Pravděpodobnost a matematická statistika – cvičení

Mirko Navara a kol.
Centrum strojového vnímání
katedra kybernetiky FEL ČVUT
http://cmp.felk.cvut.cz/~navara/psi

13. prosince 2016

Obsah

Část I
Teorie pravděpodobnosti

1 Motivační příklady

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

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

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

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

2 Kombinatorické pojmy a vzorce

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

1. 12 *finalistek soutěže krásy,*
2. 4*-členné družstvo na závod Dolomitenmann,*
3. 1000 *výherců spotřebitelské soutěže?*

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

Řešení.

1. 12 *finalistek soutěže krásy: neuspořádaný výběr bez vracení,* $\binom{n}{12}$.
2. 4*-členné družstvo na závod Dolomitenmann: uspořádaný výběr bez vracení,* $\frac{n!}{(n-4)!}$.
3. 1000 *výherců spotřebitelské soutěže: neuspořádaný výběr s vracením,* $\binom{n+1000-1}{1000}$.

Není zde zastoupen uspořádaný výběr s vracením, např. výherci prvních tří cen v literární soutěži, n^3, *a permutace s opakováním, např. počet možných způsobů rozmístění osmi bílých šachových figur (bez pěšců) na 1. řadě šachovnice,* $\frac{8!}{2!\cdot 2!\cdot 2!\cdot 1\cdot 1!} = 5040$. □

Příklad 2.2 (hypergeometrické rozdělení). *Mezi* M *výrobky je* K *vadných. Jaká je pravděpodobnost, že mezi* m *náhodně vybranými výrobky je právě* k *vadných?*

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

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

Odvodili jsme tzv. ***hypergeometrické rozdělení****.* □

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

Příklad 2.3 (pravděpodobnosti zařazení do průzkumu). *Alice a Bob žijí ve státě, který má $n = 10^7$ obyvatel. Do statistického průzkumu bude vybráno $k = 10\,000$ respondentů. Pro všechny čtyři typy výběrů vypočtěte počet všech možností výběru a pravděpodobnost, že do výběru bude vybrána (a) Alice aspoň jednou, (b) Alice i Bob, (c) Alice více než jednou.*

Řešení. ***Uspořádaný výběr bez vracení:***
Celkový počet možností $\frac{n!}{(n-k)!} \doteq 6.730 \cdot 10^{69\,997}$.
(a) Alice (stejně jako Bob) není vybrána v $\frac{(n-1)!}{(n-1-k)!} \doteq 6.723 \cdot 10^{69\,997}$ *případech, tj. s pravděpodobností* $\frac{(n-1)!}{(n-1-k)!}\,\frac{(n-k)!}{n!} = \frac{n-k}{n} = 0.999$,
je vybrána s pravděpodobností $1 - \frac{(n-1)!}{(n-1-k)!}\,\frac{(n-k)!}{n!} = 1 - \frac{n-k}{n} = \frac{k}{n} = 0.001$.
(b) Alice ani Bob nejsou vybráni v $\frac{(n-2)!}{(n-2-k)!}$ *případech, tj. s pravděpodobností* $\frac{(n-2)!}{(n-2-k)!}\,\frac{(n-k)!}{n!} = \frac{(n-k)\,(n-k-1)}{n\,(n-1)} \doteq 0.998\,001$.
Od jednotky odečteme pravděpodobnost, že není vybrána Alice, a také, že není vybrán Bob, tj. $1 - 2\,\frac{n-k}{n}$.
To jsme ale dvakrát odečetli výběry bez Alice i Boba, a musíme je jednou přičíst. Pravděpodobnost, že bude vybrána Alice i Bob, je $1 - 2\,\frac{n-k}{n} + \frac{(n-k)\,(n-k-1)}{n\,(n-1)} = \frac{k\,(k-1)}{n\,(n-1)} \doteq 9.999 \cdot 10^{-7}$.
Alternativní řešení: Alice bude vybrána s pravděpodobností $\frac{k}{n}$, *ze zbývajících obyvatel do zbytku výběru Bob s pravděpodobností* $\frac{k-1}{n-1}$.

Neuspořádaný výběr bez vracení:
Celkový počet možností $\binom{n}{k} = \frac{n!}{(n-k)!\,k!} \doteq 2.365 \cdot 10^{34\,338}$.
(a) Alice je vybrána v $\binom{n-1}{k-1} = \frac{(n-1)!}{(n-k)!\,(k-1)!} \doteq 2.365 \cdot 10^{34\,335}$ *případech, tj. s pravděpodobností* $\frac{(n-1)!}{(n-k)!\,(k-1)!}\,\frac{(n-k)!\,k!}{n!} = \frac{k}{n} = 0.001$.
(b) Alice i Bob jsou vybráni v $\binom{n-2}{k-2}$ *případech, tj. opět s pravděpodobností* $\frac{(n-2)!}{(n-k)!\,(k-2)!}\,\frac{(n-k)!\,k!}{n!} = \frac{k\,(k-1)}{n\,(n-1)} \doteq 9.999 \cdot 10^{-7}$.

Uspořádaný výběr s vracením:
Celkový počet možností $n^k = 10^{70\,000}$.
(a) Alice není vybrána v $(n-1)^k \doteq 9.99 \cdot 10^{69\,999}$ *případech, tj. s pravděpodobností* $\left(\frac{n-1}{n}\right)^k \doteq 0.999$,
je vybrána s pravděpodobností $1 - \left(\frac{n-1}{n}\right)^k \doteq 0.001$.
(b) Alice ani Bob nejsou vybráni v $(n-2)^k \doteq 9.98 \cdot 10^{69\,999}$ *případech, tj. s pravděpodobností* $\left(\frac{n-2}{n}\right)^k \doteq 0.998\,001$.
Obdobně jako u výběru bez vracení, pravděpodobnost, že bude vybrána Alice i Bob, je $1 - 2\,\left(\frac{n-1}{n}\right)^k + \left(\frac{n-2}{n}\right)^k = \frac{k\,(k-1)}{n\,(n-1)} \doteq 9.989 \cdot 10^{-7}$.
(c) Pokud je Alice vybrána právě jednou, může se to stát při k *příležitostech; zbývajících* $k-1$ *respondentů je vybráno z* $n-1$ *obyvatel, možností je* $k\,(n-1)^{k-1}$. *Pravděpodobnost, že Alice bude vybrána více než jednou, je* $1 - \left(\frac{n-1}{n}\right)^k - \frac{k\,(n-1)^{k-1}}{n^k} \doteq 4.996 \cdot 10^{-7}$.
(U výběrů bez vracení byla nulová.)

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

3 Vlastnosti pravděpodobnosti

4 Geometrická pravděpodobnost

Příklad 4.1 (**Buffonova úloha**). *Na linkovaný papír hodíme jehlu, jejíž délka je rovna vzdálenosti mezi linkami. Jaká je pravděpodobnost, že jehla protne nějakou linku?*

Řešení. *BÚNO: Délka jehly (a vzdálenost linek) je jednotková.*
Označme $x \in \langle 0, \pi/2 \rangle$ *úhel mezi linkou a jehlou a* $y \in \langle 0, 1/2 \rangle$ *vzdálenost středu jehly od nejbližší linky (za jednotku bereme vzdálenost mezi linkami). Předpokládáme, že tyto náhodné veličiny jsou nezávislé a mají*

rovnoměrná rozdělení na příslušných intervalech. Za množinu elementárních jevů vezmeme dvojrozměrný interval (obdélník) $\Omega = \langle 0, 1/2\rangle \times \langle 0, \pi/2\rangle$, na kterém máme rovnoměrné rozdělení. Jev A – protnutí linky – nastane, pokud $y < \frac{1}{2}\sin x$,

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

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

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

□

Příklad 4.2. *Na rovnoměrnou nekonečnou čtvercovou mřížku, kde vzdálenost průsečíků je a, hodíme minci o průměru b, $b < a$. Jaká je pravděpodobnost, že mince zakryje část některé z linek této mřížky?*

Příklad 4.3. *Házíme mincí na čáru; náhodná veličina X udává vzdálenost hozené mince od čáry. Její rozdělení pravděpodobnosti je dáno hustotou:*

$$f_X(x) = \begin{cases} 1 - \frac{x}{2} & \textit{pokud} \quad x \in \langle 0, 2\rangle\,, \\ 0 & \textit{pokud} \quad x > 2\,. \end{cases}$$

Náhodná veličina $Y = \frac{1}{X}$ udává výhru (zisk) z jednoho hodu. Jaké je rozdělení (střední hodnota, rozptyl) náhodné veličiny Y?

5 Kolmogorovův model pravděpodobnosti

6 Nezávislé jevy

6.1 Nezávislost dvou jevů

Příklad 6.1 (vylepšení náhodného generátoru)**.** *Alice a Bob chtějí spravedlivě vybrat jednoho z nich. Mohou si hodit mincí, ale ta je zdeformovaná, takže jsou pochyby, zda padají oba výsledky se stejnou pravděpodobností. Dohodnou se, že hodí mincí dvakrát. Alice vyhrává, pokud padnou stejné výsledky, Bob při různých výsledcích. Kdo z nich má větší naději na výhru?*

Řešení. *Líc padá s pravděpodobností $1/2 + \varepsilon$,*
rub s pravděpodobností $1/2 - \varepsilon$, kde $\varepsilon \in (-1/2, 1/2)$.
$2\times$ líc s pravděpodobností $(1/2 + \varepsilon)^2$,
$2\times$ rub s pravděpodobností $(1/2 - \varepsilon)^2$.
Alice vyhrává s pravděpodobností

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

Bob s pravděpodobností

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

Pravděpodobnost se od $1/2$ liší od $1/2$ o $h(\varepsilon) = 2\,\varepsilon^2$ namísto ε.
Např. pro $\varepsilon = 0.01$ Alice vyhrává s pravděpodobností $1/2 + 2\,\varepsilon^2 = 0.500\,2$. Udělali jsme ***ze špatného náhodného generátoru lepší****.* □

Příklad 6.2 (vylepšení náhodného generátoru 2)**.** *Vylepšete předchozí příklad.*

Řešení. *Potřebujeme více než dva hody.*
Např. při sudém počtu líců vyhrává Alice, při lichém Bob.
Pro 3 hody vyhrává Alice s pravděpodobností

$$\left(\tfrac{1}{2} - \varepsilon\right)^3 + 3\,\left(\tfrac{1}{2} - \varepsilon\right)\,\left(\tfrac{1}{2} + \varepsilon\right)^2 = \tfrac{1}{2} - 4\,\varepsilon^3\,.$$

Pro $\varepsilon = 0.01$ je to $0.499\,996$.Pro 4 hody

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

Pro velmi špatnou minci, u níž líc padá s pravděpodobností 0.9*, tj.* $\varepsilon = 0.4$*,*
provedeme např. $2^5 = 32$ *pokusů.*
Pravděpodobnost, že počet líců je sudý, se liší od 1/2 *o*

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

Takto lze vytvořit ***z velmi špatného náhodného generátoru libovolně dobrý*** *(byť pomalejší).* □

6.2 Nezávislost více jevů

Příklad 6.3. *Nezávislé jevy* A, B, C *mají po řadě pravděpodobnosti* 0.2, 0.3, 0.4*. Určete pravděpodobnost jevu* $X = (A \vee B) \wedge C$*.*

Řešení. *Pro nezávislé jevy*

$$P(A \vee B) = P(A) + P(B) - P(A) \cdot P(B) = 0.2 + 0.3 - 0.2 \cdot 0.3 = 0.44\ ,$$
$$P(X) = P((A \vee B) \wedge C) = P(A \vee B) \cdot P(C) = 0.44 \cdot 0.4 = 0.176\,.$$

□

Příklad 6.4. *Hladina je kontrolována 4 spínači dle obrázku. Při nízké hladině mají být všechny sepnuty, při vysoké vypnuty. Každý z nich (nezávisle) je s pravděpodobností* 10 % *v opačném stavu, než by měl být. Jaká je pravděpodobnost poruchy celého zapojení v sepnutém, resp. vypnutém stavu? Porovnejte s použitím jednoho spínače.*

Řešení. *Označme* p *pravděpodobnost, že spínač je sepnutý. Paralelní spojení dvou nezávislých spínačů je spojené s pravděpodobností* $q = 1 - (1-p)^2$*, sériové spojení dvou takových obvodů s pravděpodobností* $r = q^2$*.*
Sepnutý stav: $p = 0.9$*,* $q = 0.99$*,* $r = 0.9801$*, pravděpodobnost poruchy je* $1 - r = 0.0199$*.*
Vypnutý stav: $p = 0.1$*,* $q = 0.19$*,* $r = 0.0361$*, což je i pravděpodobnost poruchy.*
V obou stavech se pravděpodobnost poruchy několikanásobně snížila.

7 Podmíněná pravděpodobnost

Příklad 7.1. *U* 10% *řidičů, kteří způsobili dopravní nehodu, bylo prokázáno požití alkoholu. Rozsáhlý průzkum ukázal, že riziko nehody se požitím alkoholu zvyšuje* $7\times$*. Odhadněte, kolik procent řidičů požilo alkohol.*

Řešení. *Jevy:*
A *… požil alkohol,*
H *… způsobil nehodu.*
$P\left(A|H\right) = 0.1$*,* $P\left(H|A\right) = 7\,P\left(H|\bar{A}\right)$*.*

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

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

□

Příklad 7.2. *Požití alkoholu bylo prokázáno u* 1% *všech řidičů a u* 10% *řidičů, kteří způsobili dopravní nehodu. Kolikrát se požitím alkoholu zvyšuje riziko nehody?*

Řešení. *Jevy:*
A *… požil alkohol,*
H *… způsobil nehodu.*
$P\left(A\right) = 0.01$*,* $P\left(A|H\right) = 0.1$*.*

$$0.1 = P\left(A|H\right) = \frac{P\left(H|A\right)\,P\left(A\right)}{P\left(H|A\right)\,P\left(A\right) + P\left(H|\bar{A}\right)\,P\left(\bar{A}\right)}$$
$$= \frac{P\left(H|A\right)\,0.01}{P\left(H|A\right)\,0.01 + P\left(H|\bar{A}\right)\,0.99} = \frac{1}{1 + \frac{P\left(H|\bar{A}\right)}{P\left(H|A\right)}\,99}\,,$$

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

□

Příklad 7.3. *Když je Egon střízlivý, udělá v průměru jednu gramatickou chybu na 100 slov, když je opilý, 2× tolik. V semestrální práci o 1000 slovech měl 16 chyb. Alice soudí, že ji musel psát opilý, Bob tvrdí, že Egon byl střízlivý. Co vše můžete k jejich sporu říci, můžete-li si dovolit riziko 5 %, že váš úsudek bude chybný?*

Příklad 7.4. *V populaci je infikována 1/4 jedinců, ale jen u 2/3 infikovaných se nákaza projevuje (a u žádných neinfikovaných). Jaká je pravděpodobnost, že jedinec bez příznaků není infikovaný?*

Příklad 7.5. *Pravděpodobnost onemocnění cukrovkou je 5% u těch, jejichž rodiče tuto nemoc neměli, 10% tam, kde ji měl jeden z rodičů, a 30%, pokud měli cukrovku oba rodiče.*

1. *Jaký je rovnovážný podíl nemocných cukrovkou v populaci (stejný u generace rodičů i dětí) za předpokladu, že onemocnění otce a matky jsou nezávislé jevy?*

2. *Jestliže pacient onemocněl cukrovkou, jaká je pravděpodobnost, že tuto nemoc měl aspoň jeden z jeho rodičů, pokud předpokládáme, že v populaci je rovnovážný výskyt dle bodu 1?*

Řešení. $c = 0.05(1-c)^2 + 0.3c^2 + 0.2c(1-c)$
$c = 5.608 \times 10^{-2}$
$P(R_0|C) = \frac{0.05(1-c)^2}{c} = 0.7944$
$P(\neg R_0|C) = 1 - \frac{0.05(1-c)^2}{c} = 0.2056$ □

Příklad 7.6. *Jazykový korektor změní 99 % chybných slov na správná a 0.01 % správných na chybná. Změnil 2 % slov. Odhadněte množství chybných slov v jeho výstupu.*

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

Příklad 7.7. *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. *A ... žije,*
B ... zažil ztroskotání,
$P(A|B) = \frac{2}{3}$, $P(A|\neg B) = 1$, $P(B|A) = \frac{5}{60} = \frac{1}{12}$ *(odhad)*
Bayesova věta:

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

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

Příklad 7.8 (pozitivní test na nemoc)**.** *Test nemoci je u 1% zdravých falešně pozitivní a u 10% nemocných falešně negativní. Nemocných je v populaci 0.001. Jaká je pravděpodobnost, že pacient s pozitivním testem je nemocný?*

Řešení. *T ... pozitivní test,* $\quad$ *N ... nemocný.*
$P(N) = 0.001, \quad P(T|\overline{N}) = 0.01, \quad P(\overline{T}|N) = 0.1.$

$$\begin{aligned}
P(T) &= P(T|\overline{N}) \cdot P(\overline{N}) + \underbrace{P(T|N) \cdot P(N)}_{P(N\wedge T)} = \\
&= 0.01 \cdot (1 - 0.001) + (1 - 0.1) \cdot 0.001 = 0.010\,89\,, \\
P(N|T) &= \frac{P(N \wedge T)}{P(T)} \doteq 0.09\,.
\end{aligned}$$

□

Příklad 7.9 (výskyt nemoci v populaci)**.** *Modifikace předchozího příkladu: Nevíme, kolik nemocných je v populaci, ale víme, že pravděpodobnost pozitivního testu je* 0.02. *(Test nemoci je u* 1% *zdravých falešně pozitivní a u* 10% *nemocných falešně negativní.) Odhadněte podíl nemocných je v populaci.*

Řešení. *T ... pozitivní test,* $\quad$ *N ... nemocný.*
$P(T) = 0.02, \quad P(T|\overline{N}) = 0.01, \quad P(\overline{T}|N) = 0.1.$

$$\begin{aligned}
P(T) &= P(T|\overline{N}) \cdot P(\overline{N}) + \underbrace{P(T|N) \cdot P(N)}_{P(N\wedge T)}\,, \\
0.02 &= 0.01 \cdot (1 - P(N)) + (1 - 0.1) \cdot P(N) = 0.89\,P(N) + 0.01\,, \\
P(N) &\doteq 0.011\,236\,.
\end{aligned}$$

□

Příklad 7.10 (bayesovský odhad vstupu informačního kanálu)**.** *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)

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

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

(b)

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

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

(c)

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

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

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).* □

Příklad 7.11. [2]*A. Rodina má dvě děti, starší je dcera. Jaká je pravděpodobnost, že mají dvě dcery? B. Rodina má dvě děti, (aspoň) jedno z nich je dcera. Jaká je pravděpodobnost, že mají dvě dcery?*

Řešení. *A. Jde o pravděpodobnost, že mladší z dětí je dcera, což nastává s pravděpodobností q blízkou $1/2$, přesněji asi 0.52. (Předpokládáme, že pohlaví dětí jsou nezávislá, což je přibližně správně.)*
B. ***Pozor, nejde o stejnou úlohu jako A!*** *Pokud pro jednoduchost předpokládáme $q = 1/2$, pak předpoklad J, že rodina má aspoň 1 dceru, je splněn s pravděpodobností $P(J) = 1 - (1-q)^2 = 3/4$, ale to, že má 2 dcery, je podjev $D \subseteq J$ s pravděpodobností $P(D) = q^2 = 1/4 = P(D \wedge J)$. Podmíněná pravděpodobnost je*

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

Obecněji pro pravděpodobnost narození dívky q

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

pro $q = 0.52$

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

8 Náhodné veličiny

Příklad 8.1 (rodiny s jedním chlapcem)**.** *V zemi je rodinám povoleno mít pouze jednoho chlapce a všechny rodiny usilují o to, aby ho měly. Jaký je podíl dívek? (Pro jednoduchost zanedbáváme úmrtnost a vícečetné porody a předpokládáme, že pravděpodobnost narození chlapce i dívky je stejná.)*

Řešení. *Může se stát, že rodina má samé dívky a na chlapce dosud čeká. Prozatím to ignorujme a uvažujme rodiny, které mají chlapce (jako poslední dítě).*
Počet dívek v náhodně vybrané rodině z tohoto souboru je náhodná veličina X, jejíž hodnoty jsou nezáporná celá čísla.

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

S pravděpodobností $1/2$ *se narodil chlapec jako první dítě a* $X = 0$.
S pravděpodobností $1/4$ *se narodil chlapec jako druhé dítě a* $X = 1$.
S pravděpodobností $1/8$ *se narodil chlapec jako třetí dítě a* $X = 2$.
...
Průměrný počet dívek na jednoho chlapce je dán střední hodnotou

$$\mathrm{E}X = \sum_{n=0}^{\infty} n\, P[X = n] = \sum_{n=0}^{\infty} n\, 2^{-(n+1)} = 1\,.$$

Alternativa: Lze říci, že
$1/2$ *chlapců má nejstarší sestru,*
$1/4$ *chlapců má druhou sestru,*
$1/8$ *chlapců má třetí sestru,*
...
celkem

$$\sum_{n=0}^{\infty} 2^{-(n+1)} = 1\,.$$

Alternativa: Podle předpokladu se rodí stejně chlapců jako dívek, to vede přímo ke správnému závěru.
Problém: Zanedbávali jsme rodiny, ve kterých není chlapec. Ten se narodí ***skoro jistě****, tj. s pravděpodobností* 1*, ale budoucí chlapci mají již teď sestry bez bratrů.*
Je to jen problém definice počátku a konce pokusu, s rostoucí délkou pokusu jeho vliv klesá k nule.
Problém: Vliv by neklesal k nule, kdyby docházelo k velkému populačnímu růstu nebo poklesu.
Podmínky úlohy vylučují velký nárůst, nikoli však velký pokles. □

9 Směs náhodných veličin

Příklad 9.1. *Máme* 2 *hrací kostky, na jedné padají pouze lichá čísla* $1, 3, 5$*, na druhé pouze sudá,* $2, 4, 6$*, všechna se stejnou pravděpodobností* $1/3$*. Najděte rozdělení a střední hodnotu výsledků následujících pokusů:*

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

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

Řešení. *(a) Rozlišíme* 9 *stejně pravděpodobných možností, vedoucích k následujícím výsledkům:*

	1	3	5
2	1.5	2.5	3.5
4	2.5	3.5	4.5
6	3.5	4.5	5.5

Možné výsledky a jejich pravděpodobnosti:

1.5	2.5	3.5	4.5	5.5
1/9	2/9	3/9	2/9	1/9

Střední hodnota je

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

(b) S pravděpodobností $1/2$ *určuje výsledek první kostka, se stejnou pravděpodobností druhá; dostáváme* 6 *stejně pravděpodobných výsledků, rozdělení je stejné jako u normální hrací kostky.*

1	3	5
2	4	6

Střední hodnota je stejná,

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

□

Příklad 9.2. *V urně je* 15 *hracích kostek, z toho* 10 *správných, na nichž padají všechna čísla se stejnou pravděpodobností, a* 5 *vadných, na nichž padá šestka s pravděpodobností* $1/2$*, ostatní čísla s pravděpodobností* $1/10$*. Náhodně vybereme jednu kostku a hodíme; jaká je pravděpodobnost možných výsledků?*

Řešení. *Označme náhodné veličiny:*
U *výsledek na správné kostce,*
V *výsledek na vadné kostce,*
X *výsledek celého pokusu (směs náhodných veličin* U, V *s koeficientem* $c = 10/15 = 2/3$*).*

$$
\begin{aligned}
P[X = t] &= \frac{2}{3} P[U = t] + \frac{1}{3} P[V = t] \\
P[X = 1] &= \frac{2}{3} \frac{1}{6} + \frac{1}{3} \frac{1}{10} = \frac{13}{90} \\
P[X = 6] &= \frac{2}{3} \frac{1}{6} + \frac{1}{3} \frac{1}{2} = \frac{5}{18}
\end{aligned}
$$

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

□

10 Druhy náhodných veličin

10.1 Diskrétní náhodné veličiny

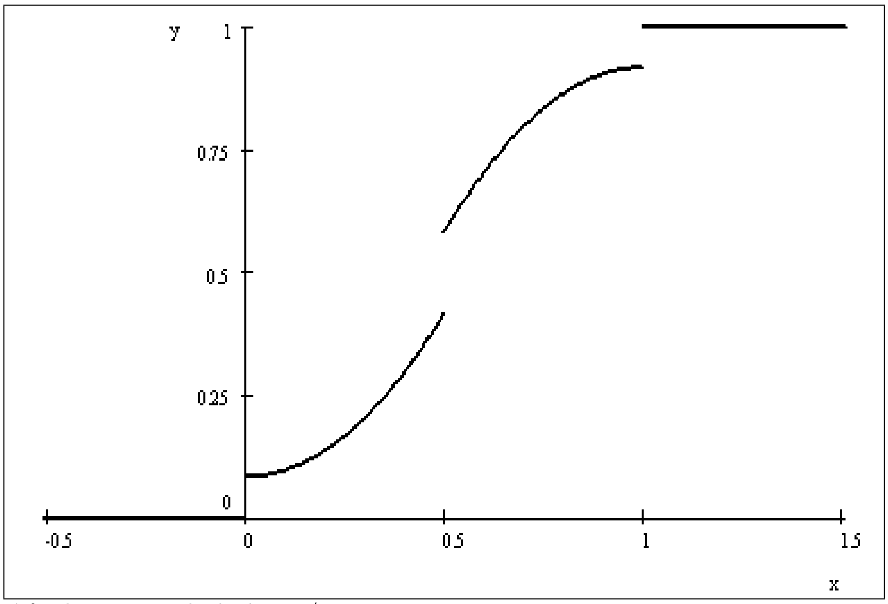
10.2 Spojité náhodné veličiny

10.3 Náhodné veličiny se smíšeným rozdělením

Příklad 10.1. *Náhodná veličina* X *má distribuční funkci*

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

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



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

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

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

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

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

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

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

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

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

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

□

Příklad 10.2. *Náhodná veličina X má alternativní rozdělení (s hodnotami $0,1$), $P[X=1] = 2/3$. Náhodná veličina Y má spojité rovnoměrné rozdělení na intervalu $\langle 0,2\rangle$. Popište rozdělení jejich směsi $Z = \mathrm{Mix}_{2/3}(X,Y)$.*

Řešení.

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

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

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

□

Příklad 10.3. *Náhodná veličina X má distribuční funkci*

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

Vyjádřete její rozdělení jako směs diskrétního a spojitého rozdělení.

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

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

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

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

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

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

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

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

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

□

11 Nezávislost náhodných veličin

12 Operace s náhodnými veličinami

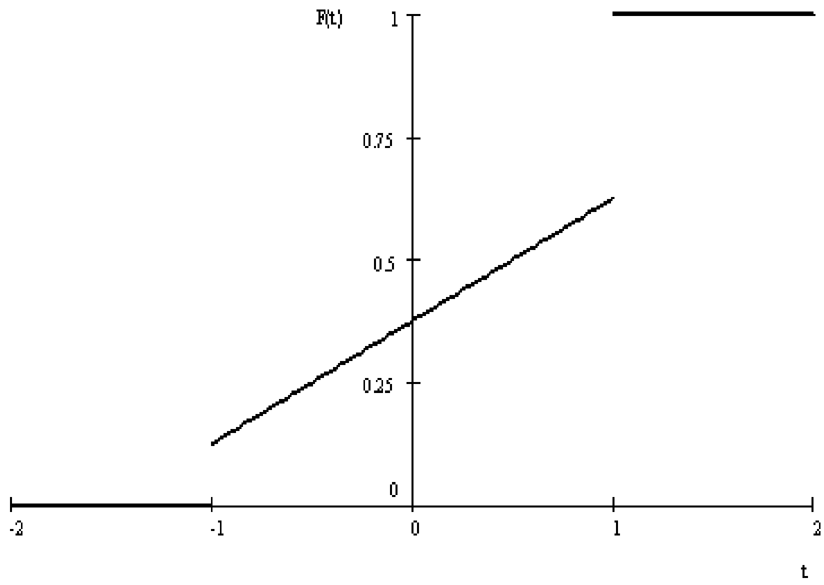
Příklad 12.1. *Náhodná veličina má spojité rovnoměrné rozdělení na intervalu $\langle -3, 5\rangle$. Zobrazte ji funkcí*

$$h(x) = \begin{cases} -1 & \textit{pro } x < -2, \\ x/2 & \textit{pro } x \in \langle -2, 2\rangle\,, \\ 1 & \textit{pro } x > 2, \end{cases}$$

výsledné rozdělení popište a znázorněte.

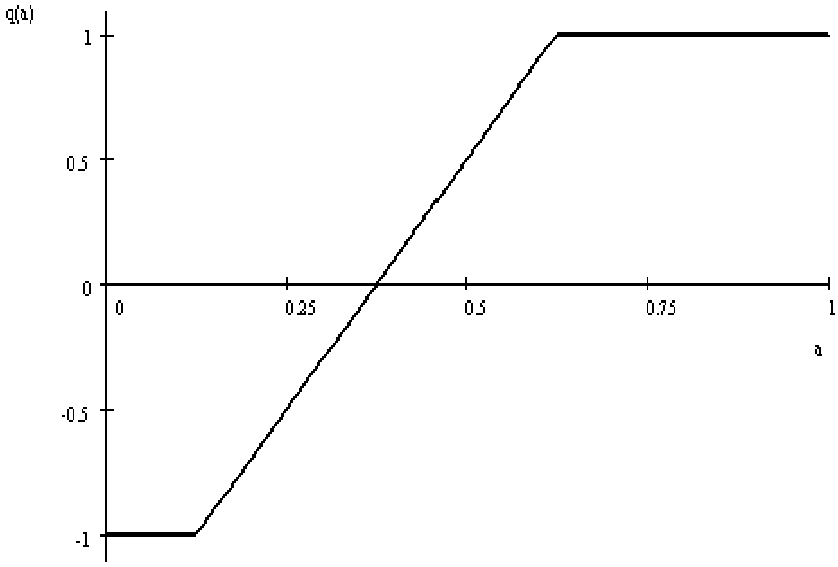
Řešení. *Výstup -1 odpovídá vstupu v intervalu $\langle -3,-2\rangle$, a má tedy pravděpodobnost $1/8$, $P[h(X) = -1] = 1/8$. Výstup 1 odpovídá vstupu v intervalu $\langle 2, 5\rangle$, a má tedy pravděpodobnost $3/8$, $P[h(X) = 1] = 3/8$. Zbývající hodnoty vedou na spojité rovnoměrné rozdělení na $\langle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

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



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

$$q_{h(X)}(a) = h(q_X(a)) = \begin{cases} -1 & \text{pro} \quad a \le 1/8, \\ 4\,a - 3/2 & \text{pro} \quad a \in (1/8, 5/8), \\ 1 & \text{pro} \quad a \ge 5/8. \end{cases}$$



□

Příklad 12.2. *Náhodné veličiny* X, Y *jsou nezávislé,* X *má spojité rovnoměrné rozdělení na intervalu* $\langle 0, 1\rangle$, Y *má alternativní rozdělení,*

$$p_Y(t) = \begin{cases} 1/2 & \text{pro } t \in \{0, 1\}, \\ 0 & \text{jinak.} \end{cases}$$

Popište a znázorněte rozdělení náhodných veličin

1. $X + Y$,
2. $\mathrm{Mix}_{1/2}(X, Y)$ *(směs* X *a* Y*)*,
3. $X + \mathrm{E}Y$.

Příklad 12.3. *Náhodné veličiny* X, Y *jsou nezávislé, mají spojité rovnoměrné rozdělení;* X *na intervalu* $\langle 0, 1\rangle$, Y *na intervalu* $\langle 1, 2\rangle$. *Popište a znázorněte rozdělení náhodných veličin*

1. $X + Y$,

2. $\mathrm{Mix}_{1/2}(X, Y)$ *(směs X a Y)*,

3. $X + \mathrm{E}Y$.

Příklad 12.4. *Náhodná veličina X má distribuční funkci*

$$F_X(t) = \begin{cases} 1 - \frac{1}{2} e^{-x-2} & \textit{pokud } x \geq -1, \\ 0 & \textit{jinak.} \end{cases}$$

Určete a znázorněte rozdělení náhodných veličin

1. $X + 2$,

2. $X/2$,

3. $-2X$.

13 Základní charakteristiky náhodných veličin

Příklad 13.1. *Náhodný vektor (X, Y) má následující parametry: $\mathrm{E}X = 10$, $\sigma_X = 5$, $\mathrm{E}Y = 150$, $\sigma_Y = 20$, $\varrho(X, Y) = 0.5$ (korelace). Stanovte střední hodnotu a rozptyl náhodných veličin $T = 2\,X + 3$, $U = 200 - Y$, $V = X + Y$.*

Příklad 13.2. *Nezávislé náhodné veličiny $X_1, X_2, \ldots, X_n$, $n \in \mathbb{N}$ mají (stejné) rovnoměrné rozdělení na intervalu $(-a, 2\,a)$, $a \in (0, \infty)$. Určete střední hodnotu a rozptyl náhodné veličiny*

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

Příklad 13.3. *V písemce jsou 2 různě obtížné otázky, studenti z nich v průměru získají $p_i \times$ celkový počet bodů za otázku, $p_i \in (0, 1\rangle$, $i = 1, 2$. Nabízejí se tři bodovací systémy:*

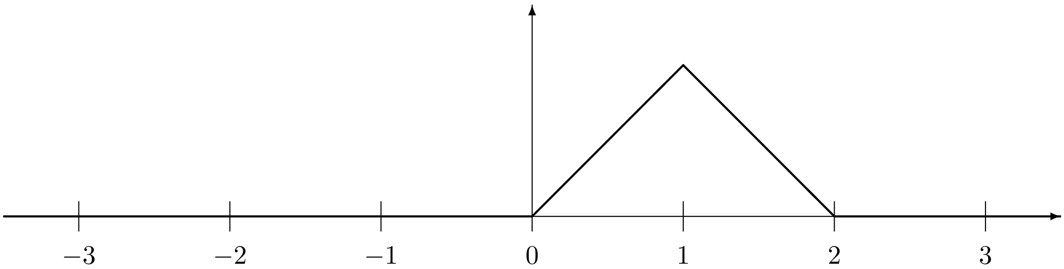
1. *všechny otázky mají stejný počet bodů,*

2. *počet bodů za i-tou otázku je úměrný $1 - p_i$.*

3. *počet bodů za i-tou otázku je nepřímo úměrný p_i.*

(Celkový počet bodů je ve věech případech stejný.) Při kterém systému získají studenti v průměru více bodů?

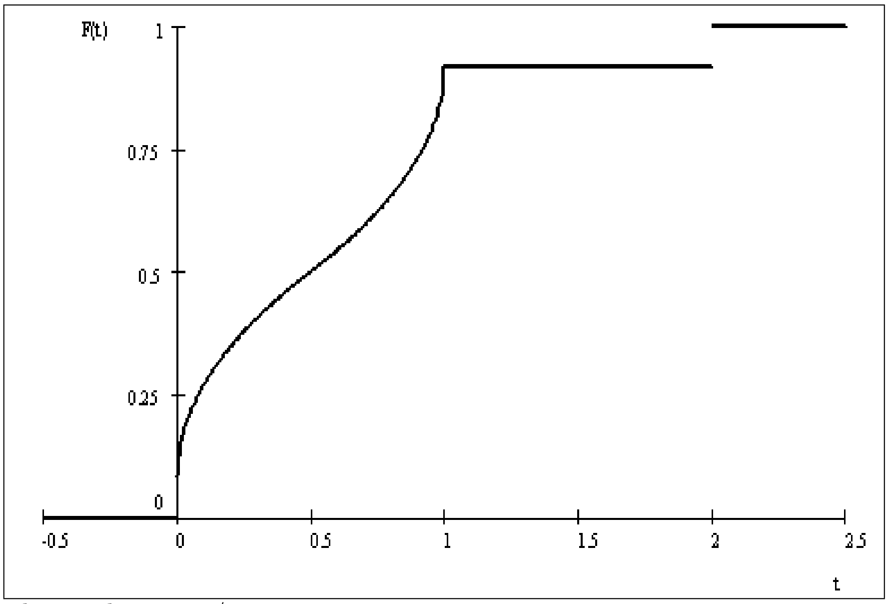
Příklad 13.4. *Náhodná veličina U má hustotu danou grafem. Určete a znázorněte hustoty a distribuční funkce náhodných veličin (a) $U - 1$, (b) $-2\,U$, (c) $\exp U$.*



Příklad 13.5. *Náhodná veličina X má distribuční funkci*

$$F_X(t) = \begin{cases} 0 & \textit{pro} \quad t < 0 \\ \frac{1+5\sqrt{2\,t}}{12} & \textit{pro} \quad 0 \leq t < 1/2 \\ \frac{11-5\sqrt{2\,(1-t)}}{12} & \textit{pro} \quad 1/2 \leq t < 1 \\ \frac{11}{12} & \textit{pro} \quad 1 \leq t < 2 \\ 1 & \textit{pro} \quad t \geq 2 \end{cases}$$

Najděte její střední hodnotu.



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

14 Náhodné vektory (vícerozměrné náhodné veličiny)

Příklad 14.1. *Dvojrozměrný náhodný vektor* (X, Y) *má pravděpodobnosti hodnot dané tabulkou:*

$Y \backslash X$	1	2
0	1/3	1/3
1	0	1/3

Vypočtěte korelaci náhodných veličin X, Y.

Řešení. $EX = \frac{5}{3}$, $EY = \frac{1}{3}$, $DX = DY = \frac{2}{9}$, $E\,(X\,Y) = \frac{2}{3}$,

$$\varrho\,(X, Y) = \frac{E\,(X\,Y) - EX\,EY}{\sigma_X\,\sigma_Y} = \frac{1}{2}\,.$$

□

Příklad 14.2. *Náhodný vektor má rovnoměrné rozdělení na trojúhelníku s vrcholy* $(0, 0)$, $(1, 0)$, $(1, 1)$. *Popište a znázorněte distribuční funkce jeho složek (marginální rozdělení).*

Řešení. *1. postup: Marginální hustoty jsou*

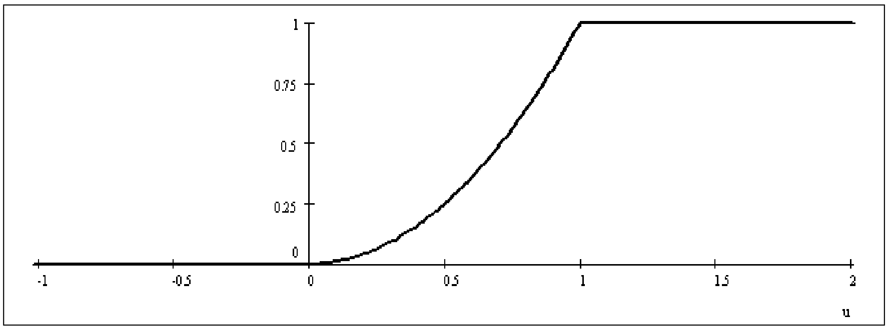
$$f_X\,(t) = \begin{cases} 2\,t & \text{pro } 0 \le t \le 1, \\ 0 & \text{jinak,} \end{cases}$$

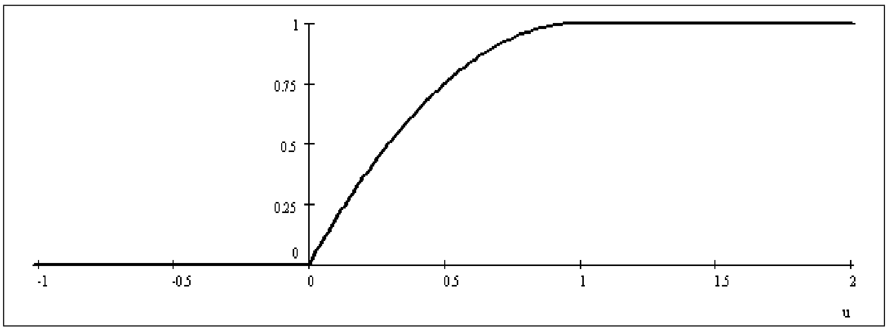
$$f_Y\,(t) = \begin{cases} 2\,(1-t) & \text{pro } 0 \le t \le 1, \\ 0 & \text{jinak,} \end{cases}$$

distribuční funkce dostaneme jejich integrací:

$$F_X\,(u) = \int_{-\infty}^{u} f_X\,(t)\,\mathrm{d}t = \begin{cases} 0 & \text{pro } u < 0, \\ u^2 & \text{pro } 0 \le u \le 1, \\ 1 & \text{pro } u > 1, \end{cases}$$

$$F_Y\,(u) = \int_{-\infty}^{u} f_Y\,(t)\,\mathrm{d}t = \begin{cases} 0 & \text{pro } u < 0, \\ 2\,u - u^2 & \text{pro } 0 \le u \le 1, \\ 1 & \text{pro } u > 1. \end{cases}$$





2. postup: Distribuční funkce je podle definice dána poměrem obsahů ploch (vesměs se jedná o trojúhelníky nebo lichoběžníky, takže nepotřebujeme integrovat a vystačíme s geometrií ze základní školy); vždy je nutno dělit obsahem celého daného trojúhelníka, což je $1/2$. *Pro* $0 \le u \le 1$ *vychází*

$$F_X(u) = \frac{\frac{u^2}{2}}{\frac{1}{2}} = u^2\,,$$

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

□

Příklad 14.3. *Náhodné veličiny* X, Y *jsou nezávislé. Určete korelaci* $\varrho(U,V)$ *náhodných veličin* $U = X + Y$, $V = X - Y$.

Příklad 14.4. *Známe korelace náhodných veličin* $\varrho(X,Y) = 0.5$, $\varrho(Y,Z) = 1/\sqrt{2}$. *Můžeme něco říci o korelaci* $\varrho(X,Z)$ *(a její existenci)?*

15 Čebyševova nerovnost, centrální limitní věta

Příklad 15.1. *Ryby mohou si vybrat ze* 2 *cest, z nichž jedna je správná (vede k potravě). Každá ryba nezávisle pozná správnou cestu s pravděpodobností* $q = 0.6$. *Jaká je pravděpodobnost, že „většinové hlasování" v hejnu* n *ryb vybere správnou cestu?*

Řešení. *Rozhodnutí jednotlivých ryb popisují nezávislé náhodné veličiny* X_j, $j = 1, \ldots, n$ *s alternativním rozdělením s parametrem* $q = 0.6$. *(Správnou cestu vyhodnocujeme jako* 1, *špatnou* 0.*) Z vlastností alternativního rozdělení*

$$\mathrm{E}X_j = q\,, \qquad \mathrm{D}X_j = q\,(1-q)\,.$$

Pro výběrový průměr

$$\mathrm{E}\overline{X} = q\,, \qquad \mathrm{D}\overline{X} = \frac{q\,(1-q)}{n}\,,$$

jeho rozdělení pro velká n *můžeme podle centrální limitní věty přibližně nahradit normálním rozdělením se stejnými parametry, tj.* $\mathrm{N}\left(q, \frac{q\,(1-q)}{n}\right)$. *Odchylku střední hodnoty od* $50\,\%$, $0.5 - q = 0.5 - 0.6 = -0.1$ *budeme měřit směrodatnou odchylkou výběrového průměru*

$$\sigma_{\overline{X}} = \sqrt{\frac{q\,(1-q)}{n}} = \sqrt{\frac{0.6 \cdot (1-0.6)}{n}} \doteq \frac{0.49}{\sqrt{n}}\,,$$

poměr $\frac{-0.1}{0.49}\sqrt{n}$ *bude argumentem distribuční funkce normovaného normálního rozdělení. Pravděpodobnost, že se hejno rozhodne chybně, je*

$$P\left[\overline{X} \le 0.5\right] \doteq F_{\mathrm{N}\left(q,\frac{q\,(1-q)}{n}\right)}(0.5-q) = \Phi\left(\frac{0.5-q}{\sigma_{\overline{X}}}\right) \doteq$$
$$\doteq \Phi\left(\frac{-0.1}{0.49}\sqrt{n}\right) \doteq \Phi\left(-0.204\,08\sqrt{n}\right).$$

Pravděpodobnost správného rozhodnutí hejna je k ní doplňková,

$$P\left[\overline{X} > 0.5\right] \doteq 1 - F_{\mathrm{N}\left(q,\frac{q\,(1-q)}{n}\right)}(0.5-q) = 1-\Phi\left(\frac{0.5-q}{\sigma_{\overline{X}}}\right) = \Phi\left(\frac{q-0.5}{\sigma_{\overline{X}}}\right) \doteq$$
$$\doteq \Phi\left(\frac{0.1}{0.49}\sqrt{n}\right) \doteq \Phi\left(0.204\sqrt{n}\right).$$

Číselné hodnoty pro několik hodnot n *udává tabulka:*

n	$0.204\sqrt{n}$	$P\left[\overline{X} > 0.5\right]$
10	0.645	0.74
100	2.04	0.98
1000	6.45	$1-6\cdot10^{-11}$

□

Příklad 15.2. *Ve vzorku je* 1 mg *uhlíku, tj. asi* $6\cdot10^{23}\cdot10^{-3}/12 = 5\cdot10^{19}$ *atomů. Z nich je přibližně* $1/10^{12}$, *tj. asi* $5\cdot10^{7}$, *atomů radioaktivního izotopu C14. Určete symetrický* 95 %*-ní intervalový odhad počtu atomů, které se rozpadnou za* 1 *rok, tj. za* 1/5730 *poločasu rozpadu. Co o tom říká Čebyševova nerovnost?*

Řešení. *Odhadujeme náhodnou veličinu* X *s rozdělením* $\mathrm{Bi}(n,p)$, $n = 5\cdot10^7$, $p = 1-1/2^{1/5730} \doteq 1.2\cdot10^{-4}$ *(=pravděpodobnost, že se atom v daném čase rozpadne),*

$$\mathrm{E}X = n\,p \doteq 6048\,,$$
$$\mathrm{D}X = n\,p\,(1-p) \doteq 6047\,,$$
$$\sigma_X = \sqrt{n\,p\,(1-p)} \doteq 78\,.$$

Při aproximaci normálním rozdělením vyjdou meze $\mathrm{E}X \pm \sigma_X\,\Phi^{-1}(0.975) \doteq 6048 \pm 78\cdot1.96 \doteq 6048\pm153$, *interval přibližně* $\langle 5895, 6201\rangle$, *relativní chyba zhruba* $153/6048 \doteq 2.5\,\%$.
Exaktní výpočet z binomického rozdělení by byl pracný a vedl by k velmi podobným výsledkům.
Čebyševova nerovnost nezohledňuje znalost rozdělení (přibližně normální) a vede na intervalový odhad s tolerancí ε *splňující nerovnost*

$$\frac{\mathrm{D}X}{\varepsilon^2} \le 0.05\,,$$
$$\varepsilon \ge \sqrt{\frac{\mathrm{D}X}{0.05}} = \frac{\sigma_X}{\sqrt{0.05}} \doteq \frac{78}{\sqrt{0.05}} \doteq 349\,,$$

meze $6048\pm349 = 6397$, *interval přibližně* $\langle 5699, 6397\rangle$, *relativní chyba zhruba* $349/6048 = 5.7\,\%$. □

Příklad 15.3. *Životnost baterie má exponenciální rozdělení se střední hodnotou* 3 *hodiny. Určete pravděpodobnost, že* 100 *baterií zajistí alespoň* 252 *hodin provozu.*

Příklad 15.4. *Na oboru má studovat* 600 *studentů, avšak fakulta smí stanovit pouze počet přijatých. Z dlouhodobé zkušenosti se ukazuje, že z přijatých studentů se zapíše asi* 2/3. *Jaké se má stanovit směrné číslo pro přijetí, aby počet zapsaných byl co největší, ale aby překročil* 600 *s pravděpodobností nejvýše* 5 %*? Jaký bude průměrný počet zapsaných studentů? Jak se úloha změní pro obor, na který má být přijato* 60 *studentů? Uveďte použité předpoklady.*

Příklad 15.5. *Alice nabídla Bobovi sázku* 1 : 1000, *že nedokáže z* 500 *hodů mincí aspoň v* 60 % *hodit líc. Bob váhá, proto Alice navíc nabízí, že Bob má* 10 *pokusů (po* 500 *hodech) a stačí, když aspoň v jednom z nich uspěje. Kurs zůstává* 1 : 1000. *Má Bob sázku přijmout?*

Řešení. *Je-li mince regulérní a $n = 500$ je počet pokusů, pak výběrový průměr má podle centrální limitní věty rozdělení přibližně $N\left(\frac{1}{2}, \frac{1}{4\,n}\right)$. Počet líců je $n\times$ větší, má tedy rozdělení přibližně $N\left(\frac{n}{2}, \frac{n}{4}\right)$. Pravděpodobnost, že Bob v jednom kole dosáhne o $10\,\%$ víc než polovinu líců, je*

$$1 - \Phi\left(\frac{0.1\,n}{\sqrt{\frac{n}{4}}}\right) = 1 - \Phi\left(0.2\,\sqrt{n}\right) \doteq 1 - \Phi\,(4.\,472) \doteq 3.\,9 \times 10^{-6}\,.$$

Při 10 opakovaných pokusech se pravděpodobnost úspěchu zvýší méně než $10\times$, sázka zůstává pro Boba velmi nevýhodná. □

Příklad 15.6. *Počet X ryb, které rybář uloví za den, je popsán Poissonovým rozdělením,*

$$p_X(k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,, \qquad k \in \{0, 1, 2, \ldots\}\,,$$

s parametrem $\lambda = 3$. Na ryby jde $n = 100\times$ za rok. Najděte (co nejmenší) symetrický interval, v němž se počet ulovených ryb za rok nachází s pravděpodobností aspoň $95\,\%$.

Řešení. $300 \pm 1.96 \cdot \sqrt{3} \cdot 10 = \langle 266.\,05, 333.\,95\rangle$ □

Část II
Základy matematické statistiky

16 Bodové odhady charakteristik rozdělení

17 Intervalové odhady charakteristik rozdělení

17.1 Intervalové odhady normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$

17.1.1 Odhad střední hodnoty při známém rozptylu σ^2

Příklad 17.1. *Rozvodné závody dodávaly elektřinu, jejíž napětí ve voltech mělo normální rozdělení* $\mathrm{N}(230, 25)$. *Nyní se jim podařilo snížit rozptyl na 10. O kolik mohou zvýšit střední hodnotu při zachování horní meze, která je překročena jen s pravděpodobností 10^{-4}?*

Příklad 17.2. *Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě 67 a 75 m a směrodatné odchylky 6 a 3 m. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál.*

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

$$1 - F_{\mathrm{N}(-8,45)}\,(0) = 1 - \Phi\left(\frac{0 - (-8)}{\sqrt{45}}\right) \doteq 1 - \Phi\,(1.\,192\,6) \doteq 1 - 0.883 = 0.117\,.$$

□

17.1.2 Odhad střední hodnoty při neznámém rozptylu

17.1.3 Odhad rozptylu a směrodatné odchylky

Příklad 17.3. *Opakovaná měření stejné koncentrace látky vedla k následujícím výsledkům:* $(0.2, 0.23, 0.21, 0.16, 0.18, 0.19, 0.14, 0.18, 0.21)$. *Najděte symetrické oboustranné $90\,\%$-ní odhady střední hodnoty, rozptylu a směrodatné odchylky.*

Řešení. *Odhadujeme parametry náhodné veličiny X z realizace rozsahu $n = 9$, jejíž statistiky jsou realizace výběrového průměru $\bar{x} \doteq 0.189$, realizace výběrového rozptylu $s_x^2 \doteq 7.6 \cdot 10^{-4}$, realizace výběrové směrodatné* *odchylky* $s_x = \sqrt{s_x^2} \doteq 2.76 \cdot 10^{-2}$. *Intervalový odhad střední hodnoty:*

$$\left\langle \bar{\boldsymbol{x}} - \frac{s_x}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(0.95), \bar{\boldsymbol{x}} + \frac{s_x}{\sqrt{n}}\, q_{\mathrm{t}(n-1)}(0.95) \right\rangle \doteq$$
$$\doteq \left\langle 0.189 - \frac{2.76 \cdot 10^{-2}}{3}\, \underbrace{q_{\mathrm{t}(8)}(0.95)}_{1.86}, 0.189 + \frac{2.76 \cdot 10^{-2}}{3}\, q_{\mathrm{t}(8)}(0.95) \right\rangle \doteq$$
$$\doteq \langle 0.172, 0.206 \rangle\,.$$

Intervalový odhad rozptylu:

$$\left\langle \frac{(n-1)\, s_x^2}{q_{\chi^2(n-1)}\,(0.95)}, \frac{(n-1)\, s_x^2}{q_{\chi^2(n-1)}\,(0.05)} \right\rangle \doteq$$
$$\doteq \left\langle \frac{8 \cdot 7.6 \cdot 10^{-4}}{\underbrace{q_{\chi^2(8)}\,(0.95)}_{15.51}}, \frac{8 \cdot 7.6 \cdot 10^{-4}}{\underbrace{q_{\chi^2(8)}\,(0.05)}_{2.73}} \right\rangle \doteq$$
$$\doteq \left\langle 3.9 \cdot 10^{-4},\, 2.2 \cdot 10^{-3} \right\rangle\,.$$

Intervalový odhad směrodatné odchylky (odmocnina z předchozího):

$$\left\langle \sqrt{\frac{(n-1)\, s_x^2}{q_{\chi^2(n-1)}\,(0.95)}}, \sqrt{\frac{(n-1)\, s_x^2}{q_{\chi^2(n-1)}\,(0.05)}} \right\rangle \doteq$$
$$\doteq \left\langle \sqrt{3.9 \cdot 10^{-4}},\, \sqrt{2.2 \cdot 10^{-3}} \right\rangle \doteq \left\langle 1.97 \cdot 10^{-2}. 4.7 \cdot 10^{-2} \right\rangle\,.$$

Všimněte si, že inervalové odhady výběrového rozptylu, resp. směrodatné odchylky nejsou symetrické kolem jejich bodových odhadů $s_x^2 \doteq 7.6 \cdot 10^{-4}$, *resp.* $s_x \doteq 2.76 \cdot 10^{-2}$. □

18 Odhad parametrů (metoda momentů, metoda maximální věrohodnosti)

18.1 Odhady diskrétních rozdělení

Příklad 18.1. *Gen se vyskytuje ve 4 variantách A, B, C, D. Model předpokládá, že B se vyskytuje 3× častěji než A a D 3× častěji než C. Odhadněte jejich pravděpodobnosti na základě zjištěných četností v tabulce.*

varianta	*A*	*B*	*C*	*D*
četnost	10	15	15	40

Příklad 18.2. *V urně je mnoho hracích kostek, z nichž některé jsou správné, některé falešné. Na falešných padá šestka s pravděpodobností* 1/2*, zbývající čísla mají stejnou pravděpodobnost. Opakovaně jsme vytáhli kostku, hodili s ní a vrátili ji zpět. Četnost výsledků udává tabulka:*

hodnota	1	2	3	4	5	6
četnost	18	20	12	15	10	25

Odhadněte, kolik procent kostek je falešných.

Řešení. *Podíl falešných kostek označme* $p \in \langle 0, 1 \rangle$.

Metoda momentů:
Střední hodnota výsledku pro správnou kostku je 3.5*, pro falešnou* 4.5*, pro směs s koeficientem* p *vychází* $3.5\,(1-p) + 4.5\,p = 3.5 + p$.
Realizace výběrového průměru je 3.54.
Srovnáním těchto dvou hodnot vyjde odhad $\hat{p} = 0.04 \in \langle 0, 1 \rangle$*, což vyhovuje zadání.*

Metoda maximální věrohodnosti:
Ve směsi rozdělení má šestka pravděpodobnost $\frac{1}{6}\,(1-p) + \frac{1}{2}\,p = \frac{1+2\,p}{6}$ *a padla* 25×*, ostatní čísla* $\frac{1}{6}\,(1-p) + \frac{1}{10}\,p = \frac{5-2\,p}{30}$ *a padla* 75× *(není třeba mezi nimi rozlišovat).*

$$L\,(p) = \left(\frac{5-2\,p}{30}\right)^{75} \cdot \left(\frac{1+2\,p}{6}\right)^{25},$$
$$\ell\,(p) = 75\,\ln(5-2\,p) + 25\,\ln(1+2\,p) - 75\,\ln 30 - 25\,\ln 6\,.$$

Maximum nastává pro $\hat{p}$ takové, že

$$\frac{\partial}{\partial \hat{p}} \ell(\hat{p}) = \frac{-150}{5-2\,\hat{p}} + \frac{50}{1+2\,\hat{p}} = 0\,,$$
$$\hat{p} = \frac{1}{4} \in \langle 0,1 \rangle\ .$$

□

Příklad 18.3. *Náhodná veličina X je směsí náhodných veličin Y, Z, jejichž pravděpodobnostní funkce jsou dány tabulkou:*

hodnota	1	2	3	4
p_Y	0.4	0.4	0.1	0.1
p_Z	0.1	0.1	0.4	0.4
pozorovaná četnost	12	13	9	6

Poslední řádek udává četnosti hodnot v realizaci náhodného výběru s rozdělením, které má náhodná veličina X. Odhadněte z nich neznámý koeficient směsi.

Řešení. *Metoda momentů:*

$$\begin{aligned} \mathrm{E}X &= w\,\mathrm{E}Y + (1-w)\ \mathrm{E}Z = \\ &= w\ (0.4\cdot 1 + 0.4\cdot 2 + 0.1\cdot 3 + 0.1\cdot 4) + \\ &+ (1-w)\ (0.1\cdot 1 + 0.1\cdot 2 + 0.4\cdot 3 + 0.4\cdot 4) = \\ &= 1.9\,w + 3.1\ (1-w) = 3.1 - 1.2w = \\ &= \frac{12\cdot 1 + 13\cdot 2 + 9\cdot 3 + 6\cdot 4}{12+13+9+6} = \frac{89}{40} = 2.225\,, \\ w &= \frac{35}{48} \doteq 0.729\,17\,. \end{aligned}$$

Vyhovuje zadání.
Metoda maximální věrohodnosti:
$0.4\,w + 0.1\ (1-w) = 0.3\,w + 0.1,\ 0.1\,w + 0.4\ (1-w) = 0.4 - 0.3w\,.$

hodnota	1	2	3	4
p_X	$0.1 + 0.3\,w$	$0.1 + 0.3\,w$	$0.4 - 0.3w$	$0.4 - 0.3w$
pozorovaná četnost	12	13	9	6

$$\begin{aligned} L(w) &= (0.1 + 0.3\,w)^{12+13} \cdot (0.4 - 0.3w)^{9+6} = (0.1 + 0.3\,w)^{25}\ (0.4 - 0.3w)^{15}\ , \\ \ell(w) &= \ln(L(w)) = 25\ \ln(0.1 + 0.3\,w) + 15\ \ln(0.4 - 0.3w)\ , \\ \ell'(w) &= \frac{7.5}{0.1 + 0.3\,w} - \frac{4.5}{0.4 - 0.3\,w} = 0\,, \\ w &= \frac{17}{24} \doteq 0.708\,33\,. \end{aligned}$$

□

Příklad 18.4. *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. $p = 1 - q - q^2$
Metoda momentů: $\mu_X = q + 2q^2$, $m_X = \frac{1\cdot 12 + 2\cdot 6}{2+12+6} = \frac{6}{5}$, $\mu_X = m_X$.
$\Rightarrow q_1 = -\frac{1}{20}\sqrt{265} - \frac{1}{4} = -1.063\,9$ *(nevyhovuje)*, $q_2 = \frac{1}{20}\sqrt{265} - \frac{1}{4} = 0.563\,94$ *(vyhovuje)*, $p = 1 - q_2 - q_2^2 = 0.118\,03$.
Metoda maximální věrohodnosti: $L(q) = 2\ln p + 12\ln q + 6\ln q^2 = 2\ln\left(1 - q - q^2\right) + 24\ln q$,
$\frac{\partial L}{\partial q}(q) = \frac{24}{q} + 2\frac{-2q-1}{1-q^2-q} = 0 \Rightarrow q_1 = -\frac{3}{2}$*(nevyhovuje)*, $q_2 = \frac{4}{7} = 0.571\,43$ *(vyhovuje)*, $p = 1 - q_2 - q_2^2 = \frac{5}{49} = 0.102\,04$. □

Příklad 18.5. *V osudí jsou 2 druhy kostek, na prvních jsou čísla $1, \ldots, 6$, na druhých pouze $1, 3, 5$, u obou druhů jsou všechny možné výsledky stejně pravděpodobné. Vytáhli jsme 20 kostek a jednou jimi hodili; četnost výsledků udává tabulka. Odhadněte, kolik z těchto kostek bylo prvního druhu.*

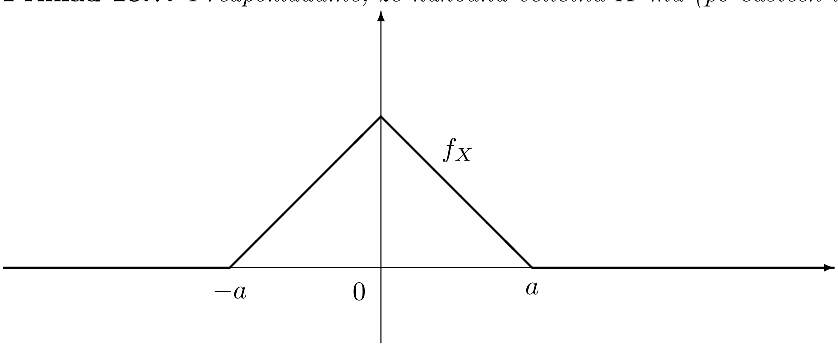
hodnota	*1*	*2*	*3*	*4*	*5*	*6*
četnost	*3*	*4*	*4*	*4*	*2*	*3*

Příklad 18.6. *Náhodná veličina nabývá výsledky $1, 2, 3$. Tabulka uvádí jejich pravděpodobnosti a pozorované četnosti. Odhadněte parametry a, b.*

hodnota	1	2	3
teoretická pravděpodobnost	$a+b$	$a+2b$	$a+3b$
četnost	10	10	20

18.2 Odhady spojitých rozdělení

Příklad 18.7. *Předpokládáme, že náhodná veličina X má (po částech lineární) hustotu dle obrázku.*



Na základě realizace

1. $\mathbf{x} = (-2, 1, 1)$
2. $\mathbf{x} = (-1, 1, 2)$

odhadněte parametr $a > 0$.

Příklad 18.8. *Předpokládáme, že náhodná veličina X má posunuté exponenciální rozdělení s hustotou*

$$f_X(t) = \begin{cases} \frac{1}{\tau} \exp\left(-\frac{t-T}{\tau}\right) & \text{pro } t \geq T, \\ 0 & \text{jinak,} \end{cases}$$

kde $\tau > 0$. Z realizace $x = (2, 3, 8, 4, 10, 3, 5)$ odhadněte parametry T, τ.

Řešení. *Metoda maximální věrohodnosti:*

$$L(T, \tau) = \ln\left(\prod_{i=1}^{n} \frac{1}{\tau} \exp\left(-\frac{x_i - T}{\tau}\right)\right) = -n \ln \tau - \frac{1}{\tau} \sum_{i=1}^{n} x_i + \frac{1}{\tau} n T,$$

pokud $T \leq \min_i x_i$ (jinak 0). To je rostoucí funkce T, takže $\hat{T} = \min_i x_i$.

$$0 = \frac{\partial L}{\partial \tau}\left(\hat{T}, \hat{\tau}\right) = -\frac{n}{\hat{\tau}} + \frac{1}{\hat{\tau}^2} \underbrace{\sum_{i=1}^{n} x_i}_{n \bar{x}} - \frac{1}{\hat{\tau}^2} n \hat{T},$$

$$\hat{\tau} = \bar{x} - \hat{T} = \bar{x} - \min_i x_i .$$

V našem případě $\hat{T} = 2$, $\hat{\tau} = 5 - 2 = 3$.

Metoda momentů:

$$\mu_X = \int_{\mathbb{R}} t\, f_X(t)\,\mathrm{d}t = \int_T^\infty \frac{t}{\tau} \exp\left(-\frac{t-T}{\tau}\right)\mathrm{d}t = (-t-\tau)\exp\left(-\frac{t-T}{\tau}\right)\Big|_{t=T}^{\infty} =$$
$$= T+\tau\,,$$
$$\mu_{X^2} = \int_{\mathbb{R}} t^2\, f_X(t)\,\mathrm{d}t = \int_T^\infty \frac{t^2}{\tau} \exp\left(-\frac{t-T}{\tau}\right)\mathrm{d}t =$$
$$= \left(-t^2 - 2\,\tau\,t - 2\,\tau^2\right)\exp\left(-\frac{t-T}{\tau}\right)\Big|_{t=T}^{\infty}$$
$$= T^2 + 2\,\tau\,T + 2\,\tau^2 = (T+\tau)^2 + \tau^2 = \mu_X^2 + \tau^2\,.$$

K těmto výsledkům lze dojít bez integrování, neboť $X = Y + T$, kde T je konstanta a Y je náhodná veličina s exponenciálním rozdělením, $\mu_Y = \tau$, $\sigma_Y^2 = \tau^2$; $\mu_X = \mu_Y + T$, $\sigma_X^2 = \sigma_Y^2$, $\mu_{X^2} = \mu_X^2 + \sigma_X^2$.
V našem případě

$$m_X = \frac{1}{n}\sum_{i=1}^n x_i = 5\,, \qquad m_{X^2} = \frac{1}{n}\sum_{i=1}^n x_i^2 = \frac{227}{7} = 32.429\,,$$

soustava rovnic

$$\hat{T} + \hat{\tau} = m_X = 5$$
$$m_X^2 + \hat{\tau}^2 = m_{X^2} = \frac{227}{7}$$

má kladné řešení $\hat{\tau} = \frac{2\sqrt{91}}{7} = 2.725\,5$, $\hat{T} = 2.274\,5$, které ovšem neodpovídá zadání, neboť $\hat{T} > x_1 = 2$, takže nalezený model nepřipouští pozorovanou hodnotu x_1 (ta by měla nulovou hustotu pravděpodobnosti). □

19 Testování hypotéz

20 Testy střední hodnoty a rozptylu

20.1 Testy střední hodnoty normálního rozdělení

20.1.1 Při známém rozptylu σ^2

20.1.2 Při neznámém rozptylu

Příklad 20.1. *Z 10 měření krevního tlaku u jednoho pacienta jsme obdrželi výběrový průměr 150 a výběrovou směrodatnou odchylku 20. Rozhodněte na hladině významnosti 5%, zda je střední hodnota krevního tlaku nejvýše 140. Za jakých předpokladů výsledek platí?*

Příklad 20.2. *Voltmetr vykázal následující četnosti chyb měření. Otestujte na hladině významnosti 1% hypotézu, že má nulovou stálou chybu. Diskutujte použité předpoklady.*

chyba [mV]	-0.2	-0.1	0	0.1	0.2	0.3
četnost chyby	2	2	10	10	5	1

20.2 Testy rozptylu normálního rozdělení

Příklad 20.3. *Do laboratoře bylo odesláno 5 stejných vzorků krve ke stanovení obsahu alkoholu. Výsledky byly: 0.8, 1, 0.6, 1.4, 0.9 promile. Posuďte na hladině významnosti 5 %, zda směrodatná odchylka měření je nejvýše 0.1 promile. Uveďte použité předpoklady.*

Řešení. *Z výběrového rozptylu vypočítáme testovací statistiku*

$$t = \frac{(n-1)\,s_{\boldsymbol{x}}^2}{\mathrm{D}X} \doteq \frac{4\cdot 0.088}{0.1^2} \doteq 35.2\,,$$

*kterou porovnáme s kvantilem $q_{\chi^2(n-1)}(1-\alpha) = q_{\chi^2(4)}(0.95) \doteq 9.49$, hypotézu **zamítáme**. Vycházíme z předpokladu, že chyby jednotlivých měření jsou nezávislé a mají všechny stejné normální rozdělení; potom má testovací statistika rozdělení $\chi^2(n-1)$.* □

Příklad 20.4. *Z 10 měření stejného napětí nám vyšla výběrová směrodatná odchylka voltmetru* 3 mV*. Posuďte na hladině významnosti 5%, zda směrodatná odchylka voltmetru je nejvýše* 2 mV*, jak uvádí výrobce. Uveďte použité předpoklady.*

Řešení. *Z výběrového rozptylu vypočítáme testovací statistiku*

$$t = \frac{(n-1)\,s_{\boldsymbol{x}}^2}{\mathrm{D}X} = \frac{81}{4} = 20.25\,,$$

kterou porovnáme s kvantilem $q_{\chi^2(n-1)}(1-\alpha) \doteq 16.92$*, hypotézu* ***zamítáme****. Vycházíme z předpokladu, že chyby jednotlivých měření jsou nezávislé a mají všechny stejné normální rozdělení; potom má testovací statistika rozdělení* $\chi^2(n-1)$*.* □

20.3 Porovnání dvou normálních rozdělení

20.3.1 Test rozptylů dvou normálních rozdělení

Příklad 20.5. *Jeden vzorek byl rozdělen na mnoho stejných částí a zaslán opakovaně k měření dvěma laboratořím. Výsledky jsou v tabulce. Posuďte na hladině významnosti 5%, zda rozptyl jejich výsledků je stejný. Uveďte použité předpoklady.*

1. laboratoř	10.1	10.3	11.1	9.7	10.4	10.8	10.4
2. laboratoř	9.8	9.6	11.3	9.3	10.5	10.7	10.2

20.3.2 Testy středních hodnot dvou normálních rozdělení se známým rozptylem σ^2

20.3.3 Testy středních hodnot dvou normálních rozdělení se (stejným) neznámým rozptylem

Příklad 20.6. *U testovací skupiny* 20 *pacientů, kterým byl podáván lék na snížení krevního tlaku, byla naměřena realizace výběrového průměru* 140 torr*, realizace výběrové směrodatné odchylky* 20 torr*. U srovnávací skupiny* 50 *pacientů, kterým lék nebyl podáván, byl naměřena realizace výběrového průměru* 150 torr*, realizace výběrové směrodatné odchylky* 15 torr*. Posuďte, zda je tím prokázána účinnost léku na hladině významnosti* 1%*. Uveďte použité předpoklady.*

Řešení. $\bar{x} = 140,\ s_x = 20,\ m = 20,$
$\bar{y} = 150,\ s_y = 15,\ n = 50,$
$H_0' : s_x^2 = s_y^2, \qquad H_1' : s_x^2 \neq s_y^2$
$\frac{s_x^2}{s_y^2} = \frac{16}{9} \doteq 1.778$ *porovnáme s*

$$q_{F(19,49)}(0.995) \doteq 2.47, \qquad q_{F(19,49)}(0.005) = \frac{1}{q_{F(49,19)}(0.995)} \doteq \frac{1}{2.96} \doteq 0.338,$$

hypotézu H_0' *o rovnosti rozptylů* ***nezamítáme***
Odhad rozptylu a směrodatné odchylky:

$$s^2 = \frac{(m-1)\ s_x^2 + (n-1)\ s_y^2}{m+n-2} \doteq 273.9, \qquad s \doteq \sqrt{273.9} \doteq 16.55,$$

$H_0 : \mathrm{E}X \geq \mathrm{E}Y, \qquad H_1 : \mathrm{E}X < \mathrm{E}Y$

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \doteq \frac{-10}{16.55\sqrt{\frac{1}{20} + \frac{1}{50}}} \doteq -2.284$$

porovnáme s $q_{t(68)}(0.01) = -q_{t(68)}(0.99) \doteq -2.38$ *a hypotézu, že lék* ***nesnižuje*** *krevní tlak,* ***nezamítáme*** *na na hladině významnosti* 1%*. (Mohli bychom ji zamítnout na hladině významnosti* 5%*, pro tu je* $q_{t(68)}(0.05) = -q_{t(68)}(0.95) \doteq -1.66$*.)*
Předpoklady: normální rozdělení (stejné uvnitř každého souboru), nezávislost, stejné rozptyly. □

Příklad 20.7. *Stejnou veličinu jsme měřili dvěma metodami, každou* 10×*. Výsledky shrnuje následující tabulka.*

	výběrový průměr	*výběrová směrodatná odchylka*
1. metoda	20	3
2. metoda	21	5

Posuďte na hladině významnosti 5 %*, zda lze považovat obě metody za stejně přesné a jejich střední hodnoty za stejné. Diskutujte použité předpoklady.*

Příklad 20.8. *V řetězcích A a B jsme koupili* 11 *balíčků cukru a jejich zvážením jsme dospěli k těmto hodnotám:*

	A	B
výběrový průměr	$0.951\,\mathrm{kg}$	$0.912\,\mathrm{kg}$
výběrový rozptyl	$0.021\,\mathrm{kg}^2$	$0.067\,\mathrm{kg}^2$
výběrová směrodatná odchylka	$0.144\,\mathrm{kg}$	$0.258\,\mathrm{kg}$

Je možné na základě těchto dat zamítnout na hladině významnosti 5 % *hypotézu, že střední hodnoty hmotnosti balíčků cukru v těchto dvou řetezcích jsou stejné?*

20.4 Testy středních hodnot dvou normálních rozdělení – párový pokus

20.4.1 Pro známý rozptyl σ^2

20.4.2 Pro neznámý rozptyl

Příklad 20.9. *U dvou benzínových stanic byly vždy v tutéž dobu sledovány ceny benzínu, výsledky jsou v tabulce:*

X	32.50	32.20	31.30	30.60	29.20	27.60	27.20	26.90	25.90	25.90	23.90	23.90
Y	32.70	32.30	31.50	30.60	29.30	27.70	27.40	26.70	26.50	25.50	24.90	23.50

Posuďte na hladině významnosti 5% *hypotézu, že benzín u stanice X není levnější. Uveďte použité předpoklady.*

Řešení. *Rozdíly cen jsou*
$\delta = (-0.2, -0.1, -0.2, 0, -0.1, -0.1, -0.2, 0.2, -0.6, 0.4, -1, 0.4)$,
$n = 12$, $\bar{\delta} = -0.125$, $s_\delta^2 \doteq 0.153$, $s_\delta \doteq 0.391$,

$$t = \frac{\bar{\delta}}{s_\delta}\sqrt{n} \doteq -1.107\,,$$

porovnáme s kvantilem $q_{t(11)}(0.05) = -q_{t(11)}(0.95) \doteq -1.80$ *a nulovou hypotézu nezamítáme.*
Předpoklady pro párový pokus: střední hodnoty náhodných veličin v obou výběrech kolísají stejně, odchylky od nich mají normální rozdělení a jsou nezávislé. □

21 χ^2-test dobré shody

Příklad 21.1. *Realizací náhodného výběru jsme dostali následující četnosti hodnot:*

hodnota	0	1	2	3	4	5
pozorovaná četnost	2	7	15	12	3	1

Posuďte na hladině významnosti 5% *hypotézu, že výběr pochází z binomického rozdělení* $\mathrm{Bi}\,(5, p)$, *kde p neznáme.*

Řešení. *Odhad p metodou momentů:* $\mathrm{E}X = 5\,p = \bar{x} = 2.25$, $p = 0.45$. *Stejný výsledek dává i metoda maximální věrohodnosti, viz [Navara: PMS, str. 180].*

hodnota k	0	1	2	3	4	5
pozorovaná četnost	2	7	15	12	3	1
teoretická četnost $\binom{5}{k}\,p^k\,(1-p)^{5-k}$	2.013	8.236	13.476	11.026	4.511	0.738

Pro $k \in \{0, 5\}$ vychází teoretická četnost příliš malá, musíme sdružit třídy:

hodnota k	$0-1$	2	3	$4-5$
pozorovaná četnost	9	15	12	4
teoretická četnost	10.2487	13.476375	11.026125	5.2488
příspěvek ke kritériu	0.152141412	0.172259464	0.086016848	0.297115806

Hodnota kritéria je 0.70753353, *porovnáme s kvantilem* $q_{\chi^2(2)}\,(0.95) \doteq 5.99$ *a hypotézu nezamítáme.* □

Příklad 21.2. *Sportovec* 25× *prohrál (*0 *bodů),* 118× *remizoval (*1 *bod) a* 123× *vyhrál (*2 *body). Posuďte na hladině významnosti* 5 %, *zda tato data vyhovují binomickému rozdělení* $\mathrm{Bi}(2, q)$, *kde $q \in \langle 0, 1\rangle$ je neznámý parametr.*

Příklad 21.3. *Tabulka uvádí, kolik z respondentů odpovědělo v průzkumu na otázku kladně, v závislosti na vzdělání. Máme důvod se domnívat, že odpověď závisí na vzdělání?*

ukončené vzdělání	*počet respondentů*	*počet kladných odpovědí*
žádné	5	1
základní	195	10
střední	450	14
vyšší střední	150	10
vysokoškolské	200	15
celkem	1000	50

Příklad 21.4. *Posuďte na hladině významnosti* 5 %*, zda data v tabulce odpovídají následujícímu pravděpodobnostnímu modelu: Každý rok je přijímán stejný počet studentů* (1200)*, z každého ročníku do dalšího postoupí* 80 %*, ostatní fakultu opustí.*

ročník	1	2	3	4	5
počet studentů	1200	860	650	530	450

21.1 χ^2-test dobré shody dvou rozdělení

21.2 χ^2-test nezávislosti dvou rozdělení

22 Korelace, její odhad a testování

22.1 Test nekorelovanosti dvou výběrů z normálních rozdělení

Příklad 22.1. *Na vzorku* 50 *pacientů byla zjištěna korelace* -0.4 *mezi tělesnou hmotností a věkem, kterého se dožili. Otestujte na hladině významnosti* 5 % *hypotézu, že zvýšená hmotnost nezkracuje život.*

Řešení.

$$t = \frac{r_{\boldsymbol{x},\boldsymbol{y}}\sqrt{n-2}}{\sqrt{1-r_{\boldsymbol{x},\boldsymbol{y}}^2}} = \frac{-0.4\sqrt{50-2}}{\sqrt{1-0.4^2}} \doteq -3.02\,,$$

porovnáme s $q_{t(48)}(0.05) = -q_{t(48)}(0.95) \doteq -1.68$ *a hypotézu* ***zamítáme****.*

23 Neparametrické testy

23.1 Znaménkový test

23.2 Wilcoxonův test (jednovýběrový)

Část III
Přílohy

24 Příklady pro opakování

Příklad 24.1. *Vysvětlete rozdíly mezi následujícími pojmy: (a) střední hodnota, (b) výběrový průměr, (c) realizace výběrového průměru.*

Řešení. *Střední hodnota nemusí existovat. Pokud existuje, je to číslo, které nám může zůstat utajeno; projevuje se pouze zprostředkovaně v realizacích náhodné veličiny a je limitou některých odhadů. Výběrový průměr je náhodná veličina vypočítaná z náhodného výběru, na rozdíl od střední hodnoty vždy existuje (pro numerické náhodné veličiny). Pokud původní rozdělení má rozptyl, je výběrový průměr nestranným konzistentním odhadem střední hodnoty, takže k ní v jistém smyslu konverguje pro rozsah výběru jdoucí do nekonečna. Realizace výběrového průměru je číslo získané z realizace náhodného výběru, sloužící k (realizaci) odhadu neznámé střední hodnoty.* □

Příklad 24.2. *K úspěšnému absolvování zkoušky je potřeba nadpoloviční počet bodů z písemky. Každý příklad je hodnocen* 0, 1*, nebo* 2 *body a student odhadl, že všechna bodová hodnocení jsou stejně pravděpodobná a nezávislá na výsledcích v ostatních příkladech. Kdy má větší šanci na úspěch, pokud bude mít zkouška* 2 *příklady, nebo* 3*?*

Příklad 24.3. *Najděte příklad nezáporné náhodné veličiny, která má střední hodnotu* 1 *a směrodatnou odchylku* 10*, nebo dokažte, že taková náhodná veličina neexistuje.*

Příklad 24.4. *Semena mají klíčivost* $p \in (0,1)$. *Jaký je optimální počet* n *semen v jamce, aby byla co nejvyšší pravděpodobnost, že vyklíčí právě jedno? Řešte obecně a pro* $p = 1/3$.

Příklad 24.5. *Náhodná veličina* X *má binomické rozdělení* $\mathrm{Bi}(2, \frac{1}{3})$, *náhodná veličina* Y *má spojité rovnoměrné rozdělení* $\mathrm{R}(0,1)$. *Popište a znározněte rozdělení náhodných veličin*

1. $Y + \mathrm{E}X$,
2. $X - \mathrm{E}Y$,
3. $-2\,X$,
4. $-2\,Y$,
5. $\mathrm{Mix}_{1/3}(X, Y)$.

Příklad 24.6. *Náhodná veličina* X *má distribuční funkci*

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

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

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

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

Změna znaménka:

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

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

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

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

□

Příklad 24.7. *Náhodná veličina* X *má alternativní rozdělení; nabývá hodnot* $0, 1$ *s pravděpodobností* $1/2$. *Náhodná veličina* Y *má rovnoměrné rozdělení na intervalu* $\langle 0, 1\rangle$. *Určete a znázorněte rozdělení náhodných veličin (a)* $2\,Y + 1$, *(b)* $\mathrm{Mix}_{2/3}\,(Y, X)$, *(c)* $X + Y$ *(návod:* X *je směsí dvou konstatních náhodných veličin).*

Příklad 24.8. *Náhodná veličina* X *má hustotu*

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

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

Příklad 24.9. *Nezávislé náhodné veličiny* X, Y, Z *mají po řadě rozdělení* $N(2,3)$, $N(0,1)$, $N(0,1)$. *Určete*

1. *rozdělení náhodné veličiny* $X + Y$,
2. *střední hodnotu směsi náhodných veličin* $\mathrm{Mix}_{1/2}(X,Y)$,
3. *rozdělení náhodné veličiny* $Y^2 + Z^2$.

Příklad 24.10. *Nezávislé náhodné veličiny* X, Y, Z *mají po řadě rozdělení* $N(2,3)$, $N(5,1)$, $N(0,1)$. *Určete*

1. *rozdělení náhodné veličiny* $X - Y$,
2. *střední hodnotu náhodné veličiny* $X \cdot Y$,
3. *rozdělení náhodné veličiny* Z^2.

Příklad 24.11. *Spojitá náhodná veličina je frekvence v* Hz. *Jaký fyzikální rozměr má její rozptyl, směrodatná odchylka, medián, dále argumenty a výsledky distribuční a kvantilové funkce a hustoty?*

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

Příklad 24.12. *Stykač má být zapnut* 8 *hodin denně. Má-li být vypnutý, je s pravděpodobností* 10 % *zapnutý, má-li být zapnutý, je s pravděpodobností* 5 % *vypnutý. (a) S jakou pravděpodobností nepracuje správně? Jaká bude tato pravděpodobnost, pokud použijeme dva nezávislé stykače a spojíme je (b) sériově, (c) paralelně?*

Příklad 24.13. *Předpokládejme, že politická strana má volební preference* 3 %. *Jaká je pravděpodobnost, že v průzkumu odhad jejích preferencí dosáhne aspoň* 5 %, *je-li rozsah výběru (a)* 500, *(b)* 1000*?*

Příklad 24.14. *Na stejném místě měříme teplotu dvěma nezávislými teploměry se směrodatnými odchylkami* 2 °C. *Ukazují* 3 °C, *resp.* 2.5 °C. *Jaké je riziko, že mrzne? Uveďte použité předpoklady.*

Řešení. *Aritmetický průměr obou údajů je* 2.75 °C *se směrodatnou odchylkou* $\sqrt{2}$ °C,

$$\Phi\left(\frac{-2.75}{\sqrt{2}}\right) \doteq 1 - \Phi\left(1.944\,54\right) \doteq 1 - 0.974 = 0.026\,.$$

□

Příklad 24.15. *Náhodná veličina* X *je počet dětí ve školním věku v jedné rodině. Předpokládáme, že má Poissonovo rozdělení s parametrem* $\lambda = 0.8$, *tj.*

$$p_X(k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,, \qquad k \in \{0,1,2,\ldots\}\,,$$
$$\mathrm{E}X = \lambda\,, \qquad \mathrm{D}X = \lambda\,.$$

Ve městě bydlí $n = 10\,000$ *rodin. Jaký počet míst ve školách bude postačovat s pravděpodobností aspoň* 95 %*? (Předpokládáme, že všechny děti chodí do školy v obci, ve které bydlí.) Uveďte použité předpoklady.*

Řešení. *1. postup: Použijeme centrální limitní větu; počet dětí má přibližně normální rozdělení* $\mathrm{N}(n\,\lambda, n\,\lambda) = \mathrm{N}(8\,000, 8\,000)$. *Výsledkem je kvantil*

$$q_{\mathrm{N}(n\,\lambda,n\,\lambda)}\left(0.95\right) = n\,\lambda + \sqrt{n\,\lambda}\,\Phi^{-1}\left(0.95\right) \doteq 8\,000 + \sqrt{8\,000}\,1.645 \doteq 8147.13\,.$$

Zaokrouhlíme nahoru; potřebujeme aspoň 8148 *míst.*
2. postup: Součet nezávislých Poissonových rozdělení má Poissonovo rozdělení, zde s parametrem $n\,\lambda = 8\,000$. *Pro intervalový odhad je nahradíme normálním rozdělením* $\mathrm{N}(8\,000, 8\,000)$, *další postup je stejný.*
Předpokládáme nezávislost počtu dětí v jednotlivých rodinách. Existence rozptylu je zaručena předpoklady. Dále považujeme počet rodin za dostatečně velký na to, abychom mohli zanedbat chybu v náhradě výsledného (Poissonova) rozdělení normálním. □

Příklad 24.16. *Za první účast na zkoušce se platí 30 EUR, za každý opravný pokus 2× více než za předešlý. Student má v každém pokusu pravděpodobnost úspěchu p. Na kolik ho v průměru zkouška přijde (v závislosti na p)?*

Řešení. *Pokus je popsán binomickým rozdělením* $\mathrm{Bi}(n,p)$, *maximalizujeme hodnotu*

$$\binom{n}{1} p^1 (1-p)^{n-1} = n\,p\,(1-p)^{n-1}$$

v závislosti na n. V reálném oboru vychází

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

Funkce je unimodální (do maxima rostoucí, pak klesající), takže optimum v oboru celých čísel nastává pro jedno ze dvou celých čísel, která jsou nejblíže této hodnotě. Pro $p = 1/3$, $n = -\frac{1}{\ln\frac{2}{3}} = 2.466\,3$
$n = 0$
$n\,p\,(1-p)^{n-1} = 0$
$n = 1$
$n\,p\,(1-p)^{n-1} = \frac{1}{3}$
$n = 2$
$n\,p\,(1-p)^{n-1} = \frac{4}{9}$
$n = 3$
$n\,p\,(1-p)^{n-1} = \frac{4}{9}$
$n = 4$
$n\,p\,(1-p)^{n-1} = \frac{32}{81}$
$\frac{\partial}{\partial n} n\,p\,(1-p)^{n-1} = p\,(1-p)^{n-1} + np\,(\ln(1-p))\,(1-p)^{n-1} = 0$, *řešení:*

$$\begin{cases} \mathbb{C} & \textit{if} \quad p = 0 \\ \left\{-\frac{1}{\ln(-p+1)}\right\} \cup \mathbb{C} \setminus \{0\} & \textit{if} \quad p = 1 \\ \left\{-\frac{1}{\ln(-p+1)}\right\} & \textit{if} \quad p \neq 0 \wedge p \neq 1 \end{cases}$$

□

Příklad 24.17. *Posuďte, který z pravděpodobnostních modelů v tabulce nejlépe odpovídá pozorovaným četnostem známek:*

známka	1	2	3	4
pravděpodobnost dle modelu A	1/4	1/4	1/4	1/4
pravděpodobnost dle modelu B	1/6	1/6	1/3	1/3
pravděpodobnost dle modelu C	1/8	1/8	1/4	1/2
četnost	20	27	70	99

Příklad 24.18. *Jaké jsou vztahy mezi nezávislostí a nekorelovaností náhodných veličin? Uveďte jeden příklad u každé kombinace těchto vlastností, která může nastat.*

Řešení. *Nezávislé náhodné veličiny jsou nekorelované, příkladů je mnoho. Příklady ostatních případů:*
Závislé a korelované: $X = Y$ *libovolné kromě konstatních.*
Závislé a nekorelované: (X,Y) *nabývá hodnot* $(-2,-1), (-1,1), (1,1), (2,-1)$ *s pravděpodobnostmi* $1/4$. *Pak* $\mathrm{E}X = \mathrm{E}Y = \mathrm{E}(X\,Y) = 0$,

$$P[X=2, Y=1] = 0 \neq P[X=2]\cdot P[Y=1] = \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8}\,.$$

□

Příklad 24.19. *Po 2/3 dní neprší. V ostatní dny má srážkový úhrn v mm přibližně logaritmickonormální rozdělení* $\mathrm{LN}(0, 2.5)$, *tj. rozdělení náhodné veličiny tvaru* $X = \exp(Y)$, *kde* Y *má rozdělení* $\mathrm{N}(0, 2.5)$. *Její hustota je*

$$f_X(u) = \begin{cases} \frac{1}{u\sqrt{5\,\pi}} \exp\left(-\frac{(\ln u)^2}{5}\right) & \textit{pro } u > 0\,, \\ 0 & \textit{jinak.} \end{cases}$$

Odhadněte, jak velký denní úhrn srážek je překročen 1× za 100 let.

Příklad 24.20. *Na desce jsou kruhové kapky. Jejich plošný obsah v* mm^2 *má rozdělení* χ^2 *s 1 stupněm volnosti. Jaké je rozdělení a medián jejich obvodu?*

Příklad 24.21. *Pokud generátor náhodných čísel je nedokonalý (dává některé výsledky s vyšší pravděpodobností než jiné), typickým technickým řešením je zpětná vazba, která to kompenzuje. Posuďte možnost uplatnění tohoto principu.*

Řešení. *Takový generátor by nebyl dobrý, jeho výsledky by byly závislé. Pokud např. náhodou vyjdou 3 stejné výsledky za sebou, má být pravděpodobnost opakování téhož výsledku v dalším pokusu stále stejná, ale zde by se snížila.* □

Příklad 24.22. *Podmínka nekorelovanosti náhodných veličin je tvaru rovnosti dvou reálných čísel, což, jak známo, je velmi neobvyklý případ. Co z toho vyplývá pro nekorelovanost náhodných veličin?*

Řešení. *Pokud jsou náhodné veličiny* ***závislé****, je pravděpodobnost, že vyjdou nekorelované, velmi malá (typicky nulová); nemůžeme to však vyhodnotit, takže nanejvýš můžeme vyvrátit hypotézu, že jsou nekorelované.*
Pokud jsou však ***nezávislé*** *(což není tak neobvyklé), pak nekorelovanost vychází z podstaty pokusu a platí přesně.*
Důsledkem je, že dostatečně přesný (rozsáhlý) test na nekorelovanost odhalí závislost náhodných veličin s vysokou pravděpodobností, ačkoli jistotu nedává ani teoreticky přesná nekorelovanost. □

Příklad 24.23. *Profesor chodí na přednášky s malým zpožděním. Zjistil, že studenti chtějí statisticky vyhodnotit toto zpoždění. Napadl ho trik: na poslední přednášku přijde hodně pozdě, čímž zvýší rozptyl a zpoždění nevyjde statisticky významné. Má tato strategie naději na úspěch? Zdůvodněte. Jaké testy mohou studenti zvolit pro svoji hypotézu?*

Literatura

[Navara: PMS] Navara, M.: *Pravděpodobnost a matematická statistika.* Skriptum ČVUT, Praha, 2007.

[Rogalewicz] Rogalewicz, V.: *Pravděpodobnost a statistika pro inženýry.* 2. přepracované vydání, Skriptum FBMI ČVUT, Praha, 2007.

[Zvára, Štěpán] Zvára, K., Štěpán, J.: *Pravděpodobnost a matematická statistika* (2. vydání). Matfyzpress, MFF UK, Praha, 2002.

[Kalina, Bacigál, Schiesslová] Kalina, M., Bacigál, T., Schiesslová, A.: Základy pravdepodobnosti a matematickej štatistiky. STU Bratislava, 2010.

[Kalina, Minarechová] Kalina, M., Minarechová, Z.: Applied Mathematics For Civil Engineers. STU Bratislava, 2015.

[Anděl: Statistické metody] Anděl, J.: *Statistické metody.* 2. vyd., Matfyzpress, Praha, 1998.

[Anděl: Matematická statistika] Anděl, J.: *Matematická statistika.* SNTL/Alfa, Praha, 1978.

[Disman] Disman, M.: *Jak se vyrábí sociologická znalost.* Karolinum, UK, Praha, 2005.

[Jaroš a kol.] Jaroš, F. a kol.: *Pravděpodobnost a statistika.* Skriptum VŠCHT, 2. vydání, Praha, 1998.

[Likeš, Machek] Likeš, J., Machek, J.: *Matematická statistika.* 2. vydání, SNTL, Praha, 1988.

[Nagy] Nagy, I.: *Pravděpodobnost a matematická statistika.* Cvičení. Skriptum FD ČVUT, Praha, 2002.

[Něničková] Něničková, A.: *Matematická statistika — cvičení.* Skriptum ČVUT, Praha, 1990.

[Riečanová a kol.] Riečanová, Z. a kol.: *Numerické metódy a matematická štatistika.* Alfa/SNTL, Bratislava, 1987.

[Riečan a kol.] Riečan, B., Lamoš, F., Lenárt, C.: *Pravdepodobnosť a matematická štatistika.* Alfa/SNTL, Bratislava, 1984.

[SH10] Schlesinger, M.I., Hlaváč, V.: *Deset přednášek z teorie statistického a strukturního rozpoznávání.* ČVUT, Praha, 1999.

[Swoboda] Swoboda, H.: *Moderní statistika.* Svoboda, Praha, 1977.

[Chatfield] Chatfield, C.: *Statistics for Technology.* 3rd ed., Chapman & Hall, London, 1992.

[Hsu] Hsu, H.P.: *Probability, Random Variables, and Random Processes.* McGraw-Hill, 1996.

[Mood a kol.] Mood, A.M., Graybill, F.A., Boes, D.C.: *Introduction to the Theory of Statistics.* 3rd ed., McGraw-Hill, 1974.

[Papoulis] Papoulis, A.: *Probability and Statistics.* Prentice-Hall, 1990.

[Papoulis, Pillai] Papoulis, A., Pillai, S.U.: *Probability, Random Variables, and Stochastic Processes.* 4th ed., McGraw-Hill, Boston, USA, 2002.

[Spiegel et al. 2000] Spiegel, M.R., Schiller, J.J., Srinivasan, R.A.: *Probability and Statistics.* McGraw-Hill, 2000.

[Wasserman] Wasserman, L.: *All of Statistics. A Concise Course in Statistical Inference.* Springer, 2004.