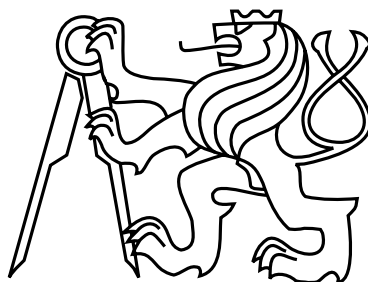


České vysoké učení technické v Praze
Fakulta elektrotechnická



X01MVT

Matematika pro výpočetní techniku (poznámky ze cvičení)

Ing. Tomáš Kroupa, PhD.

Poslední úprava: 18.1.2009

Zpracovali: k3vin, deric, kito

Studijní program: Elektrotechnika a informatika strukturovaný navazující magisterský

Obor: Výpočetní technika

leden 2009

Obsah

1	Cvičení 1	2
2	Cvičení 2	6
3	Cvičení 3	10
4	Cvičení 4	12
5	Cvičení 5	18
6	Cvičení 6	19
7	Cvičení 7	25
8	Cvičení 8	30
9	Cvičení 9	36
10	Cvičení 10	44
11	Cvičení 11	48
12	Cvičení 12	52
13	Cvičení 13	57
14	Cvičení 14	61
15	Literatura	63

Úvod

Dostaly se Vám do rukou zápisy ze cvičení z roku 2008/2009 *Ing. Tomáše Kroupy, PhD.*, který nám ochotně pomohl s jejich korekcí a kterému bychom chtěli touto cestou poděkovat. I tak mnohé příklady nejsou dotaženy do konce a nevylučujeme ani výskyt chyb. Doufáme, že Vám tento materiál pomůže v přípravě na zkoušku. Dílo je možné vylepšovat a dále šířit bez souhlasu autorů, pokud k němu budou přiloženy veškeré zdrojové kódy **.tex* souborů a obrázky (*k3vin* © 2009).

kolektiv spoluautorů :)

1 Cvičení 1

Příklad 1.1 Mějme komunikační kanál s abecedou $S(n \geq 2)$, přes který jsou přenášeny zprávy délky k . Jaká je pravděpodobnost, že dojde k přenosu zprávy, která neobsahuje $s \in S$?

- Ω = všechny možné elementární jevy
- $\Omega = \{\text{všechny } k\text{-tice prvků z } S\}, |\Omega| = n^k$
- $\mathcal{A} = \{A | A \subseteq \Omega\}$
- $P : \mathcal{A} \rightarrow \langle 0, 1 \rangle$
- A - dojde k přenosu zprávy, která neobsahuje $s \in S$

Použijeme klasický (Laplaceův) model:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{(n-1)^k}{n^k}$$

Toto však nevyhovuje, každé zprávě bychom měli přiřadit pravděpodobnost, jelikož různé symboly nemusí mít stejnou pravděpodobnost.

Příklad 1.2 Problém rytíře de Méré ¹

Je pravděpodobnější padnutí *alespoň jedné 6* ve 4 hodech (A) nebo *alespoň jedné dvojice (6, 6)* ve 24 hodech dvojicí kostek (B)?

$$\begin{array}{l} 1. \quad |\Omega_a| = 6^4 \\ \quad |\bar{A}| = 5^4 \Rightarrow \\ \quad \uparrow \\ \quad \text{žádná 6} \end{array}$$

$$\begin{aligned} P(A) &= 1 - P(\bar{A}) \\ &= 1 - \left(\frac{5}{6}\right)^4 \\ &\doteq 0,518 \end{aligned}$$

$$\begin{array}{l} 2. \quad |\Omega_b| = (6^2)^{24} = 36^{24} \\ \quad \text{zakážu jednu 6} \\ \quad \downarrow \\ \quad |\bar{B}| = 35^{24} \Rightarrow \end{array}$$

$$\begin{aligned} P(B) &= 1 - P(\bar{B}) \\ &= 1 - \left(\frac{35}{36}\right)^{24} \\ &\doteq 0,491 \end{aligned}$$

$\Rightarrow$ pravděpodobnější je první možnost.

¹Vlámský renesanční gentleman Antoine Gombaud, chevalier de Méré, 17. století.

Příklad 1.3 Mějme los o 10 políčkách. Políčka obsahují:

- 2x písmeno **V**
- 2x písmeno **P**
- 6x **nic**

Cena za los je 60 Kč. Potencionální výhra 300 Kč tehdy, pokud budou odkryta obě písmena **V** dříve než jakékoliv písmeno **P**. Koupil by si racionálně uvažující hráč tento los?

$$|\Omega| = \frac{4!}{\underbrace{2!2!}_{\uparrow}} = \binom{4}{2} = 6 \quad \text{variace s opakováním}$$

Musíme dělit, protože dané kombinace ($V_1V_2P_1P_2$ nebo $V_2V_1P_1P_2$ atd.) jsou obsaženy v tom $4!$

Prázdná políčka nehrají žádnou roli $\Rightarrow$ nepočítáme s nimi. Jediná správná bude:

$$P(V) = \frac{1}{|\Omega|} = \frac{1}{6}$$

$$\text{Střední hodnota zisku} = 240 \cdot \frac{1}{6} + (-60) \cdot \frac{5}{6} = -10$$

Střední hodnota zisku by neměla být záporná, los se tedy nevyplatí zakoupit.

Příklad 1.4 Vybíráme m výrobků z celkového počtu M , kde je obsaženo K vadných výrobků. Jaká je pravděpodobnost, že ve výběru bude K vadných?

$$\begin{aligned} |\Omega| &= \text{počet všech výběrů o velikosti } m \text{ z } M = \binom{M}{m} \\ |A_k| &= \binom{K}{k} \cdot \binom{M-K}{m-k} \\ P(A_k) &= \frac{\binom{K}{k} \binom{M-K}{m-k}}{\binom{M}{m}} \end{aligned}$$

Tomuto říkáme *hypergeometrické rozdělení*².

Příklad 1.5 Chceme jednoznačně identifikovat studenty. K dispozici máme č. 0 ... 999. Jaká je pravděpodobnost, že náhodně zvolená čísla budou jednoznačně identifikovat 150 studentů?

A ... 150-tice obsahuje všechna čísla různá

²Hypergeometrické rozdělení je rozdělení náhodné veličiny, kdy při opakování náhodného pokusu je výskyt sledovaného jevu závislý na výsledcích předcházejících pokusů. Jde tedy o pokusy, které jsou na sobě závislé. Typickým představitelem je výběr prvků bez vracení.

$$\begin{aligned}
 P(A) &= \frac{1000 \cdot 999 \dots 851}{1000^{150}} = \frac{\prod_{k=0}^{149} (1000 - k)}{\prod_{k=0}^{149} 1000} = \prod_{k=0}^{149} \left(1 - \frac{k}{1000}\right) \\
 &\approx \prod_{k=0}^{149} \left(e^{-\frac{k}{1000}}\right) = e^{-\frac{1}{1000} \cdot \sum_{k=0}^{149} k} = e^{-\frac{1}{1000} \cdot \frac{150 \cdot 149}{2}} \doteq 1,6 \cdot 10^{-5}
 \end{aligned}$$

Příklad 1.6 Máme 20 náhodně promíchaných výrobků, z nichž jsou právě 3 vadné. Kontrolu provádíme sekvenčně. Jaká je pravděpodobnost nalezení všech vadných, že:

- a) A ... nebudeme muset prohlédnout *více jak 17*?
 b) B ... budeme muset prohlédnout *právě 17*?

- a) $|\Omega| = \binom{20}{3} \Rightarrow P(A) = \frac{1 + \binom{17}{3}}{\binom{20}{3}} \doteq 0,597$
 b) $|\Omega| = \binom{20}{3} \Rightarrow P(B) = \frac{1 + \binom{16}{2}}{\binom{20}{3}} \doteq 0,106$

Příklad 1.7 Petrohradský paradox³

Potenciálnímu hráči je nabídnuto sehrát partii hry, která spočívá v tom, že se hází mincí tak dlouho, dokud nepadne *líc*. Objeví-li se líc poprvé v n -tém hození, dostane částku 2^n Kč, a tím hra končí. Poté je hráč vyzván, aby sám určil částku, kterou je ochoten vložit za možnost partii hry sehrát.

- $L = \text{líc}, R = \text{rub}$
- Nemůžu se opřít o klasický model pravděpodobnosti $\Rightarrow L, R : |\Omega| = \infty$
- $A_n \dots$ padne $(n-1)$ -krát R a nakonec L
- $n \in \mathbb{N}, \underbrace{(R, \dots, R)}_{(n-1)\text{-krát}}, L$

$$\begin{aligned}
 P(A_n) &= \frac{1}{2^n} \\
 \sum_{n=1}^{\infty} P(A_n) &= 1 \\
 \text{Střední hodnota výhry} &= \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} \\
 &= \infty
 \end{aligned}$$

³Traduje se, že tato hra byla vymyšlena v jednom petrohradském kasinu. Název je vymyšlený Nicolausem Bernoullim, aby ilustroval, jak je matematika někdy bezmocná proti lidské psychologii.

$\Rightarrow$ získáme řadu, která nekonverguje.

Z tohoto vyplývá, že člověk neutrální k riziku by měl být ochoten zaplatit *libovolnou* částku, protože opět můžeme získat *cokoliv*. Přesto je málokdo ochoten zaplatit více jak 50 Kč. A hlavně, na světě není neomezený kapitál ;-). Rigoroznější zdůvodnění je založeno na tzv. funkci užitku a její konkávnosti.

2 Cvičení 2

Příklad 2.1 Házíme n -krát kostkou ($n > 4$). Jaká je pravděpodobnost, že padly 2 šestky, 2 pětky a pak už čísla různá?

- $A \dots$ padly 2 šestky, 2 pětky, dále čísla různá od 5 a 6
- $\Omega = \{\text{všechny } n\text{-tice prvků z } \{1, \dots, 6\}\}$
- $\mathcal{A} = \exp \Omega$

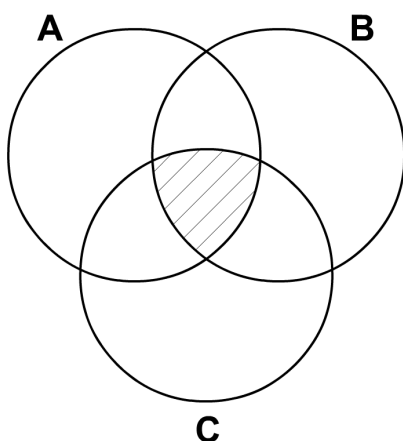
$$\begin{aligned}
 |\Omega| &= 6^n \\
 &\quad \text{umístím 2 šestky} \\
 &\quad \downarrow \\
 |A| &= \binom{n}{2} \cdot \binom{n-2}{2} \cdot 4^{n-4} \\
 &\quad \uparrow \\
 &\quad \text{umístím 2 pětky} \\
 \Rightarrow P(A) &= \frac{|A|}{|\Omega|} = \dots \text{dopočítat}
 \end{aligned}$$

Příklad 2.2 Pravděpodobnost nastání alespoň jednoho jevu

Šatnářka náhodně vydává n hostům kabáty při odchodu. Jaká je pravděpodobnost, že alespoň 1 host dostane svůj kabát?

$$A^{(n)} = A_1 \cup \dots \cup A_n, \quad A_i \dots i\text{-tý host dostane svůj kabát } (i = 1, \dots, n)$$

Odbočka: Využijeme principu inkluze a exkluze ¹:



$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(B \cap C) + P(A \cap C)] + P(A \cap B \cap C)$$

¹Využívá se tehdy, pokud se zajímáme o počty prvků jistých konečných množin; konkrétně takových, které jsou sjednocením k jiných množin.

$$|\Omega| = n!$$

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n \cdot (n-1)}$$

$$P(A_i \cap \dots \cap A_n) = \frac{1}{n!}$$

$$\Rightarrow P(A^{(n)}) = P(A_1 \cup \dots \cup A_n) = 1 - \binom{n}{2} \cdot \frac{1}{n \cdot (n-1)} + \binom{n}{3} \cdot \frac{1}{n \cdot (n-1) \cdot (n-2)} - \dots + (-1)^{n+1} \cdot \frac{1}{n!}$$

↑
střídá se dle sudosti/lichosti

n	1	2	3	4	...	7
$P(A^{(n)})$	1	0,5	0,666	0,625	...	0,6321

↑

↓

↑

↓

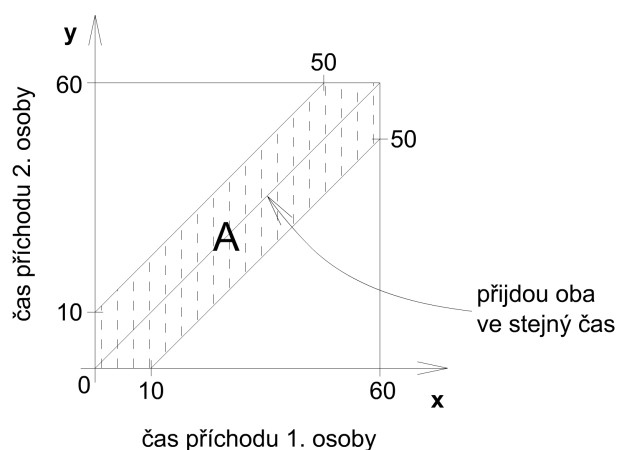
↑

osciluje

blíží se $\lim_{n \rightarrow \infty} P(A^{(n)})$

Příklad 2.3 Geometrická pravděpodobnost² - úloha o setkání

Dva lidé se chtějí sejít kdykoliv mezi 12:00 - 13:00. Budou na sebe čekat nejvýše 10min., ale max. do 13:00. Jaká je pravděpodobnost jejich setkání?



- $\Omega = \langle 0, 60 \rangle \times \langle 0, 60 \rangle$
- Nastane setkání s tolerancí 10min: $A = \{(x, y) \in \Omega \mid |x - y| \leq 10\}$

²Geometrickou pravděpodobnost můžeme považovat za zobecnění *klasické pravděpodobnosti* pro případ, že prostor elementárních jevů je nespočetný.

$$P(A) = \frac{S(A)}{S(\Omega)} = \frac{60^2 - 50^2}{60^2} = 1 - \left(\frac{5}{6}\right)^2 = 1 - \frac{25}{36} = \frac{11}{36}$$

$$P(B) = 0$$

↑
stejný čas příchozího

$\Rightarrow$ není to ale jev nemožný! Nulová pravděpodobnost, ale může nastat (usečka, jejíž plošná míra je 0). Tomuto se říká *rovnoměrné rozdělení*:

$$\begin{aligned} A, B \in \mathcal{A} \quad S(A) &= S(B) \\ &\Downarrow \\ P(A) &= P(B) \end{aligned}$$

Příklad 2.4 Nezávislost jevů

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)? (Řeště za předpokladu nezávislosti, ale zvažte jeho adekvátnost.)

$A \dots$ během týdne ztratíme data na PC, $P(A) = 0,003$

$B \dots$ během týdne ztratíme zálohu dat, $P(B) = 0,001$

Pravděpodobnost, že o data přijdeme během 1 týdne:

$$P(A \cap B) = P(A) \cdot P(B) = 3 \cdot 10^{-6}$$

↑
předpoklad nezávislosti

b) Pravděpodobnost, že o data přijdeme během 156 týdnů:

$$P(C) = 1 - (1 - 3 \cdot 10^{-6})^{156} \doteq 4,7 \cdot 10^{-4}$$

↑
týden bez ztráty dat

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.

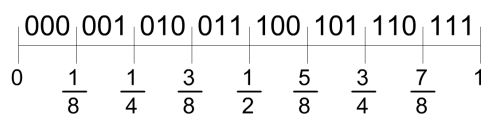
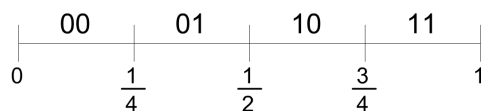
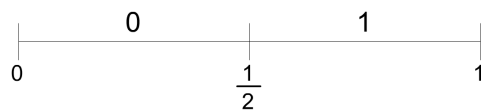
Příklad 2.5 Generujeme náhodné číslo z intervalu $(0, 1)$ pomocí hodů mincí $(0/1)$.

- $|\Omega| = (0, 1)$
- $\forall \omega \in \Omega$
- $\omega = \frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \dots + \frac{a_n}{2^n} = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \quad a_i \in \{0, 1\}$

Výběr náhodného čísla z $(0, 1)$ lze chápat jako výsledek nekonečné série hodů mincí.

- 1 .1111... $\sum_{n=1}^{\infty} \frac{1}{2^n}$

$$\bullet \quad \frac{1}{2} \quad .0111\dots \quad \sum_{n=2}^{\infty} \frac{1}{2^n}$$



Při nekonečné sérii hodů jsou na prvních n pozicích daná $x_1, \dots, x_n$ $\left(x_i \in \{0, 1\} \Rightarrow \frac{1}{2^n} \right)$

$$A = \left\{ \omega \in \Omega \mid a_1 = x_1, \dots, a_n = x_n \right\} \Rightarrow A = \left(\sum_{i=1}^n \frac{a_i}{2^i}, \sum_{i=1}^n \frac{a_i}{2^i} + \frac{1}{2^n} \right) \Rightarrow P(A) = \overset{\substack{\text{délka intervalu} \\ \downarrow}}{\frac{1}{2^n}}$$

3 Cvičení 3

Příklad 3.1 Podmíněná pravděpodobnost ¹

- 26 znaků anglické abecedy (obsahuje 5 samohlásek)
- $S \dots$ 1. znak je samohláska
- $A \dots$ 1. znak je a

- $P(A|S) \dots$ podmíněná pravděpodobnost A za podmínky S
- $|\Omega| = 26$

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{P(A)}{P(S)} = \frac{\frac{1}{26}}{\frac{5}{26}} = \frac{1}{5} \quad P(S) > 0$$

Příklad 3.2 Vzorec úplné pravděpodobnosti

$$2 \text{ zásilky} = \begin{cases} 1. \text{ zásilka obsahuje } 150 \text{ ks (z toho 5 ks vadných)} \\ 2. \text{ zásilka obsahuje } 250 \text{ ks (z toho 5 ks vadných)} \end{cases}$$

- Jaká je pravděpodobnost, že *náhodně* vybraný kus je vadný? (po smíchání obou zásilek)
- Jaká je pravděpodobnost, že *náhodně* vybraný kus z *náhodně* vybrané zásilky je vadný? (2-stupňový výběr)

- $Z_i \dots i$ -tá zásilka byla vybrána $(i \in 1, 2) \Rightarrow P(Z_1) = P(Z_2) = \frac{1}{2}$
- $A \dots$ daný výrobek je vadný $(i \in 1, 2) \quad P(A|Z_1) = \frac{1}{30}, P(A|Z_2) = \frac{1}{50}$

$$a) \Rightarrow P(A) = \frac{5+5}{250+150} = \frac{10}{400} = 0,025$$

$$b) \Rightarrow P(A) = P(A|Z_1) \cdot P(Z_1) + P(A|Z_2) \cdot P(Z_2) = \frac{1}{2} \cdot \left(\frac{1}{30} + \frac{1}{50} \right) \doteq 0,02\bar{6}$$

Příklad 3.3 Dva studenti počítají příklady.

$$A \dots \text{Alois počítal správně} \quad P(A) = \frac{1}{8}$$

$$B \dots \text{Bartoloměj počítal správně} \quad P(B) = \frac{1}{12}$$

$$N \dots \text{výsledek obou je shodný, ale nesprávný} \quad P(S|\bar{A} \cap \bar{B}) = \frac{1}{1001}$$

¹Náhodný jev určujeme vždy k určitým podmínkám. Nejsou-li na výskyt daného jevu A kladeny žádné další podmínky, potom pravděpodobnost $P(A)$ jevu A označujeme jako nepodmíněnou pravděpodobnost.

Jaká je pravděpodobnost, že výsledek obou je shodný (S) a zároveň i správný ($C, C = A \cap B$)?

Předpokládáme nezávislost, a proto:

$$\begin{aligned}
 S &= (A \cap B) \cup (A \cap \bar{A} \cap \bar{B}) \\
 P(S) &= P(C) + P(S \cap \bar{A} \cap \bar{B}) = P(A) \cdot P(B) + P(\bar{A}) \cdot P(\bar{B}) \cdot P(S|\bar{A} \cap \bar{B}) \\
 P(C|S) &= \frac{P(C \cap S)}{P(S)} = \frac{P(C)}{P(S)} = \frac{P(A \cap B)}{P(S)} \\
 &= \frac{P(A) \cdot P(B)}{P(S)} = \frac{\frac{1}{8} \cdot \frac{1}{12}}{\frac{1}{8} \cdot \frac{1}{12} + \frac{7}{8} \cdot \frac{11}{12} \cdot \frac{1}{1001}} = \frac{13}{14}
 \end{aligned}$$

Příklad 3.4 Vězeň má poslední šanci se zachránit před smrtí, a to tak, že má (pro něj co nejlépe) rozmístit 12 bílých (=život) a 12 černých (=smrt) kuliček do 2 urn. Žalářník poté vybere jednu urnu a z ní jednu kuličku.

- $A \dots$ vytažení bílé kuličky
- $U_1 \dots$ vybrání 1. urny $(1 \leq n \leq 12)$
- $U_2 \dots$ vybrání 2. urny $(0 \leq i \leq 12)$

$$P(U_1) = P(U_2) = \frac{1}{2}$$

$$P(A|U_1) = \frac{i}{n}$$

$$P(A|U_2) = \frac{12-i}{24-n}$$

$$\begin{aligned}
 P_{n,i} &= P(A) = P(A|U_1) \cdot P(U_1) + P(A|U_2) \cdot P(U_2) \\
 &= \frac{1}{2} \cdot \left(\frac{i}{n} + \frac{12-i}{24-n} \right) \Rightarrow \text{maximalizovat vzhledem k } i \text{ a } n \\
 &= \frac{1}{2} \cdot \left(\frac{1}{1} + \frac{11}{23} \right) \doteq 74\%
 \end{aligned}$$

4 Cvičení 4

Příklad 4.1 Ve městě se stala dopravní nehoda a řidič od ní odjel. Svědek tvrdí, že řidič byl v modrém taxíku. Ve městě je 85% zelených a 15% modrých taxíků. Ovlivní toto tvrzení vyšetřování?

Za těchto předpokladů ne. Pokud ale víme, že pravděpodobnost, zda mluvil pravdu je 80% a pravděpodobnost, že špatně identifikoval barvu (např. při světelných podmínkách) je 20%, vyšetřování ovlivněno bude.

- $M \dots$ taxi bylo modré
- $Z \dots$ taxi bylo zelené
- $S_1 \dots$ svědek říká, že taxi bylo modré

$$\left. \begin{array}{l} P(M) = 0,15 \\ P(Z) = 0,85 \end{array} \right\} \text{apriorní pravděpodobnosti (před dodáním dalších informací)}$$

$$\begin{aligned} P(S_1|M) &= P(\bar{S}_1|Z) = 0,8 \\ P(S_1|Z) &= P(\bar{S}_1|M) = 0,2 \end{aligned}$$

A nyní použijeme *Bayesovu větu*¹:

$$P(M|S_1) = \frac{P(S_1|M) \cdot P(M)}{\underbrace{P(S_1|M) \cdot P(M) + P(S_1|Z) \cdot P(Z)}_{P(S_1)}} = \frac{0,8 \cdot 0,15}{0,8 \cdot 0,15 + 0,2 \cdot 0,85} \doteq 0,4138$$

Mějme jinou situaci. Dva svědkové $S_{1,2}$ tvrdí, že taxík byl modrý.

$$\begin{aligned} P(S_{1,2}|M) &= ? \\ P(S_{1,2}|Z) &= ? \end{aligned}$$

Toto mohu opět vypočítat pouze za předpokladu, že oba svědkové jsou *nezávislí*.

$$\begin{aligned} P(S_{1,2}|M) &= 0,8^2 = 0,64 \\ P(S_{1,2}|Z) &= 0,2^2 = 0,04 \end{aligned}$$

Což znamená, že klesla pravděpodobnost chyby. A jdeme na *Bayese*:

$$P(M|S_{1,2}) = \frac{P(S_{1,2}|M) \cdot P(M)}{P(S_{1,2}|M) \cdot P(M) + P(S_{1,2}|Z) \cdot P(Z)} = \frac{0,64 \cdot 0,15}{0,64 \cdot 0,15 + 0,04 \cdot 0,85} \doteq 0,7385$$

$\Rightarrow$ nyní i při malém počtu modrých taxi je výpověď spolehlivá.

¹ *Thomas Bayes*, anglický duchovní, „geniální outsider“, 18. století. Bayesovské metody jsou ve statistice hojně používány. Ve stručnosti lze bayesovský přístup charakterizovat tak, že při odhadování nějaké pravděpodobnosti či při testování hypotéz vezmeme v potaz nejen data získaná právě provedeným pozorováním, ale i tzv. *informace apriorní*, tj. informace známe již před pokusem (ať už jde o informace získané nějakým dřívějším pozorováním, či o pouhou domněnku, odhad).

Příklad 4.2 Spam filter: Určete pravděpodobnost označení nevyžádané zprávy jako spam.

- pravděpodobnost že nevyžádaný e-mail je označen jako spam je 95%
- pravděpodobnost že „normální“ e-mail **není** označen jako spam je 99%
- $N \dots$ zpráva je nevyžádaná
- $S \dots$ zpráva je označena jako spam

Předpoklad: Podíl nevyžádaných e-mailů je 30%

$$\begin{aligned} P(N) &= 0.3 && \text{apriorní pravděpodobnost} \\ P(S|N) &= 0.95 \\ P(\bar{S}|\bar{N}) &= 0.99 \\ P(N|S) &= ? \end{aligned}$$

$$\begin{aligned} P(N|S) &= \frac{P(N \cap S)}{P(S)} = \frac{P(S|N)P(N)}{P(S|N)P(N) + P(S|\bar{N})P(\bar{N})} \\ &= \frac{0.95 \cdot 0.3}{\underbrace{0.95 - 0.3 + 0.01 \cdot 0.7}} = 0.976 \\ &\quad 0.292 \dots \text{pst. že vůbec něco bude označeno} \end{aligned}$$

Podíl nevyžádaných zpráv	Pravděpodobnost správného označení
30%	97.6%
1%	49%

Tabulka 4.1: Ukázka úspěšnosti spam filtru

Příklad 4.3 Máme urnu a plníme ji náhodně 10 bílými a 10 černými koulemi (s pravděpodobností $= \frac{1}{2}$). Vložíme celkem 10 koulí. Poté n -krát taháme koule z urny a vracíme.

- $B^n \dots$ všech n tažených koulí s vracením je bílých
- $B_k \dots$ v urně je k bílých koulí ($k = 0, \dots, 10$)
- $P(B_{10}|B^n) = ? \dots$ tedy pravděpodobnost, že všech 10 koulí je bílých, pokud jich n bílých vytáhneme

Pravděpodobnost, že urnu naplníme k bílými koulemi. Odpovídá to *binomickému rozdělení*².

$$\begin{aligned}
 P(B_k) &= \binom{10}{k} \cdot 2^{-10} \\
 &\quad \begin{array}{c} \uparrow \\ k \text{ z koulí je bílých} \end{array} \\
 P(B^n|B_k) &= \left(\frac{k}{10}\right)^n \\
 P(B_{10}|B^n) &= \frac{P(B^n|B_{10}) \cdot P(B_{10})}{\sum_{k=0}^n P(B^n|B_k) \cdot P(B_k)} = \frac{2^{-10}}{\sum_{k=0}^n \left(\frac{k}{10}\right)^n \cdot \binom{10}{k} \cdot 2^{-10}} \Rightarrow \\
 \lim_{n \rightarrow \infty} P(B_{10}|B^n) &= \lim_{n \rightarrow \infty} \frac{1}{\left(1 + e^{-\frac{n}{10}}\right)} = 1 \\
 P(B_{10}|B^{10}) &= 0,07 \\
 P(B_{10}|B^{20}) &= 0,37 \\
 P(B_{10}|B^{50}) &= 0,95
 \end{aligned}$$

Na tomto příkladu budeme pokračovat a ukážeme si, co je to náhodná veličina X .

- $X : \Omega \rightarrow \mathbb{R}$... například X je počet bílých koulí v urně. Zajímají nás tedy hodnoty, ne samotná funkce.

$$\begin{aligned}
 P[X = \pi] &= 0 \\
 P[X \in \{1, 2, \dots, 10\}] &= 1
 \end{aligned}$$

Pozn.: Používáme hranaté závorky, protože se bavíme o pravděpodobnosti vztažené k nějakému měření.

$$P[X = x] = \begin{cases} \binom{10}{x} \cdot 2^{-10} & x \in \{1, 2, \dots, 10\} \\ 0 & \text{jinak} \end{cases}$$

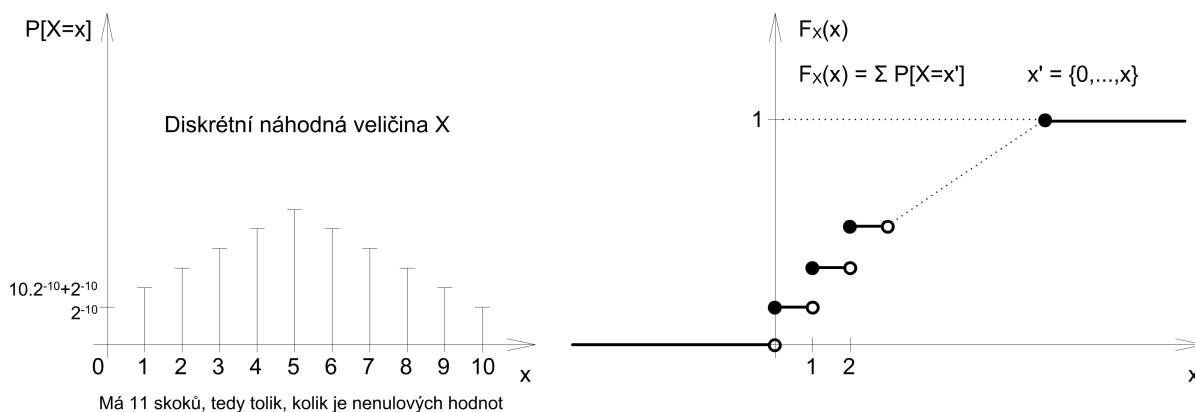
Pro zajímavost z toho uděláme graf distribuční funkce, respektive model náhodné veličiny.

- $F_X(x) = P[X \leq x]$... pro $x \in \mathbb{R}$
 $\quad \quad \quad \uparrow$
 $\quad \quad \text{zprava spojitě}$

Distribuční funkce musí dále splňovat tyto podmínky:

- je neklesající
- je zprava spojitá
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\lim_{x \rightarrow \infty} F_X(x) = 1$

²Binomické rozdělení popisuje četnost výskytu náhodného jevu v n nezávislých pokusech, v nichž má jev stále stejnou pravděpodobnost.



Pak platí:

$$\begin{aligned}
 P[X > 3] &= 1 - P[X \leq 3] = 1 - F_X(3) \\
 P[X \in (2, 7)] &= \lim_{x \rightarrow 7^-} F_X(x) - F_X(2) \\
 &\quad \uparrow \\
 &\quad \text{pst } 2+3+4+5+6+7 \\
 P[X \in O_X] &= 1 \\
 O_X &= \{x \in \mathbb{R} \mid P[X = x] = 0\} = \{0, \dots, 10\}
 \end{aligned}$$

Příklad 4.4 Na zajíce střílí *nezávisle* na sobě:

- starosta ... pravděpodobnost, že se trefí je $S = 0,2$
- myslivec ... pravděpodobnost, že se trefí je $M = 0,8$
- pytlák ... pravděpodobnost, že se trefí je $P = 0,9$

Chceme znát:

- Pravděpodobnost alespoň jednoho zásahu do zajíce (A)
- Distribuční funkci počtu zásahů

- X ... počet zásahů pro $\{0, \dots, 3\}$ pro všechna ostatní $\mathbb{R}$ čísla bude pravděpodobnost nulová
- Jinak řečeno: $P[X = x] = 0 \quad x \notin \{0, \dots, 3\}$

a) $P(A) = 1 - 0,016 = 0,984$

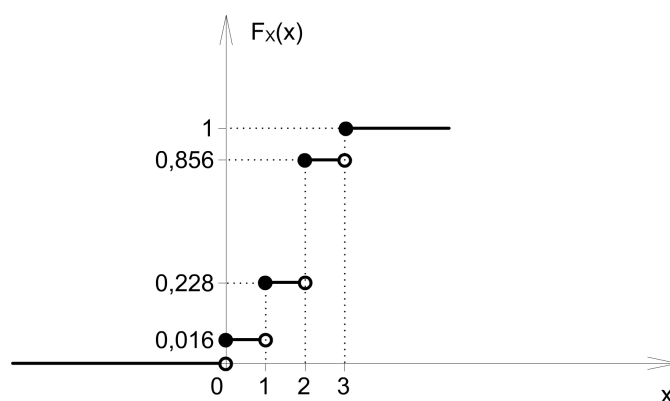
b)

$$\begin{aligned}
 O_X &= \{0, 1, 2, 3\} \rightarrow \text{Diskrétní náhodná veličina} \\
 P_X &= [X \in O_X] = 1
 \end{aligned}$$

x	$P[X = x]$
0	$(1 - S) \cdot (1 - M) \cdot (1 - P) = 0,016$
1	$S \cdot (1 - M) \cdot (1 - P) + (1 - S) \cdot M \cdot (1 - P) + (1 - S) \cdot (1 - M) \cdot P = 0,212$
2	$S \cdot M \cdot (1 - P) + (1 - S) \cdot M \cdot P + S \cdot (1 - M) \cdot P = 0,628$
3	$S \cdot M \cdot P = 0,144$

Tabulka 4.2: Tabulka pravděpodobností

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0,016 & x \in \langle 0, 1) \\ 0,228 & x \in \langle 1, 2) \\ 0,856 & x \in \langle 2, 3) \\ 1 & x \geq 3 \end{cases}$$



Pokud bychom to chtěli spočítat např. pro:

- $P[X \in (1, 3)] = F_X(3) - F_X(1) = P[X = 2] + P[X = 3]$
- $P[X > 2] = 1 - P[X \leq 2] = 1 - F_X(2)$

Příklad 4.5 Cestující přijde na nástupiště v rozmezí mezi 12:00 až 12:02 náhodně. Jaká je distribuční funkce doby příchodu?

- X ... čas příchodu

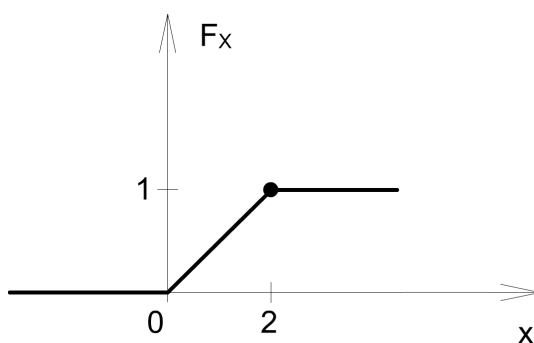
$$P[X \in \langle 0, 2 \rangle] = 1$$

$$P[X \notin \langle 0, 2 \rangle] = 0$$

$$P[X = x] = 0$$

Poslední případ vyjadřuje situaci, kdybychom čas brali jako nekonečně jemnou množinu možností, např. 12.01:03,09243862467892349846...³

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} \cdot x & x \in (0, 2) \\ 1 & x > 2 \end{cases}$$



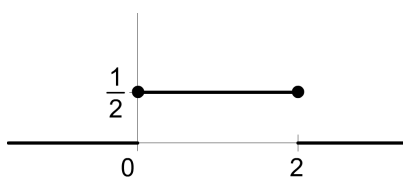
Hodnotu distribuční funkce můžeme v některých případech počítat přes hustotu f .

$$F_X(x) = \int_{-\infty}^x f_X(y) dy$$

$$O_X = \emptyset$$

$$P[X \in O_X] = 0$$

$$f_X(x) = \begin{cases} \frac{1}{2} & x \in \langle 0, 2 \rangle \\ 0 & \text{jinak} \end{cases}$$



Úlohu dále rozšíříme o pravděpodobnost toho, že je tam přesně ve 12:00. Získáme tak *Diracovo rozdělení*.

$$P[Y = 0] = 1$$

$$P[Y = y] = 0$$

³Toto je jediný správný zápis času určený českou gramatikou. Často se u nás používá čas s dvojtečkovou notací (12:01:01,0924...), ale to je anglický způsob (skoro všechno elektronické zboží je dováženo a výrobci se nevyplatí dělat speciální řadu pro českou normu)

5 Cvičení 5

Cvičení dne 30.10.2008 se nekonalo, byl přesun výuky (pondělní rozvrh).

6 Cvičení 6

Náhodné veličiny a popis jejich rozdělení:

1. **Diskrétní rozdělení** – pravděpodobnostní funkce $P[X = x]$
2. **Spojité rozdělení** – místo pravděpodobnostní funkce máme hustotu $f(x)$
3. **Smíšené rozdělení** – rozklad distribuční funkce na spojitou a diskrétní část. Pro toto rozdělení neexistuje hustota ani pravděpodobnostní funkce

Příklad 6.1 Diskretní rozdělení

Máme hrací kostku a zajímá nás nejvyšší hodnota, která padne při šesti hodech.

- $X \dots$ maximální hodnota, která padla při šesti hodech kostkou

X je diskrétní veličina a nabývá hodnot $1, \dots, 6$

$$P[X \in \{1, \dots, 6\}] = 1$$

$$P[X = 1] = \frac{1}{6^6}$$

$$P[X = 2] = \frac{2^6 - 1}{6^6}$$

$$P[X = x] = \frac{x^6 - (x-1)^6}{6^6}$$

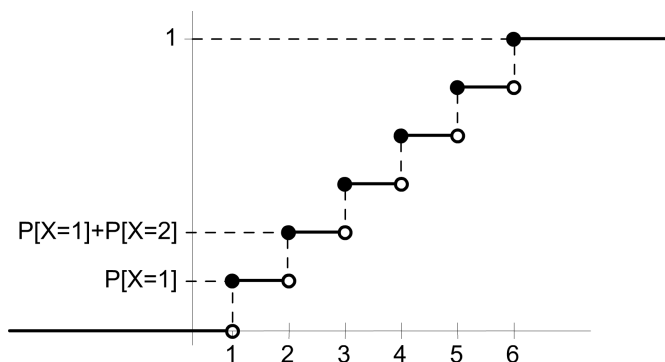
$$P[X = x] = 0$$

$$\forall x \notin \{1, \dots, 6\}$$

$$P[X = O_x] = 1$$

$$O_x = \{x \in \mathbb{R} \mid P[X = x] > 0\} = \{1, \dots, 6\}$$

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \sum_{i=1}^{\lfloor x \rfloor} P[X = i] & x \in \langle 1, 6 \rangle \\ 1 & x \geq 6 \end{cases}$$



Příklad 6.2 Diskrétní rozdělení

Komunikačním kanálem přenášíme zprávu, kterou tvoří $n = 2000$ znaků. Každý znak může být změněn s pravděpodobností $p = 10^{-3}$. Tyto chyby jsou nezávislé. Určete rozdělení počtu chyb X .

- $X \dots$ počet chyb

Jedná se o *binomické rozdělení*:

$$P[X \in \{0, \dots, 2000\}] = 1$$

$$P[X = x] = \underbrace{(10^{-3})^x \cdot (1 - 10^{-3})^{2000-x}}_{x \text{ chyb v } n \text{ znacích}} \cdot \overset{\substack{\text{všechny možnosti} \\ \downarrow}}{\binom{2000}{x}}$$

$$P[X = 0] = 1 \cdot (1 - 10^{-3})^{2000} \cdot 1$$

Spočítáme přes:

$$e^{-x} \approx 1 - x$$

$$x \approx 0$$

$\Rightarrow$

$$P[X = 0] = e^{-2000 \cdot 10^{-3}} = \frac{1}{e^2}$$

Střední hodnota *binomického rozdělení* nám pak udává počet očekávaných chyb ve zprávě (střední hodnota):

$$EX = \sum_{x=0}^{2000} x \cdot P[X = x] = n \cdot p = 2000 \cdot 10^{-3} = 2$$

A zajímá nás, jaká je pravděpodobnost, že nastane EX:

$$P[X = 2] = \binom{2000}{2} \cdot 10^{-6} \cdot (1 - 10^{-3})^{1998}$$

Pro spojitou by to bylo:

$$EX = \int_{-\infty}^{\infty} x \cdot f_x(x) dx$$

Příklad 6.3 Spojité rozdělení

Chyba měření fyzikální veličiny je náhodná hodnota x , která má *normální*, neboli *Gaussovo rozdělení* (s parametry $\mu = 0, \sigma^2 = 1$).

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \quad \forall x \in \mathbb{R}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad \forall x \in \mathbb{R}$$

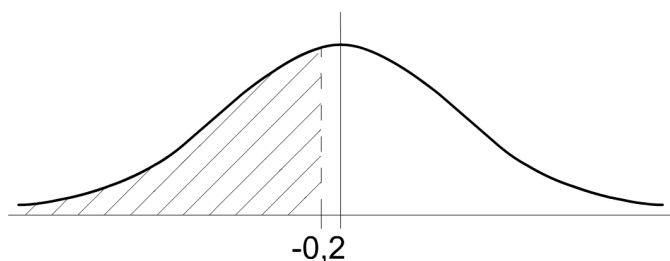
a) Pravděpodobnost, že chyba X bude menší než $-0,2$:

$$P[X \leq -0,2] = \Phi(-0,2)$$

$$\Phi(-0,2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-0,2} e^{-\frac{t^2}{2}} dt \quad \forall x \in \mathbb{R}$$

Pokud si ale uvědomíme, že pro kladná Φ známe tabulkové hodnoty (na konci skript [1][str. 229]), můžeme výpočet zjednodušit následujícím způsobem:

$$\Phi(-0,2) = 1 - \Phi(0,2) = 1 - 0,5793 = 0,4207$$



b) Pravděpodobnost, že chyba X bude v rozmezí $\langle 0,1; 0,3 \rangle$:

$$P[X \in \langle 0,1; 0,3 \rangle] = \Phi(0,3) - \Phi(0,1) = 0,0781$$

$$O_x = \{x \in \mathbb{R} \mid P[X = x] > 0\} = \emptyset$$

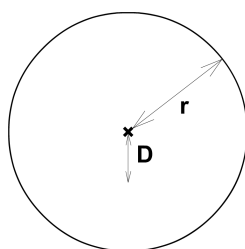
$$P[X \in O_x] = 0$$

O_x je prázdná množina, protože tam nepatří žádné veličiny, jinak by to bylo diskrétní. $\varphi(x)$ není pravděpodobnost! Vždy se bavíme o pravděpodobnosti v nějakém intervalu, ne v bodě.

Příklad 6.4 Spojité rozdělení

Střelíme náhodně na terč o poloměru $r > 0$. Výhra X je dána vzdáleností D od středu.

- $X \dots$ výhra při střelbě. Velikost výhry je dána výrazem $X = 10 \cdot (r - D)$



$$P[X \in \langle 0, 10r \rangle] = 1$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 10r \end{cases}$$

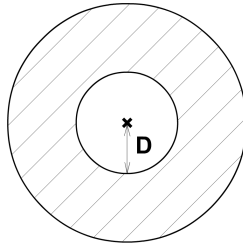
$$F_D(d) = \begin{cases} 0 & d < 0 \\ 1 & d > r \end{cases}$$

Zajímá nás distribuční funkce pro $\forall d \in \langle 0, r \rangle$ a pro $\forall x \in \langle 0, 10r \rangle$.

$$F_D(d) = P[D \leq d] = \frac{\pi \cdot d^2}{\pi \cdot r^2} = \frac{d^2}{r^2}$$

$$\begin{aligned} F_X(x) &= P[X \leq x] = P[10r - 10D \leq x] = P\left[D \geq r - \frac{x}{10}\right] = \\ &= 1 - P\left[D \leq r - \frac{x}{10}\right] = 1 - F_D\left(r - \frac{x}{10}\right) = 1 - \frac{\left(r - \frac{x}{10}\right)^2}{r^2} \end{aligned}$$

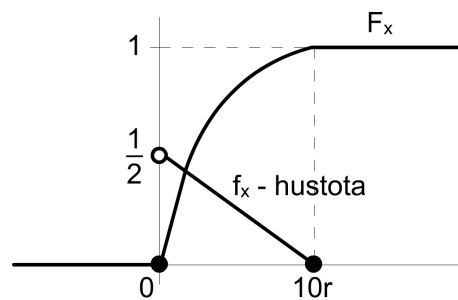
D nám udává vnitřní poloměr mezikruží.



Hustotu f_X pak vypočítáme jako derivaci distribuční funkce F_X .

$$\begin{aligned} f_X(x) &= F_X'(x) = -2 \cdot \frac{r - \frac{x}{10}}{r^2} \cdot \frac{-1}{10} \\ &= \frac{1}{3 \cdot r^2} \cdot \left(r - \frac{x}{10}\right) \quad x \in (0, 10r) \end{aligned}$$

$$f_X(x) = 0 \quad x \leq 0, x > 10r$$



Příklad 6.5 Smíšené rozdělení

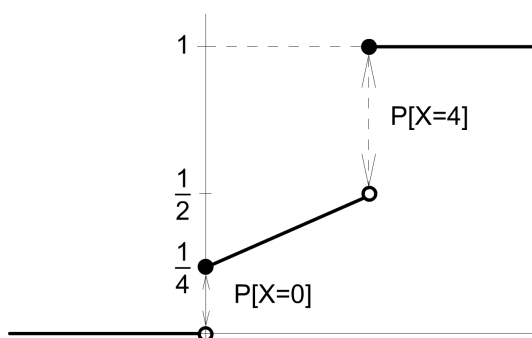
Teploměr ukazuje od 0°C do 20°C . Při nižších teplotách zůstává na nule, při vyšších na dvacítce.

- $X \dots$ výsledek měření veličiny nabývající reálných hodnot pomocí přístroje s rozsahem $\langle 0, 4 \rangle$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x+4}{16} & x \in \langle 1, 4 \rangle \\ 1 & x > 4 \end{cases}$$

$$O_x = \{x \in \mathbb{R} \mid P[X = x] > 0\} = \{0, 4\}$$

$$P[X \in O_x] = P[X = 0] + P[X = 4] = \frac{3}{4}$$



Ze 75% tedy měříme mimo rozsah teploměru. Nyní provedeme rozklad F_X na spojitou F_V a diskrétní F_U část:

$$F_X = \alpha \cdot F_U + (1 - \alpha) \cdot F_V$$

$$\alpha = P[X \in O_x] = \frac{3}{4}$$

Výpočet diskrétní složky F_U :

$$P[U = 0] = \frac{P[X = 0]}{\alpha} = \frac{1}{3}$$

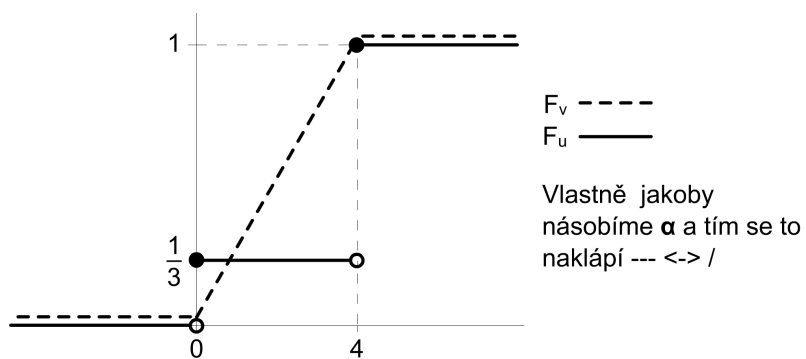
$$P[U = 4] = \frac{P[X = 4]}{\alpha} = \frac{2}{3}$$

$$F_X(x) = \begin{cases} 0 & u < 0 \\ \frac{1}{3} & x \in \langle 0, 4 \rangle \\ 1 & x > 4 \end{cases}$$

Výpočet spojitě složky F_V :

$$F_V(v) = \frac{F_X(x) - \alpha \cdot F_U(u)}{1 - \alpha}$$

$$F_X(x) = \begin{cases} 0 & u < 0 \\ \frac{\frac{v}{16} + \frac{1}{4} - \frac{3}{4} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{v}{4} & x \in (0, 4) \\ 1 & x \geq 4 \end{cases}$$



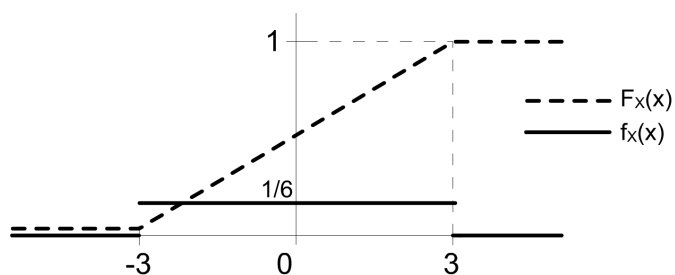
7 Cvičení 7

Příklad 7.1 Hustota pravděpodobnosti

$$f_X(x) = \begin{cases} \frac{1}{6} & x \in \langle -3, 3 \rangle \\ 0 & \text{jinak} \end{cases}$$

distribuční funkce:

$$F_X(x) = \begin{cases} 0 & x < -3 \\ \frac{1}{6}(x+3) & x \in \langle -3, 3 \rangle \\ 1 & x > 3 \end{cases}$$



$$h(x) = \begin{cases} 0 & x < 0 \\ x & x \in \langle 0, 1 \rangle \\ 1 & \text{jinak} \end{cases}$$

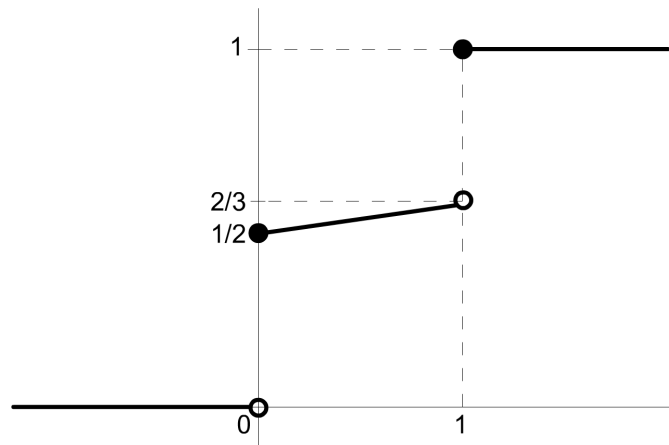
$$Y = h(X)$$

$$q_X(x) = 6x - 3 \quad x \in (0, 1)$$

Kvantilová funkce

- „opačný pohled na distribuční funkci“
- jen na otevřeném jednotkovém intervalu $[(0, 1)]$

$$F_Y(y) = P[Y \leq y] = P[h(X) \leq y] = \begin{cases} P[X \leq y] = \frac{1}{6}(y+3) & y \in (0, 1) \\ P[X < 0] = \frac{1}{2} & y = 0 \\ 1 & y \geq 1 \\ 0 & y < 0 \end{cases}$$



$$q_Y(x) = q_h(x) = h(q_X(x)) \begin{cases} 0 & x < \frac{1}{2} \\ 6x - 3 & x \in \langle \frac{1}{2}; \frac{2}{3} \rangle \\ 1 & x \geq \frac{2}{3} \end{cases}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{6} \int_{-\infty}^{\infty} x dx = 0 && \rightarrow x \text{ souřadnice těžiště grafu hustoty} \\ &= \int_0^1 q_X(x) dx = \int_0^1 (6x - 3) dx = \left[6 \frac{x^2}{2} - 3x \right]_0^1 && \rightarrow \text{pro sym. rozdělení vždy střed intervalu} \end{aligned}$$

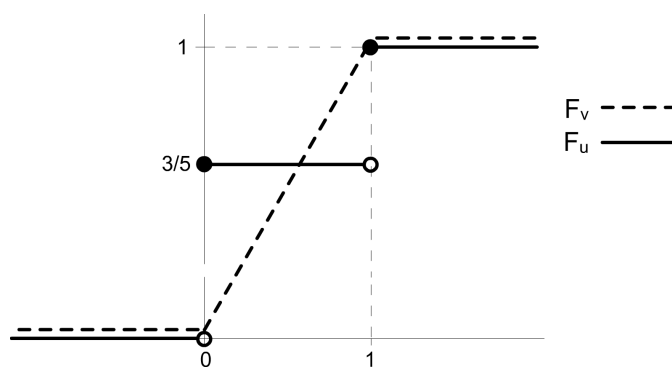
$$\begin{aligned} E(Y) &= E(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x) dx \\ &= \int_0^3 h(x) \frac{1}{6} dx = \frac{1}{6} \left(\int_0^1 x dx + \int_1^3 1 dx \right) \\ &= \frac{1}{6} \left[\frac{x^2}{2} \right]_0^1 + \frac{1}{6} \cdot 2 = \frac{1}{12} + \frac{1}{3} = \frac{5}{12} \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_0^1 q_Y(x) dx \\ &= \int_{\frac{1}{2}}^{\frac{2}{3}} (6x - 3) dx + \frac{1}{3} = \frac{5}{12} \end{aligned}$$

rozklad Y na diskrétní a spojitou část

$$F_U(u) = \begin{cases} 0, & u < 0 \\ \frac{3}{5}, & u \in \langle 0, 1 \rangle \\ 1, & u \geq 1 \end{cases}$$

$$\begin{aligned} F_V(v) &= \frac{F_Y(v) - \frac{5}{6}F_U(v)}{\frac{1}{6}} \\ &= \begin{cases} 0, & v < 0 \\ v = \frac{\frac{1}{6} \cdot (v+3) - \frac{5}{6} \cdot \frac{3}{5}}{\frac{1}{6}}, & v \in \langle 0, 1 \rangle \\ 1, & v \geq 1 \end{cases} \end{aligned}$$



$$E(Y) = \frac{5}{6} \cdot E(U) + \frac{1}{6} \cdot E(V) = \frac{5}{12}$$

jen pro smíšené rozdělení

$$E(U) = 0 \cdot P[U = 0] + 1 \cdot P[U = 1]$$

u diskrétní pravděpodobnosti

Příklad 7.2 Házíme mincí:

- X ... počet padnutí R až do prvního L
- p ... pravděpodobnost padnutí R $p \in (0, 1)$
- $E(X)$... střední hodnota počtu pokusů než padne L

$$E(X) = \sum_{x=0}^{\infty} xP[X = x]$$

geometrické rozdělení

- jednotlivé pokusy jsou nezávislé
- pravděpodobnost nastání jevu je v každém pokusu stejná
- počet úspěchů až do 1. neúspěchu

$$P[X = x] = p^x(1 - p)$$

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x p^x (1 - p) = (1 - p) \sum_{x=1}^{\infty} x p^x \\ &= (1 - p) p \underbrace{\sum_{x=1}^{\infty} x p^{x-1}}_{\frac{1}{(1-p)} x} = \frac{p}{(1 - p)} \end{aligned}$$

$$\begin{aligned} \sum_{x=0}^{\infty} p^x &= 1 + p + p^2 + \dots = \frac{1}{1 - p} \\ D(X) &= \frac{p}{(1 - p)^2} \\ \sum_{x=0}^{\infty} p^x &= \sum_{x=0}^{\infty} x p^{x-1} \end{aligned}$$

Příklad 7.3 Házíme mincí:

- $Y \dots$ počet hodů kostkou až do padnutí 6

$$\begin{aligned} P[Y = 5] &= \left(\frac{5}{6}\right)^5 \frac{1}{6} = \frac{5^5}{6^6} = 0,06 \\ E(Y) &= \frac{\frac{5}{6}}{\frac{1}{6}} = 5 \\ D(Y) &= \frac{\frac{5}{6}}{\frac{1}{30}} = 30 \\ \sigma(Y) &= \sqrt{30} \doteq 5,5 \end{aligned}$$

Příklad 7.4 Královna rodí tak dlouho děti, dokud se nenarodí syn, ale maximálně porodí 3 děti.

- X ... počet dětí v královské rodině
- pravděpodobnost narození syna je $\frac{1}{2}$

$$E(X) = ?$$

$$\begin{array}{ll} P[X = 1] = \frac{1}{2} & 1 \text{ syn} \\ P[X = 2] = \frac{1}{4} & 1 \text{ syn} + 1 \text{ dcera} \\ P[X = 3] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} & 3 \text{ dcery, nebo } 2 \text{ dcery} + 1 \text{ syn} \end{array}$$

$$\begin{aligned} E(X) &= \sum_{x \in \{1,2,3\}} x P[X = x] \\ &= P[X = 1] + 2 P[X = 2] + 3 P[X = 3] \\ &= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} \\ &= \frac{7}{4} \end{aligned}$$

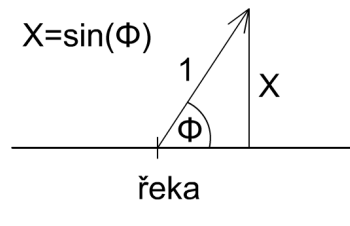
$$\begin{aligned} D(X) &= E(X^2) - (E(X))^2 \\ &= \sum_{x \in \{1,2,3\}} x^2 P[X = x] - (E(X))^2 \\ &= \frac{1}{2} + 1 + \frac{9}{4} - \frac{49}{16} = \frac{11}{16} \end{aligned}$$

směrodatná odchylka

$$\sqrt{D(X)} = \sigma(X)$$

8 Cvičení 8

Příklad 8.1 Stojíme na břehu řeky a volíme náhodně směr, kterým se vydáme (1km od řeky). Neuvažujeme variantu že bychom vstoupili do řeky, vybíráme tedy směr z intervalu $\langle 0, \pi \rangle$. Stupně v půlkruhu (tedy směry, kterými se můžeme vydat) jsou rozloženy pravidelně, jedná se o rovnoměrný model rozložení. Úhel je v našem případě náhodná veličina (a zároveň i X , protože závisí na Φ).

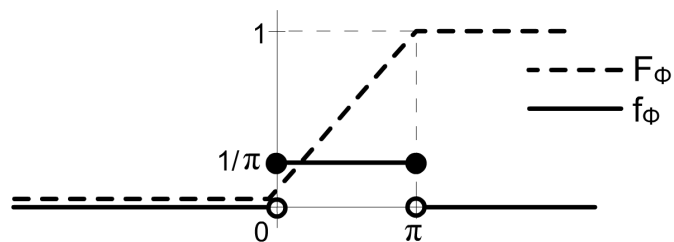


=====

$$f_{\phi}(\varphi) = \begin{cases} \frac{1}{\pi} & \varphi \in \langle 0, \pi \rangle \\ 0 & \text{jinak} \end{cases}$$

Distribuční funkce:

$$F_{\Phi} = \begin{cases} 0 & \varphi < 0 \\ \frac{\varphi}{\pi} & \varphi \in \langle 0, \pi \rangle \\ 1 & \varphi > \pi \end{cases}$$



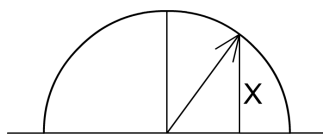
$$\begin{aligned} E(\Phi) &= \int_{-\infty}^{\infty} \varphi \cdot f_{\Phi}(\varphi) d\varphi \\ &= \frac{1}{\pi} \int_0^{\pi} \frac{\varphi^2}{2} d\varphi \\ &= \frac{\pi}{2} \end{aligned}$$

To by nám pak vyšla vzdálenost od řeky 1, což je ale špatně, protože jsme vůbec nevyužili střední hodnotu vzdálenosti $E(X)$, proto:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \underbrace{f_X(x)}_{\text{neznámá hustota}} dx \\
&= \int_{-\infty}^{\infty} \sin \varphi f_{\Phi}(\varphi) d\varphi = \int_0^{\pi} \frac{\sin \varphi}{\pi} d\varphi \\
&= \frac{1}{\pi} \left[-\cos \varphi \right]_0^{\pi} = \frac{2}{\pi}
\end{aligned}$$

Využili jsme předpokladu, že známe hodnotu původní funkce ($\sin \Phi$).

Pokud náhodně zvolíme směr, tak je to bod na jednotkové kružnici.



Příklad 8.2 Vypočítejte rozdělení počtu chycených ryb za den, pokud víte, že celkový počet ryb v rybníce je 2tis. a každá ryba má pravděpodobnost 5% na chycení. Jedná se o binomické rozdělení.

- $X \dots$ počet chycených ryb
- $X \sim B_i(n = 2000, p = 0,05)$

$\uparrow$
má rozdělení

$$P[X = x] = \binom{2000}{x} 0,05^x \cdot 0,95^{2000-x} \quad x = 0 \dots 2000$$

X má přibližně *Poissonovo rozdělení*¹

Pokud $n \rightarrow \infty$ (velký počet) a $p \approx 0$ (malá pravděpodobnost), získáme limitu výrazu:

$$\begin{aligned}
P[X = x] &= e^{-\lambda} \cdot \frac{\lambda^x}{x!} \\
E(X) &= \lambda = D(X) = n \cdot p
\end{aligned}$$

$$P[X = 0] = e^{-\lambda} = \frac{1}{e^{100}} \doteq 10^{-44}$$

$$P[X > 0] = 1 - P[X = 0] = 1 - e^{-100} \approx 1$$

využití:

¹Poissonovo rozdělení pravděpodobnosti (také označováno jako *rozdělení řídkých jevů*) má náhodná veličina, která vyjadřuje počet výskytů málo pravděpodobných, řídkých jevů v určitém časovém intervalu.

- příchod osob na určité místo v náhodnou dobu (model pro fronty)
- jako limitní případ $Bi(n, p)$
- počet meteoritů, které spadnou na určité místo

Příklad 8.3 Infocentrum navštíví v průměru 20 osob za hodinu. Jaká je pravděpodobnost, že během 15-ti minut nikdo nepřijde.

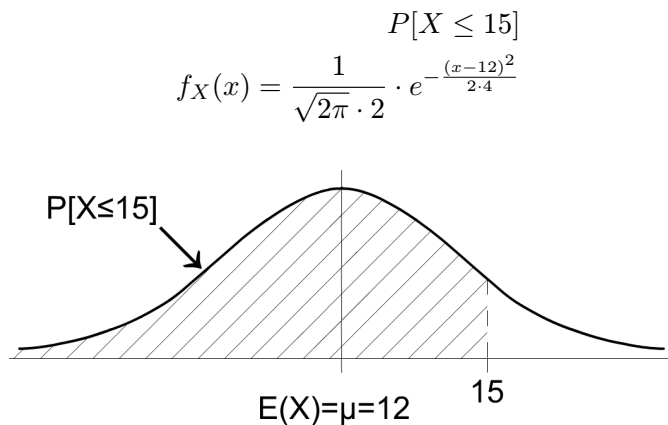
Pozn.: Takto to není dostatečné $\Rightarrow$ musíme předpokládat, že to má *Poissonovo rozdělení*.

$$\begin{array}{l}
 X \dots \text{počet osob za 15min.} \\
 P[X = 0] = e^{-\lambda} = \frac{1}{e^5} \qquad \lambda = \frac{20}{4} = 5 \\
 \uparrow \\
 \text{nikdo nepřijde}
 \end{array}$$

Příklad 8.4 Máme pár bot, jeho životnost se řídí normálním rozdělním.

- $X \dots$ životnost páru bot
- $X \sim N(\mu = 12, \sigma^2 = 4)$

Chceme znát pravděpodobnost výměny během 15 měsíců.



Musíme převést výpočet na jiný model (jiné rozdělení) $\Rightarrow$ budeme normovat veličinu (jednoduše řečeno, přesuneme 12 do 0), protože nemáme k dispozici výpočetní zařízení a také z důvodu, že u zkoušky je možné používat pro výpočty pouze tabulky, v nichž jsou hodnoty jen $\langle 0, 1 \rangle$:

$$\begin{aligned}
 \text{Normovaná veličina: } U &= \frac{X - E(X)}{\sqrt{D(X)}} \\
 E(U) &= 0 \\
 D(U) &= 1
 \end{aligned}$$

$\Rightarrow$

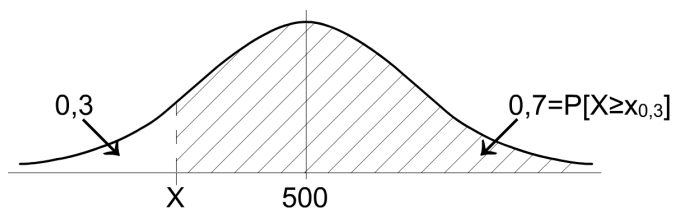
$$\begin{aligned}
 P[X \leq 15] &= P \left[\underbrace{\frac{x - 12}{2} \leq \frac{15 - 12}{2}}_{\sim N(0,1)} \right] \\
 &= P \left[U \leq \frac{3}{2} \right] = \Phi\left(\frac{3}{2}\right) \\
 &\quad \text{najdu v tabulkách} \\
 &= 0,9332
 \end{aligned}$$

Příklad 8.5 Víme, že body z testu mají následující normální rozdělení:

- $X \dots$ body z testu
- $X \sim N(\mu = 500, \sigma^2 = 100^2)$

Bylo přijato 70% uchazečů. Kolik stačilo bodů k přijetí?

Nevýhoda: není to nijak omezeno.



$P[X \geq x_{0,3}]$ hledáme 30% kvantil normálního rozdělení $N(500, 100^2)$

$\Rightarrow$ opět budeme normovat (transformujeme na jiný model):

$$\begin{aligned}
 u_\alpha &= \frac{x_\alpha - \mu}{\sigma} \\
 &\Downarrow \\
 x_{0,3} &= 100 \cdot u_{0,3} + 500
 \end{aligned}$$

$u_{0,3}$ ale není v tabulkách, je tam však hodnota pro $0,7 \Rightarrow u_{0,3} = -u_{0,7} \Rightarrow$

$$x_{0,3} = 500 - \underbrace{u_{0,7}}_{0,524} \cdot 100 \doteq 448b \quad (\text{potřebný počet bodů})$$

Příklad 8.6 Máme 2 veličiny X, Y , známe pouze střední hodnotu a rozptyl:

$$\begin{aligned}
 E(X) &= E(Y) = 2 \\
 D(X) &= 5, D(Y) = 1 \\
 Z &= 2X - Y + 3
 \end{aligned}$$

Dokážeme spočítat $E(Z)$ a $D(Z)$?

střední hodnota 3 je opět 3 ;-)

$$E(Z) = E(2X - Y + 3) = 2E(X) - E(Y) + \overset{\downarrow}{E(3)} = 5$$

$$\begin{aligned} D(Z) &= D(2X - Y + 3) \quad \text{toto můžu udělat pouze za předpokladu, že jsou nezávislé} \\ &= D(2X) + D(Y) + D(3) = 4D(X) + D(Y) + \underset{\uparrow 0}{D(3)} = 21 \end{aligned}$$

Nyní zaměňme Z za $Z_2 = XY$, potom:

$$E(Z_2) = E(XY) = E(X) \cdot E(Y) = 4$$

Příklad 8.7 Budeme generovat náhodná čísla ze dvou intervalů $\langle 0; 1 \rangle$ a $\langle 1; 2 \rangle$. Můžeme i po sečtení daných čísel číslo považovat za náhodné?

- X má rovnoměrné rozdělení na $\langle 0; 1 \rangle$
- Y má rovnoměrné rozdělení na $\langle 1; 2 \rangle$
- X a Y jsou nezávislé
- $Z = X + Y$, $\langle 1; 3 \rangle$

Musíme určit rozdělení Z a pokud bude *rovnoměrné*, můžeme číslo považovat za náhodné.

$$f_X(x) = \begin{cases} 1 & x \in \langle 0; 1 \rangle \\ 0 & \text{jinak} \end{cases}$$

$$f_Y(y) = \begin{cases} 1 & x \in \langle 1; 2 \rangle \\ 0 & \text{jinak} \end{cases}$$

f_Z je konvolucí f_X a f_Y

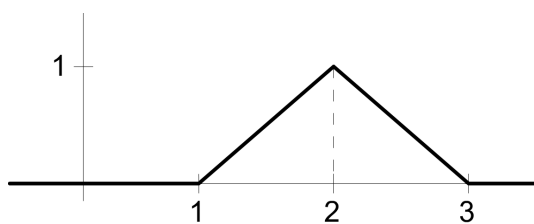
$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) dx \\ &= \int_0^1 f_Y(z-x) dx \quad \text{funkce je nulová, vyjma intervalu } \langle 0; 1 \rangle \\ &= - \int_z^{z-1} f_Y(v) dv \\ &= \int_{z-1}^z f_Y(v) dv \end{aligned}$$

$\Rightarrow$

$$f_Z(z) = \begin{cases} 0 & z < 1 \\ z - 1 & x \in \langle 1; 2 \rangle \\ 3 - z & x \in \langle 2; 3 \rangle \\ 0 & z > 3 \end{cases}$$

$$E(Z) = E(X + Y) = \frac{1}{2} + \frac{3}{2} = 2$$

$$E(X) = X_{0,5} \quad 50\% \text{ kvantil}$$



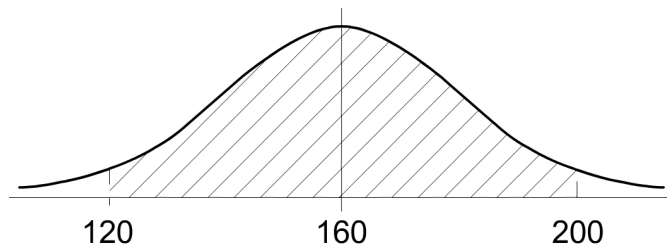
Máme tedy trojúhelníkové rozdělení a to není vhodné pro generování náhodných čísel, protože extrémní hodnoty mají menší pravděpodobnost vygenerování.

9 Cvičení 9

Příklad 9.1 Měříme a zjišťujeme pravděpodobnost výskytu výsledku v určitém intervalu.

- $X \dots$ výsledek měření $\sim N(\mu = 160, \sigma^2 = 20,4^2)$

1) $P[X \in \langle 120; 200 \rangle] = (\text{vytvoříme normovaný normální model } N(0, 1))$



$$\begin{aligned}
 &= P \left[\frac{-40}{20,4} \leq \frac{x - 160}{20,4} \leq \frac{40}{20,4} \right] \\
 &= \Phi(1,96) - \Phi(-1,96) \\
 &\quad \quad \quad \text{z tabulek} \\
 &= 2\Phi(1,96) - 1 \doteq 0,95 \\
 &\quad \quad \quad \uparrow \\
 &\text{ze symetrie stačí počítat polovinu intervalu}
 \end{aligned}$$

Y má normální rozdělení, změřím jednou, podruhé a vydělím:

$$Y = \frac{X_1 + X_2}{2}$$

$X_1, X_2 \sim N(160, 20,4^2)$
 $\uparrow$
 jsou nezávislé!

$$E(Y) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{1}{2}(E(x_1) + E(x_2)) = 160$$

Vyšla nám střední hodnota toho původního, tedy nic nového ;-)

Rozptyl:

$$\begin{aligned}
 D(Y) &= D\left(\frac{X_1 + X_2}{2}\right) \\
 &= \frac{1}{4}(D(X_1 + X_2)) \quad \text{nezávislá veličina - rozptyl součtů je součet rozptylů} \\
 &\quad \quad \quad \uparrow \\
 &\text{konstanta se vytýká s kvadrátem} \\
 &= \frac{D(X_1) + D(X_2)}{4} \\
 &= \frac{\sigma^2}{2} = \frac{(20,4)^2}{2} = 14,42^2
 \end{aligned}$$

což je menší než $\sigma^2 \Rightarrow Y \sim N\left(160, \frac{\sigma^2}{2} = \frac{20,4^2}{2}\right)$

2) $P[Y \in \langle 120; 200 \rangle]$

$$P\left[-2,77 \leq \frac{Y-160}{14,42} \leq 2,77\right] \doteq 0,994$$

Zlomek $\frac{Y-160}{14,42} \sim N(0,1)$ má také normální rozdělení, ale jiné meze.

Už jsme prakticky nemuseli dopočítávat, protože vidíme, že pravděpodobnost je větší než v prvním případě.

Příklad 9.2 Náhodný vektor a jeho sdružené rozdělení ¹

Máme 3 míčky různých barev (RGB). Náhodně je rozmístíme do tří krabiček.

Jaké jsou pravděpodobnosti rozdělení všech různých veličin? Jsou veličiny nezávislé?

- $X_1 \dots$ počet míčků v první krabičce
- $X_2 \dots$ počet obsazených krabiček (takových, kde je alespoň jeden míček)

Pokud v první krabičce budou tři míčky, pak nemůžou být obsazené 2 krabičky $\Rightarrow$ veličiny nejsou nezávislé!

$$P[X_1 = x_1, X_2 = x_2] \quad \text{sdužené rozdělení}$$

Této hodnoty nabudou políčka v tabulce, *čárka* znamená *současně*.

$3^3 = 27 \dots$ počet všech možných rozmístění tří míčků do tří krabiček

$X_2 \backslash X_1$	0	1	2	3	$P[X_2 = x_2]$
1	$\frac{2}{27}$	0	0	$\frac{1}{27}$	$\frac{3}{27}$
2	$\frac{6}{27}$	$\frac{6}{27}$	$\frac{6}{27}$	0	$\frac{18}{27}$
3	0	$\frac{6}{27}$	0	0	$\frac{6}{27}$
$P[X_1 = x_1]$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$	1

Vysvětlivky k políčkům v tabulce: *pozice* X_1, X_2

- $[0, 2], [1, 2] \dots$ stejné, až na prohození krabiček
- $[1, 3] \dots$ všechny permutace

¹Cílem je mít jeden popis pro více náhodných veličin

- a součet musí být vždy 1

$$P[X_2 = x_2] = \sum_{x_1} P[X_1 = x_1, X_2 = x_2]$$

V tabulce je *zakódováno* i to, že nejsou nezávislé, viz. např. $[1, 1] = 0$, ale násobkem (viz. definice) $\frac{3}{27} \cdot \frac{12}{27} \neq 0 \Rightarrow$

X_1, X_2 jsou nezávislé diskrétní veličiny $\Leftrightarrow$ je sdružené rozdělení rovno marginálnímu součinu.

$$P[X_1 = x_1, X_2 = x_2] = P[X_1 = x_1] \cdot P[X_2 = x_2]$$

definováno pro $\forall_{xi}, \forall_{xj}$

Také to můžeme ověřit z:

$$\begin{aligned} E(X_1) &= 1 \cdot \frac{12}{27} + 2 \cdot \frac{6}{27} + 3 \cdot \frac{1}{27} = 1 \\ E(X_2) &= 1 \cdot \frac{3}{27} + 2 \cdot \frac{18}{27} + 3 \cdot \frac{6}{27} = \frac{57}{27} = \frac{19}{9} \\ D(X_1) &= E(X_1^2) - E^2(X_1) = \frac{2}{3} \\ D(X_2) &= \frac{26}{81} \end{aligned}$$

*kovariance*²:

$$\begin{aligned} cov(X_1, X_2) &= E(X_1 X_2) - E(X_1) \cdot E(X_2) \\ &= \sum_{x_1, x_2} x_1 \cdot x_2 \cdot P[X_1 = x_1, X_2 = x_2] - \frac{19}{9} \\ &= 3 \cdot \frac{1}{27} + 2 \cdot \frac{6}{27} + 4 \cdot \frac{6}{27} + 3 \cdot \frac{6}{27} \\ &= \frac{19}{9} - \frac{19}{9} \\ &= 0 \end{aligned}$$

– součiny $x_1 \cdot x_2 \cdot P$ najdeme v tabulkách

$\Rightarrow cov(X_1, X_2) = 0$. Sice to značí, že jsou nezávislé, ale takto usuzovat nemůžeme. Kdyby X_1, X_2 byly nezávislé $\Rightarrow cov = 0$, ale obráceně to neplatí. Kdyby ale byla kovariance $\neq 0$, značí to závislost.

Příklad 9.3 Máme v osudí 3 černé a 2 bílé míče. Náhodně vybereme míč, vrátíme ho zpět a pak znovu náhodně vybereme míč.

²Kovariance je střední hodnota součinu odchylek obou náhodných veličin X, Y od jejich středních hodnot. Kovariance může nabývat hodnot z intervalu $(-\infty; \infty)$. Kovariance poskytuje informaci o intenzitě vztahu mezi dvěma veličinami.

$$X = \begin{cases} 1, & \text{1. tažený míč je bílý} \\ 0, & \text{1. tažený míč je černý} \end{cases}$$

$$Y = \begin{cases} 1, & \text{2. tažený míč je bílý} \\ 0, & \text{2. tažený míč je černý} \end{cases}$$

Jedná se o veličny, které popisují stejný typ jevu $\Rightarrow$ jsou nezávislé (protože vracíme zpět)
($P[X] = P[Y]$)

$X \backslash Y$	0	1	$P[X = x]$
0	$\frac{9}{25}$	$\frac{6}{25}$	$\frac{3}{5}$
1	$\frac{6}{25}$	$\frac{4}{25}$	$\frac{2}{5}$
$P[Y = y]$	$\frac{3}{5}$	$\frac{2}{5}$	1

$$Z = x + y \sim Bi\left(n = 2, p = \frac{2}{5}\right)$$

Součet veličin s rozdělením 0,1 $\Rightarrow$ binomické rozdělení

$$\begin{aligned} cov(X, Y) &= 0 \\ E(XY) - E(X)E(Y) &= 0 \end{aligned}$$

Nezávislost tedy slouží k usnadnění výpočtu. Když bychom měli např. 20 veličin (každá 0/1), tak pro sdružené rozdělení potřebujeme $2^{20} - 1$ políček.

Marginální distribuční funkce:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{3}{5}, & x \in \langle 0, 1 \rangle \\ 1, & x \geq 1 \end{cases}$$

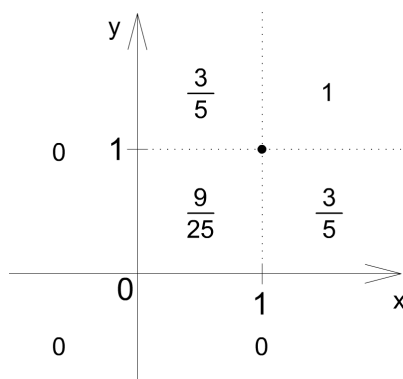
$$F_X(x) = \lim_{y \rightarrow \infty} F_{XY}(x, y)$$

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ \frac{3}{5}, & y \in \langle 0, 1 \rangle \\ 1, & \text{jinak} \end{cases}$$

Sdružená distribuční funkce:

$$\begin{aligned} F_{XY}(x, y) : \mathbb{R}^2 &\rightarrow \langle 0, 1 \rangle \\ F_{XY}(x, y) &= P[X \leq x, Y \leq y] \\ &= \sum_{x' \leq x, y' \leq y} P[X = x', Y = y'] \end{aligned}$$

$$F_{XY}(x, y) = \begin{cases} 0, & x < 0 \text{ nebo } y < 0 \\ 1, & x \geq 1, y \geq 1 \\ \frac{9}{25}, & x, y \in \langle 0, 1 \rangle \\ \frac{3}{5}, & x \geq 1, y \in \langle 0, 1 \rangle \text{ nebo } x \in \langle 0, 1 \rangle, y \geq 1 \end{cases}$$



Alternativní definice pomocí distribuční a marginální funkce

$$F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$$

$$F_{XY}(x, y) = \begin{cases} x + y, & x, y \in \langle 0, 1 \rangle \\ 0, & \text{jinak} \end{cases}$$

- marginální rozdělení
- (ne) závislost X a Y

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \begin{cases} 0, & x \notin \langle 0, 1 \rangle \\ \int_0^1 x + y dy = \left[xy + \frac{y^2}{2} \right]_0^1 = x + \frac{1}{2}, & x \in \langle 0, 1 \rangle \end{cases}$$

X a Y jsou nezávislé $\Leftrightarrow f_{XY}(x, y) \neq f_X(x)f_Y(y), \exists x, y \in \mathbb{R}$
 $\uparrow$
 nejsou nezávislé

kovariance:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy = \int_0^1 \int_0^1 xy(x+y) dx dy \\ &= \int_0^1 \left[\frac{x^3}{3}y + \frac{x^2}{2}y^2 \right]_0^1 dy = \int_0^1 \frac{y}{3} + \frac{y^2}{2} dy \\ &= \left[\frac{y^2}{6} + \frac{y^3}{6} \right]_0^1 = \frac{1}{3} \end{aligned}$$

$$E(X) - E(Y) = \int_0^1 x \cdot \left(x + \frac{1}{2} \right) dx = \left[\frac{x^3}{3} + \frac{x^2}{4} \right]_0^1 = \frac{7}{12}$$

$$\begin{aligned} D(X) = \text{cov}(X, X) &= E(X^2) - E^2(X) = \int_0^1 x^2 \cdot \left(x + \frac{1}{2} \right) dx - \left(\frac{7}{12} \right)^2 \\ &= \underbrace{\left[\frac{x^4}{4} + \frac{x^3}{6} \right]_0^1}_{\frac{10}{24}} - \frac{49}{144} = \frac{11}{144} \end{aligned}$$

korelace (znormujeme kovarianci):

$$\begin{aligned} \rho(X, Y) &\in \langle -1, 1 \rangle \\ \rho(X, Y) &= \frac{\text{cov}(X, Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} = \frac{\frac{1}{3} - \frac{49}{144}}{\frac{11}{144}} = -\frac{1}{11} = \rho(Y, X) \end{aligned}$$

Vyšel nám *koeficient korelace*. Korelace *do určité míry* měří lineární závislost $\Rightarrow$ mezi veličinami je velmi slabá lineární závislost.

Kovarianční matice:

$$\sum_{XY} \begin{bmatrix} \overset{= \text{cov}(X, X)}{\underset{\downarrow}{D(X)}} & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \underset{= \text{cov}(Y, Y)}{\overset{\uparrow}{D(Y)}} \end{bmatrix}$$

Korelační matice:

$$\rho_{XY} \begin{bmatrix} 1 & \rho(X, Y) \\ \rho(Y, X) & 1 \end{bmatrix}$$

Diagonály jsou „1“ - odpovídá korelaci veličiny sama se sebou $\rho(X, X) = 1$.

Příklad 9.4 Máme náhodný vektor dvou spojitých náhodných veličin popsaných tzv. *sdrúženou hustotou*.

$$f_{XY}(x, y) = \begin{cases} x + y, & x, y \in \langle 0; 1 \rangle \\ 0, & \text{jinak} \end{cases}$$

Aby toto platilo, obsah musí být „1“ $\Rightarrow$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$$

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(u, v) dv du$$

Musíme říct, jaká je hodnota v každém bodě. Nemáme ale tabulky, pomůžeme si tedy přes hustotu:

$$F_{XY}(x, y) = \begin{cases} 1, & x, y \geq 0 \\ 0, & x \leq 0 \text{ nebo } y < 0 \\ \frac{x^2 \cdot y + x \cdot y^2}{2}, & x, y \in \langle 0; 1 \rangle \end{cases} \quad \text{výsledek bude spojitá funkce}$$

protože:

$$\int_0^x \int_0^y (t + u) du dt \quad \text{použijeme Fubiniho větu} \Rightarrow \text{nejdříve integrujeme vnitřní a pak vnější}$$

$$\begin{aligned} &= \int_0^x \left[t \cdot u + \frac{u^2}{2} \right]_0^y dt \\ &= \int_0^x t \cdot y + \frac{y^2}{2} dt \\ &= \left[y \cdot \frac{t^2}{2} + t \cdot \frac{y^2}{2} \right]_0^x \\ &= \frac{x^2 \cdot y + x \cdot y^2}{2} \end{aligned}$$

Nyní spočítáme marginální hustotu (pořád myslíme na analogii s tabulkou):

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dy \\ &= \int_0^1 (x + y) dy \\ &= \left[x \cdot y + \frac{y^2}{2} \right]_0^1 \\ &= x + \frac{1}{2} \quad x \in \langle 0; 1 \rangle \end{aligned}$$

Počítali jsme pouze přes 1 hodnotu, stejně jako v tabulce.

$$f_Y(y) = y + \frac{1}{2} \quad y \in \langle 0; 1 \rangle$$

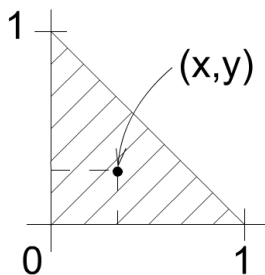
protože jsou symetrické a když:

$$\begin{array}{rcl} x \notin \langle 0; 1 \rangle & f_X(x) = 0 \\ \Rightarrow & \\ y \notin \langle 0; 1 \rangle & f_Y(y) = 0 \end{array}$$

10 Cvičení 10

Příklad 10.1 Máme trojúhelník a náhodně v něm vybereme bod.

Obě souřadnice lze považovat za realizaci náhodného vektoru spojitého rozdělení.



Máme sdružený model, sdruženou hustotu:

$$f_{XY}(x, y) = \begin{cases} 2, & (x, y) \in M \dots \text{plocha trojúhelníku} - \text{plocha čtverce je } 1, \text{ trojúhelník má } \frac{1}{2} \Rightarrow \frac{1}{\frac{1}{2}} = 2 \\ 0, & \text{jinak} \end{cases}$$

$$M = \{(x, y) \in \mathbb{R} \mid x + y \leq 1, x \geq 0, y \geq 0\}$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \begin{cases} 0, & x \notin \langle 0, 1 \rangle \\ \int_0^{1-x} 2 dy = 2 \cdot (1 - x), & x \in \langle 0, 1 \rangle \end{cases}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \cdot 2 \cdot (1 - x) dx \\ &= 2 \cdot \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = 1 - \frac{2}{3} = \frac{1}{3} \end{aligned}$$

Smíšené rozdělení náhodného vektoru

- 1) Diskrétní $X \sim$ rovnoměrné rozdělení na dvouprvkové množině $\{0, 1\}$
- 2) Spojitý $Y \sim$ rovnoměrné rozdělení na $\langle 0, 2 \rangle$

1) $P[X = 0] = P[X = 1] = \frac{1}{2}$ je to na konečné množině, proto *Laplaceův model*.

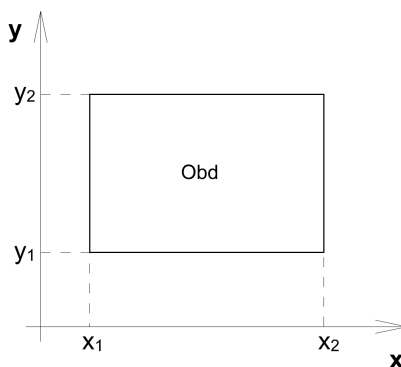
2) Distribuční funkce rovnoměrného rozdělení:

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ \frac{1}{2} \cdot y, & y \in \langle 0; 2 \rangle \\ 1, & y \geq 2 \end{cases}$$

Distribuční funkce sdruženého rozdělení (budeme předpokládat jejich nezávislost):

$$F_{XY}(x, y) = F_X(x) F_Y(y) = \begin{cases} 0 & x < 0 \text{ nebo } y < 0 \\ \frac{y}{4} & x \in \langle 0, 1 \rangle, y \in \langle 0, 2 \rangle \\ \frac{y}{2} & x \geq 1, y \in \langle 0, 2 \rangle \\ \frac{1}{2} & x \in \langle 0, 1 \rangle, y > 2 \\ 1 & \text{jinak} \end{cases}$$

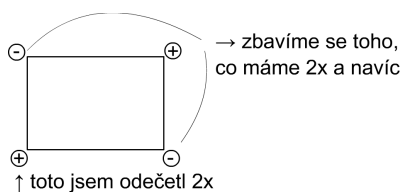
Příklad 10.2 Chceme spočítat pravděpodobnost, že náhodný vektor spadne do obdélníku. Jak to spočítáme z distribuční funkce?



plocha obdélníku:

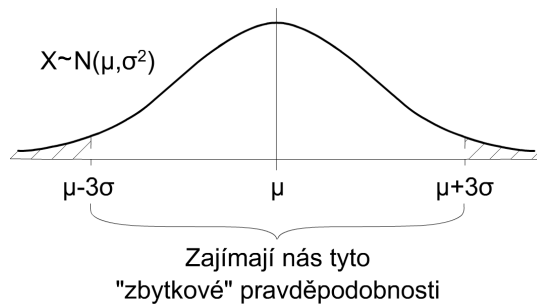
$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(x, y) \, dy \, dx$$

$$P[(X, Y) \in \text{Obd}] = F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1)$$

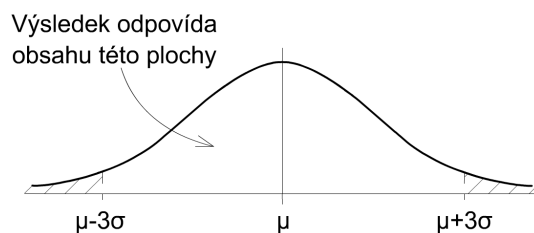


Pravidlo 3σ

Ve vzdálenosti alespoň 3σ od střední hodnoty $N(\mu, \sigma^2)$ se nachází asi 3 promile pozorování.



$$\begin{aligned}
 P[|X - \mu| < 3\sigma] &= P[-3\sigma < X - \mu < 3\sigma] \\
 &= P\left[-3 < \frac{X - \mu}{\sigma} < 3\right] \\
 &= \Phi(3) - \Phi(-3) = \underbrace{2\Phi(3)}_{=0,99865} - 1 = 0,9973 \sim N(0, 1)
 \end{aligned}$$



pokud neznáme rozdělení

$$P[|X - E(X)| < 3\sigma] \Rightarrow \text{použijeme Čebyševovu nerovnost}$$

Pro $\forall NV$, která má střední hodnotu $E(X)$ a rozptyl $D(X)$ platí:

$$\begin{aligned}
 &\forall \varepsilon > \sigma \\
 P[|(X - E(X))| < \varepsilon] &\geq 1 - \frac{D(X)}{\varepsilon^2}
 \end{aligned}$$

dolní mez – ať je tolerance jakákoliv

$$\begin{aligned}
 &\varepsilon = 3\sigma \\
 P[|X - E(X)| < 3\sigma] &\geq 1 - \frac{\sigma^2}{9\sigma^2} = \frac{8}{9} \\
 &\quad \quad \quad \uparrow \\
 &\quad \quad \quad \text{odhad meze - už nejde zlepšit} \\
 P[\sigma < X - E(X) < 3\sigma]
 \end{aligned}$$

Interpretace je následující: Pravděpodobnost, že realizace pokusu spadne do té výšece omezená 3 parametry je alespoň $\frac{8}{9}$ (dolní mez). Platí pro všechno, ale je to zbytečně pesimistické.

Příklad 10.3 Házíme 10^6 -krát symetrickou mincí. Jaká je pravděpodobnost, že četnost *líců* bude v intervalu $\langle 3 \cdot 10^5; 7 \cdot 10^5 \rangle$?

- $X \dots$ počet $L \sim Bi(n = 10^6; \rho = \frac{1}{2})$

$$P[X = x] = \binom{10^6}{x} \cdot \frac{1}{2^{10^6}}$$

$$P[X \in \langle 3 \cdot 10^5; 7 \cdot 10^5 \rangle] = \sum_{x=3 \cdot 10^5}^{7 \cdot 10^5} \binom{10^6}{x} \cdot \frac{1}{2^{10^6}} \quad \Leftarrow \text{ ale moc velké}$$

Použijeme proto hrubý odhad dle *Čebyševovy nerovnosti*:

$$E(X) = n \cdot p = 5 \cdot 10^5 \quad \varepsilon = 2 \cdot 10^5 + 1$$

$$D(X) = n \cdot p \cdot (1 - p) = \frac{10^6}{4} = 2 \cdot 5 \cdot 10^5$$

Z *Čebyševovy nerovnosti* $\Rightarrow$

$$P[|X - 5 \cdot 10^5| < 2 \cdot 10^5 + 1] \geq 1 - \frac{2 \cdot 5 \cdot 10^5}{(2 \cdot 10^5 + 1)^2} \doteq 0,999994 \quad \text{dolní odhad}$$

11 Cvičení 11

Příklad 11.1 $\frac{S_n}{n}$ je relativní četnost líců v n hodech mincí. Kolik musíme provést hodů, abychom se v odhadu $\frac{1}{2}$ padnutých líců lišili maximálně o 0,05 s pravděpodobností nejméně 0,95?

- $\bar{X} \dots$ výběrový průměr = $\frac{S_n}{n}$

Věta *Kolik musíme provést...* říká něco o pravděpodobnosti, proto:

$$P \left[\left| \frac{S_n}{n} - \frac{1}{2} \right| < 0,05 \right] \geq 0,95$$

Využijeme Čebyševovu nerovnost:

$$\varepsilon = 0,05$$

$$E \left(\frac{S_n}{n} \right) \stackrel{?}{=} \frac{1}{2}$$

A protože výběrový průměr je nestranný odhad $\Rightarrow$

$$E \left(\frac{S_n}{n} \right) = \frac{E(S_n)}{n} = \frac{\frac{n}{2}}{n} = \frac{1}{2}$$

$$S_n \sim B_i \left(n, \frac{1}{2} \right)$$

$\uparrow$
 binomické rozdělení

Nyní už máme vše připravené pro Čebyševovu nerovnost, protože podle ní platí:

$$P \left[\left| \frac{S_n}{n} - \frac{1}{2} \right| < 0,05 \right] \geq 0,95 = 1 - \frac{D \left(\frac{S_n}{n} \right)}{\varepsilon^2}$$

Chceme vyjádřit n :

$$D \left(\frac{S_n}{n} \right) = \frac{D(S_n)}{n^2} = \frac{\frac{n}{4}}{n^2} = \frac{1}{4n}$$

Po dosazení máme výraz

$$P \left[\left| \frac{S_n}{n} - \frac{1}{2} \right| < 0,05 \right] \geq 0,95 = 1 - \frac{1}{4n \cdot (0,05)^2}$$

$$0,05 = \frac{1}{0,001 \cdot n}$$

$$n = 2000$$

Musíme provést alespoň 2000 hodů, abychom splnili 95% pravděpodobnost, že počet líců se bude lišit maximálně o 0,05 od jedné poloviny hodů.

Nicméně *Čebyševova nerovnost* je pesimistická, a proto si nyní ukážeme jiný způsob řešení, který „srazí“ počet pokusů na nižší hodnotu. Řešení je založeno na *centrální limitní větě* (de Moivre-Laplace ¹), zjednodušeně řečeno: když budeme mít posloupnost veličin, které mají binomické rozdělení, máme stále stejnou minci, ale provádíme s ní velký počet pokusů:

$$S_n \sim B_i(n, p) \quad n \rightarrow \infty$$

$$\Rightarrow \lim_{n \rightarrow \infty} P \left[\frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \leq x \right] = \Phi(x)$$

Podle centrální limitní věty bude graf pravděpodobnosti (Distribuční funkce) dostatečně aproximativní.

Výraz $P \left[\left| \frac{S_n}{n} - \frac{1}{2} \right| < 0,05 \right]$ chceme převést tak, abychom ho mohli dosadit do uvedené limity.

$$P \left[\left| \frac{S_n}{n} - \frac{1}{2} \right| < 0,05 \right] \quad \text{celou nerovnost vynásobím } n$$

$$P \left[\left| S_n - \frac{n}{2} \right| < 0,05 \cdot n \right] \quad \text{dále dělím směrodatnou odchylkou } \frac{\sqrt{n}}{2}$$

$$P \left[\left| \frac{S_n - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \right| < \frac{0,05 \cdot n}{\frac{\sqrt{n}}{2}} \right] = P \left[\left| \frac{S_n - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \right| < 0,1 \cdot \sqrt{n} \right]$$

$$\Phi(0,1 \cdot \sqrt{n}) - \Phi(-0,1 \cdot \sqrt{n}) \quad \text{aproximujeme dle CLV}$$

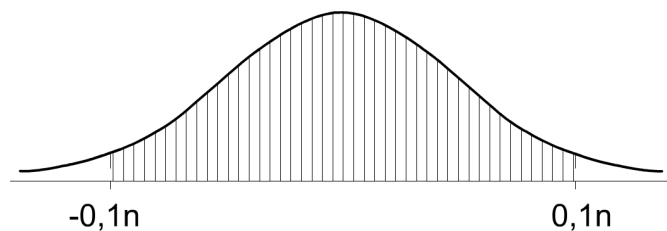
$$2\Phi(0,1 \cdot \sqrt{n}) - 1 \geq 0,95$$

$$\Phi(0,1 \cdot \sqrt{n}) \geq 0,975$$

$$0,1 \cdot \sqrt{n} \geq \Phi^{-1}(0,975) = u_{0,975}$$

$$n \geq 100 \cdot u_{0,975}^2 = 385$$

$$(0,1^2)^{-1}$$



Toto potvrzuje, jak je *Čebyševova nerovnost* pesimistická, nyní nám stačí provést pouze 385 hodů, ale pro *Čebyševovu nerovnost* byla dolní mez pevně daná, zatímco CLV pracuje s aproximací (v nekonečnu).

Příklad 11.2 Kolik respondentů je třeba oslovit při průzkumu, abychom odhadli volební výsledek s pravděpodobností α a odchylkou menší než ε .

-
- $n \dots$ počet respondentů = ?

¹ Jedná se nejjednodušší vyjádření CLV. Podle *Moivreovy-Laplaceovy* věty při velkém počtu nezávislých pokusů konverguje binomické rozdělení k rozdělení normálnímu.

- $p \dots$ neznámá pravděpodobnost (preference 1 strany, jejíž výsledek nás zajímá)
- $S_n \dots$ počet respondentů preferujících danou stranu
- $\frac{S_n}{n} \dots ?$

Zajímá nás $\frac{S_n}{n}$, což je opět $\bar{X}$ = výběrový průměr (ve statistické terminologii) $\Rightarrow$ je to bodový odhad parametru p :

$$P \left[\left| \frac{S_n}{n} - p \right| \leq \varepsilon \right] \geq \alpha$$

Zajímá nás tedy interval spolehlivosti na hladině α pro parametr p , takže výraz $P \left[\left| \frac{S_n}{n} - p \right| < \varepsilon \right]$ chceme převést do tvaru centrální limitní věty:

$$\begin{aligned} P \left[\left| \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \right| \leq \frac{\varepsilon \cdot n}{\sqrt{n \cdot p \cdot (1-p)}} \right] &\doteq \\ &\doteq 2\Phi \left(\frac{\varepsilon \cdot \sqrt{n}}{\sqrt{p \cdot (1-p)}} \right) - 1 \geq \alpha \end{aligned}$$

Násobili jsme n a dělili směrodatnou odchylkou (přepokládáme, že je nenulová).
V reálném světě ε známe, ale nevíme p , které musíme odhadnout:

$$p \cdot (1-p) \rightarrow \sqrt{p \cdot (1-p)} \leq \frac{1}{2} \leftarrow \text{spodní mez}$$

$$\begin{aligned} 2\Phi \left(\frac{\varepsilon \cdot \sqrt{n}}{\frac{1}{2}} \right) - 1 &\geq \alpha \\ \Phi(2 \cdot \varepsilon \cdot \sqrt{n}) &\geq \frac{\alpha + 1}{2} \\ 2 \cdot \varepsilon \cdot \sqrt{n} &\geq \Phi^{-1} \left(\frac{\alpha + 1}{2} \right) \cdot u_{\frac{\alpha+1}{2}} \\ n &\geq \frac{1}{4\varepsilon^2} \cdot u_{\frac{\alpha+1}{2}}^2 \end{aligned}$$

Tzn., že kvantil roste s α , která odpovídá požadované přesnosti. Pro $\varepsilon = 0,01$ a $\alpha = 0,90$ vychází $n \geq 6764$, tedy musíme oslovit 6764 respondentů (*normální rozdělení*).

Pro *Čebyševovu nerovnost* by byl výsledek vyšší než 25 000 respondentů.

Pro zpřesnění výpočtu můžeme lépe analyzovat parametr p - např. z výsledku minulých voleb. Budeme brát, že $p \leq 0,3$.

$$2\Phi \left(\frac{\varepsilon \cdot \sqrt{n}}{\sqrt{0,21}} \right) - 1 \quad \text{dopočítáme jako před tím} \Rightarrow$$

$$\begin{aligned} n &\geq 0,21 \cdot \frac{1}{\varepsilon^2} \cdot n_{\frac{\alpha+1}{2}}^2 \\ n &\geq 5676 \end{aligned}$$

Dostáváme menší dolní mez, o více než 1000.

Příklad 11.3 Sečteme 300 čísel zaokrouhlených na jedno desetinné místo. Jaká je pravděpodobnost, že chyba při zaokrouhlení je menší než 1?

- $X \dots$ chyba při zaokrouhlení ($i = 1, \dots, 300$)

$$P[|X_1 + \dots + X_{300}| \leq 1]$$

Nejdříve musíme určit rozdělení našeho X_i . Předpokládáme rovnoměrné rozdělení na intervalu $-0,05$ až $0,05$ a předpokládáme, že chyby při rozdělení jsou nezávislé.

$$f_{X_i} = \begin{cases} \frac{1}{0,01} = 10 & x \in \langle -0,05; 0,05 \rangle \\ 0 & \text{jinak} \end{cases}$$

Těch $0,1$ je délka intervalu.

Mohli bychom 300 krát počítat konvoluci, ale použijeme CLV (trošku jinou než v minulém příkladě, tahle je *obecnější*). $X_1, \dots, X_n$ je posloupnost i.i.d.² veličin se střední hodnotou $\mu = E(X_i)$ a rozptylem $D(X_i) = \sigma^2$:

$$\lim_{n \rightarrow \infty} P \left[\frac{\sum_{i=1}^n X_i - n \cdot \mu}{\sqrt{n} \cdot \sigma} \right] = \Phi(x)$$

$$D \left(\sum_{i=1}^{300} X_i \right) = n \cdot \sigma^2$$

$$E \left(\sum_{i=1}^{300} X_i \right) = \sum_{i=1}^{300} E(X_i) = 0$$

$$D(X_i) = E(X_i^2) = \int_{-0,05}^{0,05} x^2 \cdot 10 dx = \frac{0,01}{12} = \frac{1}{1200} \Rightarrow$$

$$P \left[\frac{\left| \sum_{i=1}^{300} X_i \right|}{\sqrt{300} \cdot \sqrt{\frac{1}{1200}}} \leq \frac{1}{\sqrt{300} \cdot \sqrt{\frac{1}{1200}}} \right] \doteq \Phi(2) - \Phi(-2)$$

$$2 \cdot \Phi(2) - 1 \doteq 0,955$$

$$\frac{1}{\sqrt{\frac{300}{1200}}}$$

²V anglické literatuře i.i.d. = *Independent and Identically Distributed*. Náhodné vzorky definujeme jako sekvenci navzájem nezávislých náhodných veličin se stejným rozdělením.

12 Cvičení 12

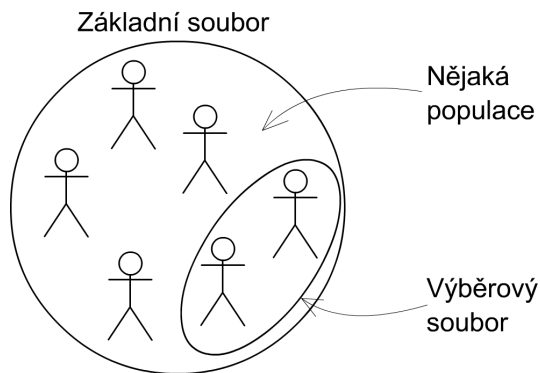
Příklad 12.1 Velikost platů ve firmě

Máme náhodný výběr o rozsahu $n = 9$, výsledek:

i	1	2	3	4	5	6	7	8	9
tis. Kč	12	15	21	25	28	32	35	48	62

1. Jaké pravděpodobnostní rozdělení odpovídá sledované situaci? Jaký je vhodný model pro rozdělení mezd?
2. Jaká je střední hodnota platu, rozptyl atd.?

Mějme nějaký *základní soubor* obsahující *nějakou populaci*. My ovšem máme k dispozici jen část $\Rightarrow$ výběrový soubor¹.



Z tohoto *výběrového souboru* získáme nějaká data:

$DATA(x_1, \dots, x_n) \in \mathbb{R}^n$ je to nějaká n -tice, v našem případě 9-tice

$\Rightarrow$ chceme postihnout všechna data, proto máme náhodný výběr $(X_1, \dots, X_n)$ je i.i.d
 $\Rightarrow$ statistika ... funkce náhodného výběru $\hat{\Theta}(X_1, \dots, X_n)$, často vyjádřeno číslem ϑ , kterému říkáme *realizace odhadu*.

Proto můžu **2.** řešit několika způsoby:

Střední hodnota $\vartheta = E(X)$:

- a) Pomocí výběrového průměru (prostý aritmetický)

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ &= \frac{1}{9} \sum_{i=1}^9 x_i \\ &\doteq 30\,900,-\end{aligned}$$

¹ *Výběrový soubor* = soubor jednotek (pro něž jsou k dispozici data nebo pro něž tato data získáváme), který zastupuje *základní soubor* ve výzkumu. Výsledky pro něj zjištěné mají být zobecnitelné na základní soubor.

- b) Pomocí výběrového mediánu ²

$$\hat{q}(0, 5) = 28\,000,-$$

- c) Pomocí „uříznutého“ průměru ³

$$\bar{x} = 29\,140,-$$

Rozptyl:

- a) Odhadujeme ho pomocí výběrového rozptylu

$$\begin{aligned} s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{8} \sum_{i=1}^9 (x_i - 30,9)^2 \\ &= 253,6 \end{aligned}$$

Dělíme ale jen 8-krát, protože se chci trefit do skutečné leč neznámé hodnoty.

- b)

$$\widehat{\sum_x^2} = \frac{n-1}{n} \cdot s_x^2 \quad \text{toto je menší než výběrový rozptyl}$$

Výběrový rozptyl a výběrový průměr jsou nestranné odhady:

$$\begin{aligned} E(s_x^2) &= D(X) \\ E(\bar{X}) &= E(X) \end{aligned}$$

1.

Platy představují *lognormální rozdělení* ⁴

Místo hodnot ale mohu vzít interval hodnot. Sem může padnout hodnota s touto pravděpodobností:

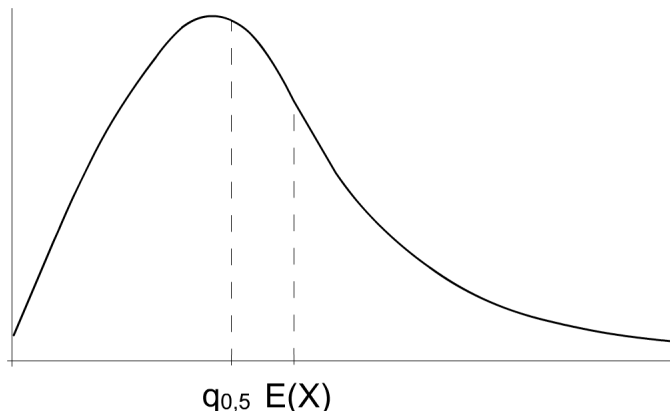
$$P[\vartheta \in I] \geq 1 - \alpha$$

$\Rightarrow$ na což navazuje další příklad.

² *Výběrový medián* je jiným odhadem střední hodnoty. Při lichém počtu pozorování je to prostřední hodnota v uspořádaném výběru, máme-li sudý počet pozorování, použijeme průměr dvou prostředních hodnot.

³ *Trimmed Mean, uříznutý*, nebo také *useknutý* průměr = určitá část největších a nejmenších hodnot se do výpočtu nezahrne. Např. desetiprocentní uřezaný průměr znamená, že se vynechá 10% nejnižších výsledků a 10% nejvyšších výsledků a ze zbytku se počítá prostý průměr. Obvykle se volí 5%, 10% nebo 25%-ní useknutý průměr. Jde o snahu nezahrnout do výpočtu extrémní hodnoty. Odstraňuje nedostatky prostého průměru.

⁴ Je to rozdělení, jež není symetrické. Je zřejmé, že silně nesymetrické rozdělení nemůže mít ty vlastnosti, kterými se vyznačuje normální rozdělení. Můžeme však uvažovat *lognormální rozdělení*.



Příklad 12.2 Pro dopravu ze Strahova na Karlovo náměstí byl použit $n = 18$ -krát autobus a tramvaj. Z toho bylo vypočteno:

	Střední hodnota	Směrodatná odchylka
Autobus	$\bar{x}_a = 18,83$	$s_a = 2\,706$
Tramvaj	$\bar{x}_t = 20,90$	$s_t = 1\,348$

Zdá se, že je lepší jezdit autobusem - potvrdí se tento předpoklad?

Zajímá nás to spíše z dlouhodobého hlediska, tedy interval ($\alpha = 0,05$). Jaká bude maximální doba cestování, kterou nepřekročíme s pravděpodobností 95%?

$\Rightarrow$ použijeme intervalový odhad střední hodnoty (μ) a zajímá nás horní mez (za předpokladu normality doby cestování).

$$I = \left(-\infty; \bar{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1,1-\alpha} \right)$$

kde:

- $\bar{x} \dots$ je střední hodnota
- $s \dots$ je směrodatná odchylka
- $\sqrt{n} \dots$ je odmocnina z počtu pozorování
- $t_{n-1,1-\alpha} \dots$ je kvantil *Studentova rozdělení*⁵, najdeme v tabulkách

$\Rightarrow$ pro oba dopravní prostředky platí:

$$I_{bus} = \left(-\infty; 18,83 + \frac{2706}{\sqrt{18}} \cdot 1,74 \right) = (-\infty; 19,93)$$

$$I_{tram} = \left(-\infty; 20,90 + \frac{1348}{\sqrt{18}} \cdot 1,74 \right) = (-\infty; 21,45)$$

⁵Označováno též jako *t rozdělení*. Název Studentovo rozdělení vznikl podle pána, který se nejmenoval Student, jak by se dalo očekávat. Tento pán se jmenoval William Sealy Gosset a nebyl povoláním statistikem, jak by se dalo očekávat. Byl povoláním chemik a začátkem 20. století pracoval v anglickém pivovaru Guinness. A toto rozdělení nevymyslel proto, že by se nudil, ale proto, že potřeboval vyvozovat na základě velmi malých vzorků použitelné závěry. A k těmto účelům se toto rozdělení používá dodnes.

$\Rightarrow$ s pravděpodobností 95% nikdy nebudeme cestovat busem více než 19,93min $\Rightarrow$ bus je i tak *výhodnější*.

ale:

1. Ze Strahova nejedí tramvaj ;-)
2. Pripouštíme i záporný čas
3. Možné závislosti
 - Cestujeme (měříme data) ve stejnou dobu
 - Na trase se může stát nehoda

Příklad 12.3 Výška 10-ti letých chlapců

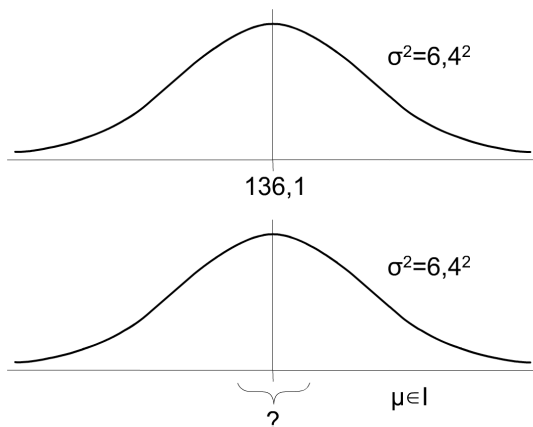
V roce 1961 bylo naměřeno $n = 15$ výšek:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
v	130	140	136	141	139	133	149	151	139	136	138	142	127	139	147

Z roku 1951 máme k dispozici tyto údaje:

- Střední hodnota $\mu = 136,1$
- Směrodatná odchylka $\sigma_o = 6,4$

Můžeme na základě těchto dat usuzovat, že se průměrná výška v těchto letech změnila?



Jsou předpoklady, které je slušné říci před tím, než začneme počítat:

- Výška má normální rozdělení
- Střední hodnota se změnit může, ale rozptyl ne $\rightarrow$ variabilita se nemění

Opět např. $\alpha = 0,05$ a interval spolehlivosti pro neznámou střední hodnotu μ při známém rozptylu σ^2 :

$$P \left[\mu \in \left\langle \bar{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{n}}; \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{n}} \right\rangle \right] \geq 1 - \alpha = 0,95$$

$$\begin{aligned} I &= \left\langle 139,1 - \frac{6,4}{\sqrt{15}} \cdot 1,96; 139,1 + \frac{6,4}{\sqrt{15}} \cdot 1,96 \right\rangle \\ &= \langle 135,9; 142,4 \rangle \end{aligned}$$

$\Rightarrow$ střední hodnota se nezměnila, spadá do intervalu. Proto je lepší použít intervalový odhad, dává nám lepší informaci, nespokojíme se jen s jednou hodnotou.

Pozn.: Jak volíme kvantil?

1. **Studentovo** = neznáme rozptyl
2. **Normální** = rozptyl známe

13 Cvičení 13

Metody odhadů parametrů:

- Metoda momentů
- Metoda maximální věrohodnosti *bývá častější*

$\bar{\vartheta} = (\vartheta_1, \dots, \vartheta_k) \in \Pi \subseteq \mathbb{R}^k$ vektor k reálných čísel

Na základě realizace náhodného výběru $(x_1, \dots, x_n) \in \mathbb{R}^n$ chceme odhadnout neznámé parametry ϑ_i

Příklad 13.1 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 .

Tedy:

- $n = 20$
- $p, q \in \langle 0; 1 \rangle$ a jsou to zároveň *omezující podmínky*
- $x_1 = 0, x_2 = 0$
- $x_3 = 1, x_4 = 1, \dots, x_{14} = 1$
- $x_{15} = 2, \dots, x_{20} = 2$
- p, q, q^2 jsou ϑ
- empirické rozdělení, např. pro 0 je to $\frac{1}{10}$

$p = 1 - q - q^2$ součet musí být 1 $\Rightarrow$ stačí odhadnout q

$$\Pi \in \mathbb{R}^3$$

$$\Pi = \{(p, q, q^2) \in \mathbb{R}^3 \mid 0 \leq p, q \leq 1, p + q + q^2 = 1\}$$

což je parametrická množina splňující nějaká omezení.

$\Rightarrow$ vybereme bod z Π , který odpovídá jistému pravděpodobnostnímu modelu.

1) Metoda momentů:

Předpokládáme rovnost výrazů:

$$EX = q + 2 \cdot q^2 = m_x = \frac{0 \cdot 2 + 1 \cdot 12 + 2 \cdot 6}{20} = \frac{6}{5}$$

Dostáváme kvadratickou rovnici o 1 neznámé

$$q + 2 \cdot q^2 = \frac{6}{5}$$

kteřá má 2 řešení

$$q_1 = -\frac{1}{20}\sqrt{265} - \frac{1}{4} \doteq -1,0639 \quad q_2 = \frac{1}{20}\sqrt{265} - \frac{1}{4} \doteq 0,56394$$

q_1 nevyhovuje zadání (nepatří do parametrické množiny). q_2 ale ano $\Rightarrow p = 1 - q_2 - q_2^2 \doteq 0,11803$

2) Metoda maximální věrohodnosti:

$$l(q) = 2 \ln p + 12 \ln q + 6 \ln q^2 = 2 \ln (1 - q - q^2) + 24 \ln q$$

$$\frac{\partial}{\partial q} l(q) = \frac{24}{q} + 2 \frac{-2q - 1}{1 - q - q^2}$$

Výraz *položíme* 0 a vyjde nám opět kvadratická rovnice s řešeními $q_1 = -\frac{3}{2}$ (nevyhovuje), $q_2 = \frac{4}{7} \doteq 0,57143$ (vyhovuje) a $p = 1 - q_2 - q_2^2 = \frac{5}{49} \doteq 0,10204$.

Odhad je u obou metod odlišný a často tomu tak bude.

Příklad 13.2 Tabulka udává pravděpodobnosti a pozorované četnosti výsledků hodů nepravděpodobnou 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

Součet pravděpodobností musí být 1 $\Rightarrow a = \frac{1}{6}$. Zbývá určit $b \in \left\langle -\frac{1}{6}; \frac{1}{6} \right\rangle$.

1) Metoda momentů:

Spočítáme střední hodnotu, co má teoreticky vyjít:

$$EX = 5b + \frac{7}{2} = \bar{x} = \frac{1 \cdot 2 + 2 \cdot 5 + 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 6 + 6 \cdot 10}{30} = \frac{21}{5}$$

$$\Rightarrow b = \frac{7}{50} = 0,14 \quad \text{což vyhovuje podmínkám}$$

2) Metoda maximální věrohodnosti:

$$L(b) = P[X_1 = x_1, \dots, X_{30} = x_{30} ; b]$$

$$= \left(\frac{1}{6} - b\right)^2 \cdot \left(\frac{1}{6}\right)^{18} \cdot \left(\frac{1}{6} + b\right)^{10}$$

$$l(b) = \ln(L(b)) = 2 \ln \left(\frac{1}{6} - b\right) + 18 \ln \frac{1}{6} + 10 \ln \left(\frac{1}{6} + b\right)$$

$$\frac{\partial}{\partial b} l(b) = -2 \frac{1}{\frac{1}{6} - b} + 10 \frac{1}{\frac{1}{6} + b}$$

$b = \frac{1}{9}$, což vyhovuje podmínkám.

Příklad 13.3 Odhad v rozdělení Bi(1,p)

Např. pravděpodobnost p padnutí 6 v 1 hod. Cílem je najít odhad parametru pomocí obou metod.

Musíme opět předpokládat, že jsme provedli náhodný výběr a jeho realizace je vektor $(x_1, \dots, x_n)$, $x_i \in \{0, 1\}$, tedy padla/nepadla 6.

1) Metoda momentů:

$E(X) = p$ nyní tedy stačí najít řešení pravé strany

$$\Rightarrow \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

2) Metoda maximální věrohodnosti:

Spočítáme věrohodnostní funkci

$$L(p) = p^x \cdot (1-p)^{n-x}$$

$$x = \sum_{i=1}^n x_i \quad \text{počet jedniček} \quad l(p) = x \ln p + (n-x) \ln (1-p)$$

$$\frac{\partial}{\partial p} l(p) = \frac{x}{p} - \frac{n-x}{1-p} = 0$$

$$x(1-p) - p(n-x) = 0$$

$$x - pn = 0$$

$$p = \frac{x}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

$\Rightarrow \bar{x}$ výběrový průměr je odhadem p pomocí metody momentů i maximální věrohodnosti.

Příklad 13.4 Odhad v biexponenciálním rozdělení

$f(x) = ce^{-a|x|}$, $x \in \mathbb{R}$, $a > 0$, $c = ?$

$$n = 10 \Rightarrow \bar{x} = 2 \quad s_x^2 = 4$$

Určíme $c \in \mathbb{R}$ z podmínky $\int_{-\infty}^{\infty} f_X(x) dx = 1$

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= c \int_{-\infty}^0 e^{ax} dx + c \int_0^{\infty} e^{-ax} dx \\ &= 2c \int_0^{\infty} e^{-ax} dx \\ &= 2c \left[\frac{e^{-ax}}{-a} \right]_0^{\infty} = \frac{2c}{a} = 1 \end{aligned}$$

$$\Rightarrow c = \frac{a}{2}$$

Metoda momentů:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = 0 \neq 2 \Rightarrow \text{použiji 2.obecný moment } E(X^2)$$

↑
funkce je sudá

$$\begin{aligned}
 E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
 &= \frac{a}{2} \int_{-\infty}^{\infty} x^2 e^{-a|x|} dx \\
 &= \frac{a}{2} \int_{-\infty}^0 x^2 e^{ax} dx + \frac{a}{2} \int_0^{\infty} x^2 e^{-ax} dx \\
 &= a \int_0^{\infty} x^2 e^{-ax} dx \quad y = ax \\
 &= a \int_0^{\infty} \frac{y^2}{a^2} e^{-y \frac{1}{a}} dy \\
 &= \frac{1}{a^2} \underbrace{\int_0^{\infty} y^2 e^{-y} dy}_2 \\
 &= \frac{2}{a^2} \leftarrow \text{levá strana rovnice}
 \end{aligned}$$

Výběrový 2. obecný moment:

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n x_i^2 &= \text{průměr z kvadrátu pozorování} \\
 &= \frac{n-1}{n} s_x^2 + \bar{x}^2 \\
 &= \frac{9}{10} \cdot 4 + 2^2 = \frac{38}{5}
 \end{aligned}$$

$$\Rightarrow a \doteq 0,51$$

14 Cvičení 14

Příklad 14.1 Výška 10-ti letých chlapců

Navazujeme na příklad z 12. cvičení.

$n = 15$

V roce 1951 $\mu_0 = 136,1$ $\sigma = 6,4$

V roce 1961 považujeme rozptyl za známý

Provedeme test o střední hodnotě normálního rozdělení se známým rozptylem na hladině $\alpha = 0,05$

$H_0 : \mu = 146,1$ (v roce 1961) nulová hypotéza

Testovací kritérium:

$$\frac{\bar{x} - 136,1}{\sigma_0} \sqrt{n}$$

Máme kvantil $u_{0,975} = 1,96$

Testovací kritérium $<$ kvantil $\Rightarrow H_0$ nezamítáme.

Příklad 14.2 Test parametru p rozdělení $Bi(1, p)$

Máme $n = 2000$ pacientů, kteří jsou po operaci bypassu. Z toho je 835 bělochů.

H_0 : zda dochází k rasové diskriminaci. $p = 0,5$

H_1 : $p \neq 0,5$

$\bar{x} = \frac{835}{2000} = 0,4175$ je realizací odhadu parametru p

$\bar{X} \dots$ výběrový průměr je nestranným odhadem p

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n} E\left(\sum x_i\right) \\ &= \frac{1}{n} \sum \underbrace{E(x_i)}_p \\ &= \frac{1}{n} np \\ &= p \\ D(\bar{X}) &= D\left(\frac{\sum x_i}{n}\right) \\ &= \frac{1}{n^2} D\left(\sum x_i\right) \\ &= \frac{1}{n^2} \sum \underbrace{D(x_i)}_{p(1-p)} \\ &= \frac{p(1-p)}{n} \end{aligned}$$

$\bar{X}$ má podle *centrální limitní věty* rozdělení $N\left(p, \frac{p(1-p)}{n}\right)$

Testovací kritérium:

$$U = \frac{\overset{0,5}{\downarrow} \bar{x} - p_0}{\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}} = -7,5 \quad \text{což je velmi malá šance, aby to vyšlo}$$

$$|U| > u_{0,975}$$

Příklad 14.3 χ^2 -test dobré shody

Testujeme shodu empirického rozdělení s rozdělením diskrétním

Máme k tříd a $n_1 + \dots + n_k = n$. V našem případě $n = 2000$

Z online generátoru (hod kostkou) jsem získali následující četnosti:

$$n_1 = 331$$

$$n_2 = 317$$

$$n_3 = 344$$

$$n_4 = 340$$

$$n_5 = 324$$

$$n_6 = 344$$

Teoretická pravděpodobnost by pro i -tou třídu měla být $p_i = \frac{1}{6}$, za předpokladu férové kostky.

H_0 : empirické rozdělení je rovnoměrné

Testovací statistika:

$$\begin{aligned} \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} &= \frac{(331 - 333, \bar{3})^2}{333, \bar{3}} + \dots + \frac{(344 - 333, \bar{3})^2}{333, \bar{3}} \\ &= 1,897 \end{aligned}$$

Kvantil:

$$\begin{aligned} \chi^2_{1-\alpha}(k-1) & \quad (k-1)\text{-stupňů volnosti} \\ \chi^2_{0,95}(5) &= 11,07 \end{aligned}$$

$$1,897 < 11,07 \Rightarrow H_0 \text{ nezamítáme.}$$

15 Literatura

- [1] *Prof. Ing. Mirko Navara, DrSc.: Pravděpodobnost a matematická statistika.*
Vydavatelství ČVUT, Praha 2007 (první vydání)
- [2] *Doc. Vladimír Rogalewicz, CSc.: Pravděpodobnost a statistika pro inženýry.*
Vydavatelství ČVUT, Praha 1998
- [3] *Stránky předmětu Matematika pro výpočetní techniku*
<http://cmp.felk.cvut.cz/~navara/MVT>