozdělení (získanou tím, že z realizace $(x_1, \ldots, x_n)$ ponecháme pouze ty hodnoty, které *jsou* v D). Následně odhadneme spojitou složku směsi metodou maximální věrohodnosti pro realizaci *spojitého* rozdělení (získanou tím, že z realizace $(x_1, \ldots, x_n)$ ponecháme pouze ty hodnoty, které *nejsou* v D). Tímto postupem se můžeme dopracovat k přijatelnému odhadu parametrů původního smíšeného rozdělení. Nemůžeme však přímo tvrdit, že je to maximálně věrohodný odhad, neboť nemáme takovou definici věrohodnosti, která by vystihovala vlastnosti získaného odhadu.

Příklad 11.2.8: Metodou maximální věrohodnosti se odhaduje množství srážek, které se modeluje smíšeným rozdělením (Diracovo v nule a tzv. rozdělení gamma).[1]

Použitelnost metody maximální věrohodnosti

I zde můžeme narazit na řadu problémů typických pro řešení soustavy (nelineárních) rovnic:

1. Maximum neexistuje. To není častý případ, neboť *spojitá* funkce na *uzavřené omezené* množině má globální maximum. Tyto podmínky mohou být porušeny tím, že věrohodnost nezávisí na parametrech spojitě, že parametrický prostor není omezený nebo není uzavřený. I v takovém případě někdy aspoň máme odhad, jehož věrohodnost je blízká maximu.
2. Je více než jedno maximum, máme více stejně dobrých odhadů. To nemusí vadit, pokud různé optimální hodnoty parametrů popisují stejné rozdělení. V tom případě máme jediné optimální rozdělení, a pokud parametry samy nejsou pro nás důležité, nemusí nám vadit jejich nejednoznačnost. (S takovým případem se setkáme v příkladu 11.3.9.)
3. Je jediné řešení, ale je obtížné je nalézt. (Lokální extrémy nemusí být globální.)
4. Hodnoty věrohodnosti mohou být velmi malé. Pro veliké soubory dat dostáváme součin velkého počtu činitelů, typicky menších než jedna (aspoň v diskrétním případě). V programech bývá potřebné ošetřit podtečení. Tato námitka poněkud zpochybňuje motivaci věrohodnosti – vybíráme největší pravděpodobnost z mnoha, které jsou všechny velmi blízké nule [Schlesinger, Hlaváč 1999].
5. *Metodu maximální věrohodnosti nelze použít pro smíšené rozdělení.* Výjimkou je výše uvedený postup, který však nelze formulovat pomocí jediného reálného kritéria, obdoby věrohodnosti.

Z výhod metody maximální věrohodnosti uveďme aspoň následující:

1. Hledání maxima je o něco snazší než řešení soustavy rovnic. Při iteračních metodách alespoň máme vodítko, nakolik se mění kritérium.
2. Různým datům je dán společný význam. I když se jednotlivé parametry měří třeba v jiných jednotkách, vše je převedeno na pravděpodobnost nebo hustotu pravděpodobnosti, která je bezrozměrná a umožňuje srovnání.
3. Metodu maximální věrohodnosti lze použít i na nenumerická data.

11.3 Příklady a cvičení na odhad parametrů

V následujících příkladech používáme někdy značení zjednodušené oproti předchozímu odvození, např. nevyznačujeme závislost veličin na parametrech nebo nerozlišujeme parametr a jeho odhad.

[1] Gneiting, T., Balabdaoui, F., Raftery, A.E.: *Probabilistic Forecasts, Calibration and Sharpness.* Technical Report no. 483, Department of Statistics, University of Washington, 2005, `http://www.stat.washington.edu/www/research/reports/2005/tr483.pdf`

Odhady diskrétních rozdělení

Cvičení 11.3.1: 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: Odhadujeme jen jeden parametr, neboť $p = 1 - q - q^2$.

Metoda momentů: Předpokládáme rovnost výrazů

$$\mathrm{E}X = q + 2\,q^2 = m_X = \frac{1 \cdot 12 + 2 \cdot 6}{2 + 12 + 6} = \frac{6}{5}\,.$$

Dostáváme kvadratickou rovnici

$$q + 2\,q^2 = \frac{6}{5}\,,$$

která má 2 řešení

$$q_1 = -\frac{1}{20}\sqrt{265} - \frac{1}{4} \doteq -1.063\,9\,, \qquad q_2 = \frac{1}{20}\sqrt{265} - \frac{1}{4} \doteq 0.563\,94\,.$$

Z nich q_1 nevyhovuje zadání, q_2 vyhovuje, $p = 1 - q_2 - q_2^2 \doteq 0.118\,03$.

Metoda maximální věrohodnosti:

$$\ell(q) = 2\,\ln p + 12\,\ln q + 6\,\ln q^2 = 2\,\ln\left(1 - q - q^2\right) + 24\,\ln q\,,$$
$$\frac{\partial}{\partial q}\ell(q) = \frac{24}{q} + 2\,\frac{-2\,q - 1}{1 - q^2 - q}\,,$$

což je opět kvadratická rovnice s řešeními $q_1 = -\frac{3}{2}$ (nevyhovuje), $q_2 = \frac{4}{7} \doteq 0.571\,43$ (vyhovuje), $p = 1 - q_2 - q_2^2 = \frac{5}{49} \doteq 0.102\,04$. □

Cvičení 11.3.2: V osudí jsou 2 druhy kostek, na prvních jsou čísla $1, \ldots, 6$, na druhých pouze $2, 4, 6$, u obou druhů jsou všechny možné výsledky stejně pravděpodobné. Vytáhli jsme 20 kostek a jednou jimi hodili: četnost výsledků v případech (a), (b) udává tabulka. Odhadněte, kolik z těchto kostek bylo prvního druhu.

hodnota	1	2	3	4	5	6
četnost (a)	3	3	4	4	2	4
četnost (b)	3	4	4	4	2	3

Řešení: Označme k počet kostek prvního druhu. Popsaný pokus můžeme považovat za náhodný výběr rozsahu 20 z rozdělení, které je směsí rozdělení pro oba druhy, a to s vahami $k/20$, $1 - k/20$.

Metoda momentů: Hledáme jeden parametr, zkusíme vystačit s prvním momentem, tj. střední hodnotou. Střední hodnota jednoho hodu je pro normální kostku 3.5, pro druhý druh 4, pro směs

$$\frac{k}{20}\,3.5 + \left(1 - \frac{k}{20}\right)\,4 = 4 - 0.025\,k\,.$$

(a) Rovnice

$$4 - 0.025\,\widehat{k} = 71/20 \doteq 3.55$$

má jediné řešení $\widehat{k} = 18$, a to vyhovuje.

(b) Rovnice

$$4 - 0.025\,\widehat{k} = 67/20 \doteq 3.35$$

má jediné řešení $\widehat{k} = 26$, které nevyhovuje zadání. Střední hodnota je v mezích $\langle 3.5, 4\rangle$, ale výběrový průměr může vyjít jakákoli hodnota v mezích $\langle 1, 6\rangle$. Metoda momentů v tomto případě žádný odhad nedává.

Metoda maximální věrohodnosti: V dané směsi rozdělení mají výsledky $1, 3, 5$ pravděpodobnost

$$\frac{k}{20}\,\frac{1}{6} = \frac{k}{120}$$

a výsledky $2, 4, 6$ pravděpodobnost

$$\frac{k}{20}\,\frac{1}{6} + \left(1 - \frac{k}{20}\right)\frac{1}{3} = \frac{1}{3} - \frac{k}{120} = \frac{40-k}{120}\,.$$

Stačí rozlišovat počet lichých a sudých výsledků. V případě (a) i (b) je 9 lichých a 11 sudých hodnot, takže pro oba případy dostaneme stejný odhad.

$$L\,(k) = \left(\frac{k}{120}\right)^9 \cdot \left(\frac{40-k}{120}\right)^{11}\,,$$

$$\ell\,(k) = \ln L\,(k) = 9\,\ln k + 11\,\ln(40-k) - 20\,\ln 120\,,$$

$$\frac{\partial}{\partial k}\ell\,(k) = \frac{9}{k} - \frac{11}{40-k}\,.$$

Tato derivace je nulová a věrohodnost maximální pro odhad $\widehat{k} = 18$.

V případě (a) daly obě metody stejný výsledek, ale to je náhoda, neboť výsledek metody maximální věrohodnosti záležel jen na počtu lichých a sudých výsledků, nikoli na výběrovém průměru. Proto tato metoda dala stejný výsledek i v případě (b), kdy metoda momentů selhala. □

Cvičení 11.3.3: Tabulka udává pravděpodobnosti a pozorované četnosti výsledků hodů nepravidelnou kostkou; odhadněte neznámé parametry a, b.

výsledek i	1	2	3	4	5	6
pravděpodobnost p_i	$a-b$	a	a	a	a	$a+b$
četnost n_i	2	5	4	3	6	10

Řešení: Součet pravděpodobností je 1, tedy $a = 1/6$. Zbývá určit $b \in \langle -1/6, 1/6\rangle$. Přitom $n = \sum_i n_i = 30$.

Metoda momentů:

$$\mathrm{E}X = \sum_i i\,p_i = \frac{7}{2} + 5\,b = 3\,\frac{1}{2} + 5\,b = \bar{x} = \frac{1}{n}\sum_i i\,n_i =$$

$$= \frac{2\cdot 1 + 5\cdot 2 + 4\cdot 3 + 3\cdot 4 + 6\cdot 5 + 10\cdot 6}{30} = \frac{21}{5} = 4.2\,,$$

$$b = \frac{7}{50} = 0.14 \in \left\langle -\tfrac{1}{6}, \tfrac{1}{6}\right\rangle\,.$$

Metoda maximální věrohodnosti:

$$\ell(b) = 2\,\ln\left(\tfrac{1}{6} - b\right) + (5+4+3+6)\,\ln\tfrac{1}{6} + 10\,\ln\left(\tfrac{1}{6} + b\right)\,,$$

$$0 = \frac{\partial}{\partial b}\ell(b) = \frac{-2}{\frac{1}{6} - b} + \frac{10}{\frac{1}{6} + b}\,,$$

$$b = \tfrac{1}{9} \in \left\langle -\tfrac{1}{6}, \tfrac{1}{6}\right\rangle\,.$$

□

Cvičení 11.3.4: Náhodná veličina nabývá hodnot $j \in \{0,1,2\}$ s pravděpodobnostmi $a + bj$. Odhadněte parametry a, b z četností dle tabulky:

hodnota	0	1	2
četnost	17	15	8

Řešení: Jelikož součet pravděpodobností musí být jednotkový, dostáváme rovnici $3\,a + 3\,b = 1$, tj. $b = 1/3 - a$. Zbývá odhadnout jediný parametr.

Metoda momentů: Střední hodnota je

$$(a+b) + 2\,(a+2\,b) = 3\,a + 5\,b = \frac{5}{3} - 2\,a\,,$$

její odhad – výběrový průměr

$$\frac{1}{17+15+8}\,(15 + 8\cdot 2) = \frac{31}{40}\,.$$

Rovnice $\frac{5}{3} - 2\,\widehat{a} = \frac{31}{40}$ má řešení $\widehat{a} = \frac{107}{240} \doteq 0.445\,833$, $\widehat{b} = \frac{-9}{80} \doteq -0.112\,5$.

Metoda maximální věrohodnosti:

$$\ell(a) = 17\,\ln a + 15\,\ln(a+b) + 8\,\ln(a+2b) = 17\,\ln a + 15\,\ln\tfrac{1}{3} + 8\,\ln\left(\tfrac{2}{3} - a\right)\,,$$

derivace

$$\frac{\partial}{\partial a}\ell(a) = \frac{17}{a} - \frac{8}{\frac{2}{3} - a}$$

je nulová pro $\widehat{a} = \frac{34}{75} \doteq 0.453\,333$, $\widehat{b} = \frac{-3}{25} = -0.12$. □

Následující klíčový výsledek řeší jednoduše řadu úloh, je však potřeba dát pozor na předpoklady:

Věta 11.3.5: *Pokud na* diskrétní *rozdělení* nejsou kladeny žádné omezující podmínky, *pak empirické rozdělení je jeho odhadem podle metody momentů i maximálně věrohodným odhadem.*

Důkaz: Označme X náhodnou veličinu s hledaným diskrétním rozdělením a $\boldsymbol{x} = (x_1, \ldots, x_n)$ realizaci náhodného výběru, na jejímž základě provedeme odhad parametrů rozdělení.

Za možné hodnoty rozdělení budeme považovat všechny ty, které se vyskytly v realizaci, a žádné jiné. (Vynechat samozřejmě žádnou hodnotu nesmíme, jinak by náš model pozorovanou realizaci nedovoloval. Ponecháváme na čtenáři ověření, že kdybychom uvažovali ještě další hodnoty, jejich pravděpodobnosti by při stejném postupu vyšly nulové.) Označme $u_1, \ldots, u_k$ $(k \leq n)$ hodnoty, které se vyskytly v realizaci $\boldsymbol{x}$; platí pro ně *množinová* rovnost

$$\{u_1, \ldots, u_k\} = \{x_1, \ldots, x_n\}\,.$$

Zbývá odhadnout pravděpodobnosti hodnot u_i, $i = 1, \ldots, k$:

$$q_i = p_X(u_i)\,.$$

Protože o $u_1, \ldots, u_k$ máme jasno, za hledané parametry považujeme pouze $q_1, \ldots, q_k \in \langle 0, 1\rangle$; jsou vázány jedinou podmínkou

$$1 - \sum_{i=1}^{k} q_i = 0\,.$$

Jelikož v realizaci nezáleží na pořadí, budeme pracovat s četnostmi. Označme n_i četnost a $r_i = n_i/n$ relativní četnost hodnoty u_i v realizaci $\boldsymbol{x}$. Připomeňme, že empirické rozdělení $\mathrm{Emp}(\boldsymbol{x})$ nabývá hodnot $u_1, \ldots, u_k$ s pravděpodobnostmi po řadě $r_1, \ldots, r_k$.

Metoda maximální věrohodnosti: Věrohodnost

$$L(q_1, \ldots, q_k) = \prod_{j=1}^{n} p_X(x_j) = \prod_{i=1}^{k} (p_X(u_i))^{n_i} = \prod_{i=1}^{k} q_i^{n_i}$$

je polynom, jehož maximum lze najít přímo, navzdory jeho vysokému stupni n. Pohodlnější však bude přechod k logaritmu

$$\ell(q_1, \ldots, q_k) = \sum_{i=1}^{k} n_i \ln q_i .$$

(Tento výraz není definován, je-li některé $q_i = 0$, jeho limita pro $q_i \to 0$ je $-\infty$. Vyloučením tohoto případu se zřejmě nepřipravujeme o maximum.) Máme najít maximum této funkce za podmínky jednotkového součtu neznámých; použijeme metodu Lagrangeových multiplikátorů (podrobnosti viz např. [Apl. mat. 1978]), tj. přičteme c-násobek této podmínky a hledáme globální maximum funkce

$$h(c, q_1, \ldots, q_k) = \sum_{i=1}^{k} n_i \ln q_i + c\left(1 - \sum_{i=1}^{k} q_i\right),$$

$$\frac{\partial}{\partial q_s} h(c, q_1, \ldots, q_k) = \frac{n_s}{q_s} - c .$$

Pro maximálně věrohodný odhad $\widehat{q}_1, \ldots, \widehat{q}_k$ budou tyto parciální derivace nulové pro všechna $s = 1, \ldots, k$, takže hodnota

$$\frac{n_s}{\widehat{q}_s} = c$$

bude nezávislá na s. Určíme ji z omezující podmínky

$$1 = \sum_{s=1}^{k} \widehat{q}_s = \frac{1}{c} \sum_{s=1}^{k} n_s = \frac{n}{c} ,$$

$$\frac{n_s}{\widehat{q}_s} = c = n .$$

Dostáváme

$$\widehat{q}_s = \frac{n_s}{n} = r_s$$

(pravděpodobnost empirického rozdělení) jako jediné řešení.

Metoda momentů: Pro $s = 1, \ldots, k$ vychází s-tý moment z teoretického modelu

$$u_i^s = \sum_{i=1}^{k} q_i u_i^s ,$$

jeho odhad z realizace $\boldsymbol{x}$ je

$$\frac{1}{n} \sum_{j=1}^{n} x_j^s = \frac{1}{n} \sum_{i=1}^{k} n_i u_i^s = \sum_{i=1}^{k} r_i u_i^s .$$

Hledáme odhad $\widehat{q}_1, \ldots, \widehat{q}_k$ (s trochou nepřesnosti označený stejně jako u předchozí metody, ač by obecně mohl vyjít jinak), pro který se oba výrazy rovnají, tj.

$$\sum_{i=1}^{k} \widehat{q}_i \, u_i^s = \sum_{i=1}^{k} r_i \, u_i^s \,,$$

pro všechna $s = 1, \ldots, k$. To je soustava k lineárních rovnic pro k neznámých $\widehat{q}_1, \ldots, \widehat{q}_k$. Řešením je zřejmě $\widehat{q}_i = r_i$, což odpovídá empirickému rozdělení. Je to jediné řešení, neboť matice soustavy

$$\begin{bmatrix} u_1^1 & u_2^1 & \cdots & u_k^1 \\ u_1^2 & u_2^2 & \cdots & u_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ u_1^k & u_2^k & \cdots & u_k^k \end{bmatrix}$$

je tzv. Vandermondova matice [Olšák 2007], která je regulární, právě když čísla $u_1, \ldots, u_k$ jsou navzájem různá. □

Poznámka 11.3.6: Za podmínek věty 11.3.5 je empirické rozdělení maximálně věrohodným odhadem i pro kvalitativní náhodné veličiny. Empirické rozdělení by vyšlo i metodou momentů, pokud bychom si nenumerické hodnoty očíslovali. Výsledek je nezávislý na tom, jaká čísla jsme hodnotám přiřadili. (Jediný požadavek je, že různým hodnotám musíme přiřadit různá čísla. Uděláme-li to nevhodně, může výpočet momentů být zatížen velkými numerickými chybami, ale teoretické řešení je stejné a momenty díky větě 11.3.5 ani nepotřebujeme počítat.) V tomto případě *výjimečně* můžeme použít metodu momentů k odhadu rozdělení kvalititativní náhodné veličiny, neboť výsledek nezáleží na zvoleném očíslování hodnot.

Poznámka 11.3.7: Odhad empirickým rozdělením je velmi přirozený, dospěl by k němu i laik „selským rozumem". Odborník navíc ví, kdy je tento postup korektní. Věta 11.3.5 platí pouze za předpokladu, že na diskrétní rozdělení nejsou kladena žádná omezení. Např. ji nelze použít v předchozích příkladech, kde hledaná rozdělení byla speciálních typů. Proto obě metody daly různé výsledky. Často bychom nepoužitelnost věty 11.3.5 poznali už z toho, že empirické rozdělení nemusí omezující podmínky splňovat.

Věta 11.3.5 dává snadné a nepřekvapivé řešení řady příkladů, které zde pro jejich jednoduchost neuvádíme. Odkazujeme na citovanou literaturu, kde je jim naopak věnováno překvapivě mnoho místa.

Poznámka 11.3.8: Snaha zobecnit větu 11.3.5 na diskrétní rozdělení s nekonečným (spočetným) počtem hodnot nemá praktický význam, neboť z konečně mnoha údajů nemůžeme odhadnout nekonečně mnoho parametrů.

Příklad 11.3.9: * Odhadněte parametry m, q binomického rozdělení $\mathrm{Bi}(m, q)$ (a) obecně, (b) na základě realizace náhodného výběru $\boldsymbol{x} = (1, 0, 3, 1, 2, 2, 5, 3, 3, 4)$.

Řešení: (a) Odhadujeme parametry na základě realizace náhodného výběru $\boldsymbol{x} = (x_1, \ldots, x_n)$. Nejprve vyšetřeme obor, v němž se mohou nalézat hodnoty parametrů. Zřejmě $q \in \langle 0, 1 \rangle$, zatímco $m \in \mathbb{N}$. Navíc musí být $m \geq \max_j x_j$, neboť jinak by se největší výsledek v realizaci nemohl vyskytnout.

Metoda momentů: Pro stanovení dvou neznámých parametrů můžeme porovnat první dva obecné momenty s jejich odhady a vyřešit vzniklou soustavu rovnic. (Podrobný postup lze najít v [Rogalewicz 2000].) Nemusí nám vyjít odhad q z intervalu $\langle 0, 1 \rangle$, ale zejména je velmi nepravděpodobné, že by pro m vyšel odhad celočíselný, jak je požadováno. Proto je nutno metodu momentů zamítnout jako nevhodnou.

Metoda maximální věrohodnosti: Věrohodnost

$$L\,(m, q) = \prod_{j=1}^{n} \binom{m}{x_j} q^{x_j} \, (1 - q)^{m - x_j}$$

by pro $q \in \{0,1\}$ byla nulová, *takže její maximum hledáme pro* $q \in (0,1)$ *a můžeme přejít k logaritmům:*

$$\ell(m,q) = \sum_{j=1}^{n} \ln \binom{m}{x_j} + \sum_{j=1}^{n} x_j \ln q + \sum_{j=1}^{n} (m - x_j) \ln(1-q) =$$

$$= \sum_{j=1}^{n} \ln \binom{m}{x_j} + n\,\bar{\boldsymbol{x}} \ln q + n\,(m - \bar{\boldsymbol{x}}) \ln(1-q)\,.$$

Pokud by vyšlo $\bar{\boldsymbol{x}} = 0$ (všechny hodnoty v realizaci nulové), pak maximum nastává pro $\widehat{q} = 0$ a $\widehat{m}$ libovolné. I když $\widehat{m}$ neurčíme, je $\mathrm{Bi}(\widehat{m}, 0)$ stále stejné rozdělení, totiž Diracovo v 0. *(Máme zde příklad toho, že více hodnot parametru může určovat stejné rozdělení. Záleží nyní na tom, zda cílem bylo určit rozdělení, což lze, nebo hodnotu parametru, což nelze.)*

V dalším předpokládáme $\bar{\boldsymbol{x}} > 0$. Pak bude parciální derivace

$$\frac{\partial}{\partial q}\ell(m,q) = n\left(\frac{\bar{\boldsymbol{x}}}{q} - \frac{m - \bar{\boldsymbol{x}}}{1-q}\right)$$

pro maximálně věrohodné odhady $\widehat{m}, \widehat{q}$ nulová,

$$\frac{\bar{\boldsymbol{x}}}{\widehat{q}} = \frac{\widehat{m} - \bar{\boldsymbol{x}}}{1 - \widehat{q}}\,,$$

řešením je

$$\widehat{q} = \frac{\bar{\boldsymbol{x}}}{\widehat{m}}\,,$$

což nepřekvapí, neboť pak je realizace výběrového průměru rovna střední hodnotě,

$$\bar{\boldsymbol{x}} = \widehat{q}\,\widehat{m}\,.$$

Zbývá určit odhad $\widehat{m} \in \mathbb{N}$, $\widehat{m} \geq \max_j x_j$, který maximalizuje výraz

$$\ell\left(\widehat{m}, \frac{\bar{\boldsymbol{x}}}{\widehat{m}}\right) = \sum_{j=1}^{n} \ln \binom{\widehat{m}}{x_j} + n\,\bar{\boldsymbol{x}} \ln \frac{\bar{\boldsymbol{x}}}{\widehat{m}} + n\,(\widehat{m} - \bar{\boldsymbol{x}}) \ln\left(1 - \frac{\bar{\boldsymbol{x}}}{\widehat{m}}\right) =$$

$$= \sum_{j=1}^{n} \ln \binom{\widehat{m}}{x_j} - n\,\widehat{m} \ln \widehat{m} + n\,(\widehat{m} - \bar{\boldsymbol{x}}) \ln(\widehat{m} - \bar{\boldsymbol{x}}) + n\,\bar{\boldsymbol{x}} \ln \bar{\boldsymbol{x}}\,.$$

I když poslední člen na $\widehat{m}$ nezávisí, je to obecně obtížné. Pro konkrétní zadání lze maximum najít numericky.

(b) Máme $\bar{\boldsymbol{x}} = 2.4$, $\max_j x_j = 5$, tedy $\widehat{m} \geq 5$. Složitost výpočtu všech uvedených vzorců je zhruba stejná, můžeme maximalizovat přímo věrohodnost $L\left(\widehat{m}, \frac{\bar{\boldsymbol{x}}}{\widehat{m}}\right) = g(\widehat{m})$. Podle očekávání má jediné maximum, pro menší hodnoty roste, pro větší klesá. Po zběžném průzkumu $g(5) = 1.156 \cdot 10^{-8}$, $g(10) = 2.16 \cdot 10^{-8}$, $g(18) = 2.173 \cdot 10^{-8}$, $g(31) = 2.126 \cdot 10^{-8}$ víme, že maximum je v intervalu $\langle 10, 31\rangle$. Další hodnoty nás dovedou k maximu: $g(23) = 2.152 \cdot 10^{-8}$, $g(15) = 2.185 \cdot 10^{-8}$, $g(13) = 2.187\,3 \cdot 10^{-8}$, $g(12) = 2.185 \cdot 10^{-8}$, $g(14) = 2.186\,8 \cdot 10^{-8}$. Až nyní můžeme *říci, že maximálně* věrohodný odhad je $\widehat{m} = 13$, $\widehat{q} = 2.4/13 \doteq 0.185$.

Volba kroků při numerickém řešení nebyla nahodilá; od chvíle, kdy jsme znali horní mez, byla optimální. Je založena na tom, že rozdíly hodnot jsou Fibonacciho čísla. Zdůvodnění lze najít např. v [Schlesinger, Hlaváč 1999]. □

Poznámka 11.3.10: Předchozí příklad ukázal několik typických aspektů aplikace metod odhadu parametrů na obtížnější úlohy. Mj. jsme viděli, že hodnoty věrohodnosti mohou být velmi malé i pro nevelké výběry. Také je obvyklé, že maximálně věrohodný odhad je potřeba dopočítat numerickými optimalizačními metodami a že je to velmi pracné. Pro řadu častých úloh byly vyvinuty speciální algoritmy, např. *EM algoritmus*. Některé takové úlohy z rozpoznávání jsou pěkně rozebrány v [Schlesinger, Hlaváč 1999].

Cvičení 11.3.11: Na základě tabulky četností

hodnota	VV	VM	MV	MM
četnost	20	20	10	5

odhadněte váhu w směsi $X = Mix_w(Y, Z)$, kde Y, Z jsou diskrétní náhodné veličiny s následujícími pravděpodobnostními funkcemi:

hodnota	VV	VM	MV	MM
p_Y	3/8	3/8	1/8	1/8
p_Z	3/8	1/8	3/8	1/8

Cvičení 11.3.12: Nezávislé náhodné veličiny X, Y mají alternativní rozdělení s následujícími pravděpodobnostmi:

i	0	1
$P[X = i]$	$1 - a$	a
$P[Y = i]$	$1 - b$	b

Odhadněte parametry a, b na základě realizace veličiny $Z = X + Y$, u níž jsme pozorovali následující četnosti:

hodnota Z	0	1	2
četnost	10	5	0

Cvičení 11.3.13: Odhadněte parametr w geometrického rozdělení

$$p_i = (1 - w)\, w^i\,, \qquad i = 0, 1, 2, \ldots$$

na základě realizace s následujícími četnostmi výsledků:

hodnota	0	1	2	3
četnost	20	10	7	3

Cvičení 11.3.14: Odhadněte parametry a, b rozdělení náhodné veličiny na základě teoretických pravděpodobností a pozorovaných četností dle tabulky:

hodnota	0	1	2
pravděpodobnost	b	a	a^2
četnost	30	30	30

Odhady spojitých rozdělení

Tvrzení 11.3.15: *Pro normální rozdělení* $\mathrm{N}(\mu, \sigma^2)$ *je odhad parametrů* μ *a* $r = \sigma^2$ *na základě realizace náhodného výběru* $\boldsymbol{x} = (x_1, \ldots, x_n)$ *podle metody momentů i metody maximální věrohodnosti stejný, a to*

$$\widehat{\mu} = \bar{\boldsymbol{x}} = \frac{1}{n} \sum_{j=1}^{n} x_j\,, \qquad \widehat{r} = \widehat{\sigma}^2_X = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}})^2\,. \tag{11.1}$$

(Výsledek $\widehat{\sigma}^2_X$ *je realizace odhadu ze vzorce (9.4).)*

Důkaz: Označme X náhodnou veličinu s hledaným rozdělením $\mathrm{N}(\mu, \sigma^2)$. Jedním z parametrů je rozptyl $r = \sigma^2$, tento zápis by se nám mohl plést, až budeme podle tohoto parametru derivovat, proto výjimečně použijeme jednoduchý symbol r. (Mohli bychom také za parametr považovat směrodatnou odchylku σ a došli bychom ke stejným závěrům s o něco větším úsilím.)

Metoda momentů: Pro stanovení dvou parametrů použijeme první dva obecné momenty,

$$\mathrm{E}X = \mu\,, \qquad \mathrm{E}X^2 = (\mathrm{E}X)^2 + \mathrm{D}X = \mu^2 + \sigma^2 = \mu^2 + r\,,$$

které položíme rovny příslušným výběrovým momentům. Řešením budou hodnoty $\widehat{\mu}, \widehat{r}$, splňující rovnice

$$\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_j\,,$$

$$\widehat{\mu}^2 + \widehat{r} = \frac{1}{n}\sum_{j=1}^{n} x_j^2\,.$$

Odtud $\widehat{\mu} = \bar{x}$,

$$\widehat{r} = \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \widehat{\mu}^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \bar{x}^2\,,$$

což je požadovaný výsledek, neboť

$$\widehat{\sigma}_X^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \frac{2}{n}\,\bar{x}\sum_{j=1}^{n} x_j + \frac{1}{n}\sum_{j=1}^{n}\bar{x}^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2 - 2\,\bar{x}^2 + \bar{x}^2 =$$

$$= \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \bar{x}^2\,.$$

Metoda maximální věrohodnosti:

$$\Lambda(\mu, r) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\,\pi\,r}}\,\exp\left(\frac{-(x_j-\mu)^2}{2\,r}\right),$$

$$\lambda(\mu, r) = \ln\Lambda(\mu, r) = \frac{-1}{2\,r}\sum_{j=1}^{n}(x_j-\mu)^2 - \frac{n}{2}\ln r - \frac{n}{2}\ln 2\,\pi\,,$$

$$\frac{\partial}{\partial\mu}\lambda(\mu, r) = \frac{1}{r}\sum_{j=1}^{n}(x_j-\mu) = \frac{1}{r}\left(\sum_{j=1}^{n} x_j - n\,\mu\right) = \frac{n}{r}\,(\bar{x}-\mu)\,,$$

$$\frac{\partial}{\partial r}\lambda(\mu, r) = \frac{1}{2\,r^2}\sum_{j=1}^{n}(x_j-\mu)^2 - \frac{n}{2\,r} = \frac{n}{2\,r^2}\left(\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2 - r\right).$$

Maximum nastává pro hodnoty z (11.1). □

Poznámka 11.3.16: Na právě uvedeném výsledku je zvláštní to, že jsme oběma metodami dostali pro rozptyl vychýlený odhad ze vzorce (9.4), nikoli výběrový rozptyl, který je nestranný. Z toho je vidět, že oba odhady rozptylu mají své opodstatnění.

Cvičení 11.3.17: Náhodná veličina X má *biexponenciální* rozdělení s hustotou

$$f_X(t) = c\,\exp(-a\,|t|)\,,$$

kde $a > 0$. *(Toto rozdělení se používá např. pro popis jasu fotografie.)* Metodou momentů odhadněte parametry na základě realizace rozsahu $n = 10$, z níž jsme vypočítali $\bar{x} = 2$, $s_x^2 = 4$. Můžete využít vztahy (C.8).

Řešení: Z podmínky

$$\int_{-\infty}^{\infty} f_X(t)\,\mathrm{d}t = 2\,c\int_0^{\infty}\exp(-a\,t)\,\mathrm{d}t = 2\,c\,\frac{\exp(-a\,t)}{-a}\bigg|_{t=0}^{\infty} = \frac{2\,c}{a} = 1$$

dostáváme $c = a/2$.

První moment je teoreticky 0 a nic nám o parametrech neříká, použijeme výběrový 2. obecný moment,

$$\int_{-\infty}^{\infty} t^2 f_X(t)\,\mathrm{d}t = a\int_{-\infty}^{\infty} t^2 \exp(-a\,t)\,\mathrm{d}t = \frac{1}{a^2}\int_0^{\infty} y^2 \exp(-y)\,\mathrm{d}y = \frac{2}{a^2},$$

který položíme rovný realizaci 2. obecného momentu,

$$m_2 = \frac{1}{n}\sum_i x_i^2 = \frac{1}{n}\left(\sum_i (x_i - \bar{x})^2 + \sum_i \bar{x}^2\right) = \frac{n-1}{n}\,s_{\boldsymbol{x}}^2 + \bar{x}^2 = \frac{9}{10}\,4 + 2^2 = \frac{38}{5}.$$

Z rovnice

$$\frac{2}{a^2} = \frac{38}{5}$$

dostaneme kladné řešení

$$a = \frac{1}{19}\sqrt{95} = 0.512\,99, \qquad c = \frac{a}{2} = \frac{1}{38}\sqrt{95} = 0.256\,49.$$

Pokud bychom odhadli 2. obecný moment výrazem

$$m_2 = s_{\boldsymbol{x}}^2 + \bar{x}^2 = 4 + 2^2 = 8$$

(což není nestranný odhad), pak bychom z rovnice $\frac{2}{a^2} = 8$ *dostali kladné řešení* $a = \frac{1}{2}$, $c = \frac{a}{2} = \frac{1}{4}$. □

12. Testování hypotéz

12.1 Základní pojmy a principy testování hypotéz

(Doporučená literatura: [Jaroš 1998].)

Vyložíme obecnou metodiku testování hypotéz na jednodušším speciálním případě. Zde použité značení a terminologie jsou standardními dorozumívacími prostředky v této oblasti.

Příklad 12.1.1: V r. 2006 v ČR vzbudilo rozruch výběrové řízení s následujícím průběhem: Z 16 přihlášených firem bylo v 5 kolech (pro 5 zakázek) vylosováno do užšího výběru vždy 5 firem, jejichž nabídky byly dále posuzovány. Jedna z firem se zúčastnila prostřednictvím čtyř filiálek, tj. 4 z 16 podaných přihlášek byly od jejích dceřinných společností. Žádná z nich nebyla vybrána v žádném z 5 kol. Samozřejmě je takový výsledek losování možný, nicméně je velmi nepravděpodobný. Firma proto zažalovala zadavatele, neboť považuje losování za zmanipulované.[1]

Formulace otázky a rizika chyb

Máme posoudit hypotézu o hodnotě nějakého parametru ϑ spojitého rozdělení. Použijeme k tomu **testovací statistiku** neboli **kritérium** T, které na parametru závisí. Předpokládejme nejprve pro jednoduchost, že parametr ϑ nabývá pouze dvou hodnot, 0 pro „normální" populaci, 1 pro „anomální" prvky. Otázka je, z které z těchto skupin je proveden náhodný výběr $\boldsymbol{X}$. Máme rozhodnout o hodnotě parametru, tj. mezi dvěma hypotézami:

- **Nulová hypotéza** H_0: Výběr je z normální populace.
- **Alternativní hypotéza** H_1: Výběr je z anomální populace.

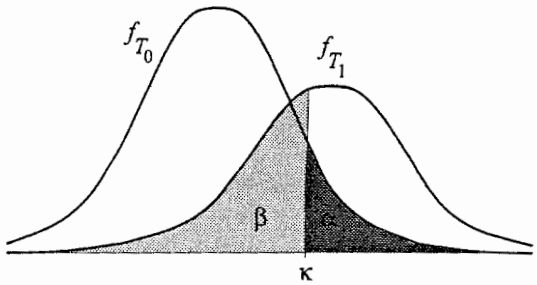
Předpokládejme, že obě skupiny mají známá rozdělení, která se liší středními hodnotami μ_0, μ_1, kde $\mu_0 < \mu_1$, což použijeme k testu. Dále předpokládáme, že na intervalu (μ_0, μ_1) je hustota rozdělení normální populace nerostoucí a hustota rozdělení anomální skupiny neklesající (obr. 12.1). Za kritérium použijeme (ne nutně) výběrový průměr $T = \overline{\boldsymbol{X}}$. Zvolíme **kritickou hodnotu** $\kappa \in (\mu_0, \mu_1)$ a výběr vyhodnotíme jako normální (klasifikujeme 0) pro $T \leq \kappa$, jako anomální (klasifikujeme 1) pro $T > \kappa$. Kromě správných rozhodnutí hrozí následující chyby:

1. **Chyba prvního druhu** (angl. *type I error*): normální skupina je klasifikována jako anomální, což nastane s pravděpodobností $\alpha(\kappa)$.
2. **Chyba druhého druhu** (angl. *type II error*): anomální skupina je klasifikována jako normální, což nastane s pravděpodobností $\beta(\kappa)$.

Příklad 12.1.2: Máme zastavit používání léku pro podezření z nežádoucích účinků? Uvedené pojmy mají následující významy:

- Nulová hypotéza H_0: Výrobce je nevinen, riziko se nezvyšuje.
- Alternativní hypotéza H_1: Výrobce je vinen, riziko se zvyšuje.
- Chyba prvního druhu: Zamítneme nulovou hypotézu, která platí (obviníme nevinného).
- Chyba druhého druhu: Nezamítneme nulovou hypotézu, která neplatí (osvobodíme vinného).

[1] Informace laskavě poskytl Prof. RNDr. J. Štěpán, DrSc., který se spolupracovníky vypracoval znalecký posudek a referoval o něm v televizi.



Obrázek 12.1. Pravděpodobnost chyby prvního a druhého druhu (α a β)

Jestliže z 10 000 uživatelů léku zemřelo do roka 171 pacientů a v kontrolní skupině 10 000 lidí 170, pak to neznamená, že lék má na svědomí život jednoho pacienta. Kdyby však zemřelo 1700 pacientů, bylo by to důvodem k zákazu léku. Otázka je, kde mezi těmito extrémními případy máme stanovit mez, od níž budeme považovat výsledky za (statisticky) významné a škodlivost za prokázanou.

Volbou kritické hodnoty (přísnosti kritéria) snižujeme riziko jedné chyby na úkor zvýšení rizika druhé chyby. Škody způsobené chybami různých druhů mohou být zásadně odlišné. *Jediný způsob, jak snížit pravděpodobnost obou typů chyb, je zvětšit rozsah výběru* (pokud to situace dovoluje).

V našem příkladu je funkce α nerostoucí a β neklesající v závislosti na κ, což dovoluje navrhnout mnoho kritérií pro volbu meze κ, např.:

- rovnost pravděpodobností obou chyb, $\alpha(\kappa) = \beta(\kappa)$,
- minimalizace součtu pravděpodobností chyb $\alpha(\kappa) + \beta(\kappa)$,
- minimalizace funkce pravděpodobností chyb $e(\alpha(\kappa), \beta(\kappa))$ (tzv. **výplatní funkce**, která zohledňuje odlišné škody způsobené jednotlivými typy chyb, např. $a\,\alpha(\kappa) + b\,\beta(\kappa)$),
- $\alpha(\kappa)$ rovno předem zvolené malé hodnotě.

Většinou se používá poslední možnost. Je výpočetně méně náročná (vede na snazší úlohu). Hlavně však pro její uplatnění nepotřebujeme znát rozdělení anomální skupiny, které bychom někdy těžko definovali.

Příklad 12.1.3: Můžeme poměrně snadno určit rozdělení krevního tlaku u zdravých lidí. Je však těžké stanovit, jaké je rozdělení u „osob s vysokým krevním tlakem".

Obvykle máme více než dvě možné hodnoty parametru a rizika chyb na nich závisí, což test komplikuje. I proto se osvědčuje následující jednoduché východisko, které se stalo standardním postupem v testování hypotéz. Kritickou hodnotu testu κ stanovíme takovou, aby chyba prvního druhu nastávala se stanovenou pravděpodobností $\alpha \in \mathbb{R}$ zvanou **hladina významnosti**, též **hladina testu** (nebo s menší pravděpodobností, nelze-li dosáhnout rovnost). Jestliže nulové *hypotéze vyhovuje více hodnot parametru*, požadujeme splnění této nerovnosti při *všech* z nich, tj. *při nejnepříznivějších podmínkách.*

Poznámka 12.1.4: Pro srovnání výsledků různých výzkumů je žádoucí, aby používaly pokud možno stejnou hladinu významnosti. Podle tradice v příslušném oboru se nejčastěji užívají hladiny významnosti 1 % nebo 5 % (rozhodně mnohem menší než 1/2). Hladina významnosti 5 % je obvyklá v oborech, kde je omezená opakovatelnost pokusů a mnoho neodstranitelných rušivých vlivů, jako v lékařství, biologii, sociologii, ekonomii apod. Tam se výsledky na hladině významnosti 1 % dosahují zřídka a zmiňují jako obzvlášť vydařené. Naproti tomu v technických oborech, kde často lze experimenty mnohokrát opakovat za dobře definovaných podmínek, se používá hladina významnosti 1 % a někdy i menší, pokud je prověření zvlášť důležité (např. z důvodů spolehlivosti).

Vyhodnocení testu

Hodnoty kritéria přesahující kritickou hodnotu (odpovídající výsledkům málo pravděpodobným při platnosti nulové hypotézy) považujeme za **statisticky významné** a v tom případě *nulovou hypotézu zamítáme*. V opačném případě řekneme, že *nulovou hypotézu nezamítáme*. Někdy se říká, že ji „přijímáme", což budí dojem, že jsme dokázali její platnost. To není přesné, neboť připouštíme, že ji zamítne někdo jiný na základě důkladnějších testů, ale my jsme pro to neshledali důvod. Rozhodně by se neměla používat matoucí formulace (se kterou se lze v literatuře setkat), že nulovou hypotézu „potvrzujeme". To statistika ani nemůže.

Jestliže nezamítáme nulovou hypotézu, pak tím *nezamítáme ani alternativní hypotézu*. Dochází zde k nesymetrii ve vyhodnocení testu. Nulovou hypotézu zamítáme nebo nezamítáme. K alternativní hypotéze se nevyjadřujeme. Tato nesymetrie má původ v nesymetrickém kritériu pro nastavení kritické hodnoty – je dáno pravděpodobností chyby prvního druhu bez ohledu na chyby druhého druhu.

Příklad 12.1.5: V době vydání tohoto skripta je k dispozici řada studií o škodlivosti záření mobilních telefonů pro lidský organismus. Nulová hypotéza „mobilní telefony neškodí" většinou nebyla zamítnuta. Jen ojedinělé výzkumy ji zamítají, což se při jejich velkém počtu přirozeně stává. (Při hladině významnosti 5 % lze očekávat, že zhruba 5 % testů nulovou hypotézu zamítne, i kdyby platila.) Nikdy nebude možno dokázat, že mobilní telefony neškodí (potvrdit nulovou hypotézu). Pokud se škodlivost mnoha výzkumy nepodaří zamítnout, pak ji přijmeme v tom smyslu, že se budeme chovat, jako by mobilní telefony neškodily. Rozhodně ji nelze experimentálně potvrdit. Proto nezamítáme ani alternativní hypotézu, že mobilní telefony škodí. Vždy zůstává otevřená možnost, že pozdější výzkumy nějakou formu škodlivosti prokáží. (Otázka mimo rámec statistiky je, zda případná škodlivost bude shledána tak závažnou, aby ovlivnila naše jednání.)

Naopak, pokud bude nulová hypotéza zamítnuta (a to nejen výjimečně), pak bude škodlivost považována za prokázanou.

Příklad 12.1.6: Losovací zařízení pro loterie musí projít náročným testováním. Kdyby se ukázalo, že výsledky nejsou stejně pravděpodobné, loterie by nebyla povolena. Když v testech uspěje, přesto nelze dokázat, že rozdělení výsledků je rovnoměrné.

Poznámka 12.1.7: Statistická významnost neznamená významnost praktickou. Shromážděním velkého počtu údajů bychom např. mohli prokázat *statisticky významný* rozdíl v cenách benzínu u různých prodejců o 5 haléřů na litr. Takový rozdíl ale *není podstatný* pro naše rozhodování. Podobně může v budoucnu dopadnout i problém škodlivosti mobilních telefonů z příkladu 12.1.5.

Poznámka 12.1.8: Naopak při nedostatečném rozsahu souboru se může stát, že ani *podstatný* rozdíl hodnot *není statisticky významný* (nepostačuje k zamítnutí nulové hypotézy).

Slovníček

Pro srozumitelnost zde vystačíme s minimem termínů, nicméně pro porozumění jiným textům uvádíme další související pojmy. Často se jimi místo skutečných pravděpodobností (často neznámých) označují jejich *odhady* založené na experimentech s kontrolní skupinou, což pro jednoduchost nebudeme rozlišovat ani zde.

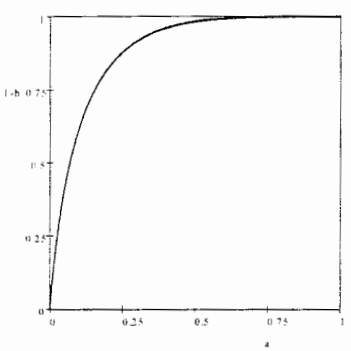
Pravděpodobnost α chyby prvního druhu se označuje FPR (angl. *false positive rate*), pravděpodobnost β chyby druhého druhu FNR (angl. *false negative rate*). Správné klasifikace jsou anglicky označovány jako *true positives* a *true negatives*. Číslo $1 - \alpha$ je **specificita testu** (angl. *specificity*), $1 - \beta$ je **síla testu** neboli **senzitivita** (angl. *power*, *sensitivity*, *recall*, *hit rate*).

Předchozí pojmy jsou nezávislé na tom, s jakou apriorní pravděpodobností se vyskytují normální a anomální případy. Pokud toto zohledníme (jako v bayesovském rozhodování), pak definujeme **přesnost testu** (angl. *precision*, *positive predictive value*) jako poměr správně klasifikovaných pozitivních případů ke všem pozitivním. Podobně *negative predictive value* je poměr správně klasifikovaných negativních případů ke všem negativním. Ve snaze vyjádřit obě kvality se zavádí *F-measure* jako harmonický průměr přesnosti a senzitivity:

$$\frac{1}{\frac{1}{2}\left(\frac{1}{přesnost}+\frac{1}{senzitivita}\right)}$$

Příklad 12.1.9: V příkladu 2.5.17 jsme měli test nemoci, který je u 1 % zdravých falešně pozitivní (α, FPR) a u 10 % nemocných falešně negativní (β, FNR). Specificita je $1-\alpha = 99\,\%$, senzitivita (síla testu) $1-\beta = 90\,\%$. Za předpokladu, že podíl nemocných v populaci je 0.001, jsme odvodili, že přesnost testu je 8.26 %.

Problém volby kritické hodnoty ozřejmí křivka, která zobrazuje závislost pravděpodobnosti chyby prvního druhu (vodorovně) a síly testu (svisle), přičemž parametrem křivky je kritická hodnota. Tato křivka se v angličtině nazývá *receiver operating characteristic*, zkráceně *ROC curve*. Pro rostoucí kritickou hodnotu probíhá obvykle z bodu $(0,0)$ do bodu $(1,1)$ nad osou prvého kvadrantu. (Kdyby byla pod touto osou, bylo by lépe klasifikovat opačně.) Je žádoucí, abychom volbou kritické hodnoty vybrali na této křivce bod co nejblíže bodu $(0,1)$, který reprezentuje bezchybnou klasifikaci. Nicméně při obvyklé metodice prostě vybereme bod, v němž se pravděpodobnost chyby prvního druhu rovná zvolenému číslu (tj. s danou vodorovnou souřadnicí). Tyto křivky dovolují graficky porovnat účinnost různých testů.[2] Typický průběh je na obr. 12.2.



Obrázek 12.2. Typický průběh *ROC* křivky

Formulace hypotéz

V literatuře se rozlišuje

- **jednoduchá hypotéza**: nulové hypotéze odpovídá jediná hodnota parametru,
- **složená hypotéza**: nulové hypotéze odpovídá více hodnot parametru,

a dále

- **jednoduchá alternativa**: alternativní hypotéze odpovídá jediná hodnota parametru,
- **složená alternativa**: alternativní hypotéze odpovídá více hodnot parametru.

U složené hypotézy požadujeme, aby pravděpodobnost chyby prvního druhu byla nejvýše α pro všechny hodnoty parametru vyhovující nulové hypotéze. V úvodu jsme uvažovali jednoduchou hypotézu a jednoduchou alternativu.

Příklad 12.1.10: Testujeme, zda lék změnil hodnoty krevního tlaku skupiny pacientů. Jednoduchá hypotéza a složená alternativa: H_0: tlak je stejný jako dříve, H_1: tlak se změnil. Složená hypotéza a složená alternativa: H_0: tlak se nezvýšil, H_1: tlak se zvýšil.

[2] Poprvé se začaly používat za 2. světové války k vyhodnocování radarových signálů.

Nulová a alternativní hypotéza se rozhodně musí navzájem vylučovat. Často se formulují tak, že nejsou navzájem svými negacemi a nepokrývají prostor všech možných hodnot parametru. Vzniká tím chaos (viz většina ostatní literatury). Snadno se mu vyhneme (a zpravidla o nic nepřijdeme), když budeme formulovat nulovou hypotézu jako negaci alternativní hypotézy.

Je-li např. $H_1 : \vartheta > c$, pak nevolíme $H_0 : \vartheta = c$, ale $H_0 : \vartheta \leq c$. (Největší riziko chyby prvního druhu obykle odpovídá případu $\vartheta = c$, takže postup je stejný.)

Při zde probíraných testech hypotéz vystačíme s výsledky z kapitoly 3.7. Nulovou hypotézu zamítneme, právě když hodnota kritéria získaná z realizace nepadne do intervalu spolehlivosti pro koeficient spolehlivosti $1 - \alpha$. To znamená, že kritická hodnota je mezí intervalového odhadu.

Obráceně se můžeme ptát, při jaké mezní hladině významnosti by pozorovaná hodnota byla kritická; tomu říkáme **dosažená významnost**, též **dosažená hladina testu** (angl. *significance*, *P-value*). Čím nižší dosažená významnost, tím významnější výsledek. Pro vyhodnocení statistické významnosti stačí porovnat dosaženou významnost s předem zvolenou hladinou významnosti testu. Téměř všechny programy dávají za výsledek dosaženou významnost (obvykle se značí P), takže hladinu významnosti není třeba předem zadat; navíc se dovíme, jak daleko jsme byli od této hladiny. (Např. hodnota 0.012 říká, že nemůžeme zamítnout nulovou hypotézu na hladině významnosti 1 %, ale chybělo k tomu málo.) Tento postup dříve nepoužíval hlavně proto, že vyžaduje stanovení hodnot distribučních funkcí příslušných rozdělení, k čemuž by byly potřeba mnohem rozsáhlejší statistické tabulky. Místo toho je v tabulkách jen několik používaných kvantilů. S rozvojem výpočetní techniky tato výhoda ztratila na významu, neboť numerické algoritmy jsou zhruba stejně náročné pro oba přístupy.

Typický tvar testu

Testovací statistiku T se známým rozdělením (přesněji její realizaci t) porovnáváme s kvantily příslušného rozdělení a zamítneme při extrémních hodnotách (nepravděpodobných při platnosti nulové hypotézy):

H_0	H_1	zamítáme pro	dosažená významnost
$\vartheta \leq c$	$\vartheta > c$	$t > q_T(1-\alpha)$	$1 - F_T(t)$
$\vartheta \geq c$	$\vartheta < c$	$t < q_T(\alpha)$	$F_T(t)$
$\vartheta = c$	$\vartheta \neq c$	$t > q_T(1-\alpha/2)$ nebo $t < q_T(\alpha/2)$	$2\min(F_T(t), 1 - F_T(t))$

V literatuře se setkáme i s následujícími případy hypotéz, které se však řeší stejně jako první dva výše uvedené:

H_0	H_1
$\vartheta = c$	$\vartheta > c$
$\vartheta = c$	$\vartheta < c$

Poznámka 12.1.11: Často by si uživatel přál testovat hypotézy $\vartheta > c$ nebo $\vartheta < c$. Mohl by se dokonce ptát, zda $\vartheta \neq c$, tedy jestli se zkoumaný parametr *liší* od dané hodnoty. Takovou podmínku nelze zadat jako nulovou hypotézu. Musíme totiž vycházet z toho, jaký model *je*, nikoli, jaký *není*. Z parametrů vyhovujících nulové hypotéze se odvozuje rozdělení testovací statistiky a její kritické hodnoty. Kdybychom nulovou hypotézou řekli, že $\vartheta \neq c$, tj. $\vartheta \in (-\infty, c) \cup (c, \infty)$, nebylo by to nic platné. (Alternativní hypotéza by byla $\vartheta = c$.) Pokud testovací statistika závisí na parametru ϑ spojitě, pak její alternativní hodnota pro $\vartheta = c$ je byla limitou případů, které vyhovují nulové hypotéze, a není možné obě hypotézy tímto kritériem oddělit.

Nezbývá tedy než volit $\vartheta = c$ za nulovou hypotézu a $\vartheta \neq c$ za alternativní. To má za následek, že možná budeme moci zamítnout hypotézu $\vartheta = c$, ale *nikdy* nebudeme moci zamítnout hypotézu $\vartheta \neq c$. Má to svoji logiku: přesnou rovnost (reálných čísel) nelze prakticky dokázat, protože hodnoty se mohou lišit o méně, než dovoluje odhalit přesnost naší metody. Z téhož důvodu nelze testovat nulové hypotézy $\vartheta > c$ nebo $\vartheta < c$.

Obecný tvar testu

Za rámcem této učebnice zůstává situace, s níž se ale čtenář může snadno v budoucnu setkat: hodnoty parametru nejsou uspořádané.

Příklad 12.1.12: Z obrázku máme posoudit hypotézu o tom, z kterého místa byl pořízen. Omezíme se na směr pohledu, který popíšeme bodem na jednotkové sféře v trojrozměrném prostoru. Jak má vypadat oblast, pro niž zamítáme nulovou hypotézu?

V takových situacích nám intervalový odhad nepomůže. Východiskem bude pravděpodobnost realizace v závislosti na parametru, tedy *věrohodnost*, viz kapitola 11.2. Zaměříme se zde pouze na případ spojitého rozdělení, kdy věrohodnost je určena *hustotou* rozdělení v závislosti na parametru.

Pokud známe *apriorní* pravděpodobnosti (nebo hustoty) hodnot parametru, pak můžeme použít aparát podmíněných pravděpodobností a *bayesovské rozhodování*. Pokud o rozdělení parametrů nic nevíme, východiskem pro rozhodování bude věrohodnost. Formulujeme nulovou a alternativní hypotézu. Poté vymezíme tzv. **kritickou oblast** testu, tj. množinu výsledků, při nichž zamítáme nulovou hypotézu. V této oblasti by se měl výsledek testu nacházet s malou pravděpodobností (hladina významnosti) při platnosti nulové hypotézy a s velkou pravděpodobností (síla testu) při platnosti alternativní hypotézy.

Návod nám dává tzv. **Neymanovo-Pearsonovo lemma**. Podle něj pro úvodní úlohu, kdy parametr může nabývat jen dvou hodnot, docílíme nejsilnější test při dané hladině významnosti α volbou takové kritické oblasti, jejíž pravděpodobnost za platnosti nulové hypotézy je α a k níž existuje takové $c \in \mathbb{R}$, že poměr věrohodností odpovídajících alternativní a nulové hypotéze je větší nebo roven c právě v této oblasti. Kritická oblast je tedy charakterizována *poměrem věrohodností* (angl. *likelihood ratio*) obou případů. To dává návod pro její nalezení.

Předchozí typický tvar testu je velmi speciálním případem tohoto obecnějšího principu, nicméně pokrývá řadu běžně užívaných metod. Obdobně se navrhují i testy pro reálné parametry, pokud hustoty mají více než jedno maximum.

12.2 Testy střední hodnoty a rozptylu

V těchto testech použijeme již dříve odvozené intervalové odhady, proto uvádíme postupy bez podrobného zdůvodnění.

12.2.1 Testy střední hodnoty normálního rozdělení

Při *známém* rozptylu σ^2

Příklad 12.2.1: Máme podezření, že měřicí přístroj se známou přesností má systematickou chybu větší než povolenou. Jak to prokážeme?

Testovací statistiku (kritérium)

$$T = \frac{\overline{X} - c}{\sigma}\sqrt{n}$$

porovnáváme s kvantily *normovaného normálního rozdělení*:

H_0	zamítáme pro	dosažená významnost
$\mu \leq c$	$t > \Phi^{-1}(1-\alpha)$	$1 - \Phi(t)$
$\mu \geq c$	$t < -\Phi^{-1}(1-\alpha) = \Phi^{-1}(\alpha)$	$\Phi(t)$
$\mu = c$	$\lvert t\rvert > \Phi^{-1}(1-\alpha/2)$	$2\min(\Phi(t), 1-\Phi(t))$

Při *neznámém* rozptylu

Příklad 12.2.2: Máme podezření, že měřicí přístroj s neznámou přesností má systematickou chybu větší než povolenou. Jak to prokážeme?

Testovací statistiku

$$T = \frac{\overline{X} - c}{S_X} \sqrt{n}$$

porovnáváme s kvantily *Studentova rozdělení* s $n-1$ stupni volnosti:

H_0	zamítáme pro	dosažená významnost
$\mu \leq c$	$t > q_{t(n-1)}(1-\alpha)$	$1 - F_{t(n-1)}(t)$
$\mu \geq c$	$t < -q_{t(n-1)}(1-\alpha)$	$F_{t(n-1)}(t)$
$\mu = c$	$\lvert t\rvert > q_{t(n-1)}(1-\alpha/2)$	$2 \min\left(F_{t(n-1)}(t), 1 - F_{t(n-1)}(t)\right)$

Cvičení 12.2.3: Měřením $n = 10$ zdrojů stejnosměrného napětí o nominální hodnotě $U_n = 5\ V$ jsme obdrželi výběrový průměr $\overline{X} = 5.3\ V$ a výběrovou směrodatnou odchylku $S_X = 0.3\ V$. Předpokládáme, že jejich chyby jsou nezávislé a mají normální rozdělení; chyby měření zanedbáváme. Posuďte na hladině významnosti 1 % hypotézu, že střední hodnota napětí zdrojů je rovna jejich nominální hodnotě. Posuďte adekvátnost předpokladů.

Tři takové zdroje napětí spojíme do série. Odhadněte rozdělení výsledného napětí.

V dalším pokusu jeden ze zdrojů připojíme k rezistoru $R = 10\ \Omega$. Odhadněte rozdělení a střední hodnotu výkonu.

Řešení: Testovací statistiku $T = \frac{\overline{X}-U_n}{S_X}\sqrt{n} = \sqrt{10} \doteq 3.162\,3$ porovnáme s kvantily $q_{t(n-1)}(0.995) \doteq 3.2498$, $q_{t(n-1)}(0.005) = -q_{t(n-1)}(0.995) \doteq -3.2498$ a hypotézu *nezamítáme.*

Rozhodně je problematický předpoklad nezávislosti chyb; mohou být závislé jak kvůli systematickým chybám ve výrobě, tak vlivem společných vnějších podmínek.

Řešení: Pro napětí U_i z i-tého zdroje použijeme nestranné odhady $\widehat{\mathrm{E}U_i} = \overline{X} = 5.3\,V$ a $\widehat{\mathrm{D}U_i} = S_X^2 = 0.09\,V^2$, tedy U_i/V (kde V značí volt) má normální rozdělení odhadnuté jako $\mathrm{N}\left(\overline{X}, S_X^2\right) = \mathrm{N}\,(5.3, 0.09)$. Součet 3 nezávislých náhodných veličin s tímto rozdělením má rozdělení $\mathrm{N}\left(3\,\overline{X}, 3S_X^2\right) = \mathrm{N}\,(15.9, 0.27)$. □

Rozdělení veličiny U_i/V lze ekvivalentně popsat jako $\overline{X} + S_X\,C_i = 5.3 + 0.3\,C_i$, kde C_i má $\mathrm{N}\,(0,1)$. Pro výkon ve wattech, P_i/W, dostáváme

$$P_i = \frac{U_i^2}{R} \qquad (R \text{ je konstanta, } U_i^2 \text{ je náhodná veličina}),$$

$$P_i/W = \frac{(5.3 + 0.3\,C_i)^2}{10} \doteq 2.809 + 0.318\,C_i + 0.009\,C_i^2$$
$$= 2.809 + 0.318\,F + 0.009\,G\,,$$

kde $F = C_i$ má rozdělení $\mathrm{N}\,(0,1)$, $G = C_i^2$ má rozdělení χ^2 s 1 stupněm volnosti (podle definice rozdělení χ^2) a F, G jsou nezávislé (díky nezávislosti statistik $\overline{X}, S_X^2$); jejich střední hodnoty jsou 0 a 1, takže

$$\mathrm{E}P_i \doteq 2.809 + 0.009 = 2.818\ W\,.$$

Alternativní řešení třetí části bez druhé: Náhodná veličina U_i^2 má střední hodnotu

$$\mathrm{E}U_i^2 = (\mathrm{E}U_i)^2 + \mathrm{D}U_i\,,$$

oba sčítance nahradíme nestrannými odhady

$$\widehat{EU_i^2} = \overline{X}^2 + S_X^2 = 5.3^2 + 0.09 = 28.18\ V\,,$$
$$\widehat{EP} = \frac{1}{R}\,\widehat{EU_i^2} = \frac{28.18}{10} = 2.818\ W\,.$$

□

Cvičení 12.2.4: Obvyklý výskyt nemoci je 16 případů na 100 000 obyvatel za rok. Je podezření, že určitý lék riziko nemoci zvyšuje. Jak velký vzorek uživatelů léku potřebujeme sledovat 1 rok, abychom mohli toto podezření potvrdit na hladině významnosti 5 %, kdyby výskyt nemoci vyšel dvojnásobný? Diskutujte předpoklady.

Řešení: Nulová hypotéza: riziko se nezvyšuje. Za tohoto předpokladu má výskyt v souboru rozsahu n binomické rozdělení $\mathrm{Bi}(n, q)$, kde $q = 16/100\,000$. To pomocí centrální limitní věty přibližně nahradíme normálním rozdělením se stejnou střední hodnotou $\mu = n\,q$ a rozptylem $\sigma^2 = n\,q\,(1-q)$. Zajímá nás, kdy kritická hodnota odpovídá výběrovému průměru $2\,\mu = 2\,n\,q$. Testovací statistika $T = \frac{\overline{X}-\mu}{\sigma}$ má přibližně rozdělení $\mathrm{N}(0,1)$ a nabývá kritické hodnoty $\Phi^{-1}(0.95) = 1.64$ pro $\overline{X} = 2\,\mu$, tj. pro

$$T = \frac{\mu}{\sigma} = \frac{n\,q}{\sqrt{n\,q\,(1-q)}} = \sqrt{\frac{n\,q}{1-q}} \doteq \sqrt{n\,q}$$

Z rovnice $\sqrt{n\,q} \doteq 1.64$ dostáváme $n \doteq \frac{1.64^2}{q} = 16810$. Velký rozsah výběru zdánlivě opravňuje použití centrální limitní věty, ale teoretická četnost je $\frac{16810}{100000} \cdot 16 = 2.689\,6 \doteq 3$. Bylo by tedy žádoucí ověřit výsledek (mnohem pracnější) analýzou binomického rozdělení; alespoň však víme, kterého, což se nedalo bez centrální limitní věty odhadnout. I tak klade přesný (i numerický) výpočet značné nároky na výpočetní techniku, než ověříme, že

$$1 - \sum_{k=0}^{5} \binom{16810}{k} q^k\,(1-q)^{16810-k} \doteq 0.0559\,,$$
$$1 - \sum_{k=0}^{6} \binom{16810}{k} q^k\,(1-q)^{16810-k} \doteq 0.0202\,,$$

takže 6 výskytů u 16810 lidí stačí k zamítnutí nulové hypotézy, ve shodě s předchozím postupem, který dal odhad kritické hodnoty $2 \cdot \frac{16810}{100000} \cdot 16 \doteq 5.379\,2$.

(Je to příklad výpočtu, kdy na špatně položenou otázku, např. kolik je

$$\sum_{k=7}^{16810} \binom{16810}{k} q^k\,(1-q)^{16810-k}$$

nedostaneme v přijatelném čase odpověď ani od pokročilé techniky. Hledat mez 16810 zkusmo by bylo ještě zdlouhavější.) □

Cvičení 12.2.5: Náhodná veličina X je efektivní hodnota střídavého napětí. Posuďte na hladině významnosti 5 % hypotézu, že střední hodnota náhodné veličiny X je nejvýše 230 V, jestliže z 20 nezávislých měření vyšla realizace výběrového průměru 233 V a realizace výběrové směrodatné odchylky 5 V.

Cvičení 12.2.6: Náhodná veličina X je odpor rezistorů. 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.

Cvičení 12.2.7: Náhodná veličina X má alternativní rozdělení s parametrem q. Podle nulové hypotézy je $q = 1/2$. Jaký by musel být rozsah výběru, aby realizace výběrového průměru mimo interval $\langle 0{,}45,\ 0{,}55\rangle$ dovolovala zamítnout nulovou hypotézu na hladině významnosti 1 %?

Cvičení 12.2.8: Posuďte na hladině významnosti $\alpha = 1\,\%$ hypotézu, že výsledky hodů mincí mají pravděpodobnost 1/2, jestliže (a) při 200 hodech padl líc $80\times$, (b) při 100 hodech padl líc $40\times$.

Řešení: (a) zamítáme, (b) nezamítáme (ačkoli v obou případech šlo o 40 % výsledků). □

Cvičení 12.2.9: Při 100 hodech kostkou padla $25\times$ šestka. Posuďte na hladině významnosti $\alpha = 5\,\%$ hypotézu, že šestka padá s pravděpodobností 1/6.

Cvičení 12.2.10: Z 10 měření krevního tlaku u jednoho pacienta jsme obdrželi výběrový průměr 160 a výběrovou směrodatnou odchylku 30. Rozhodněte na hladině významnosti 5 %, zda je střední hodnota krevního tlaku nejvýše 140. Za jakých předpokladů výsledek platí?

Cvičení 12.2.11: Měřením odporu 10 rezistorů jsme obdrželi výběrový průměr $102\,\Omega$ a výběrovou směrodatnou odchylku $2\,\Omega$. Posuďte na hladině významnosti 1 % hypotézu, že střední hodnota odporu rezistorů je rovna jejich nominální hodnotě $100\,\Omega$.

Cvičení 12.2.12: Průzkumem byla zjištěna účinnost výběru mýta 94 %. Jak velký musel být rozsah výběru, abychom na hladině významnosti 1 % mohli zamítnout, že tato účinnost je aspoň 95 %, jak požaduje smlouva?

12.2.2 Testy rozptylu normálního rozdělení

Příklad 12.2.13: Máme podezření, že měřicí přístroj má přesnost menší než povolenou. Jak to prokážeme?

Testovací statistiku

$$T = \frac{(n-1)\,S_{\boldsymbol{X}}^2}{c}$$

porovnáváme s kvantily χ^2-*rozdělení* s $n-1$ stupni volnosti:

H_0	zamítáme pro	dosažená významnost
$\sigma^2 \le c$	$t > q_{\chi^2(n-1)}(1-\alpha)$	$1 - F_{\chi^2(n-1)}(t)$
$\sigma^2 \ge c$	$t < q_{\chi^2(n-1)}(\alpha)$	$F_{\chi^2(n-1)}(t)$
$\sigma^2 = c$	$t < q_{\chi^2(n-1)}(\alpha/2)$ nebo $t > q_{\chi^2(n-1)}(1-\alpha/2)$	$2\min\left(F_{\chi^2(n-1)}(t), 1 - F_{\chi^2(n-1)}(t)\right)$

Cvičení 12.2.14: Z $n = 21$ měření stálého napětí stejným voltmetrem jsme dostali realizaci výběrové směrodatné odchylky $s_{\boldsymbol{x}} = 1.5\,mV$. Posuďte na hladině významnosti $\alpha = 1\,\%$ hypotézu, že náhodná chyba voltmetru má směrodatnou odchylku nejvýše $\sigma_X = 1\ mV$, kterou uvádí výrobce. Uveďte použité předpoklady.

Řešení: Z výběrového rozptylu vypočítáme testovací statistiku

$$t = \frac{(n-1)\,s_{\boldsymbol{x}}^2}{\mathrm{D}X} = 45\,,$$

kterou porovnáme s kvantilem $q_{\chi^2(20)}(0.99) \doteq 37.57$ a 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 12.2.15: Měřicí přístroj má mít směrodatnou odchylku nejvýše 0.1. Ověřte tuto hypotézu na 5 % hladině významnosti na základě opakovaných měření téže hodnoty s četnostmi dle tabulky. Uveďte použité předpoklady.

hodnota	9.9	10.0	10.1	10.3	10.4
četnost	1	5	5	2	1

12.2.3 Porovnání dvou normálních rozdělení

Příklad 12.2.16: Domníváme se, že náš program je rychlejší než konkurenční. Jak to prokážeme?

Předpokládáme dva náhodné výběry, $(X_1, \ldots, X_m)$ z rozdělení $\mathrm{N}(\mathrm{E}X, \mathrm{D}X)$ a $(Y_1, \ldots, Y_n)$ z rozdělení $\mathrm{N}(\mathrm{E}Y, \mathrm{D}Y)$, které jsou *nezávislé*, tj. náhodné veličiny $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ jsou nezávislé. Chceme testovat hypotézy porovnávající neznámé parametry těchto rozdělení.

Test rozptylu dvou normálních rozdělení

Pokud jsou oba rozptyly stejné, $\mathrm{D}X = \mathrm{D}Y = \sigma^2$, pak by mělo vyjít $S_{\mathbf{X}}^2 \doteq S_{\mathbf{Y}}^2$. Testovací statistikou je poměr

$$T = \frac{S_{\mathbf{X}}^2}{S_{\mathbf{Y}}^2}.$$

Ten má F-rozdělení (Fisherovo-Snedecorovo rozdělení) $\mathrm{F}(\xi, \eta)$ s ξ a η stupni volnosti, neboli rozdělení náhodné veličiny

$$F = \frac{\frac{U}{\xi}}{\frac{V}{\eta}},$$

kde U, V jsou *nezávislé* náhodné veličiny s rozdělením $\chi^2(\xi)$, resp. $\chi^2(\eta)$. To ověříme dosazením:

$$U = \frac{(m-1)\,S_{\mathbf{X}}^2}{\sigma^2} \text{ má } \chi^2(m-1)\,,$$

$$V = \frac{(n-1)\,S_{\mathbf{Y}}^2}{\sigma^2} \text{ má } \chi^2(n-1)\,,$$

$$\xi = m-1\,,$$

$$\eta = n-1\,,$$

$$F = \frac{\frac{U}{\xi}}{\frac{V}{\eta}} = \frac{\frac{(m-1)\,S_{\mathbf{X}}^2}{(m-1)\,\sigma^2}}{\frac{(n-1)\,S_{\mathbf{Y}}^2}{(n-1)\,\sigma^2}} = \frac{S_{\mathbf{X}}^2}{S_{\mathbf{Y}}^2} = T\,.$$

Testujeme realizaci t na rozdělení $\mathrm{F}(m-1, n-1)$:

H_0	zamítáme pro	dosažená významnost
$\mathrm{D}X \leq \mathrm{D}Y$	$t > q_{\mathrm{F}(m-1,n-1)}(1-\alpha)$	$1 - F_{\mathrm{F}(m-1,n-1)}(t)$
$\mathrm{D}X \geq \mathrm{D}Y$	$t < q_{\mathrm{F}(m-1,n-1)}(\alpha)$	$F_{\mathrm{F}(m-1,n-1)}(t)$
$\mathrm{D}X = \mathrm{D}Y$	$t < q_{\mathrm{F}(m-1,n-1)}(\alpha/2)$ nebo $t > q_{\mathrm{F}(m-1,n-1)}(1-\alpha/2)$	$2\min(F_{\mathrm{F}(m-1,n-1)}(t), 1 - F_{\mathrm{F}(m-1,n-1)}(t))$

Pro každou hladinu významnosti potřebujeme dvojdimenzionální tabulku kvantilů indexovanou ξ, η, což je mnoho dat; obvykle je tabelována jen polovina, druhou je třeba dopočítat podle vzorce

$$q_{\mathrm{F}(\xi,\eta)}(\beta) = \frac{1}{q_{\mathrm{F}(\eta,\xi)}(1-\beta)}$$

(pozor na opačné pořadí indexů!) nebo uvažovat $\frac{S_{\mathbf{Y}}^2}{S_{\mathbf{X}}^2}$ místo $\frac{S_{\mathbf{X}}^2}{S_{\mathbf{Y}}^2}$.

Testy středních hodnost dvou normálních rozdělení se známým rozptylem σ^2

Parametry výběrových průměrů mají následující rozdělení:

$$\begin{aligned} \overline{\boldsymbol{X}}_m &\text{ má } \mathrm{N}\left(\mathrm{E}X, \sigma^2/m\right) , \\ \overline{\boldsymbol{Y}}_n &\text{ má } \mathrm{N}\left(\mathrm{E}Y, \sigma^2/n\right) , \\ \overline{\boldsymbol{X}}_m - \overline{\boldsymbol{Y}}_n &\text{ má } \mathrm{N}\left(\mathrm{E}X - \mathrm{E}Y, \sigma^2\,(1/m + 1/n)\right) . \end{aligned}$$

Za předpokladu $\mathrm{E}X = \mathrm{E}Y$ testovací statistika

$$T = \frac{\overline{\boldsymbol{X}}_m - \overline{\boldsymbol{Y}}_n}{\sigma\sqrt{1/m + 1/n}} \text{ má } \mathrm{N}(0,1) .$$

Testujeme její realizaci t na $\mathrm{N}(0,1)$ (viz kapitola 12.2.1).

Testy středních hodnost dvou normálních rozdělení se (stejným) neznámým rozptylem

Předpokládáme, že $\mathrm{D}X = \mathrm{D}Y = \sigma^2$. Proto bychom nejprve měli ověřit předpoklad rovnosti rozptylů (dle kapitoly 12.2.3).

Poznámka 12.2.17: Ve skutečnosti nemůžeme předpoklad ověřit, jedině vyvrátit; pokusíme se o to testováním nulové hypotézy o rovnosti rozptylů. Pokud tuto hypotézu nevyvrátíme, pokračujeme dále; v opačném případě bychom museli použít podstatně složitější postup, který zájemci najdou např. v [Mood et al. 1974].

Pro jednu neznámou hodnotu σ^2 máme dva odhady $S^2_{\boldsymbol{X}}, S^2_{\boldsymbol{Y}}$. Abychom maximálně využili tuto informaci, použijeme jejich průměr. Abychom vystihli závislost na rozsahu výběru, přikládáme jim váhu zhruba úměrnou rozsahu výběru; ve skutečnosti zmenšenou o 1, tj. $m-1, n-1$,

$$S^2 = \frac{(m-1)\,S^2_{\boldsymbol{X}} + (n-1)\,S^2_{\boldsymbol{Y}}}{m+n-2} .$$

Tato volba má tu výhodu, že dovedeme stanovit rozdělení váženého průměru:

$$\begin{aligned} \frac{(m-1)\,S^2_{\boldsymbol{X}}}{\sigma^2} &\text{ má } \chi^2(m-1) , \\ \frac{(n-1)\,S^2_{\boldsymbol{Y}}}{\sigma^2} &\text{ má } \chi^2(n-1) , \\ \frac{(m-1)\,S^2_{\boldsymbol{X}} + (n-1)\,S^2_{\boldsymbol{Y}}}{\sigma^2} &\text{ má } \chi^2(m+n-2) , \end{aligned}$$

a tedy střední hodnotu $m+n-2$, takže

$$\frac{(m-1)\,S^2_{\boldsymbol{X}} + (n-1)\,S^2_{\boldsymbol{Y}}}{(m+n-2)\,\sigma^2} = \frac{S^2}{\sigma^2}$$

má střední hodnotu 1 a S^2 je nestranný odhad rozptylu σ^2. Pro další výpočet použijeme odhad směrodatné odchylky

$$S = \sqrt{\frac{(m-1)\,S^2_{\boldsymbol{X}} + (n-1)\,S^2_{\boldsymbol{Y}}}{m+n-2}} .$$

Výběrové průměry mají následující rozdělení:

$$\begin{aligned} \overline{\boldsymbol{X}}_m &\text{ má } \mathrm{N}\left(\mathrm{E}X, \sigma^2/m\right) , \\ \overline{\boldsymbol{Y}}_n &\text{ má } \mathrm{N}\left(\mathrm{E}Y, \sigma^2/n\right) , \\ \overline{\boldsymbol{X}}_m - \overline{\boldsymbol{Y}}_n &\text{ má } \mathrm{N}\left(\mathrm{E}X - \mathrm{E}Y, \sigma^2\,(1/m + 1/n)\right) . \end{aligned}$$

Za předpokladu $EX = EY$ pak

$$\frac{\overline{X}_m - \overline{Y}_n}{\sigma\sqrt{1/m + 1/n}} \text{ má } \mathrm{N}(0,1)\,,$$

$$\frac{(m+n-2)\,S^2}{\sigma^2} = \frac{(m-1)\,S_X^2 + (n-1)\,S_Y^2}{\sigma^2} \text{ má } \chi^2(m+n-2)\,,$$

$$T = \frac{\overline{X}_m - \overline{Y}_n}{S\sqrt{1/m+1/n}} = \frac{\frac{\overline{X}_m - \overline{Y}_n}{\sigma\sqrt{1/m+1/n}}}{\sqrt{\frac{S^2}{\sigma^2}}} \text{ má } \mathrm{t}(m+n-2)\,.$$

Testujeme realizaci t na Studentovo rozdělení $\mathrm{t}(m+n-2)$ (viz kapitola 12.2.1).

Cvičení 12.2.18: Z realizací náhodných veličin X, Y z výběrů rozsahu $m = 11$, resp. $n = 21$, jsme stanovili realizace odhadů $\bar{x} = 10$, $\bar{y} = 12$, $s_x = 2$, $s_y = 3$. Posuďte na hladině významnosti 5 % hypotézu, že střední hodnoty náhodných veličin X, Y jsou stejné. Uveďte a dle možnosti zkontrolujte použité předpoklady.

Řešení: Test rovnosti rozptylů:

$$\frac{s_x^2}{s_y^2} = \frac{4}{9} \doteq 0.444$$

porovnáme s kvantily $q_{\mathrm{F}(10,20)}(0.975) \doteq 2.77$ a $q_{\mathrm{F}(10,20)}(0.025) = \frac{1}{q_{\mathrm{F}(20,10)}(0.975)} = \frac{1}{3.42} \doteq 0.292\,40$ a nulovou hypotézu nezamítáme.

Rozptyl odhadneme hodnotou

$$s^2 = \frac{(m-1)\,s_x^2 + (n-1)\,s_y^2}{m+n-2} = \frac{22}{3} \doteq 7.333\,,$$

směrodatnou odchylku

$$s = \sqrt{\frac{(m-1)\,s_x^2 + (n-1)\,s_y^2}{m+n-2}} = \sqrt{\frac{22}{3}} \doteq 2.708\,.$$

Test rovnosti středních hodnot:

$$\frac{\bar{x} - \bar{y}}{s\sqrt{1/m+1/n}} = -\frac{3}{4}\sqrt{7} \doteq -1.984$$

porovnáme s kvantily $q_{\mathrm{t}(30)}(0.975) = -q_{\mathrm{t}(30)}(0.025) \doteq 2.042$ a nulovou hypotézu nezamítáme.

Použité předpoklady: nezávislé náhodné veličiny s normálním rozdělením a stejným (neznámým) rozptylem. □

Cvičení 12.2.19: Stejnou veličinu jsme měřili dvěma metodami, každou 10×. Výsledky shrnuje následující tabulka.

	výběrový průměr	výběrová směrodatná odchylka
1. metoda	16	5
2. metoda	11	3

Posuďte na hladině významnosti 5 %, zda lze považovat obě metody za stejně přesné a jejich střední hodnoty za stejné. Diskutujte použité předpoklady.

Cvičení 12.2.20: Dva náhodné výběry mají následující parametry:

rozsah	výběrový průměr	výběrová směrodatná odchylka
11	183	10
21	170	15

Otestujte na hladině významnosti 5 % hypotézu, že výběry pocházejí z rozdělení se stejnou střední hodnotou. Jaké předpoklady jste použili?

Cvičení 12.2.21: U náhodných výběrů $N = 15$ fotbalistů, resp. $M = 25$ hokejistů, jsme z naměřených hodnot hmotnosti stanovili výběrový průměr a výběrovou směrodatnou odchylku $\bar{x} = 85\ kg$, $s_x = 16\ kg$, resp. $\bar{y} = 95\ kg$, $s_y = 10\ kg$. Posuďte na hladině významnosti 5 % hypotézu, že střední hodnota obou náhodných veličin je stejná.

Cvičení 12.2.22: U dvou skupin pacientů byly naměřeny následující údaje o krevním tlaku [torr]:

skupina	počet pacientů	výběrový průměr	výběrová směr. odchylka
1	11	130	10
2	21	140	15

Na hladině významnosti $\alpha = 5\,\%$ posuďte hypotézu, že střední hodnoty v obou skupinách jsou stejné. Diskutujte použité předpoklady.

12.2.4 Testy středních hodnost dvou normálních rozdělení – párový pokus

Příklad 12.2.23: Máme porovnat průměrnou teplotu na dvou místech. Standardní test středních hodnot dvou normálních rozdělení (z předchozí kapitoly) je slabý kvůli velkému rozptylu, který však má společnou příčinu a projevuje se proto synchronně v obou výběrech. Zde výběry *nejsou navzájem nezávislé*; měříme vždy obě veličiny současně.

Poznámka 12.2.24: Tato tématika je jednoduchá na použití, ale bývá často špatně interpretována. Zde se výklad opírá o [Schlesinger, Hlaváč 1999] a o konzultaci s Prof. M. Schlesingerem.

Předpokládáme, že pro $j = 1, \ldots, n$ mají náhodné veličiny X_j, Y_j normální rozdělení $\mathrm{N}(\mu_j, \sigma^2)$ se stálým rozptylem σ^2 a *proměnnými* středními hodnotami $\mu_j = \mathrm{E}X_j = \mathrm{E}Y_j$. Pro popis můžeme použít náhodné veličiny $U_j = X_j - \mu_j, V_j = Y_j - \mu_j$, které *jsou nezávislé* a mají rozdělení $\mathrm{N}(0, \sigma^2)$. Náhodné veličiny $\boldsymbol{\Delta}_j = X_j - Y_j = U_j - V_j$ jsou nezávislé a mají rozdělení $\mathrm{N}(0, 2\sigma^2)$. Výběrový průměr $\overline{\boldsymbol{\Delta}}$ má rozdělení $\mathrm{N}\left(0, 2\sigma^2/n\right)$. Pro výpočet použijeme realizace $\delta_j = x_j - y_j$ náhodných veličin Δ_j.

Poznámka 12.2.25: Předpoklady jdou za rámec definice náhodného výběru, neboť náhodné veličiny X_j, Y_j mají rozdělení závislé na j. Teprve náhodné vektory (U_j, V_j) mají rozdělení nezávislé na j a tvoří dvojrozměrný náhodný výběr $((U_1, V_1), \ldots, (U_n, V_n))$ dle smyslu pozn. 8.2.5.

Pro *známý* rozptyl σ^2

Neznámé parametry sdruženého rozdělení jsou $\mu_1, \ldots, \mu_n$, ale nepotřebujeme je.
Dle kapitoly 12.2.1 (pro $c = 0$) testujeme statistiku

$$T = \frac{\overline{\boldsymbol{\Delta}}}{\sigma}\sqrt{\frac{n}{2}} = \frac{\overline{\boldsymbol{X}} - \overline{\boldsymbol{Y}}}{\sigma}\sqrt{\frac{n}{2}}$$

na rozdělení $\mathrm{N}(0, 1)$.

Pro *neznámý* rozptyl

Neznámé parametry sdruženého rozdělení jsou $\Theta = (\sigma^2, \mu_1, \ldots, \mu_n)$, potřebujeme z nich pouze $\sigma^2 = \mathrm{D}X$.
Můžeme pracovat přímo s výběrem $(\Delta_1, \ldots, \Delta_n)$ z normálního rozdělení. Dle kapitoly 12.2.1 (pro $c = 0$) testujeme

$$T = \frac{\overline{\Delta}}{S_\Delta}\sqrt{n}$$

na Studentovo rozdělení $\mathrm{t}(n-1)$.

Poznámka 12.2.26: Z cvičných důvodů odvoďme maximálně věrohodný odhad parametrů:

$$\begin{aligned}
\Lambda(\Theta) &= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x_j-\mu_j)^2}{2\sigma^2}\right) \cdot \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y_j-\mu_j)^2}{2\sigma^2}\right), \\
\lambda(\Theta) &= -\sum_{j=1}^{n} \frac{(x_j-\mu_j)^2}{2\sigma^2} - \sum_{j=1}^{n} \frac{(y_j-\mu_j)^2}{2\sigma^2} - 2n\ln\sigma - 2n\ln\sqrt{2\pi}, \\
0 = \frac{\partial}{\partial\widehat{\mu_j}}\lambda(\widehat{\vartheta}) &= \frac{\partial}{\partial\widehat{\mu_j}}\left(-\frac{(x_j-\widehat{\mu_j})^2}{2\widehat{\sigma^2}} - \frac{(y_j-\widehat{\mu_j})^2}{2\widehat{\sigma^2}}\right) \\
&= \frac{1}{\widehat{\sigma^2}}\left((x_j-\widehat{\mu_j}) + (y_j-\widehat{\mu_j})\right) = \frac{1}{\widehat{\sigma^2}}\left(x_j + y_j - 2\widehat{\mu_j}\right), \\
\widehat{\mu_j} &= \frac{x_j+y_j}{2}, \qquad j = 1, \ldots, n.
\end{aligned}$$

Odhady $\widehat{\mu_j}$, $(j = 1, \ldots, n)$ *nejsou konzistentní.* Po jejich dosazení

$$\begin{aligned}
\lambda(\widehat{\vartheta}) &= -\sum_{j=1}^{n} \frac{(x_j-y_j)^2}{4\widehat{\sigma^2}} - n\ln\widehat{\sigma^2} - 2n\ln\sqrt{2\pi}, \\
0 = \frac{\partial}{\partial\widehat{\sigma^2}}\lambda(\widehat{\vartheta}) &= \sum_{j=1}^{n} \frac{(x_j-y_j)^2}{4\widehat{\sigma^2}^2} - \frac{2n}{\widehat{\sigma^2}}, \\
\widehat{\sigma^2} &= \frac{1}{2n}\sum_{j=1}^{n}(x_j-y_j)^2 = \frac{1}{2n}\sum_{j=1}^{n}\delta_j^2.
\end{aligned}$$

Odhad $\widehat{\sigma^2}$ *je konzistentní.*

Testy hypotéz o charakteristikách rozdělení jsou velmi častým nástrojem pro objektivní posouzení výsledků v téměř všech vědeckých disciplínách, stejně jako v praxi při posuzování kvality a spolehlivosti. Pro správné použití potřebujeme předpoklady o typu rozdělení a nezávislosti veličin; někdy bývá jejich splnění problematické, ale obecný postup bez nich neschůdný. V takových případech metodu použijeme, ale musíme uvážit, jaký důsledek může mít porušení předpokladů na učiněné závěry.

Cvičení 12.2.27: Jsou dány výběrové průměry a výběrové rozptyly dvou náhodných výběrů z normálních rozdělení. Jak provedete párový pokus pro test rovnosti středních hodnot?

Řešení: To nelze, párový pokus předpokládá, že data jsou pořizována ve dvojicích a bez tohoto přiřazení jej nelze uvedeným postupem vyhodnotit. □

Cvičení 12.2.28: Teploty v technologickém procesu jsme měřili současně dvěma metodami, získali jsme následující dvojice hodnot:

1. metoda	1123	1212	1303	1398	1456	1500
2. metoda	1143	1234	1307	1400	1450	1500

Posuďte na hladině významnosti 5 % hypotézu, že střední hodnota obou metod je stejná.

Řešení: Toto je typická situace pro párový pokus. Předpokládáme, že střední hodnota je pro obě proměnné stejná, ale kolísá. Předpokládáme normální rozdělení odchylek od ní.

Rozdíly naměřených hodnot jsou:

−20	−22	−4	−2	6	0

Realizace výběrového průměru je -7, výběrového rozptylu 129.2, výběrové směrodatné odchylky 11.36. Realizaci testovací statistiky

$$\frac{\bar{\delta}}{s_\delta}\sqrt{n} = \frac{-7}{11.36}\sqrt{6} \doteq -1.51$$

porovnáme s kvantily $q_{t(5)}(0.025) = -q_{t(5)}(0.975) \doteq -2.57$, hypotézu nezamítáme. □

Cvičení 12.2.29: Na dvou místech jsme měřili současně teploty s následujícími výsledky:

teplota1	10	11.5	12	14	13	9.5	8	10	11
teplota2	12	13	14	14.5	12	10	9	10	9

Posuďte na hladině významnosti 5 % hypotézu, že na druhém místě není tepleji než na prvním. Uveďte použité předpoklady.

Řešení: Jedná se o párový pokus; rozdíly teplot jsou $\boldsymbol{\delta} = (-2, -1.5, -2, -0.5, 1, -0.5, -1, 0, 2)$, $n = 9$, $\bar{\delta} = -0.5$, $s_\delta^2 \doteq 2.09375$, $s_\delta \doteq 1.44698$,

$$t = \frac{\bar{\delta}}{s_\delta}\sqrt{n} \doteq -1.0367,$$

porovnáme s kvantilem $q_{t(8)}(0.05) = -q_{t(8)}(0.95) \doteq -1.86$ a nulovou hypotézu nezamítáme. Předpoklady pro párový pokus: střední hodnoty náhodných veličin v obou výběrech kolísají stejně, odchylky od nich mají normální rozdělení a jsou nezávislé. □

Cvičení 12.2.30: Pacient si měřil teplotu vždy současně dvěma teploměry, hodnoty jsou v tabulce:

teplota1	37.5	38.2	38.3	38.6	38.2	37.6	37.2	36.9	36.6
teplota2	37.7	38.3	38.5	38.6	38.3	37.7	37.4	36.7	36.5

Posuďte na hladině významnosti 5 % hypotézu, že mezi průměrnými údaji teploměrů není rozdíl.

Řešení: Rozdíly teplot jsou $\boldsymbol{\delta} = (-0.2, -0.1, -0.2, 0, -0.1, -0.1, -0.2, 0.2, 0.1)$, $n = 9$, $\bar{\delta} \doteq -0.0667$, $s_\delta^2 \doteq 0.02$, $s_\delta \doteq 0.141$,

$$t = \frac{\bar{\delta}}{s_\delta}\sqrt{n} \doteq -1.41\,,$$

porovnáme s kvantilem $q_{t(8)}(0.025) = -q_{t(8)}(0.975) \doteq -2.306$ a nulovou hypotézu nezamítáme. □

12.3 χ^2-test dobré shody

Příklad 12.3.1: Máme podezření, že výsledky generátoru náhodných čísel neodpovídají rovnoměrnému rozdělení. Jak to prokážeme?

Až dosud jsme předpokládali, že typ rozdělení je znám až na několik parametrů, které máme testovat. Testy dobré shody slouží k testování hypotézy, že náhodná veličina má předpokládané rozdělení. Protože umíme hypotézy pouze zamítat, nikdy nepotvrdíme, že takové rozdělení opravdu má; máme jen naději, že se nám v případě, kdy se od předpokládaného rozdělení značně liší, podaří tuto hypotézu zamítnout.

Asi nejčastěji používaným testem dobré shody je χ^2-test. Ve své původní podobě je určen k testování *diskrétního rozdělení*. Chceme-li jej použít pro spojité rozdělení, musíme je napřed *vhodným způsobem diskretizovat*.

Nulová hypotéza je, že náhodná veličina má diskrétní rozdělení s k hodnotami, jejichž pravděpodobnosti jsou $p_1, \ldots, p_k$. Testujeme ji pomocí realizace náhodného výběru rozsahu n. Není důležité pořadí výsledků, pouze jejich *četnosti* n_i, resp. *relativní četnosti* n_i/n $(i = 1, \ldots, k)$. Porovnáváme četnosti n_i s *teoretickými četnostmi* np_i. Testovací statistikou je

$$T = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} . \tag{12.1}$$

Test je založen na tom, že rozdělení této statistiky se pro $n \to \infty$ blíží rozdělení $\chi^2(k-1)$. Dosažená významnost je $1 - F_{\chi^2(n-1)}(t)$. Nulovou hypotézu zamítáme pro $t > q_{\chi^2(k-1)}(1-\alpha)$, tj. $1 - F_{\chi^2(n-1)}(t) < \alpha$.

Poznámka 12.3.2: Počet $k-1$ stupňů volnosti má intuitivní význam. Testujeme diskrétní rozdělení s k hodnotami. To je popsáno k pravděpodobnostmi, jejichž součet je 1. Tomu odpovídá $k-1$ volitelných parametrů.

Předchozí postup má několik důležitých modifikací, které jsou nutné pro jeho správné použití.

Příklad 12.3.3: Bob 100× vsadil v loterii a nic nevyhrál. Z toho usoudil, že průměrná výhra v loterii je nulová.

Problém 12.3.4: Testujeme pomocí rozdělení, kterému se skutečné jen limitně blíží. Tím se dopouštíme blíže neurčené dodatečné chyby. Proto je potřeba, aby rozsah výběru byl „dostatečně veliký". To je ovšem relativní pojem.

Řešení: Teoretické četnosti tříd tedy nesmí být příliš malé. Nelze říci žádnou přesnou mez; obvykle se požaduje, aby každá třída měla teoretickou četnost aspoň 5. Vychází-li teoretická četnost některých tříd příliš malá, sloučíme je s jinými třídami. Ty by měly být pokud možno „blízké"; nebylo by moc informativní vytvořit např. třídu „jedna, nebo aspoň 5". □

Příklad 12.3.5: Uvažujme test rozdělení mnohačetných porodů, tj. kolik se narodilo jednotlivých dětí, dvojčat, trojčat atd. Paterčat je tak málo, že je těžko můžeme popsat rozdělením χ^2 i při značném rozsahu souboru. Přirozeně to řešíme tím, že je sdružíme do nové „hodnoty", např. třídy „tři a více" tak, aby tato byla dostatečně početná.

Pro testy dobré shody potřebujeme velký rozsah souboru. Uvažme, jak závisí výsledek χ^2-testu na počtu dat.

Tvrzení 12.3.6: *Pokud v χ^2-testu dobré shody zvětšíme rozsah souboru r-krát a poměr četností zůstane zachován, hodnota kritéria se zvýší r-krát.*

Důkaz: Jmenovatelé ve výrazu (12.1) se zvýší r-krát, čitatelé r^2-krát. □

Důsledkem je, že pokud rozdělení neodpovídá předpokládanému a rozsah výběru je velký, setkáváme se s hodnotami testovací statistiky, které převyšují kritické hodnoty o mnoho řádů. Nemusí to tedy být hrubá chyba, ale obvyklé chování metody pro velké soubory dat. Znamená to, že pomocí rozsáhlých dat můžeme zamítnout nulovou hypotézu s vysokým koeficientem spolehlivosti i při rozdělení málo odlišném od předpokládaného.

Poznámka 12.3.7: Dalším důsledkem je, že nemůžeme použít data např. v tisících obyvatel (snížili bychom tím hodnotu kritéria 1000×), protože potřebujeme posuzovat zvlášť každý subjekt, který může být nezávisle zařazen do třídy. Pokud např. známe procentuální zastoupení zkoumaných obejktů ve třídách, ale neznáme jejich celkový počet, χ^2-test *nemůžeme použít.*

Problém 12.3.8: Zkoumané rozdělení může záviset na parametrech, které neznáme.

Příklad 12.3.9: Otázka může znít, zda rozdělení dat odpovídá *nějakému* normálnímu rozdělení. V příkladu 12.3.5 (rozdělení mnohačetných porodů) bychom se mohli ptát, zda jej vystihuje *nějaké* geometrické či Poissonovo rozdělení.

Chybějící parametry lze odhadnout některou z dříve popsaných metod. Je ovšem nutno rozlišit dva případy. Pokud parametry odhadneme na základě *jiného* náhodného výběru, postup se nemění. To však nebývá obvyklé, neboť je nehospodárné použít část dat pouze k odhadu parametrů a *jinou* část dat k testu dobré shody. Proto je častý druhý případ, kdy parametry odhadneme na základě *stejného* náhodného výběru, který používáme k testu dobré shody. Tím jsme však změnili podmínky, neboť dosahujeme větší shody, než je adekvátní.

Příklad 12.3.10: Předpokládejme, že odhadnutými parametry jsou pravděpodobnosti některých hodnot. Kdybychom si mohli zvolit $k-1$ pravděpodobností, mohli bychom (v duchu poznámky 12.3.2) volit libovolné rozdělení s k hodnotami a dostat úplnou shodu pozorovaných četností s teoretickými. Přišli jsme o všech $k-1$ stupňů volnosti. Obdobná situace je typická i při volbě jiných $k-1$ parametrů.

Řešení (problému 12.3.8): Volbou c parametrů jsme snížili počet stupňů volnosti o c. Tomu odpovídá stejný postup s jedinou změnou: testujeme stejnou statistiku na rozdělení $\chi^2(k-1-c)$. *(Ovšem pouze tehdy, pokud k odhadu parametrů i testu dobré shody byla použita stejná data.)* □

Problém 12.3.11: Chceme testovat shodu se *spojitým* nebo *smíšeným* rozdělením.

Řešení: Rozdělení napřed diskretizujeme, tj. všechny možné výsledky rozdělíme do k disjunktních tříd. Třídy volíme tak, aby všechny teoretické četnosti byly dostatečně velké, nejlépe všechny zhruba stejné. To docílíme např. tak, že hodnoty náhodné veličiny X rozdělíme do intervalů

$$(-\infty, q_X(1/k)), (q_X(1/k), q_X(2/k)), (q_X(2/k), q_X(3/k)), \ldots, (q_X((k-1)/k), \infty).$$

□

Poznámka 12.3.12: Prvky v jedné třídě mají být „blízké", jinak bychom snížili sílu testu. Nebylo by vhodné rozdělit např. hodnoty z $\langle 0,1\rangle$ do tříd podle třetího desetinného místa. Tím bychom napodobili generátor náhodných čísel a dostali bychom přibližně rovnoměrné rozdělení do tříd, téměř nezávisle na původním rozdělení (o němž bychom se málo dověděli).

Poznámka 12.3.13: Test χ^2 vůbec *nelze použít* na data, která jsou spojitá, např. na otázku, zda je spotřeba benzínu v jednotlivých měsících stejná. Není totiž jasné, v jakých jednotkách bychom ji měli měřit, a výsledek na tom závisí, viz pozn. 12.3.7.

Podstatou χ^2-testu je předpoklad, že jsou zde diskrétní objekty, které jsou náhodně zařazeny do tříd (tvořících disjunktní rozklad množiny všech možných výsledků). Takovými objekty nemohou být ani litry ani barely benzínu, nanejvýš jednotliví zákazníci (o těch bychom mohli testovat hypotézu, zda přijíždějí rovnoměrně).

Poznámka 12.3.14: Výjimečně se v χ^2-testu používá i *dolní* intervalový odhad kritéria, kdy nulovou hypotézu zamítáme pro $t < q_{\chi^2(k-1)}(\alpha)$. Takový test použijeme tehdy, když chceme odhalit nápadnou shodu údajů s předpokládaným rozdělením. Například některé generátory pseudonáhodných čísel během své periody dají každý možný výsledek *právě jednou*; pak se rozdělení až nápadně podobá rovnoměrnému. Skutečné rovnoměrné rozdělení s nezávislými výsledky by tak dobrou shodu nedalo a tento test to odhalí.

Cvičení 12.3.15: Realizací náhodné veličiny X jsme dostali následující četnosti výsledků:

hodnota	0	1	2	3	4	5	6
pozorovaná četnost	29	15	10	5	3	0	2

Posuďte na hladině významnosti 5 % hypotézu, že náhodná veličina X má geometrické rozdělení s parametrem $q = 1/2$, $p_X(i) = (1-q)\,q^i$.

Řešení: Rozsah výběru je $n = 64$.

hodnota	0	1	2	3	4	5	≥ 6
pozorovaná četnost	29	15	10	5	3	0	2
teoretická pravděpodobnost	1/2	1/4	1/8	1/16	1/32	1/64	1/64
teoretická četnost	32	16	8	4	2	1	1

Pro dosažení teoretické četnosti aspoň 5 potřebujeme sloučit poslední skupiny:

hodnota	0	1	2	≥ 3
pozorovaná četnost	29	15	10	10
teoretická pravděpodobnost	1/2	1/4	1/8	1/8
teoretická četnost	32	16	8	8
příspěvek ke kritériu χ^2	0.281	0.0625	0.5	0.5

Hodnotu kritéria $0.281 + 0.0625 + 0.5 + 0.5 = 1.3435$ porovnáme s kvantilem $q_{\chi^2(3)}(0.95) \doteq 7.815$ a nulovou hypotézu nezamítáme. □

Cvičení 12.3.16: Náhodná veličina X nabývá hodnot $1, 2, 3, 4, 5, 6$. Její realizací jsme dostali následující četnosti výsledků:

hodnota	1	2	3	4	5	6
pozorovaná četnost	2	3	5	13	10	12

Posuďte na hladině významnosti 5 % hypotézu, že náhodná veličina X má rozdělení, při němž pravděpodobnost je úměrná hodnotě, tj. $P[X = k] = c\,k$, kde c je konstanta.

Řešení: Konstantu c určíme z rovnice (nikoli z dat):

$$1 = \sum_k P[X = k] = 21\,c\,,$$

$c = 1/21$. Sloučíme první dvě skupiny, jinak by jejich teoretické četnosti byly příliš malé:

hodnota	$1-2$	3	4	5	6	celkem
pozorovaná četnost	5	5	13	10	12	45
teoretická pravděpodobnost	1/7	1/7	4/21	5/21	2/7	1
teoretická četnost	6.429	6.429	8.5711	10.714	12.857	45
příspěvek ke kritériu	0.317	0.317	2.288	0.048	0.057	3.028

Porovnáme hodnotu kritéria 3.028 s kvantilem $q_{\chi^2(4)}(0.95) \doteq 9.488$ a hypotézu nezamítáme. □

Cvičení 12.3.17: Realizací náhodné veličiny X jsme dostali následující četnosti výsledků:

hodnota	0	1	2	3	4	5	6
pozorovaná četnost	30	13	10	5	2	3	1

Posuďte na hladině významnosti 5 % hypotézu, že náhodná veličina X má geometrické rozdělení, $P[X = k] = c\,2^{-k}$, kde c je konstanta a $k = 0, 1, 2, 3, \ldots$.

Řešení: Konstantu c určíme z rovnice (nikoli z dat):

$$1 = \sum_k P[X = k] = 2c,$$

$c = 1/2$. Sloučíme výsledky „3 a více" do jedné skupiny, aby teoretické četnosti nebyly příliš malé:

hodnota	0	1	2	3 a více	celkem
pozorovaná četnost	30	13	10	1	64
teoretická pravděpodobnost	1/2	1/4	1/8	1/8	1
teoretická četnost	32	16	8	8	64
příspěvek ke kritériu	0.125	0.5625	0.5	1.125	2.3125

Porovnáme hodnotu kritéria 2.3125 s kvantilem $q_{\chi^2(3)}(0.95) \doteq 7.81$ a hypotézu nezamítáme. □

Cvičení 12.3.18: Agentura F zjistila u výběrového souboru četnosti volebních preferencí dle tabulky. Posuďte na 5 % hladině významnosti hypotézu, že preference odpovídají předpovědi agentury C (viz tabulka)[3]:

četnost dle F	187	192	82	61	58	94
pravděpodobnost dle C	0.308	0.242	0.121	0.058	0.086	0.185

Cvičení 12.3.19: Tabulka udává rozdělení (podmíněné) pravděpodobnosti, že volič strany zastoupené v parlamentu volil danou stranu. Posuďte na 5 % hladině významnosti hypotézu, že stejné rozdělení mají i poslanci.[4]

relativní preference	0.376	0.344	0.136	0.077	0.067
počet poslanců	81	74	26	13	6

Řešení: Doplníme tabulku (poslední sloupec uvádí celkový údaj):

relativní preference	0.376	0.344	0.136	0.077	0.067	1
počet poslanců	81	74	26	13	6	200
teor. četnost	75.2	68.8	27.2	15.4	13.4	200
příspěvek k χ^2	0.447	0.393	0.052	0.374	4.086	*5.353*

Hodnotu kritéria 5.353 porovnáme s kvantilem $q_{\chi^2(4)}(0.95) \doteq 9.4877$ a hypotézu nezamítáme (poněkud překvapivý závěr vzhledem k tomu, že poslední dvě strany mají téměř stejnou podporu u voličů, ale poslední má více než 2× méně poslanců). □

Cvičení 12.3.20: Posuďte na hladině významnosti 5 %, zda data dle následující tabulky četností odpovídají binomickému rozdělení s parametry $n = 7$, $p = 1/2$.

hodnota	0	1	2	3	4	5	6	7
četnost	3	10	15	35	40	15	5	5

Cvičení 12.3.21: Posuďte na hladině významnosti $\alpha = 5\,\%$ hypotézu, že data s následujícími četnostmi

hodnota	0	1	2	3	4	5	6	7
četnost	50	63	50	20	10	5	1	1

pocházejí (a) z Poissonova rozdělení s parametrem $w = 1.5$, (b) z jakéhokoli Poissonova rozdělení.

[3] Údaje z volebních předpovědí agentur Factum a SC&C pro parlamentní volby v ČR 2006.
[4] Údaje z parlamentních voleb v ČR 2006.

Cvičení 12.3.22: Posuďte na hladině významnosti $\alpha = 5\,\%$ hypotézu, že data s následujícími četnostmi

hodnota	0	1	2	3	4	5	6	7
četnost	40	30	10	10	7	1	0	2

pocházejí (a) z geometrického rozdělení s parametrem $w = 0{,}5$, (b) z jakéhokoli geometrického rozdělení.

12.3.1 χ^2-test dobré shody dvou rozdělení

(Dle [Mood et al. 1974].)

Příklad 12.3.23: Máme otestovat rozšířenou domněnku, že auta vyrobená v pondělí nebo v pátek mají více závad a menší spolehlivost než auta vyrobená v jiné dny. *(Za nulovou hypotézu zvolíme, že tomu tak není; pokud nulovou hypotézu zamítneme, původní podezření se potvrdí.)*

Příklad 12.3.24: Dvě agentury odhadují volební preference. Máme posoudit, zda rozdíly jejich odhadů lze vysvětlit statistickou chybou. Máme tedy testovat hypotézu, že oba náhodné výběry pocházejí ze *stejného neznámého* rozdělení.

(Podle předchozí kapitoly nemůžeme porovnávat preference stran, pro které hlasovalo velmi málo respondentů; je potřeba sdružit strany do skupin, které mají dostatečně velkou četnost.)

Zatímco v předchozí kapitole jsme testovali shodu náhodného výběru *s daným rozdělením*, nyní porovnáváme *s jiným náhodným výběrem*, který dává jen zprostředkovanou informaci o *neznámém rozdělení*. Vycházíme z realizací dvou náhodných výběrů $\boldsymbol{x} = (x_1, \ldots, x_m)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$. Mohli bychom jeden výběr testovat na shodu s druhým výběrem, např. testovat $\boldsymbol{x}$ na shodu s empirickým rozdělením $\mathrm{Emp}\,(\boldsymbol{y})$ (nebo naopak) postupem z předchozí kapitoly jako v příkladech 12.3.18 a 12.3.19. Je však možno testovat přímo shodu obou náhodných výběrů. V tom případě testujeme nulovou hypotézu, že dvě diskrétní náhodné veličiny mají stejné diskrétní rozdělení. Toto rozdělení z nich odhadneme, poté otestujeme shodu obou výběrů s odhadnutým rozdělením.

Rozsahy výběrů jsou m, resp. n; četnost hodnoty j označme m_j, resp. n_j $(j = 1, \ldots, k)$. Předpokládáme rozdělení s neznámými teoretickými pravděpodobnostmi p_j $(j = 1, \ldots, k)$. O rozdělení testovacích statistik z χ^2-testu dobré shody víme následující:

$$\sum_{j=1}^{k} \frac{(m_j - m\,p_j)^2}{m\,p_j} \text{ se blíží } \chi^2(k-1)\,,$$

$$\sum_{j=1}^{k} \frac{(n_j - n\,p_j)^2}{n\,p_j} \text{ se blíží } \chi^2(k-1)\,,$$

$$T = \sum_{j=1}^{k} \frac{(m_j - m\,p_j)^2}{m\,p_j} + \sum_{j=1}^{k} \frac{(n_j - n\,p_j)^2}{n\,p_j} \text{ se blíží } \chi^2(2\,(k-1))\,.$$

Neznámé parametry p_j odhadneme metodou maximální věrohodnosti,

$$p_j = \frac{m_j + n_j}{m+n}\,;$$

z nich je $k-1$ nezávislých (neboť $\sum_{j=1}^{k} p_j = 1$), takže výsledný počet stupňů volnosti je

$$2\,(k-1) - (k-1) = k-1\,.$$

Testujeme realizaci t na rozdělení $\chi^2(k-1)$. Nulovou hypotézu zamítáme pro $t > q_{\chi^2(k-1)}(1-\alpha)$, pro dosaženou významnost $1 - F_{\chi^2(k-1)}(t) < \alpha$.

Praktičtější vzorec pro výpočet téže testovací statistiky je

$$t = (1/m + 1/n) \sum_{j=1}^{k} \frac{(m_j - m\,p_j)^2}{p_j}\,.$$

Jeho výhodou je, že sčítáme jen přes jeden z výběrů, druhý se projevil prostřednictvím odhadů pravděpodobností p_j.

Cvičení 12.3.25: Z daných údajů o účasti a úspěšnosti studentů v jednotlivých termínech zkoušek posuďte na hladině významnosti 5 % hypotézu, že pravděpodobnost úspěchu u zkoušky je konstanta nezávislá na termínu.[5]

termín	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	celkem
přišli	9	7	11	7	12	13	15	8	9	16	7	18	9	8	149
udělali	4	5	10	3	9	7	8	5	5	14	4	16	9	6	105

12.3.2 χ^2-test nezávislosti dvou rozdělení

(Dle [Likeš, Machek 1988].)

Příklad 12.3.26: Máme podezření, že dva po sobě následující výsledky generátoru náhodných čísel jsou závislé. Jak to prokážeme?

Často jsme používali předpoklad nezávislosti dvou náhodných veličin. I ten lze testovat vhodně modifikovaným χ^2-testem. Nulová hypotéza je, že dva náhodné výběry pocházejí z nezávislých diskrétních rozdělení (která neznáme).

Opět nám stačí četnosti. Předpokládejme, že první náhodný výběr nabývá k hodnot s pravděpodobnostmi $p_1, \ldots, p_k$, druhý m hodnot s pravděpodobnostmi $q_1, \ldots, q_m$. Realizace dvojrozměrného náhodného výběru $((x_1, y_1), \ldots, (x_n, y_n))$ obsahuje dvojice realizací náhodných veličin X, Y; z výsledků nás zajímají opět pouze četnosti n_{ij} $(i = 1, \ldots, k;\ j = 1, \ldots, m)$. Ty bývají uspořádány do tzv. **kontingenční tabulky**. Počet tříd je $k\,m$.

Za předpokladu nezávislosti jsou pravděpodobnosti výsledků $p_i\,q_j$ $(i = 1, \ldots, k;\ j = 1, \ldots, m)$, rozdělení testovací statistiky

$$T = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(n_{ij} - n\,p_i\,q_j)^2}{n\,p_i\,q_j}$$

se blíží $\chi^2(k\,m - 1)$. Neznámé parametry p_i, q_j odhadneme pomocí maxima věrohodnosti,

$$p_i = \frac{1}{n} \sum_{j=1}^{m} n_{ij}, \qquad q_j = \frac{1}{n} \sum_{i=1}^{k} n_{ij}\,;$$

z nich je jen $(k-1) + (m-1)$ nezávislých (neboť $\sum_{i=1}^{k} p_i = 1$, $\sum_{j=1}^{m} q_j = 1$), takže výsledný počet stupňů volnosti je $k\,m - 1 - (k-1) - (m-1) = (k-1)(m-1)$. Testujeme realizaci t na $\chi^2((k-1)\,(m-1))$. Nulovou hypotézu zamítáme pro

$$t > q_{\chi^2((k-1)\,(m-1))}(1-\alpha)\,, \qquad \text{tj. } 1 - F_{\chi^2((k-1)\,(m-1))}(t) < \alpha\,.$$

Testy dobré shody dovolují zjistit, jaké má náhodná veličina rozdělení (přesněji, jaké rozdělení *nemá*, neboť přímé potvrzení statistické metody nedovolují). Lze je použít i k testům souvislostí mezi dvěma výběry. To může být součástí ověření předpokladů jiných metod nebo přímo cílem výzkumu, který hledá vhodný pravděpodobnostní model pro daný pokus.

[5] Jedná se o skutečná data z předmětu Matematika 6F v roce 2004.

12.4 Korelace, její odhad a testování

(Dle [Likeš, Machek 1988].)

Příklad 12.4.1: Máme posoudit, zda výskyt rakoviny kůže souvisí s tloušťkou ozónové vrstvy.

Často si klademe otázku, zda dvě náhodné veličiny spolu souvisí. Jejich závislost může mít mnoho podob, jednu z nich prokážeme testem korelace.

Na základě dvojrozměrného náhodného výběru $((X_1, Y_1), \ldots, (X_n, Y_n))$ (viz poz. 8.2.5) vypočteme **výběrový koeficient korelace**

$$R_{X,Y} = \frac{\sum_{j=1}^{n} (X_j - \overline{X})\,(Y_j - \overline{Y})}{\sqrt{\left(\sum_{j=1}^{n} (X_j - \overline{X})^2\right)\left(\sum_{j=1}^{n} (Y_j - \overline{Y})^2\right)}}\,.$$

Jeho realizace

$$r_{\boldsymbol{x},\boldsymbol{y}} = \frac{\sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}})\,(y_j - \bar{\boldsymbol{y}})}{\sqrt{\left(\sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}})^2\right)\left(\sum_{j=1}^{n} (y_j - \bar{\boldsymbol{y}})^2\right)}}$$

je číslo z intervalu $\langle -1, 1\rangle$, neboť je to kosinus úhlu vektorů

$$(x_1 - \bar{\boldsymbol{x}}, \ldots, x_n - \bar{\boldsymbol{x}})\,, (y_1 - \bar{\boldsymbol{y}}, \ldots, y_n - \bar{\boldsymbol{y}}) \in \mathbb{R}^n\,.$$

Výběrový koeficient korelace se používá k odhadu (koeficientu) korelace.

Pro výpočet se používá následující vzorec, který je jednoprůchodový:

Věta 12.4.2:

$$r_{\boldsymbol{x},\boldsymbol{y}} = \frac{n \sum_{j=1}^{n} x_j\, y_j - \left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} y_j\right)}{\sqrt{\left(n \sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2\right)\left(n \sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2\right)}} = \tag{12.2}$$

$$= \frac{n}{n-1}\,\frac{\frac{1}{n}\sum_{j=1}^{n} x_j\, y_j - \bar{\boldsymbol{x}}\,\bar{\boldsymbol{y}}}{s_{\boldsymbol{x}}\, s_{\boldsymbol{y}}}\,.$$

Důkaz:

$$\sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}})\,(y_j - \bar{\boldsymbol{y}}) = \sum_{j=1}^{n} x_j\, y_j - \bar{\boldsymbol{x}} \sum_{j=1}^{n} y_j - \bar{\boldsymbol{y}} \sum_{j=1}^{n} x_j + n\,\bar{\boldsymbol{x}}\,\bar{\boldsymbol{y}} =$$

$$= \sum_{j=1}^{n} x_j\, y_j - n\,\bar{\boldsymbol{x}}\,\bar{\boldsymbol{y}} = \sum_{j=1}^{n} x_j\, y_j - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} y_j\right),$$

$$\sum_{j=1}^{n} (x_j - \bar{\boldsymbol{x}})^2 = \sum_{j=1}^{n} x_j^2 - 2\,\bar{\boldsymbol{x}} \sum_{j=1}^{n} x_j + n\,\bar{\boldsymbol{x}}^2 = \sum_{j=1}^{n} x_j^2 - n\,\bar{\boldsymbol{x}}^2 = \sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2$$

a podobně pro y. □

K použití vzorce (12.2) stačí střádat v 6 registrech následující údaje:

$$n\,,\ \sum_{j=1}^{n} x_j\,,\ \sum_{j=1}^{n} y_j\,,\ \sum_{j=1}^{n} x_j^2\,,\ \sum_{j=1}^{n} y_j^2\,,\ \sum_{j=1}^{n} x_j\,y_j\,.$$

Takto se počítá výběrový koeficient korelace i v kalkulátorech, které nemají paměť na uložení všech dat.

Příklad 12.4.3: Doby ledové v minulosti významně korelovaly s poklesy koncentrací kysličníku uhličitého v atmosféře. Nicméně dosud si odborníci netroufají vyjadřovat se k tomu, který z těchto jevů je příčinou druhého (navíc nemusí být pravda ani jedno z toho).

Příklad 12.4.4: Průzkumy z vyspělých zemí ukazují, že bohatí lidé se dožívají vyššího věku. Není jasné, zda je to tím, že si dopřávají lepší výživu a zdravotní péči, nebo se na jejich bohatství podílí lepší zdraví. Také může být společná příčina v prostředí, z něhož lidé pocházejí, může korelovat péče o majetek i zdraví. Je obtížné navrhnout (a ještě obtížnější provést) test, který by tyto vlivy rozlišil. Nicméně korelace ukazuje, že *nějaká* souvislost (velmi pravděpodobně) existuje.

Příklad 12.4.5: Vyskytly se práce, potvrzující zvýšený počet sebevražd v letech maxima sluneční aktivity. Ty se však nepotvrdily při zkoumání za větší časové období. Původní potvrzení závislosti způsobilo maximum sluneční aktivity v roce 1929, současně s hospodářskou krizí, kterou o rok později následovala vlna sebevražd.[6]

Testy korelace jsou vděčným zdrojem zajímavých postřehů a vedly k poznání pozoruhodných souvislostí. Nicméně nedávají odpověď, zda statisticky významná korelace dvou jevů může být způsobena tím, že jeden je příčinou druhého, nebo naopak, nebo jsou oba vyvolány společnou příčinou. Navíc se může stát, že pouze oba měly např. podobný časový průběh a nemáme dost dat na to, aby se ukázalo, že to byla jen náhoda. Rozhodnout mezi těmito možnostmi může být velmi obtížné a statistika pro to nedává přímo použitelné nástroje.

12.4.1 Test nekorelovanosti dvou výběrů z normálních rozdělení

Předpokládáme, že dvojrozměrná náhodná veličina (X, Y) má (dvojrozměrné) normální rozdělení. Na základě dvojrozměrného náhodného výběru rozsahu n, $n \geq 3$, testujeme nulovou hypotézu, že X, Y jsou nekorelované, tj. korelace $\varrho_{X,Y} = 0$.

Testovací statistika je odvozena od výběrového koeficientu korelace:

$$T = \frac{R_{\boldsymbol{X},\boldsymbol{Y}}\sqrt{n-2}}{\sqrt{1-R^2_{\boldsymbol{X},\boldsymbol{Y}}}}\,.$$

Použitá nelineární transformace

$$u \mapsto \frac{u\sqrt{n-2}}{\sqrt{1-u^2}}$$

zajišťuje, že za předpokladu nekorelovanosti má testovací statistika Studentovo rozdělení $\mathrm{t}(n-2)$, na které ji testujeme dle kapitoly 12.2.1.

Testy korelace jsou nejběžnějším (ale zdaleka ne jediným) nástrojem pro odhalení závislosti dvou náhodných veličin. S rozvojem výpočetní techniky lze snadno testovat závislosti velmi mnoha parametrů, což vedlo k řadě překvapivých poznatků. (Je ovšem nutno je ověřovat opakovanými testy, neboť z metodiky testování hypotéz a pravděpodobnosti chyby 1. druhu vyplývá, že z velkého počtu testů na nezávislost některé zamítneme neprávem.)

[6] Dvořák, J.: *Země, lidé a katastrofy*. Naše vojsko, Praha, 1987.

Cvičení 12.4.6: Nulová hypotéza je, že dvě náhodné veličiny jsou nekorelované. Na výběru rozsahu $n_1 = 30$ vyšla korelace 0.15. Poté bylo provedeno ověření na výběru rozsahu $n_2 = 300$ se stejným výsledkem. Posuďte statistickou významnost výsledků.

Řešení: Testujeme hodnotu

$$\frac{0.15\sqrt{n_j - 2}}{\sqrt{1 - 0.15^2}}$$

na rozdělení $\mathrm{t}(n_j - 2)$, $j = 1, 2$. V prvním případě je kritérium 0.803 a kritická hodnota na 5 %-ní hladině významnosti $q_{\mathrm{t}(28)}(0.975) \doteq 2.05$, nulovou hypotézu nezamítáme. Ve druhém případě je kritérium 2.619, kritickou hodnotu $q_{\mathrm{t}(298)}(0.975)$ odhadneme z normálního rozdělení, $q_{\mathrm{N}(0,1)}(0.975) = \Phi^{-1}(0.975) \doteq 1.96$, překročena je i kritická hodnota na 1 %-ní hladině významnosti $\Phi^{-1}(0.995) \doteq 2.58$, nulovou hypotézu zamítáme. Protože $\Phi(2.619) \doteq 0.9956$, je dosažená významnost $2\,(1 - 0.9956) = 0.0088 < 1\,\%$. V prvním případě bychom potřebovali hodnotu $F_{\mathrm{t}(28)}(0.803) \doteq 0.786$ (která není v tabulkách), dosažená významnost byla $2\,(1 - 0.786) = 0.428$. Nicméně i z odhadu pomocí normálního rozdělení $\Phi(0.803) \doteq 0.788$ je zřejmé, že první výsledek zdaleka nebyl statisticky významný.

Cvičení 12.4.7: Máme posoudit, zda existuje souvislost mezi počtem obyvatel na jednoho lékaře a dětskou úmrtností (počet úmrtí na 1000 dětí do 4 let). K dispozici je tabulka údajů z $n = 41$ asijských zemí z r. 2000.[7]

Řešení: Metoda předpokládá normální rozdělení. Údaje jsou však zaručeně kladné a mají velký rozptyl, takže jejich rozdělení není normální. Mohli bychom snad předpokládat, že se jedná o logaritmickonormální rozdělení. V tom případě logaritmy z těchto údajů budou mít normální rozdělení a metodu na ně můžeme použít. (Není podstatné, jaký základ logaritmů použijeme, neboť tomu odpovídá po zlogaritmování lineární transformace, která nemá na korelaci vliv.) V posledních dvou sloupcích jsou dekadické logaritmy původních dat.

Vypočítali jsme realizace výběrových průměrů a výběrových směrodatných odchylek obou veličin (viz tabulka). Kromě toho potřebujeme

$$\frac{1}{n}\sum_{j=1}^{n} x_j\, y_j \doteq 5.638\,.$$

Dosazením do vzorce dostaneme

$$r_{x,y} = \frac{n}{n-1}\cdot\frac{\frac{1}{n}\sum_{j=1}^{n} x_j\, y_j - \bar{x}\,\bar{y}}{s_x\, s_y} \doteq \frac{41}{40}\cdot\frac{5.638 - 3.365\cdot 1.644}{0.5\cdot 0.427} \doteq 0.51\,.$$

Transformované kritérium

$$\frac{0.51\sqrt{41-2}}{\sqrt{1-0.51^2}} \doteq 3.7$$

převyšuje i kritickou hodnotu $q_{\mathrm{t}(39)}(0.9995) \doteq q_{\mathrm{t}(40)}(0.9995) \doteq 3.55$ pro hladinu významnosti 0.1 %, dosažená významnost vychází přibližně $6.6\cdot 10^{-4}$. Nulovou hypotézu, že dětská úmrtnost má nulovou korelaci s počtem lékařů, můžeme tedy zamítnout, a to přesto, že o správnosti některých údajů lze pochybovat (srovnejte např. Irák a Kuvajt).

[7] Data dle Peška, M.: Semestrální práce z Matematiky 6F, FEL ČVUT, 2001, převzata z `http://www.mujweb.cz/www/jpdepot`

Tabulka 12.1. Údaje o počtu lékařů a dětské úmrtnosti

Země	u: obyv. na 1 lékaře	v: dětská úmrtnost	$x = \log u$	$y = \log v$
Irák	177	122	2.248	2.086
Mongolsko	371	71	2.569	1.851
Libanon	537	40	2.730	1.602
Japonsko	556	6	2.745	0.778
Kypr	568	10	2.754	1.000
Jordánsko	615	25	2.789	1.398
Čína	637	47	2.804	1.672
Katar	667	21	2.824	1.322
Singapur	709	4	2.851	0.602
Saúdská Arábie	749	30	2.874	1.477
Bahrajn	760	22	2.881	1.342
Jižní Korea	817	7	2.912	0.845
Turecko	916	47	2.962	1.672
Omán	1131	18	3.053	1.255
Spojené arabské emiráty	1190	18	3.076	1.255
Sýrie	1198	34	3.078	1.531
Pákistán	1930	136	3.286	2.134
Taiwan	2146	32	3.332	1.505
Malajsie	2226	13	3.348	1.114
Vietnam	2287	44	3.359	1.643
Indie	2459	111	3.391	2.045
Severní Korea (KLDR)	2625	43	3.419	1.633
Írán	3000	37	3.477	1.568
Laos	4280	128	3.631	2.107
Thajsko	4288	38	3.632	1.580
Hong Kong	4861	68	3.687	1.833
Bangladéš	4970	112	3.696	2.049
Kuvajt	5000	14	3.699	1.146
Brunej	5340	82	3.728	1.914
Macao	5632	93	3.751	1.968
Indonésie	5959	71	3.775	1.851
Bhútán	6250	127	3.796	2.104
Srí Lanka	6843	19	3.835	1.279
Afghánistán	7001	257	3.845	2.410
Izrael	7355	73	3.867	1.863
Filipíny	8424	38	3.926	1.580
Kambodža	9544	170	3.980	2.230
Jemen	9570	105	3.981	2.021
Barma	12756	150	4.106	2.176
Nepál	13568	116	4.133	2.064
Maledivy	14333	76	4.156	1.881
výb. průměr	**4006**	**65.2**	**3.365**	**1.644**
výb. směr. odchylka	**3849**	**54.7**	**0.500**	**0.427**

12.5 Neparametrické testy

Někdy se setkáme s rozdělením, které se nepodobá žádnému z běžně používaných rozdělení, takže je neumíme jednoduše popsat. I přesto se můžeme pokusit o testy některých vlastností rozdělení. Takové **testy** nazýváme **neparametrické**. Nevyžadují znalost typu rozdělení, jsou však slabší, neboť vycházejí ze slabších předpokladů. Zde vybrané testy nám také poslouží jako ukázky, že ne vždy musí být rozdělení statistiky spojité.

12.5.1 Znaménkový test

V tomto testu rozlišujeme pouze znaménko odchylky od zvolené hodnoty c. Tím ztrácíme kvantitativní informaci a tedy i možnost testovat např. střední hodnotu. Místo ní testujeme medián $q_X(1/2)$.

Příklad 12.5.1: Ptáme se respondentů, zda je jejich plat větší nebo menší než zvolené číslo c. Pokud je c medián, lze očekávat, že obě odpovědi budou přibližně stejně časté. (Pro jednoduchost neuvažujeme, že může nastat rovnost.)

Před znaménkovým testem z výběru předem vyloučíme nulové odchylky. Při platnosti nulové hypotézy $q_X(1/2) = c$ by kladné i záporné odchylky měly být stejně pravděpodobné. Testovací statistikou T je počet kladných odchylek, který by měl mít binomické rozdělení $\mathrm{Bin}(n, 1/2)$. Nulovou hypotézu zamítáme, pokud realizace t nebude v intervalu

$$\left\langle q_{\mathrm{Bin}\left(n,\frac{1}{2}\right)}(\alpha/2)\,;\, q_{\mathrm{Bin}\left(n,\frac{1}{2}\right)}(1-\alpha/2)\right\rangle .$$

Pro jednostranné testy bychom dostali intervaly

$$\left(-\infty, q_{\mathrm{Bin}\left(n,\frac{1}{2}\right)}(1-\alpha)\right\rangle ,$$
$$\left\langle q_{\mathrm{Bin}\left(n,\frac{1}{2}\right)}(\alpha)\,,\,\infty\right) .$$

Výpočet kvantilů je pracný, ale kritické hodnoty jsou tabelovány (v závislosti na n a α). Dosažená významnost se počítá o trochu snáze.

Pro velká n používáme centrální limitní větu a testujeme

$$T_0 = \frac{2T - n}{\sqrt{n}}$$

na $\mathrm{N}(0,1)$.

Příklad 12.5.2: Odhad smrtelné dávky látky lze těžko testovat jinak než znaménkovým testem. Totéž platí pro některé další destruktivní testy.

Poznámka 12.5.3: Znaménkový test je jednoduchý nejen výpočetně, ale i z hlediska pořizování údajů. Např. na otázku na výši platu by mnozí respondenti odmítli odpovědět. Pokud však chceme zjistit medián platů, můžeme jej odhadnout a svůj odhad prověřit znaménkovým testem jako v příkladu 12.5.1. Lze očekávat, že na takovou otázku by nám odpovědělo více oslovených respondentů a neměli by s ní problém ti, jejichž plat je vysoce nadprůměrný.

12.5.2 Wilcoxonův test (jednovýběrový)

Tímto testem lze posoudit hypotézu, že náhodná veličina má rozdělení symetrické kolem zvolené hodnoty c. (V tom případě je c mediánem i střední hodnotou.) Přitom nezáleží na tom, jaké symetrické rozdělení to je.

Z realizace $(x_1, \ldots, x_n)$ vypočteme posloupnost $(z_1, \ldots, z_n)$, kde $z_j = x_j - c$. Seřadíme ji vzestupně podle absolutních hodnot $|z_j| = |x_j - c|$, čímž j-tému prvku přiřadíme pořadí r_j. Je-li

více stejných rozdílů, přiřadíme jim stejné pořadí rovné aritmetickému průměru odpovídajících pořadí. Testovací statistikou je

$$T_1 = \sum_{j:z_j>0} r_j \,.$$

Její realizaci porovnáme s tabulkou kritických hodnot pro tento test. Často se vyskytuje modifikace, kdy kritériem je

$$T_2 = \min\Big(\sum_{j:z_j>0} r_j, \sum_{j:z_j<0} r_j \Big) \,,$$

v tabulkách je vždy uvedeno, ke které variantě se vztahují.

Wilcoxonův test je příkladem celé řady testů, které používají data seřazená podle velikosti a záleží jen na jejich pořadí v posloupnosti. Podobné principy se používají v testech parametrů pořadových náhodných veličin.

Příklad 12.5.4: Ještě v polovině 19. století postihovaly Londýn epidemie cholery. Lékař John Snow (1813–1858) si všiml ulice, v níž na jedné straně bylo mnoho případů, na druhé jediný. To bylo velmi nepravděpodobné, vyjdeme-li z tehdy převládajícího názoru, že choleru způsobuje špatné ovzduší, zápach a mlhy. Snow dospěl k závěru, že příčinou je špatná voda, neboť každá strana ulice byla zásobována vodou z jiného zdroje. Po uzavření podezřelé pumpy případů cholery ubylo. Konečné řešení přinesla teprve inovace kanalizačního systému.

Neškodí si připomenout, že ještě v tak nedávné době neměli lidé tušení o příčinách nemocí, které je sužovaly. Zde stačil k odhalení závislosti úsudek a prokázal by ji velmi jednoduchý test. Mnohé jiné závislosti nejsou tak zjevné a jejich odhalení si žádá propracovanější statistické metody, z nichž některé jsme zde popsali. Ty mají svůj podíl na tom, že dnes o příčinách jevů víme mnohem víc a můžeme účelným jednáním předcházet nesnázím.

Část III

Přílohy

A. Příklady pro opakování

Zde jsou příklady, které se tématicky váží k více kapitolám, takže nemohly být zařazeny jen k jednomu tématu. Mohou sloužit k opakování látky a procvičení souvislostí. Bohatou zásobu cvičení lze najít v [Hsu 1996, Spiegel et al. 2000]. Analýze pokročilejších příkladů je věnována publikace [Anděl 2000].

Cvičení A.1.1: Je možné, aby dva jevy byly současně nezávislé a neslučitelné?

Řešení: Pokud existují dva takové jevy A, B, pak $P(A \cup B)$ vychází z neslučitelnosti $P(A)+P(B)$ a z nezávislosti $P(A)+P(B)-P(A)\cdot P(B)$, tedy $P(A)\cdot P(B)=0$, tj. aspoň jeden z jevů A, B má pravděpodobnost 0. Dle tvrzení 1.5.9 je tato podmínka spolu s neslučitelností i postačující; libovolné dva neslučitelné jevy, z nichž aspoň jeden má nulovou pravděpodobnost, jsou nezávislé. (Jev s nulovou pravděpodobností nemusí být nemožný a nemusí být neslučitelný s libovolným jiným jevem, takže neslučitelnost nelze v učiněném závěru vynechat.) □

Cvičení A.1.2: Náhodná veličina X má exponenciální rozdělení s hustotou

$$f_X(u) = \begin{cases} \mathrm{e}^{-u} & \text{pro } u \geq 0, \\ 0 & \text{jinak.} \end{cases}$$

Určete střední hodnotu, rozptyl, a znázorněte rozdělení náhodných veličin (a) $X+2$, (b) $-2X$, (c) X^2.

Cvičení A.1.3: Rozhodněte, zda může existovat náhodná veličina X taková, že $q_X(0.9) < \mathrm{E}X$ (90 %-ní kvantil je menší než střední hodnota).

Řešení: To samozřejmě lze, protože kvantil vypovídá jen o pořadových vlastnostech rozdělení, nikoli o vzdálených hodnotách, které zásadně ovlivňují střední hodnotu. Příkladem může být alternativní rozdělení s hodnotami $0, 1$, nabývající 0 s pravděpodobností větší než 0.9. Pak 90 %-ní kvantil je 0, střední hodnota je kladná. □

Cvičení A.1.4: Známe-li medián náhodné veličiny X, co lze říci o mediánu jejího kvadrátu, X^2?

Řešení: Pokud X nemění znaménko a kvantilová funkce je v $1/2$ spojitá, pak je to kvadrát mediánu,

$$q_{X^2}\left(\tfrac{1}{2}\right) = \left(q_X\left(\tfrac{1}{2}\right)\right)^2,$$

neboť kvadratická funkce zúžená na obor hodnot X je monotónní. □

Cvičení A.1.5: Náhodná veličina X nabývá hodnot z intervalu $\langle 0, 100\rangle$, její medián je $q_X(1/2) = 70$. Co lze říci o její střední hodnotě a rozptylu?

Řešení: Jelikož $X \leq 70$ s pravděpodobností aspoň $1/2$ a ve zbývajících případech je $X \leq 100$, musí být $\mathrm{E}X \leq (70+100)/2 = 85$. Podobně dostaneme dolní odhad $\mathrm{E}X \geq 70/2 = 35$. Střední hodnota je integrál kvantilové funkce, která je omezena po částech konstantními funkcemi s uvedenými integrály.

Rozptyl bude zřejmě maximální, bude-li náhodná veličina nabývat co možno nejčastěji jen své extrémní hodnoty, 0 a 100, a to s pravděpodobnostmi danými střední hodnotou. Podle vzorce

pro rozptyl alternativního rozdělení (zde vynásobeného 100) je bez omezení mediánu rozptyl maximálně $100^2\ (1 - \mathrm{E}X)\ \mathrm{E}X$, největší může být pro $\mathrm{E}X = 1/2$, a to $100^2/4 = 2500$. V tom případě by podmínka na medián nebyla splněna, ta však má jen lokální vliv. Můžeme vzít např. rozdělení s distribuční funkcí

$$F_X(u) = \begin{cases} 0 & \text{pro } u < 0, \\ \frac{\exp(k\,(u-70))}{2} & \text{pro } 0 \le u < 70, \\ 1 - \frac{\exp(k\,(70-u))}{2} & \text{pro } 70 \le u < 100, \\ 1 & \text{pro } u \ge 100. \end{cases}$$

Zvolíme-li kladnou konstantu k dostatečně velkou, rozptyl bude libovolně blízký vypočtené mezi, vždy však $\mathrm{D}X < 2500$. □

Cvičení A.1.6: Náhodná veličina X nabývá hodnot z intervalu $\langle 0, 100\rangle$ a má kvantil $q_X(0.7) = 50$. Co lze říci o její střední hodnotě a rozptylu?

Cvičení A.1.7: Náhodná veličina X nabývá hodnot z intervalu $\langle 0, 10\rangle$, její střední hodnota je $\mathrm{E}X = 1.5$ a medián $q_X(1/2) = 4$. Co lze říci o jejím rozptylu?

Řešení: Podle výsledku příkladu A.1.5 taková náhodná veličina neexistuje. □

Cvičení A.1.9: Náhodné veličiny X, Y jsou nezávislé, mají spojitá rovnoměrná rozdělení; X na intervalu $\langle 0, 1\rangle$, Y na intervalu $\langle 1, 2\rangle$. Popište a znázorněte rozdělení náhodných veličin (a) $X+Y$, (b) $\mathrm{Mix}_{1/2}(X, Y)$ (c) $X + \mathrm{E}Y$.

Cvičení A.1.10: Opakujte příklad A.1.9, má-li náhodná veličina Y alternativní rozdělení,

$$p_Y(t) = \begin{cases} 1/2 & \text{pro } t \in \{0, 1\}, \\ 0 & \text{jinak.} \end{cases}$$

Cvičení A.1.11: Náhodné veličiny X, Y jsou nezávislé, mají diskrétní rovnoměrná rozdělení; X na $\{0, 1\}$, Y na $\{0, 1, 2\}$. Popište a znázorněte rozdělení náhodných veličin (a) $X/2 + Y/2$, (b) $\mathrm{Mix}_{1/2}(X, Y)$, (c) $X/2 + \mathrm{E}Y/2$.

Řešení: (a) $X/2 + Y/2$ nabývá hodnot 0, 1/2, 1, 3/2 s pravděpodobnostmi po řadě 1/6, 1/3, 1/3, 1/6 (pouze zde potřebujeme nezávislost), (b) $\mathrm{Mix}_{1/2}(X, Y)$ nabývá hodnot 0, 1, 2 s pravděpodobnostmi po řadě 5/12, 5/12, 1/6, (c) jelikož $\mathrm{E}Y = 1$, $X/2 + \mathrm{E}Y/2$ nabývá hodnot 1/2, 1 s pravděpodobností 1/2. □

Cvičení A.1.12: Najděte 95 %-ní horní intervalový odhad náhodné veličiny Z, která má rozdělení $\chi^2(500)$. Jelikož toto rozdělení není v tabulkách, použijte k jeho aproximaci centrální limitní větu.

Cvičení A.1.13: Nezávislé náhodné veličiny X, Y, Z, W mají po řadě normální rozdělení $\mathrm{N}(2, 3)$, $\mathrm{N}(0, 1), \mathrm{N}(0, 1), \mathrm{N}(5, 1)$. Určete (a) rozdělení náhodných veličin $X + Y$, $X - W$, Z^2, $Y^2 + Z^2$, (b) střední hodnotu směsi náhodných veličin $\mathrm{Mix}_{1/2}(X, Y)$, (c) střední hodnotu náhodné veličiny $X \cdot W$.

Cvičení A.1.14: Nezávislé náhodné veličiny A, B mají normované normální rozdělení $\mathrm{N}(0, 1)$. Určete a znázorněte (a) $3A + 2$, (b) $A + B$, (c) $A^2 + B^2$.

Cvičení A.1.15: Odvoďte charakteristickou funkci rozdělení $\chi^2(\eta)$ (vzorec (B.1)).

Řešení: Podle věty 4.7.14 a definice rozdělení χ^2 je

$$\Psi_{\chi^2(\eta)} = \Psi^{\eta}_{\chi^2(1)}\,.$$

Stačí najít charakteristickou funkci pro jednu hodnotu η. Pro $\eta = 1$ bychom si museli poradit s určitým integrálem, jemuž příslušná primitivní funkce je transcendentní. Snazší je případ $\eta = 2$, což je exponenciální rozdělení (viz tvrzení B.2.2), které má charakteristickou funkci

$$\Psi_{\chi^2(2)}(\omega) = \frac{1}{1 - 2\,\mathrm{i}\,\omega}\,.$$

□

Cvičení A.1.16: Kolik protipříkladů je třeba najít, abychom mohli na hladině významnosti 5 % zamítnout hypotézu, že (alespoň) „99 % studentů nebude statistiku potřebovat“[1]? Studentů bylo 280.

Řešení: Jde (v mezním případě vyhovujícím nulové hypotéze) o náhodnou veličinu X s binomickým rozdělením s parametry $n = 280$, $p = 0.01$.

Pomocí centrální limitní věty: $\mathrm{E}X = np = 2.8$, $\mathrm{D}X = np(1-p) = 2.772$, $\sigma_X = \sqrt{np(1-p)} \doteq 1.664\,9$. Jednostranný horní odhad je (s užitím kvantilu $\Phi^{-1}(0.95) \doteq 1.645$)

$$\mathrm{E}X + \sigma_X\,\Phi^{-1}(0.95) \doteq 2.8 + 1.664\,9 \cdot 1.645 \doteq 5.538\,8\,,$$

takže k vyvrácení nulové hypotézy by mělo stačit nalezení 6 studentů z tohoto ročníku, kteří statistiku potřebují. Použití centrální limitní věty je však zpochybněno malým počtem studentů v této skupině.

Přesný výpočet: Pravděpodobnost výsledku méně než t je

$$F_X(t) = \sum_{k=0}^{t-1} \binom{n}{k} p^k\,(1-p)^{n-k}\,,$$

konkrétně $1 - F_X(6) \doteq 6.414\,6 \cdot 10^{-2}$, $1 - F_X(7) \doteq 2.375\,9 \cdot 10^{-2}$, $1 - F_X(8) \doteq 7.791\,0 \cdot 10^{-3}$, takže nalezení 6 studentů nestačí, 7 ano, 8 by stačilo i na hladině významnosti 1 %. □

Cvičení A.1.17: V populaci je pravděpodobnost q, že osoba má srpkovou anémii. Ta je způsobena vadným genem, který každé dítě zdědí s pravděpodobností 1/2. Pokud má dítě po obou rodičích vadné geny, nepřežije; pokud má jeden gen vadný, má opět srpkovou anémii. Jaká je pravděpodobnost, že žijící jedinec v další generaci má srpkovou anémii? Znázorněte závislost na q. Za jakých předpokladů výsledek platí?

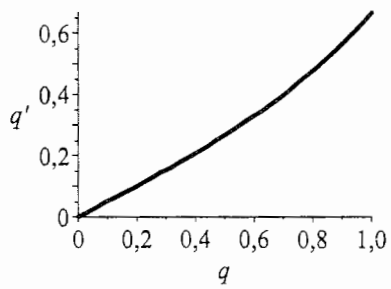
Řešení: Dítě má s pravděpodobností q^2 oba rodiče nemocné a s pravděpodobností $q^2/4$ zdědí oba vadné geny, pak ale nepřežije a nepočítá se. Pravděpodobnost, že má oba rodiče nemocné a zdědí jeden vadný gen, je $q^2/2$. Pravděpodobnost, že má jednoho rodiče nemocného, je $q\,(1-q)$, a že v tom případě zdědí jeden vadný gen, $q\,(1-q)\,/2$. Prozkoumali jsme dvě situace, které se navzájem vylučují, jejich pravděpodobnost je tedy součet $q^2/2 + q\,(1-q)\,/2 = q/2$. Zajímá nás však pravděpodobnost podmíněná přežitím, které má (z této příčiny) pravděpodobnost $1 - q^2/4$, výsledkem je poměr

$$q' = \frac{q/2}{1 - q^2/4} = \frac{-2\,q}{q^2 - 4}$$

(viz obr. A.1).

To vše platí za předpokladů, že rodiče jsou nositeli vadného genu nezávisle na sobě (to asi ano) a že srpková anémie neovlivňuje přežití dítěte. Poslední předpoklad je pochybný, protože tato nemoc je v normálních podmínkách handicapem. Naše analýza by byla součástí složitějšího modelu a aspoň ukazuje, že by srpková anémie brzy téměř vymizela, kdyby nepřinášela i nějakou výhodu. Tou je, že způsobuje odolnost proti malárii, a proto se v některých zemích vyskytuje často. □

[1] Odpověď studenta v dotazníku k hodnocení výuky Matematiky 6F, 2005.



Obrázek A.1. Závislost výskytu vadného genu v následující generaci

B. Základní typy rozdělení

Není-li řečeno jinak, značí v této kapitole X náhodnou veličinu s rozdělením popisovaným v příslušném odstavci.

B.1 Diskrétní rozdělení

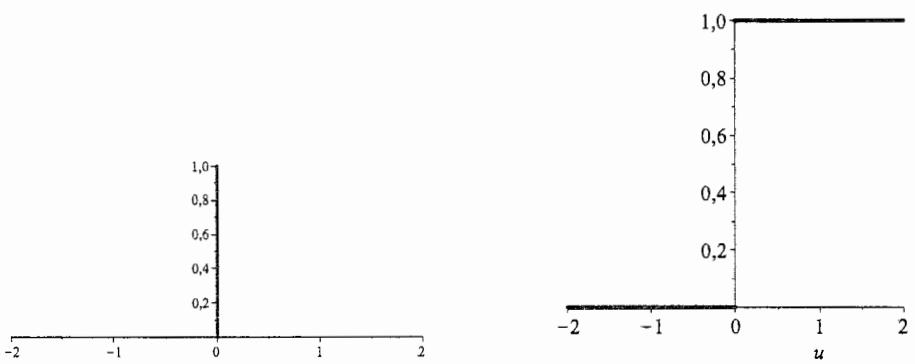
Diracovo rozdělení má jediný možný výsledek $r \in \mathbb{R}$.

$$p_X(r) = 1\,, \quad F_X(u) = \begin{cases} 0 & \text{pro } u < r\,, \\ 1 & \text{pro } u \geq r\,. \end{cases} \quad q_X(\alpha) = r\,,$$

$$\mathrm{E}X = r\,, \quad \mathrm{D}X = 0\,, \quad \Psi_X(\omega) = \mathrm{e}^{\mathrm{i}\omega r}\,.$$

Všechna diskrétní rozdělení jsou směsi Diracových rozdělení.

Parametr r má stejný fyzikální rozměr jako příslušná náhodná veličina.



Obrázek B.1. Pravděpodobnostní a distribuční funkce Diracova rozdělení pro $r = 0$

Alternativní rozdělení (též **Bernoulliovo**) má dva možné výsledky (je směsí dvou Diracových rozdělení). Pokud možné výsledky jsou $0, 1$, kde 1 má pravděpodobnost q, dostáváme

$$p_X(1) = q\,, \quad p_X(0) = 1 - q\,,$$

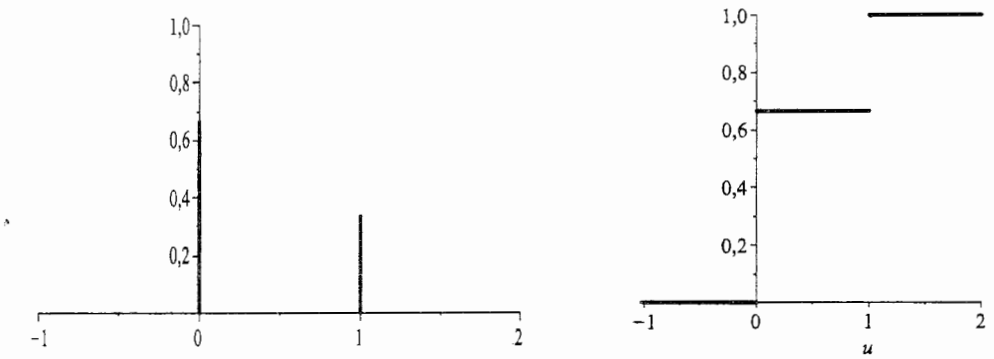
$$\mathrm{E}X = q\,, \quad \mathrm{D}X = q\,(1-q)\,, \quad \Psi_X(\omega) = (1-q) + \mathrm{e}^{\mathrm{i}\omega}\,q\,.$$

Parametr $q \in \langle 0, 1\rangle$ (pravděpodobnost) je bezrozměrný.

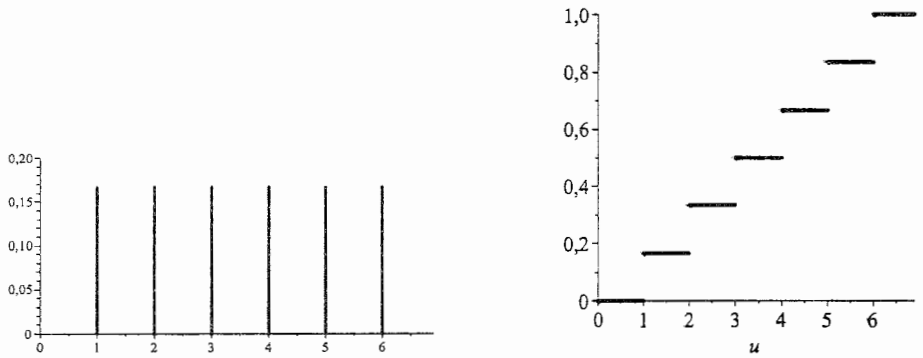
Rovnoměrné diskrétní rozdělení s m možnými výsledky je takové, že všechny *možné* výsledky mají stejnou pravděpodobnost $p_X(k) = 1/m$. Speciálně pro obor hodnot $\{1, 2, \ldots, m\}$ má charakteristiky

$$\mathrm{E}X = \frac{m+1}{2}\,, \quad \mathrm{D}X = \frac{1}{12}\,(m+1)\,(m-1)\,, \quad \Psi_X(\omega) = \frac{\mathrm{e}^{\mathrm{i}\omega m} - 1}{(1 - \mathrm{e}^{-\mathrm{i}\omega})\,m}\,.$$

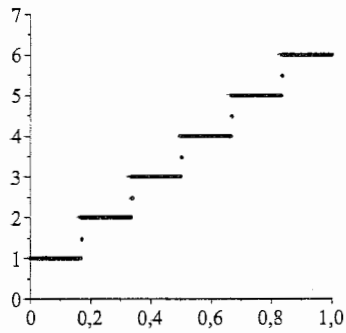
Parametr $m \in \mathbb{N}$ (počet) je bezrozměrný.



Obrázek B.2. Pravděpodobnostní a distribuční funkce alternativního rozdělení pro $q = 1/3$



Obrázek B.3. Pravděpodobnostní a distribuční funkce rovnoměrného rozdělení na $\{1, 2, \ldots, 6\}$



Obrázek B.4. Kvantilová funkce rovnoměrného rozdělení na $\{1, 2, \ldots, 6\}$

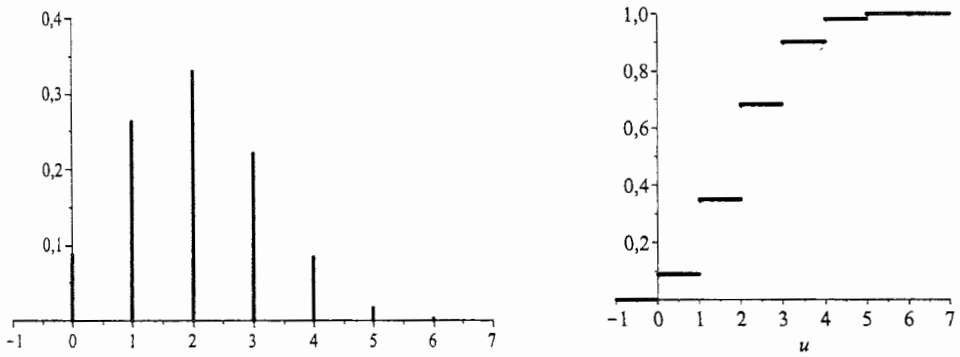
Binomické rozdělení $\mathrm{Bi}(m, q)$ vyjadřuje počet úspěchů z m nezávislých pokusů, je-li v každém stejná pravděpodobnost úspěchu $q \in \langle 0, 1 \rangle$. (Je to součet m alternativních rozdělení.)

$$p_X(k) = \binom{m}{k} q^k (1-q)^{m-k}, \qquad k \in \{0, 1, 2, \ldots, m\},$$

$$\mathrm{E}X = m\,q, \qquad \mathrm{D}X = m\,q\,(1-q), \qquad \Psi_X(\omega) = \left((1-q) + \mathrm{e}^{\mathrm{i}\omega}\,q\right)^m.$$

Parametry $m \in \mathbb{N}$ (počet) a $q \in \langle 0, 1 \rangle$ (pravděpodobnost) jsou bezrozměrné.
Výpočetní složitost výpočtu $p_X(k)$ je $O(k)$, celého rozdělení $O(m^2)$.

Poissonovo rozdělení $\mathrm{Po}(\lambda)$ má charakteristiky



Obrázek B.5. Pravděpodobnostní a distribuční funkce binomického rozdělení Bi(6, 1/3)

$$p_X(k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,, \qquad k \in \{0,1,2,\ldots\}\,,$$
$$\mathrm{E}X = \lambda\,, \qquad \mathrm{D}X = \lambda\,, \qquad \Psi_X(\omega) = \exp\left(\left(\mathrm{e}^{\mathrm{i}\omega} - 1\right)\lambda\right)\,.$$

Vychází jako limitní případ binomického rozdělení:

Příklad B.1.1: Rybář chce odhadnout rozdělení počtu ryb, které chytí za den. Pokud každá ryba v rybníce má stejnou pravděpodobnost q, že bude chycena, je to binomické rozdělení $\mathrm{Bi}(m,q)$, kde m je počet ryb v rybníce, což obvykle je velké neznámé číslo. Spíše můžeme odhadnout střední hodnotu (např. podle mnoha předchozích výsledků).

Abychom se zbavili neznámého parametru m v binomickém rozdělení $\mathrm{Bi}(m,q)$. přejdeme k limitě pro $m \to \infty$ při konstantní střední hodnotě $\lambda = m\,q > 0$, tj. $q = \lambda/m \to 0$, a dostaneme Poissonovo rozdělení $\mathrm{Po}(\lambda)$:[1]

$$p_X(k) = \binom{m}{k} q^k\,(1-q)^{m-k} = \frac{m\,(m-1)\ldots(m-(k-1))}{k!}\left(\frac{\lambda}{m}\right)^k\left(1-\frac{\lambda}{m}\right)^{m-k} =$$
$$= \frac{\lambda^k}{k!}\,\underbrace{1\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{k-1}{m}\right)}_{\to 1}\,\underbrace{\left(1-\frac{\lambda}{m}\right)^{-k}}_{\to 1}\,\underbrace{\left(1-\frac{\lambda}{m}\right)^{m}}_{\to \mathrm{e}^{-\lambda}} \to \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,.$$

Součet dvou *nezávislých* náhodných veličin s Poissonovým rozdělením má Poissonovo rozdělení (cvičení 4.7.22). Pro $\lambda \to \infty$ se rozdělení veličiny $\operatorname{norm} X = \frac{X-\lambda}{\sqrt{\lambda}}$ blíží $\mathrm{N}(0,1)$, takže pro velké hodnoty parametru λ můžeme Poissonovo rozdělení $\mathrm{Po}(\lambda)$ přibližně nahradit normálním $\mathrm{N}(\lambda,\lambda)$. *Střední hodnota Poissonova rozdělení se rovná rozptylu, což výjimečně lze říci navzdory poznámce 4.4.2, neboť se jedná vždy o bezrozměrné celočíselné náhodné veličiny (počet výskytů jevu).*

Parametr $\lambda \in (0,\infty)$ je bezrozměrný.

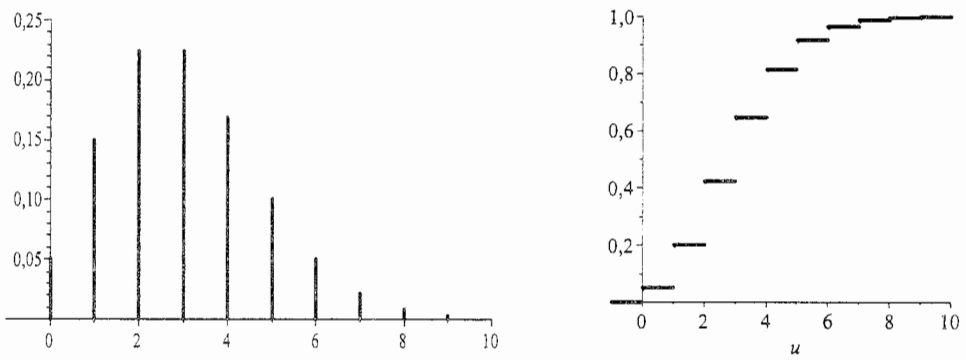
Jednotlivé pravděpodobnosti se počítají snáze než u binomického rozdělení (ale je jich nekonečně mnoho).

Geometrické rozdělení je rozdělení počtu neúspěchů do prvního úspěchu, má-li každý z nezávislých pokusů alternativní rozdělení se stejnou pravděpodobností neúspěchu $q \in (0,1)$.

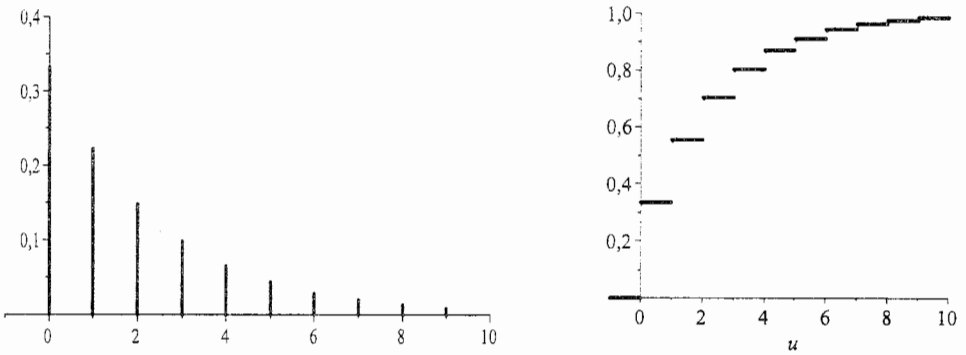
$$p_X(k) = q^k\,(1-q)\,, \qquad k \in \{0,1,2,\ldots\}\,,$$
$$\mathrm{E}X = \frac{q}{1-q}\,, \qquad \mathrm{D}X = \frac{q}{(1-q)^2}\,, \qquad \Psi_X(\omega) = \frac{1-q}{1-\mathrm{e}^{\mathrm{i}\omega}\,q}\,.$$

Parametr $q \in (0,\infty)$ je bezrozměrný.

[1] Navzdory uvedené analogii se rozdělení jmenuje podle S.D. Poissona (1781–1840), nikoli podle francouzského *poisson*=ryba.



Obrázek B.6. Pravděpodobnostní a distribuční funkce Poissonova rozdělení Po(3)



Obrázek B.7. Pravděpodobnostní a distribuční funkce geometrického rozdělení pro $q = 2/3$

Hypergeometrické rozdělení udává pravděpodobnost výskytu k „anomálních" objektů mezi m objekty, které byly náhodně vybrány z M objektů, mezi nimiž je K anomálních. Odvodili jsme je v příkladu 1.3.11.

$$p_X(k) = \frac{\binom{K}{k}\binom{M-K}{m-k}}{\binom{M}{m}}, \qquad k \in \{0, 1, 2, \ldots, m\},$$

$$\mathrm{E}X = \frac{m\,K}{M}, \qquad \mathrm{D}X = \frac{m\,K\,(M-K)\,(M-m)}{M^2\,(M-1)}.$$

Limita hypergeometrického rozdělení pro $M \to \infty$ při konstantním $\frac{K}{M} = q$, tj. $\frac{M-K}{M} = 1 - q$, vede na binomické rozdělení $\mathrm{Bi}(m, q)$, jak plyne z důsledku 1.3.10:

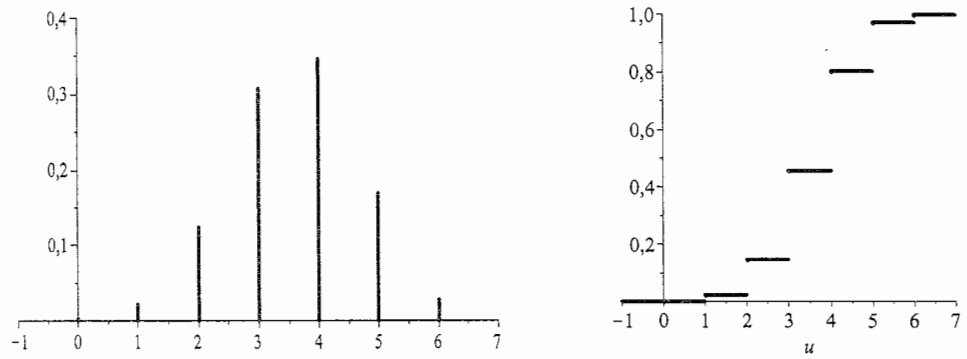
$$p_X(k) = \frac{\binom{K}{k}\binom{M-K}{m-k}}{\binom{M}{m}} \to \frac{\frac{K^k}{k!} \cdot \frac{(M-K)^{m-k}}{(m-k)!}}{\frac{M^m}{m!}} =$$

$$= \frac{m!}{k!\,(m-k)!} \cdot \frac{K^k}{M^k} \cdot \frac{(M-K)^{m-k}}{M^{m-k}} = \binom{m}{k} q^k\,(1-q)^{m-k}.$$

Parametry $m, K, M \in \mathbb{N}$ $(1 \le m \le K \le M)$ jsou bezrozměrné.
Výpočetní složitost výpočtu $p_X(k)$ je $O(m)$, celého rozdělení $O(m^2)$.

B.2 Spojitá rozdělení

Rovnoměrné spojité rozdělení $\mathrm{R}(a, b)$ má náhodná veličina X, jestliže nabývá hodnot z intervalu $\langle a, b\rangle$ a pravděpodobnost, že padne do intervalu $\langle c, d\rangle \subseteq \langle a, b\rangle$ je přímo úměrná jeho délce.

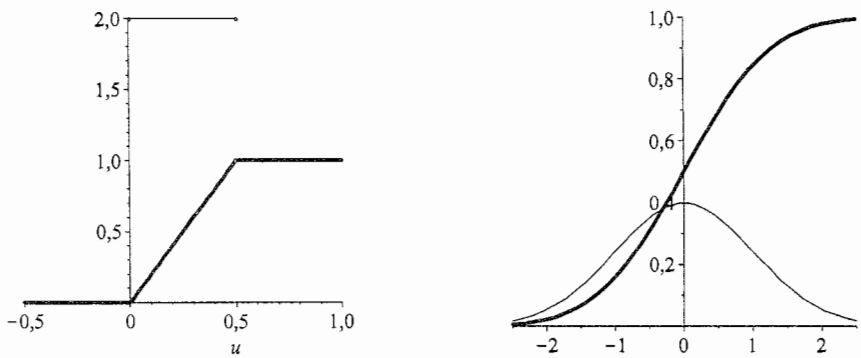


Obrázek B.8. Pravděpodobnostní a distribuční funkce hypergeometrického rozdělení pro $M = 25$, $K = 15$, $k = 6$

$$f_X(u) = \begin{cases} \frac{1}{b-a} & \text{pro } u \in \langle a, b \rangle \,, \\ 0 & \text{jinak,} \end{cases} \qquad F_X(u) = \begin{cases} \frac{u-a}{b-a} & \text{pro } u \in \langle a, b \rangle \,, \\ 0 & \text{pro } u < a \,, \\ 1 & \text{pro } u > b \,, \end{cases}$$

$$q_X(\alpha) = a + (b-a)\,\alpha \,, \qquad \Psi_X(\omega) = \frac{\exp(\mathrm{i}\,\omega\,b) - \exp(\mathrm{i}\,\omega\,a)}{\mathrm{i}\,\omega\,(b-a)} \,,$$

$$\mathrm{E}X = \frac{a+b}{2} \,, \qquad \mathrm{D}X = \frac{1}{12}\,(b-a)^2 \,.$$



Obrázek B.9. Hustota a distribuční funkce rovnoměrného rozdělení R $(0, 1/2)$ (vlevo) a normovaného normálního rozdělení N $(0, 1)$ (vpravo)

Normální rozdělení $\mathrm{N}(\mu, \sigma^2)$ (též **Gaussovo**) má následující charakteristiky:

$$f_{\mathrm{N}(\mu,\sigma^2)}(u) = \frac{1}{\sigma\sqrt{2\,\pi}} \exp\left(\frac{-(u-\mu)^2}{2\,\sigma^2}\right) \,,$$

$$\mathrm{E}\,\mathrm{N}(\mu, \sigma^2) = \mu \,, \qquad \mathrm{D}\,\mathrm{N}(\mu, \sigma^2) = \sigma^2 \,,$$

$$\Psi_{\mathrm{N}(\mu,\sigma^2)}(\omega) = \exp\left(\mathrm{i}\,\mu\,\omega - \frac{1}{2}\,\sigma^2\,\omega^2\right) \,.$$

Speciálně **normované normální rozdělení** $\mathrm{N}(0, 1)$ má hustotu

$$\varphi(u) = f_{\mathrm{N}(0,1)}(u) = \frac{1}{\sqrt{2\,\pi}} \exp\left(\frac{-u^2}{2}\right) \,.$$

Distribuční funkce je transcendentní (Gaussův integrál)

$$\Phi(u) = F_{\mathrm{N}(0,1)}(u) = \int_{-\infty}^{u} \varphi(v)\,\mathrm{d}v = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{u} \exp\left(\frac{-v^2}{2}\right) \mathrm{d}v \,.$$

Rovněž kvantilová funkce $\Phi^{-1} = q_{\mathrm{N}(0,1)}$ je transcendentní. Kvantil $\Phi^{-1}(\alpha)$ se často značí u_α. Normální rozdělení je symetrické kolem nuly, takže

$$f_{\mathrm{N}(\mu,\sigma^2)}(-u) = f_{\mathrm{N}(\mu,\sigma^2)}(u)\,, \quad F_{\mathrm{N}(\mu,\sigma^2)}(-u) = 1 - F_{\mathrm{N}(\mu,\sigma^2)}(u)\,, \quad q_{\mathrm{N}(\mu,\sigma^2)}(1-\alpha) = -q_{\mathrm{N}(\mu,\sigma^2)}(\alpha)\,,$$
$$\varphi(-u) = \varphi(u)\,, \quad \Phi(-u) = 1 - \Phi(u)\,, \quad \Phi^{-1}(1-\alpha) = \Phi^{-1}(\alpha)\,.$$

Normální rozdělení díky centrální limitní větě 6.3.1 dává přibližnou aproximaci součtu velkého počtu *nezávislých* náhodných veličin. Popisuje vliv mnoha chyb, které jsou nezávislé a sčítají se.

Parametry $\mu, \sigma \in \mathbb{R}$ $(\sigma > 0)$ jsou stejného fyzikálního rozměru jako příslušná náhodná veličina, rozměr rozptylu σ^2 je jeho kvadrátem.

Logaritmickonormální rozdělení $\mathrm{LN}(\mu, \sigma^2)$ má náhodná veličina tvaru $X = \exp(Y)$, kde Y má rozdělení $\mathrm{N}(\mu, \sigma^2)$.

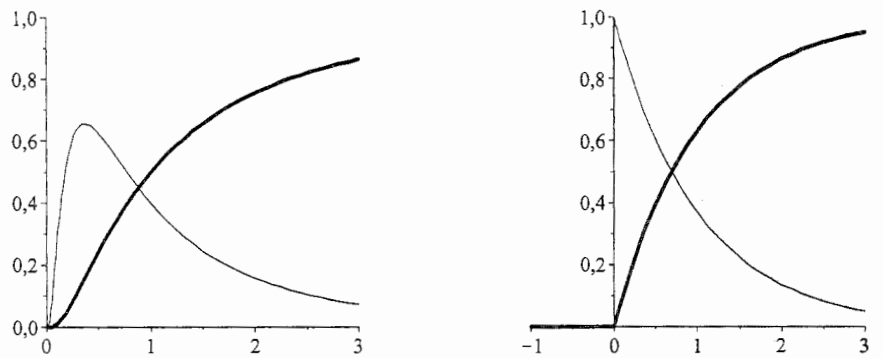
$$f_X(u) = \begin{cases} \frac{1}{u}\,\varphi(\ln u) = \frac{1}{u\,\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln u - \mu)^2}{2\sigma^2}\right) & \text{pro } u > 0\,, \\ 0 & \text{jinak,} \end{cases}$$

$$F_X(u) = \begin{cases} \Phi(\ln u) & \text{pro } u > 0\,, \\ 0 & \text{jinak,} \end{cases}$$

$$\mathrm{E}X = \exp\left(\mu + \frac{\sigma^2}{2}\right)\,, \qquad \mathrm{D}X = \left(\exp\left(2\,\mu + \sigma^2\right)\right)\left(\exp\left(\sigma^2\right) - 1\right)\,.$$

Jako je normální rozdělení modelem vlivu mnoha chyb, které jsou nezávislé a *sčítají* se, logaritmickonormální rozdělení modeluje vliv mnoha chyb, které jsou nezávislé a *násobí* se. Přispívají tedy k relativní místo k absolutní chybě. Logaritmickonormální rozdělení nabývá pouze kladných hodnot a není symetrické.

Parametry $\mu, \sigma \in \mathbb{R}$ $(\sigma > 0)$ jsou bezrozměrné, stejně jako argument u (jinak bychom nemohli počítat logaritmy a exponenciály).



Obrázek B.10. Hustota a distribuční funkce logaritmickonormálního rozdělení LN (0, 1) (vlevo) a exponenciálního rozdělení Ex (1) (vpravo)

Exponenciální rozdělení $\mathrm{Ex}(\tau)$ má náhodná veličina X, jestliže nabývá pouze nezáporných hodnot a má hustotu

$$f_X(u) = \begin{cases} \frac{1}{\tau} \exp\left(-\frac{u}{\tau}\right) & \text{pro } u > 0\,, \\ 0 & \text{jinak.} \end{cases}$$

Další charakteristiky:

$$F_X(u) = \begin{cases} 1 - \exp\left(-\frac{u}{\tau}\right) & \text{pro } u > 0\,, \\ 0 & \text{jinak,} \end{cases} \qquad q_X(\alpha) = -\tau \ln(1-\alpha)\,,$$

$$\mathrm{E}X = \tau\,, \qquad \mathrm{D}X = \tau^2\,, \qquad \Psi_X(\omega) = \frac{1}{1 - \mathrm{i}\,\omega\,\tau}\,.$$

Exponenciální rozdělení je typické pro systémy, které „nemají paměť". Např. rozpad radioaktivních izotopů nezávisí na předchozím průběhu, proto doba do rozpadu se řídí exponenciálním rozdělením, kde časová konstanta τ se častěji převádí na *poločas rozpadu* $\tau \ln 2$. Je to jediné rozdělení s touto vlastností (určené jednoznačně až na parametr τ). Stejný model se osvědčuje pro životnost polovodičových součástek. Naopak se nehodí pro životnost dílů, které podléhají opotřebení, takže po čase nejsou stejné a nemají stejnou prognózu.

Všechna exponenciální rozdělení jsou násobky jediného:

Tvrzení B.2.1: *Jestliže náhodná veličina X má exponenciální rozdělení* $\mathrm{Ex}(1)$, *pak náhodná veličina τX má exponenciální rozdělení* $\mathrm{Ex}(\tau)$.

Parametr $\tau \in (0, \infty)$ má stejný fyzikálního rozměr jako příslušná náhodná veličina; pokud je to čas, má význam časové konstanty. Často se uvádí exponenciální rozdělení v závislosti na parametru $w = 1/\tau$ (jehož fyzikální rozměr je převrácená hodnota rozměru náhodné veličiny, např. místo času frekvence); pak má hustota tvar

$$f_X(u) = \begin{cases} w \exp(-w\,u) & \text{pro } u > 0\,, \\ 0 & \text{jinak.} \end{cases}$$

Rozdělení χ^2 s η stupni volnosti[2] (značíme $\chi^2(\eta)$) je rozdělení náhodné veličiny

$$X = \sum_{j=1}^{\eta} U_j^2\,,$$

kde U_j jsou *nezávislé* náhodné veličiny s rozdělením $\mathrm{N}(0,1)$. Hustota je

$$f_X(y) = \begin{cases} c(\eta)\, y^{\frac{\eta}{2}-1}\, \mathrm{e}^{\frac{-y}{2}} & \text{pro } y > 0, \\ 0 & \text{jinak,} \end{cases}$$

kde

$$c(\eta) = \frac{1}{2^{\frac{\eta}{2}}\, \Gamma\left(\frac{\eta}{2}\right)}\,.$$

Další charakteristiky:

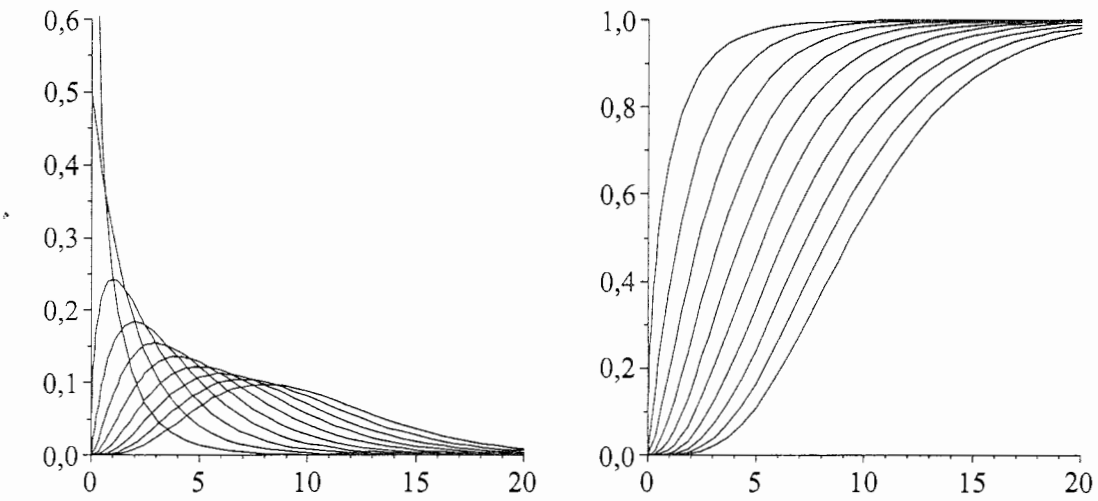
$$\mathrm{E}X = \eta\,, \qquad \mathrm{D}X = 2\,\eta\,, \qquad \Psi_X(\omega) = (1 - 2\,\mathrm{i}\,\omega)^{-\frac{\eta}{2}}\,. \tag{B.1}$$

Kvantil $q_{\chi^2(\eta)}(\alpha)$ se často značí $\chi^2_\alpha(\eta)$.

Tvrzení B.2.2: *Pro $\eta = 2$ se $\chi^2(2)$ shoduje s exponenciálním rozdělením* $\mathrm{Ex}(2)$.

Věta B.2.3: *Nechť X, Y jsou* nezávislé *náhodné veličiny s rozdělením $\chi^2(\xi)$, resp. $\chi^2(\eta)$. Pak $X + Y$ má rozdělení $\chi^2(\xi + \eta)$.*

[2] Čteme „chí kvadrát".



Obrázek B.11. Hustoty a distribuční funkce rozdělení $\chi^2(\eta)$, $\eta = 1, \ldots, 10$

Pro velký počet stupňů volnosti η lze rozdělení $\chi^2(\eta)$ aproximovat normálním rozdělením $\mathrm{N}(\eta, 2\eta)$, přesnější je však náhrada výrazem

$$Y = \frac{1}{2}\left(Z + \sqrt{2\eta - 1}\right)^2 ,$$

kde Z má rozdělení $\mathrm{N}(0,1)$. To znamená, že náhodná veličina

$$\sqrt{2X} - \sqrt{2\eta - 1}$$

má přibližně rozdělení $\mathrm{N}(0,1)$ [Spiegel et al. 2000].

Parametr $\eta \in \mathbb{N}$ (počet stupňů volnosti) je bezrozměrný.

Studentovo rozdělení neboli **t-rozdělení** $\mathrm{t}(\eta)$ **s** η **stupni volnosti**[3] má náhodná veličina

$$X = \frac{U}{\sqrt{\frac{V}{\eta}}} ,$$

kde U má rozdělení $\mathrm{N}(0,1)$, V má rozdělení $\chi^2(\eta)$ a U, V jsou nezávislé. Jeho hustota je

$$f_X(x) = c(\eta)\left(1 + \frac{x^2}{\eta}\right)^{\frac{1-\eta}{2}} ,$$

kde

$$c(\eta) = \frac{\Gamma\left(\frac{1+\eta}{2}\right)}{\sqrt{\eta\pi}\,\Gamma\left(\frac{\eta}{2}\right)} .$$

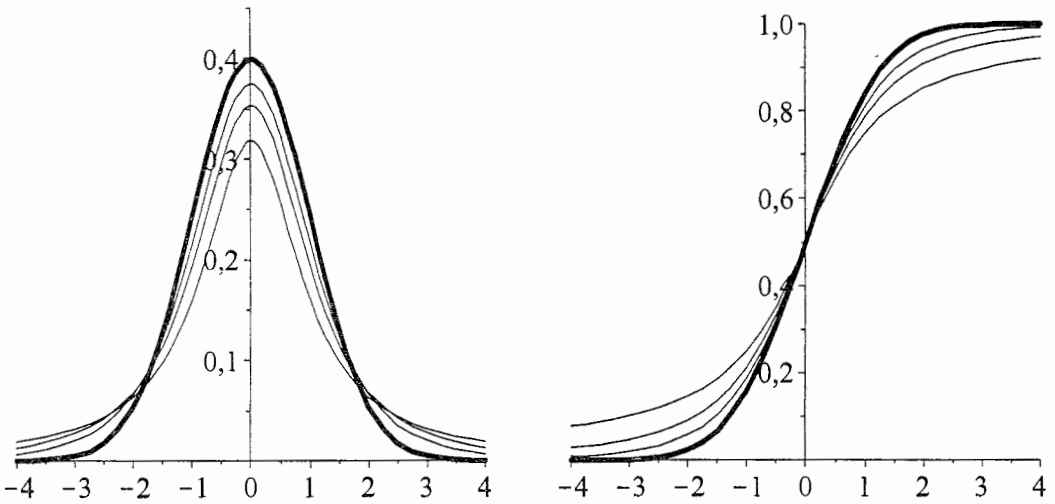
Další charakteristiky:

$$\mathrm{E}X = 0 , \qquad \mathrm{D}X = \frac{\eta}{\eta - 2} ,$$

pokud $\eta > 2$. *(Studentovo rozdělení by se mohlo normalizovat, ale nestalo se to zvykem.)*

Studentovo rozdělení je symetrické kolem nuly, takže

$$f_{\mathrm{t}(\eta)}(-u) = f_{\mathrm{t}(\eta)}(u) , \qquad F_{\mathrm{t}(\eta)}(-u) = 1 - F_{\mathrm{t}(\eta)}(u) , \qquad q_{\mathrm{t}(\eta)}(1-\alpha) = -q_{\mathrm{t}(\eta)}(\alpha) .$$



Obrázek B.12. Hustoty a distribuční funkce rozdělení t(1), t(2), t(4) a N(0,1) (tučně)

Studentovo rozdělení má statistika

$$T = \frac{\sqrt{n}}{S_{\boldsymbol{X}}} \left(\overline{\boldsymbol{X}} - \mu\right)$$

používaná při testech střední hodnoty normálního rozdělení při neznámém rozptylu, viz kapitola 12.2.1. Vyskytuje se i v dalších testech, někdy zprostředkovaně přes vhodnou transformaci kritéria, jako u testu korelace v kapitole 12.4.1.

Pro velký počet stupňů volnosti se Studentovo rozdělení nahrazuje normovaným normálním rozdělením N(0,1).

Kvantil $q_{\mathrm{t}(\eta)}(\alpha)$ se často značí $t_\alpha(\eta)$.

Parametr $\eta \in \mathbb{N}$ (počet stupňů volnosti) je bezrozměrný.

F-rozdělení (též **Fisherovo-Snedecorovo rozdělení**) $\mathrm{F}(\xi,\eta)$ **s** ξ **a** η **stupni volnosti** je rozdělení náhodné veličiny

$$X = \frac{\frac{U}{\xi}}{\frac{V}{\eta}},$$

kde U, V jsou *nezávislé* náhodné veličiny s rozdělením $\chi^2(\xi)$, resp. $\chi^2(\eta)$. Hustota pro $x > 0$ je

$$f_X(x) = c(\xi,\eta)\; x^{\frac{\xi}{2}-1} \left(1 + \frac{\xi}{\eta}\, x\right)^{-\frac{\xi+\eta}{2}} = c(\xi,\eta) \sqrt{\frac{x^{\xi-2}}{\left(1 + \frac{\xi}{\eta}\, x\right)^{\xi+\eta}}},$$

kde

$$c(\xi,\eta) = \frac{\Gamma\left(\frac{\xi+\eta}{2}\right)}{\Gamma\left(\frac{\xi}{2}\right)\Gamma\left(\frac{\eta}{2}\right)} \left(\frac{\xi}{\eta}\right)^{\frac{\xi}{2}}.$$

Další charakteristiky:

$$\mathrm{E}X = \frac{\eta}{\eta-2} \qquad \mathrm{D}X = \frac{2\,\eta^2\,(\xi+\eta-2)}{\xi\,(\eta-4)\,(\eta-2)^2},$$

pokud výrazy ve jmenovatelích jsou kladné [Spiegel et al. 2000].

[3] „Student" je pseudonym, pod kterým publikoval své práce W. Gossett.

Kvantilová funkce na intervalu $(0, 1/2)$ se dá vypočítat z hodnot na $(1/2, 1)$ (nebo naopak; pozor na opačné pořadí stupňů volnosti):

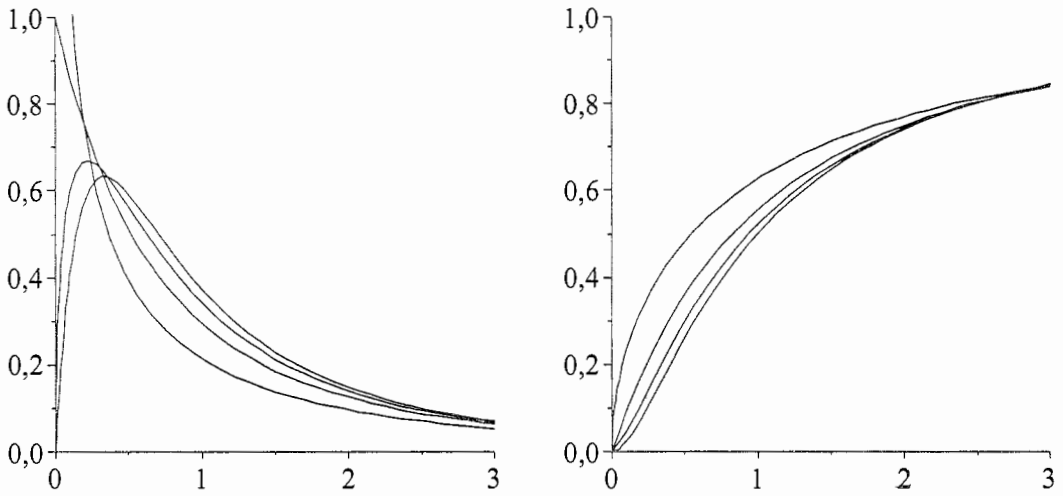
$$q_{\mathrm{F}(\xi,\eta)}(\alpha) = \frac{1}{q_{\mathrm{F}(\eta,\xi)}(1-\alpha)} .$$

Krajní případy jsou [Spiegel et al. 2000]:

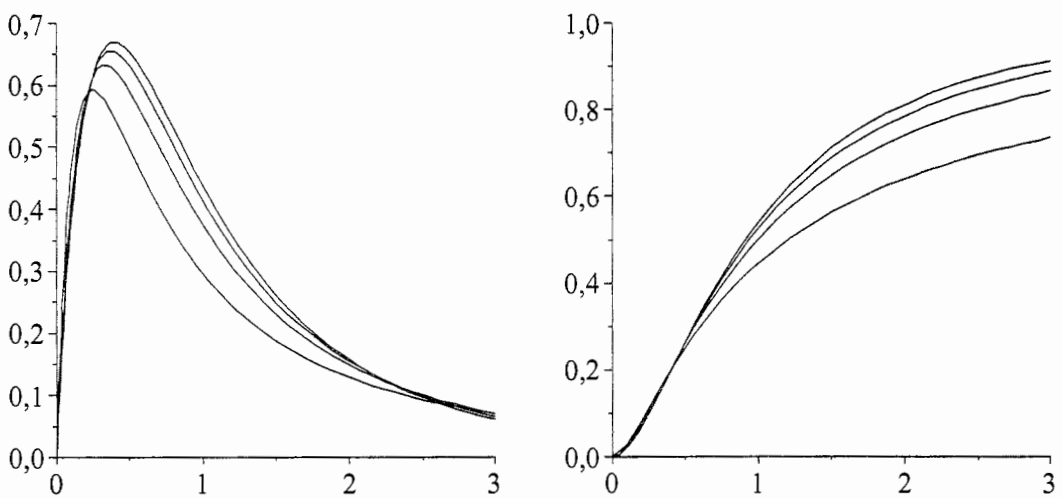
$$q_{\mathrm{F}(1,\eta)}(1-\alpha) = \left(q_{\mathrm{t}(\eta)}\left(1-\alpha/2\right)\right)^2 , \qquad \lim_{\eta\to\infty} q_{\mathrm{F}(\xi,\eta)}(\alpha) = \frac{1}{\xi}\, q_{\chi^2(\xi)}\left(\alpha\right) .$$

Kvantil $q_{\mathrm{F}(\xi,\eta)}\left(\alpha\right)$ se často značí $F_\alpha\left(\xi,\eta\right)$.

Parametry $\xi, \eta \in \mathbb{N}$ (počty stupňů volnosti) jsou bezrozměrné.



Obrázek B.13. Hustoty a distribuční funkce rozdělení $\mathrm{F}(1,4)$, $\mathrm{F}(2,4)$, $\mathrm{F}(3,4)$, $\mathrm{F}(4,4)$



Obrázek B.14. Hustoty a distribuční funkce rozdělení $\mathrm{F}(4,2)$, $\mathrm{F}(4,4)$, $\mathrm{F}(4,6)$, $\mathrm{F}(4,8)$

C. Vybrané vzorce z matematické analýzy

I když by čtenář měl být schopen si sám najít prostředky pro provedení některých obtížnějších výpočtů, pro usnadnění zde uvádíme některé skutečnosti z matematické analýzy, které byly využity v důkazech a ve cvičeních.

Konečné součty

$$\sum_{j=1}^{k-1} j\,(k-j) = \frac{1}{6}\left(k^3 - k\right) = \frac{1}{6}\,k\,(k-1)\,(k+1) \tag{C.1}$$

$$\sum_{j=1}^{k-1} j^2\,(k-j) = \frac{1}{12}\left(k^4 - k^2\right) = \frac{1}{12}\,k^2\,(k-1)\,(k+1)$$

Nekonečné součty

$$\sum_{j=1}^{\infty} \frac{1}{j} = \infty \tag{C.2}$$

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{1}{6}\,\pi^2 \tag{C.3}$$

$$\sum_{j=1}^{\infty} \frac{1}{j^3} = \zeta\,(3) \doteq 1.202 \tag{C.4}$$

$$\sum_{j=1}^{\infty} \frac{1}{j^4} = \frac{1}{90}\,\pi^4 \tag{C.5}$$

$$\sum_{m=0}^{\infty} \frac{w^m}{m!} = \mathrm{e}^w \tag{C.6}$$

(z Taylorova rozvoje exponenciální funkce se středem 0, vyhodnoceného v bodě w)

Speciální funkce

$$\Gamma\,(t) = \int_0^{\infty} u^{t-1}\,\mathrm{e}^{-u}\,\mathrm{d}u\,,$$

kde $t > 0$. Pro $t \in \mathbb{N}$:

$$\Gamma\,(t) = (t-1)!\,.$$

Limity

$$\lim_{n\to\infty} \left(1 + \frac{t}{n}\right)^n = \mathrm{e}^t \tag{C.7}$$

Integrály

$$\int_0^\infty t \exp(-t)\, \mathrm{d}t = 1 \tag{C.8}$$

$$\int_0^\infty t^2 \exp(-t)\, \mathrm{d}t = 2$$

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}t^2}\, \mathrm{d}t = \sqrt{2\pi} \tag{C.9}$$

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(t-\mathrm{i}\omega)^2}\, \mathrm{d}t = \sqrt{2\pi} \tag{C.10}$$

Konvoluce

$$(f * g)(u) = \int_{-\infty}^{\infty} f(v)\, g(u-v)\, \mathrm{d}v$$

$$\begin{aligned} f * g &= g * f \\ f * (g * h) &= (f * g) * h \\ f * (g + h) &= f * g + f * h \\ (a f) * g &= a\,(f * g) \qquad (a \in \mathbb{R}) \end{aligned}$$

Přibližné vzorce

Stirlingův vzorec: Pro velká n je

$$n! \doteq n^n \mathrm{e}^{-n} \sqrt{2\pi n} \tag{C.11}$$

v tom smyslu, že podíl obou výrazů konverguje k 1 pro $n \to \infty$.

D. Statistické tabulky

I když dnes každá knihovna matematických programů obsahuje i distribuční a další funkce běžných rozdělení a význam statistických tabulek upadá, uvádíme je zde, jednak pro řešení cvičení nezávisle na výpočetní technice, jednak pro představu čtenáře o probíraných rozděleních. Samostatné tabulky jsou obsahem skripta [Hála, Jarušková 2000]. Pěkně zpracované statistické tabulky jsou na `http://www.statsoft.com/textbook/sttable.html`.

Poznámka D.1.1: 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 D.1. Kvantilová funkce normovaného normálního rozdělení

u	$\Phi^{-1}(u)$	u	$\Phi^{-1}(u)$
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 D.2. Distribuční funkce normovaného normálního rozdělení

u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$	u	$\Phi(u)$
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 D.3. Kvantily χ^2-rozdělení $q_{\chi^2(\eta)}(\alpha)$ (η = počet stupňů volnosti)

η \ α	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 D.4. Kvantily t-rozdělení $q_{t(\eta)}(\alpha)$ (η = počet stupňů volnosti)

η \ α	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 D.5. 0.95-kvantily F-rozdělení $q_{F(\xi,\eta)}(0.95)$ (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 D.6. 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 D.7. 0.99-kvantily F-rozdělení $q_{F(\xi,\eta)}(0.99)$ (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 D.8. 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

[Apl. mat. 1978] kol.: *Aplikovaná matematika.* Oborové encyklopedie SNTL, Praha, 1978.

[Anděl 1998] Anděl, J.: *Statistické metody.* 2. vyd., Matfyzpress, Praha, 1998.

[Anděl 1978] Anděl, J.: *Matematická statistika.* SNTL/Alfa, Praha, 1978.

[Anděl 2000] Anděl, J.: *Matematika náhody.* Matfyzpress, Praha, 2000.

[Budinský, Dohnal 1991] Budinský, P., Dohnal, G.: *Tématické okruhy a celky z předmětu Matematika IV.* Doplňkové skriptum FSI ČVUT, Praha, 1991.

[Chatfield 1992] Chatfield, C.: *Statistics for Technology.* 3rd ed., Chapman & Hall, London, 1992.

[Disman 2005] Disman, M.: *Jak se vyrábí sociologická znalost.* Karolinum, UK, Praha, 2005.

[Press et al. 1986] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: *Numerical Recipes (The Art of Scientific Computing).* Cambridge University Press, Cambridge, 1986.

[Hála, Jarušková 2000] Hála, M., Jarušková, D.: *Pravděpodobnost a matematická statistika 11.* Tabulky. Skriptum FSV ČVUT, Praha, 2000.

[Jaglom 1964] Jaglom, A.M., Jaglom, I.M.: *Pravděpodobnost a informace.* Nakladatelství Československé akademie věd. Praha 1964.

[Jaroš 1998] Jaroš, F. a kol.: *Pravděpodobnost a statistika.* Skriptum VŠCHT, 2. vydání, Praha, 1998.

[Jarušková 2000] Jarušková, D.: *Pravděpodobnost a matematická statistika 12.* Skriptum FSV ČVUT, Praha, 2000.

[Jarušková, Hála 2002] Jarušková, D., Hála, M.: *Pravděpodobnost a matematická statistika 12.* Příklady. Skriptum FSV ČVUT, Praha, 2002.

[Hindls a kol. 2004] Hindls R., Hronová S., Seger J.: *Statistika pro ekonomy.* Professional Publishing, Praha, 2004.

[Hsu 1996] Hsu, H.P.: *Probability, Random Variables, and Random Processes.* McGraw-Hill, 1996.

[Knuth 1997] Knuth, D.E.: *The Art of Computer Programming.* Addison Wesley, Boston, 1997.

[Likeš, Machek 1988] Likeš, J., Machek, J.: *Matematická statistika.* 2. vydání, SNTL, Praha, 1988.

[Mood et al. 1974] Mood, A.M., Graybill, F.A., Boes, D.C.: *Introduction to the Theory of Statistics.* 3rd ed., McGraw-Hill, 1974.

[Nagy 2002] Nagy, I.: *Pravděpodobnost a matematická statistika.* Cvičení. Skriptum FD ČVUT, Praha, 2002.

[Němčíková 1990] Němčíková, A.: *Matematická statistika — cvičení.* Skriptum FEL ČVUT, Praha, 1990.

[Novovičová 2002] Novovičová, J.: *Pravděpodobnost a matematická statistika.* Skriptum FD ČVUT, Praha, 2002.

[Olšák 2007] Olšák, P.: *Úvod do algebry, zejména lineární.* Skriptum FEL ČVUT, Praha, 2007.

[Papoulis 1990] Papoulis, A.: *Probability and Statistics.* Prentice-Hall, 1990.

[Pavlík 2005] Pavlík, J.: *Aplikovaná statistika.* Skriptum VŠCHT, Praha, 2005.

[Rényi 1966] Rényi, A.: *Calcul des probabilités.* Dunod, Paris, 1966.

[Rényi 1972] Rényi, A.: *Teorie pravděpodobnosti.* Academia, Praha, 1972 *(český překlad).*

[Riečanová a kol. 1987] Riečanová, Z., Horváth, J., Olejček, V., Volauf, P.: *Numerické metódy a matematická štatistika.* Alfa/SNTL, Bratislava, 1987.

[Riečan a kol. 1984] Riečan, B., Lamoš, F., Lenárt, C.: *Pravdepodobnosť a matematická štatistika.* Alfa/SNTL, Bratislava, 1984.

[Rogalewicz 2000] Rogalewicz, V.: *Pravděpodobnost a statistika pro inženýry.* Skriptum FEL ČVUT, dotisk 1. vydání, Praha, 2000.

[Rozanov 1971] Rozanov, Ju.A.: *Slučajnyje processy.* Nauka, Moskva, 1971.

[Schlesinger, Hlaváč 1999] Schlesinger, M.I., Hlaváč, V.: *Deset přednášek z teorie statistického a strukturního rozpoznávání.* ČVUT, Praha, 1999.

[Sikorski 1969] Sikorski, R.: *Boolean Algebras.* 3rd ed., Springer-Verlag, Berlin, 1969.

[Spiegel et al. 2000] Spiegel, M.R., Schiller, J.J., Srinivasan, R.A.: *Probability and Statistics.* McGraw-Hill, 2000.

[Štěpán 1987] Štěpán, J.: *Teorie pravděpodobnosti. Matematické základy.* Academia, Praha, 1987.

[Swoboda 1977] Swoboda, H.: *Moderní statistika.* Svoboda, Praha, 1977.

[Tutubalin 1978] Tutubalin, V.N.: *Teorie pravděpodobnosti.* SNTL, Praha, 1978.

[Volauf 1988] Volauf, P.: *Matematická štatistika (zbierka príkladov).* Alfa, Bratislava, 1988.

[Wikipedia] `http://en.wikipedia.org`

[Zvára, Štěpán 2002] Zvára, K., Štěpán, J.: *Pravděpodobnost a matematická statistika.* 2. vydání, Matfyzpress, MFF UK, Praha, 2002.

Rejstřík