Matematika 4B

Prof. RNDr. Jan Hamhalter, CSc.

katedra matematiky FEL ČVUT

e-mail: hamhalte@math.feld.cvut.cz

tel: 224353587

**web:
http://math.feld.cvut.cz//hamhalte**

11. ledna 2007

16:56

- V.Rogalewicz: Pravděpodobnost a statistika pro inženýry, skripta, Vydavatelství ČVUT, 1998.
- K.Zvára a J.Štěpán: Pravděpodobnost a matematická statistika, matfyzpress, Praha 2002.
- J.Anděl: Matematika náhody, matfyzpress, Praha 2003.
- J.Anděl: Statistické metody, matfyzpress, Praha 2003.
- V.Dupač a M.Hušková: Pravděpodobnost a matematická statistika, Nakladatelství Karolinum, 1999.
- Z.Prášková: Základy náhodných procesů II, Nakladatelství Karolinum, 2004.
- A. Rényi: Teorie pravděpodobnosti, Academia, Praha 1972.

1 Historie a podstata teorie pravděpodobnosti

teorie pravděpodobnosti = matematika náhody, systémy s nedostatkem informace

svět 19. století – deterministický systém, hodinový stroj
svět 20. století – svět náhody (evoluce není možná bez náhody, mikrosvět se řídí pravděpodobnostními zákony, teorie chaosu, apod.)

• *Úloha o rozdělení sázky*
Pochází od Arabů. Nedávno objevena v rukopise z r. 1380.

Dva hráči hrají sérii partií. Výsledky jednotlivých her jsou nezávislé. Vyhrává ten kdo poprvé zvítězí v šesti partiích. Pravděpodobnost výhry je pro každého hráče stejná, t.j. $1/2$. Hra je přerušena ve chvíli kdy hráč A vyhrál 5x a hráč B 3x. Jak si rozdělí výhru?

Úloha byla vyřešena nezávisle Pascalem a Fermatem (1654).

Všechny možnosti pokračování (hra bude trvat nejvýše tři další partie):

$$\begin{array}{llll} AAA & AAB & ABA & ABB \\ BAA & BAB & BBA & \boxed{BBB} \end{array}$$

Pouze v jednom případě vítězí B, pravděpodobnost výhry hráče B je 1:8, výhra by se měla rozdělit v poměru 7:1.

• Huygens (1657) : On Reasoning in Games of Dice

• Laplace (1812): Analytic Theory of Probabilities

nestačí kombinatorické metody, je třeba uvažovat nekonečné soubory možností

statistická fyzika, Brownův pohyb, teorie míry a integrace

• A. Kolmogorov (1930): Axiomatické základy teorie pravděpodobnosti

• současný stav a perspektivy: nové obory založené na pravděpodobnostním přístupu – kvantová teorie informace, teorie her v ekonomii, teorie chaosu, ...

2 Pravděpodobnostní prostor

pravděpodobnostní model má dvě komponenty:

- struktura náhodných jevů
- pravděpodobnost jako kvantitativní funkce na jevech

2.1. Příklad. *Střelba na terč*
Ω = kruh o poloměru r
náhodné jevy = podmnožiny Ω

$$\text{pravděpodobnost}(A) = \frac{\text{obsah}(A)}{\pi r^2}\,.$$

2.2. Příklad. *Sportka*

$$\Omega = \{\{1,2,3,4,5,6\},\{1,3,4,6,2,17\},\dots\} = \\ \{ \text{ šestiprvkové podmnožiny množiny } \{1,2,\dots,49\}\}$$

tyto šestice tvoří elementární jevy s pravděpodobností

$$\frac{1}{\binom{49}{6}} = \frac{1}{13983816} = 0,7151138242 \cdot 10^{-7}$$

jev= podmnožina Ω

pravděpodobnost jevu $A \subset \Omega$.

$$P(A) = \frac{\text{velikost}(A)}{\binom{49}{6}}.$$

konkrétní výpočet v tomto modelu – spočtěte pravděpodobnost, že uhodnete (právě) tři čísla.

$$P(A) = \frac{\binom{6}{3} \cdot \binom{43}{3}}{\binom{49}{6}} = 0,1765040387$$

náhodné jevy musíme umět kombinovat
„jev A nebo jev B“, ...

2.3. Definice. Nechť Ω je neprázdná množina. Systém $\mathcal{A}$ podmnožin množiny Ω se nazývá *σ-algebra náhodných jevů*, jestliže platí

(i) $\Omega \in \mathcal{A}$.

(ii) Jestliže $A_1, A_2, \ldots$ jsou množiny v $\mathcal{A}$, pak $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

(iii) Je-li $A \in \mathcal{A}$, pak $A^c = \Omega \setminus A \in \mathcal{A}$

Terminologie:
$A^c \ldots$ *opačný jev* k jevu A
A, B jsou *navzájem vylučující se (disjunktní) jevy* jestliže $A \cap B = \emptyset$

2.4. Tvrzení. *Je-li $\mathcal{A}$ σ-algebra podmnožin Ω pak*

(i) $A_1, A_2, \ldots \in \mathcal{A} \Longrightarrow \bigcap_{i=1}^{\infty} A_i \in A.$

(ii) $A, B \in \mathcal{A} \Longrightarrow A \cap B^c \in \mathcal{A}.$

Důkaz:
(i) $A_1^c, A_2^c, \ldots \in \mathcal{A}, \Longrightarrow \bigcup_{i=1}^{\infty} A_i^c \in \mathcal{A} \Longrightarrow$
(de Morganova pravidla)

$$\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}\,.$$

(ii) $A, B^c \in \mathcal{A} \Longrightarrow A \cap B^c \in \mathcal{A}\,.$

pravděpodobnost modeluje relativní četnost, měla by respektovat stejná pravidla jako počet prvků množiny

2.5. Definice. Předpokládejme, že $\mathcal{A}$ je σ-algebra podmnožin množiny Ω. *Pravděpodobnost* P je zobrazení

$$P : \mathcal{A} \to [0,1]\,,$$

pro které platí

(i) $P(\Omega) = 1$

(ii) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$,
jestliže $A_1, A_2, ...$ jsou navzájem disjunktní množiny v $\mathcal{A}$.

Trojice $(\Omega, \mathcal{A}, P)$ se nazývá *pravděpodobnostní prostor*.

Základní vlastnosti pravděpodobnosti

(i) $A \cap B = \emptyset \Longrightarrow P(A \cup B) = P(A) + P(B)$

(ii) $P(\emptyset) = 0$
$(\Longleftarrow P(\Omega) + P(\emptyset) = P(\Omega))$

(iii) $P(A^c) = 1 - P(A)$
$(\Longleftarrow P(A) + P(A^c) = P(\Omega) = 1.)$

(iv) $A \subset B \Longrightarrow P(B \cap A^c) = P(B) - P(A)$
$(\Longleftarrow P(B) = P(A) + P(B \cap A^c))$

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

odvození: $X = A \cap (A \cap B)^c\,, \qquad Y = B \cap (A \cap B)^c$

$$P(X) + P(Y) + P(A \cap B) = P(A \cup B)$$
$$P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) = P(A \cup B)$$
$$P(A) + P(B) = P(A \cup B) + P(A \cap B)$$

Tato základní pravidla jsou často užitečná při konkrétních výpočtech.

2.6. Příklad. Určete pravděpodobnost, že při tahu Sportky bude vylosováno buďto číslo 7 nebo číslo 20.

Řešení: $A \ldots$ taženo číslo 7, $B \ldots$ taženo číslo 20.

$$P(A) = P(B) = \frac{\binom{48}{5}}{\binom{49}{6}}$$

$$P(A \cap B) = \frac{\binom{47}{4}}{\binom{49}{6}}$$

Tedy

$$P(A \cup B) = 2 \cdot \frac{\binom{48}{5}}{\binom{49}{6}} - \frac{\binom{47}{4}}{\binom{49}{6}} = \frac{13}{56} = 0,232148571\,.$$

Důležité typy pravděpodobnostních prostorů:

- klasický pravděpodobnostní prostor
- konečný pravděpodobnostní prostor
- diskrétní nekonečný pravděpodobnostní prostor
- geometrický pravděpodobnostní prostor

$$\Omega = \{\omega_1, \dots, \omega_n\}$$
$$\mathcal{A} = \text{všechny podmnožiny množiny} \quad \Omega$$
$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n}\,.$$

V tomto modelu mají elementární jevy stejnou šanci $= \frac{1}{n}$.

Někdy je těžké nalézt dobrý model slovní úlohy, často se setkáme se složitou kombinatorikou.

2.7. Příklad. Hodíme n krát mincí, rub i líc v jednom hodu mají stejnou šanci, tj. $\frac{1}{2}$. Jaká je pravděpodobnost že padne právě k krát líc?

Řešení: elementární jevy – posloupnosti nul a jedniček délky n kódující výsledky hodů.

$$|\Omega| = 2^n$$

$$\Omega = \{\omega_1, \dots, \omega_{2^n}\}$$

A_k .... posloupnost obsahuje právě k jedniček.

$$|A_k| = \binom{n}{k}$$

Tedy

$$P(A_k) = \frac{1}{2^n}\binom{n}{k}.$$

Poznámka: Hodíme n krát mincí, kde n je sudé. Jaká je pravděpodobnost, že padne stejný počet nul jako jedniček?

$$P(A_{n/2}) = \frac{1}{2^n}\binom{n}{n/2}$$

Pomocí tzv. Stirlingova vzorce lze dokázat, že

$$P(A_{n/2}) \approx \frac{1}{\sqrt{\pi n/2}} \to 0 \text{ pro } n \to \infty\,.$$

2.8. Příklad. Narozeninový problém
Jaká je pravděpodobnost, že ve třídě s n žáky se najde dvojice mající narozeniny ve stejný den? ($n \leq 365$).

Řešení:

$$\Omega = \{\text{posloupnosti délky } n \\ \text{prvků množiny } \{1, 2, \dots, 365\}\}$$

Elementární jevy kódují den narozenin 1. až n-tého žáka.

A ... sledovaný jev
A^c ... jev opačný, všichni mají narozeniny v jiný den.

$$|\Omega| = 365^n$$

$$|A^c| = 365 \cdot 364 \cdot 363 \cdots (365 - n + 1)$$

$$P(A) = 1 - \frac{365 \cdot 364 \cdot 363 \cdots (365 - n + 1)}{365^n} = \\ = 1 - \prod_{j=1}^{n-1} \left(1 - \frac{j}{365}\right)$$

Nečekané numerické hodnoty:
již pro $n = 23$ je $P(A) > 1/2$,
pro $n = 56$ je $P(A) = 0,99$.

$$\begin{aligned}\Omega &= \{\omega_1, \dots, \omega_n\} \\ \mathcal{A} &= \text{ všechny podmnožiny } \Omega \\ p_1, \dots, p_n &> 0 \dots \text{ váhy} \\ \sum_{i=1}^{n} p_i &= 1 \\ P(\{\omega_i\}) &= p_i\end{aligned}$$

• Z toho vyplývá že

$$P(A) = \sum_{\{i | \omega_i \in A\}} p_i \,,$$

pro všechny $A \subset \Omega$.

• Klasický pravděpodobnostní prostor je speciálním případem, ve kterém jsou všechny váhy stejné:

$$p_1 = p_2 = \cdots = p_n = \frac{1}{n} \,.$$

2.9. Příklad. $\Omega = \{\omega_1, \omega_2, \omega_3\}$

$$p_1 = P(\{\omega_1\}) = \frac{1}{2}$$
$$p_2 = P(\{\omega_2\}) = \frac{1}{4}$$
$$p_3 = P(\{\omega_3\}) = \frac{1}{4}$$

ω_1	ω_2	ω_3
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

Je tedy např.

$$P(\{\omega_1, \omega_2\}) = p_1 + p_2 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}\,.$$

Bernoulliovo schéma

Máme jev A (zdar) s pravděpodobností $0 < p < 1$
a jev B (nezdar) s pravděpodobností $0 < 1 - p < 1$.
V náhodném pokusu nastane právě jeden z jevů A a B s příslušnou pravděpodobností. Provedeme sérii n těchto náhodných pokusů, jejichž výsledky se navzájem neovlivňují.

Možné výstupy pro $n = 4$: $ABAA$, $BBBA$, $\ldots$
kódovány posloupnostmi 0 a 1: 1011, 0001, $\ldots$

Elementární jevy —- posloupnosti nul a jedniček délky n

Nezávislost znamená, že pravděpodobnosti se násobí:

$$P(1011) = p \cdot (1-p) \cdot p \cdot p = p^3 \cdot (1-p)$$
$$P(0001) = (1-p) \cdot (1-p) \cdot (1-p) \cdot p = p \cdot (1-p)^3$$

To nás vede k následujícímu modelu:

$\Omega=$ všechny posloupnosti nul a jedniček délky n

$|\Omega| = 2^n$.

$P(\text{posloupnost }) = p^{\text{počet 1}} \cdot (1-p)^{\text{počet 0}}$
Ověříme, že součet vah je 1:

$$\sum_{i=1}^{2^n} p_i = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} =$$

$$= (p + (1-p))^n = 1.$$

Důležitý je jev, A_k, že v sérii n pokusů nastane jev A právě k krát.

$|A_k| = \binom{n}{k}$.

$$\boxed{P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}.}$$

Konkrétní příklady: hod mincí, hod kostkou, ankety, statistické šetření, apod.

2.10. Příklad. Terč zasáhneme s pravděpodobností 1/3. Jaká je pravděpodobnost, že se dvakrát strefíme při čtyřech pokusech.

$$P(A_2) = \binom{4}{2} \frac{1}{3^2} \left(\frac{2}{3}\right)^2 = \frac{48}{161} = 0,2963\,.$$

$$\Omega = \{\omega_1, \omega_2, \dots\}$$
$$\mathcal{A} = \text{ všechny podmnožiny } \Omega$$
$$(p_n)_{n=1}^{\infty} \dots \text{ posloupnost vah}$$
$$\sum_{n=1}^{\infty} p_n = 1, p_n \geq 0$$
$$P(\{\omega_n\}) = p_n \text{ pro } n = 1, 2, \dots$$

- Z toho vyplývá, že

$$P(A) = \sum_{\{n | \omega_n \in A\}} p_n \,,$$

pro všechny $A \subset \Omega$.

Poissonův zákon

$\Omega = \{\omega_0, \omega_1, \omega_2 \ldots\}$
$\lambda > 0$ parametr

$$\boxed{p_n = \frac{\lambda^n}{n!}\, e^{-\lambda}}\,, n = 0, 1, \ldots$$

Ověříme korektnost zadání:

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}\, e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda}\, e^{\lambda} = 1\,.$$

2.11. Příklad. Za danou časovou jednotku volá na ústřednu průměrně $\lambda > 0$ účastníků. Pravděpodobnost p_n, že zavolá právě n účastníků se řídí Poissonovým zákonem:

$$p_n = \frac{\lambda^n}{n!}\, e^{-\lambda}$$

Pravděpodobnost, že zavolá alespoň někdo je $1 - e^{-\lambda}$.

pravděpodobnost je dána geometrickou kvantitou (délka, obsah, objem)

$$\Omega \subset \mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \dots ,$$

$$0 < \text{ velikost } (\Omega) < \infty ,$$

$$P(A) = \frac{\text{velikost } (A)}{\text{velikost } (\Omega)}$$

pro $A \subset \Omega$.

2.12. Příklad. Terč má poloměr 30 cm. Jaká je pravděpodobnost, že se trefíme do středu o poloměru 5cm ?

Řešení:

$$p = \frac{25\pi}{900\pi} = 0,02777\dots .$$

2.13. Příklad. Buffonova úloha
V rovině je dán systém rovnoběžek majících vzdálenost d. Na rovinu hodíme jehlu o velikosti l, $l < d$. Jaká je pravděpodobnost, že protne některou rovnoběžku?

Řešení: Polohu jehly vůči rovnoběžné síti popíšeme dvěma parametry:

$x \ldots$ vzdálenost středu jehly od nejbližší rovnoběžky

$$x \in < 0, \frac{d}{2} >$$

$\varphi \ldots$ úhel, který jehla svírá s rovnoběžnou sítí

$$\varphi \in < 0, \pi > \ .$$

Podmínka protnutí:

$$\frac{l}{2} \sin \varphi > x$$

$$P = \frac{1}{\pi \, d/2} \int\limits_0^\pi \frac{l}{2} \sin \varphi d\varphi = \frac{l}{\pi d} \Big[-\cos \varphi \Big]_0^\pi =$$
$$= \frac{2l}{\pi d} \, .$$

Další vlastnosti pravděpodobnosti:

Princip inkluze a exkluze

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

Mějme nyní tři jevy $A, B, C \in \mathcal{A}$.

$$
\begin{aligned}
&P(A \cup B \cup C) = P[(A \cup B) \cup C] = \\
&\quad P(A \cup B) + P(C) - P((A \cap C) \cup (B \cap C)) \\
&\quad = P(A) + P(B) - P(A \cap B) + P(C) \\
&\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\
&\quad = P(A) + P(B) + P(C) \\
&\quad - P(A \cap C) - P(B \cap C) - P(A \cap B) \\
&\qquad\qquad\qquad\qquad\qquad + P(A \cap B \cap C)
\end{aligned}
$$

Zobecnění se dá dokázat indukcí:

2.14. Věta. Princip inkluze a exkluze
Předpokládejme, že $A_1, \dots, A_n \in \mathcal{A}$, kde $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Pak platí

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{1 \le i \le n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j)$$
$$+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(\cap_{i=1}^{n} A_i)\,.$$

2.15. Příklad. Roztržitá šatnářka
n hostů restaurace si přichází odložit svůj kabát. Šatnářka vydává kabáty chaoticky. Jaká je pravděpodobnost, že alespoň jeden z hostů dostane svůj kabát?

Řešení: Elementární jevy jsou permutace n prvkové množiny.

$$\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix}$$

Všechny mají stejnou pravděpodobnost, tj. $\frac{1}{n!}$.

$$A = \{(k_1, \dots, k_n) \mid \text{ existuje } 1 \le i \le n \text{ tak že } k_i = i\}\,.$$

$$A = A_1 \cup A_2 \cup \dots \cup A_n\,,$$

kde

$$A_i = \{(k_1, \dots, k_n) \mid k_i = i\}$$

(i-tý host je v pořádku)

$$\begin{aligned} P(A_i) &= \frac{(n-1)!}{n!} \\ P(A_i \cap A_j) &= \frac{(n-2)!}{n!}\,, i \neq j \\ P(A_1 \cap A_2 \cap \dots \cap A_n) &= \frac{1}{n!} \end{aligned}$$

$$P(A) = 1 - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \cdots + (-1)^{n+1}\frac{1}{n!}$$
$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!}$$

numerické hodnoty:

n	1	2	3	4	5	6	7
$P(A)$	1	0,5	0,6667	0,625	0,6333	0,6319	0,6321

asymptoticky:

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots$$

$$\lim_{n\to\infty} P(A) = 1 - e^{-1} = 0,6321....$$

Zacházení s nekonečnými posloupnostmi jevů, spojitost pravděpodobnosti:

2.16. Věta. *Předpokládejme, že* $(\Omega, \mathcal{A}, P)$ *je pravděpodobnostní prostor.*

(i) *Je-li* $A_1 \subset A_2 \subset \cdots$ *pro* $A_1, A_2, \cdots$ *z* $\mathcal{A}$, *pak*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} P(A_i)$$

(ii) *Je-li* $A_1 \supset A_2 \supset \cdots$ *pro* $A_1, A_2, \cdots$ *z* $\mathcal{A}$, *pak*

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} P(A_i)$$

Důkaz: (A_n) splňuje (i)

$$A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots \cup (A_n \setminus A_{n-1})$$

je disjunktní sjednocení. Pak

$$P(A_n) = P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \\ + \cdots + P(A_n \setminus A_{n-1})\,.$$

Dále platí

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \overbrace{P(A_1) + P(A_2 \setminus A_1) + \cdots + P(A_n \setminus A_{n-1})}^{P(A_n)} + \cdots =$$

$$= \lim_{n\to\infty} P(A_n)$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

(ii) obdobným způsobem, nebo z (i) přechodem k množinovému komplementu.

3 Nezávislé jevy a podmíněná pravděpodobnost

Pravděpodobnostní model se mění dostaneme-li částečnou informaci o systému. Víme, že nastal jev B. Pak pravděpodobnost, že nastane jev A je

$$\frac{P(A \cap B)}{P(B)}\,.$$

3.1. Definice. Je dán pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a $B \in \mathcal{A}$ s $P(B) > 0$. *Podmíněná pravděpodobnost náhodného jevu* A za podmínky B je definována jako

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

Tedy

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

3.2. Příklad. Skříňka má tři zásuvky. V první jsou dvě zlaté mince, ve druhé zlatá a stříbrná mince a ve třetí dvě stříbrné mince.

$$\boxed{\bullet z_1 \qquad \bullet\ z_2}$$

$$\boxed{\bullet z_3 \qquad \circ\ s_1}$$

$$\boxed{\circ s_2 \qquad \circ\ s_3}$$

Náhodně jsme vybrali zásuvku a náhodně z ní vytáhli minci. Tažená mince je stříbrná. Jaká je pravděpodobnost, že druhá mince ve vytažené zásuvce je zlatá?
Řešení: naivní odpověď 1/2 není správná.
Dvě fáze náhodného procesu:
1. volba zásuvky 2. volba mince

$$\Omega = \{(1, z_1), (1, z_2), (2, z_3), (2, s_1), (3, s_2), (3, s_3), \}$$

Všechny tyto jevy mají stejnou šanci, tj. 1/6.

Z ... v otevřené zásuvce je zlatá mince
S ... vyjmuli jsme stříbrnou minci (tento jev nastal).
Hledáme $p = P(Z|S)$

$$P(S) = P\{(2, s_1), (3, s_2), (3, s_3)\} = \frac{3}{6} = \frac{1}{2}\,.$$

$$P(Z \cap S) = P\{(2, s_1)\} = 1/6\,.$$

$$p = \frac{1/6}{3/6} = \frac{1}{3}\,.$$

Formální vlastnosti podmíněné pravděpodobnosti:

(i) $P(B|B) = 1$

(ii) $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$
jsou-li $A_1, A_2 \in \mathcal{A}$ disjunktní.

3.3. Tvrzení. *Je-li* $(\Omega, \mathcal{A}, P)$ *pravděpodobnostní prostor a* $B \in \mathcal{A}$ *s* $P(B) > 0$, *pak* $(\Omega, \mathcal{A}, P((\cdot|B))$ *je také pravděpodobnostní prostor.*

V pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P((\cdot|B))$ mají jevy disjunktní s B nulovou pravděpodobnost, podmnožiny v B mají pravděpodobnost normovanou pravděpodobností jevu B.

Nezávislé jevy jsou jevy jejichž podmíněné pravděpodobnosti se neovlivňují:

$P(A), P(B) > 0$

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

3.4. Definice. Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Jevy A a $B \in \mathcal{A}$ nazýváme *nezávislé,* jestliže

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}\,.$$

3.5. Příklad. Dvakrát hodíme mincí. Všechny výsledky jsou stejně pravděpodobné. Ukažte, že výsledky v prvním a druhém hodu jsou nezávislé.

Řešení: $R\ldots$ rub, $L\ \ldots$ líc

$$\Omega = \{RL, RR, LR, LL\}$$

$$P(\{RL, RR\}) = \frac{1}{2}$$
$$P(\{RL, LL\}) = \frac{1}{2}$$
$$P(\{RL\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}\,.$$

Obecnější definice:

3.6. Definice. Jevy $A_1, \dots, A_n$ v pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ jsou *nezávislé*, jestliže

$$\boxed{P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})}$$

pro všechna $i_1 < i_2 < \cdots < i_k$, $k \leq n$.

Bernoulliovo schéma (revisited)

Výsledky pokusů v Bernoulliově schématu jsou nezávislé jevy. Bernoulliovo schéma tedy můžeme chápat jako sérii nezávislých pokusů se dvěma možnými výsledky, které mají doplňkovou pravděpodobnost.

3.7. Příklad. Elektrický obvod znázorněný na obrázku je náhodně přerušován pěti nezávislými spínači. V jedné větvi jsou tři spínače a ve druhé dva. Jaká je pravděpodobnost že obvodem prochází proud? Každý spínač je přerušen s pravděpodobností 1/2.

Řešení: A_i … i-tý vypínač je sepnut

$$
\begin{aligned}
p = P[(A_1 \cap A_2 \cap A_3) \cup (A_4 \cap A_5)] =\\
= P(A_1 \cap A_2 \cap A_3) + P(A_4 \cap A_5)\\
- P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) =\\
= \frac{1}{2^3} + \frac{1}{2^2} - \frac{1}{2^5} = \frac{11}{2^5} = 0,34375\,.
\end{aligned}
$$

3.8. Tvrzení. *Jsou-li jevy $A_1, \dots, A_n$ v pravděpodobnostním prostoru nezávislé, pak jsou nezávislé i jevy $A_1^c, A_2, \dots, A_n$.*

Důkaz:

$$
\begin{aligned}
P(A_1^c \cap A_2 \cap \dots \cap A_n) = P(A_2 \cap A_3 \cap \dots \cap A_n) -\\
- P(A_1 \cap A_2 \cap \dots \cap A_n) =\\
P(A_2)P(A_3) \cdots P(A_n) - P(A_1)P(A_2) \cdots P(A_n) =\\
= (1 - P(A_1))P(A_2) \cdots P(A_n) =\\
= P(A_1^c)P(A_2)P(A_3) \cdots P(A_n)\,.
\end{aligned}
$$

Důsledek: Nahradíme-li v nezávislém systému jevů některé jevy jejich opakem, dostaneme opět nezávislý systém.

Situace: $A_1, \dots, A_n$ disjunktní jevy, takové že

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1$$

a $P(A_i) > 0$ pro všechna i. Pak

$$P((A_1 \cup A_2 \cup \cdots \cup A_n)^c) = 0\,.$$

Pro každé $B \in \mathcal{A}$ máme

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + \\ &\cdots + P(B \cap A_n) + 0 = \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)\,. \end{aligned}$$

3.9. Definice. Posloupnost $A_1, \dots, A_n$ disjunktních jevů v pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ se nazývá *úplný systém jevů* jestliže $A_1, \dots, A_n$ jsou disjunktní, $P(A_i) > 0$ pro všechna i a

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1\,.$$

3.10. Věta. (Věta o úplné pravděpodobnosti) *Předpokládejme že $A_1, \dots, A_n$ je úplný systém jevů v pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ takový, že $P(A_i) > 0$ pro všechna $i = 1, \dots n$. Pro každé $B \in \mathcal{A}$ platí*

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \\ + \cdots + P(B|A_n)P(A_n)$$

„$P(B)$ je kombinace pravděpodobností $P(A_1), \dots, P(A_n)$ s váhami danými podmíněnými pravděpodobnostmi"

3.11. Příklad. V urně č.1 je 50 černých a 60 bílých kuliček. V urně č.2 je 60 černých a 50 bílých kuliček. Hodíme si hrací kostkou. Padne-li šestka vybereme urnu č. 1. V opačném případě urnu č.2. Z vybrané urny vybereme náhodně kuličku. Jaká je pravděpodobnost, že je bílá?

Řešení:

$$\boxed{50\bullet \quad 60\circ} \qquad \boxed{60\bullet \quad 50\circ} \tag{1}$$

A_1...padne šestka $\qquad$ A_2...nepadne šestka

$P(A_1) = 1/6$ $\qquad$ $P(A_2) = 5/6$

B ... vytažená kulička je bílá

$$P(B|A_1) = \frac{60}{110} \qquad P(B|A_2) = \frac{50}{110}\,.$$

$$p = \frac{60}{110}\cdot\frac{1}{6} + \frac{50}{110}\cdot\frac{5}{6} = \frac{1}{11} + \frac{25}{66} = \frac{31}{66} = \\ = 0,4697.$$

Bayesův vzorec
Bayes (1761) ... stejné apriorní pravděpodobnosti
Laplace (1774) ... obecný případ

věta o úplné pravděpodobnosti:
$P(A_i), P(B|A_i) \to P(B)$
nyní určíme $P(A_i|B)$:

3.12. Věta. Bayesův vzorec
Je-li $A_1, \dots, A_n$ úplný systém jevů v pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ a $B \in \mathcal{A}$ s $P(B) > 0$ pak

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

pro všechna $j = 1, 2, \dots, n$.

Důkaz:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

---

vstup:
$P(A_1), P(A_2), \dots P(A_n)$ ... apriorní pravděpodobnosti
$P(B|A_1), P(B|A_2), \dots P(B|A_n)$ ... podmíněné pravděpodobnosti
B ... přináší novou informaci o stavu systému

výstup: $P(A_1|B), P(A_2|B), \dots P(A_n|B)$ —— upřesněná informace

3.13. Příklad. Máme dvě krabice s bílými a černými kuličkami. V krabici č. 1 je jedna bílá a devět černých kuliček. V krabici č.2 je jedna černá a devět bílých kuliček.

$\boxed{1 \circ \quad 9 \bullet}$	č.1
$\boxed{1 \bullet \quad 9 \circ}$	č.2

Za plentou byla vylosována jedna z krabic. Náhodně jsme z ní vytáhli jednu kuličku. Byla bíla. Jaká je pravděpodobnost, že máme před sebou krabici č.1. ?

$A_1 \ldots$ první krabice $\qquad P(A_1) = \frac{1}{2}$

$A_2 \ldots$ druhá krabice $\qquad P(A_2) = \frac{1}{2}$

$B \ldots$ tažena bílá kulička $\qquad P(B|A_1) = \frac{1}{10}$

$$P(B|A_2) = \frac{9}{10}\,.$$

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$
$$= \frac{\frac{1}{10}\cdot\frac{1}{2}}{\frac{1}{10}\cdot\frac{1}{2}+\frac{9}{10}\cdot\frac{1}{2}} = \frac{\frac{1}{10}\cdot\frac{1}{2}}{\frac{1}{2}} = \frac{1}{10}\,.$$

3.14. Příklad. Diagnóza nemoci

Senzitivita testu: 0,95 (tj. má-li osoba AIDS je test pozitivní v 95% případů)

specificita testu: 0,95 (tj. nemá-li osoba AIDS je test negativní v 95% případů)

prevalence nemoci: 0,005 (tj. 0,5% populace je nakaženo)

Jaká je pravděpodobnost, že osoba s pozitivním testem je nakažena virem HIV?

(Naivní odpověď 0,95 je úplně mimo.)

AIDS	NE-AIDS
0,005	0,995

$+$, - .... výsledky testu

$$P(+|AIDS) = 0,95 \quad P(+|NEAIDS) = 0,05$$
$$P(-|AIDS) = 0,05 \quad P(-|NEAIDS) = 0,95$$

$$\begin{aligned} P(AIDS|+) &= \\ &= \frac{P(+|AIDS)P(AIDS)}{P(+|AIDS)P(AIDS) + P(+|NEAIDS)P(NEAIDS)} \\ &= \frac{0,95 \cdot 0,005}{0,95 \cdot 0,005 + 0,05 \cdot 0,995} = 0,087\,. \end{aligned}$$

apriorní pravděpodobnosti $\rightarrow$ aposteriorní pravděpodobnosti
$(0,005, 0,995) \rightarrow (0,087, 0,913)\,.$

Test opakujeme znovu. Testovaná osoba je opět pozitivní.
Jaká je nyní pravděpodobnost že má AIDS ?
Opakujeme postup s

$$P(AIDS) = 0,087 \qquad P(NEAIDS) = 0,913\,.$$

numerické výsledky:
i ... počet pozitivních testů,
P_i ... pravděpodobnost, že daná osoba má AIDS.

i	0	1	2	3	4	5
P_i	0,005	0,087	0,645	0,972	0,998	0,9992

4 Náhodná veličina

• Zajímá nás pouze sledovaná numerická veličina, nikoliv celý pravděpodobnostní prostor: počet zákazníků, cena akcie, hodnota měření napětí, ...

• Podstatné je stanovit pravděpodobnost, že náhodná veličina má hodnoty v daném rozmezí.

Značení:
I ... interval na reálné ose, zahrnujeme i jednobodové množiny.

$X : \Omega \to \mathbb{R}$ ... funkce definovaná na množině Ω.

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

4.1. Definice. Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Funkce X definovaná na Ω se nazývá náhodná veličina jestliže

$$[X \in I] \in \mathcal{A}$$

pro všechny intervaly $I \subset \mathbb{R}$.

Všechny funkce na konečném nebo diskrétním pravděpodobnostním prostoru jsou náhodné veličiny.

Náhodné veličiny popisujeme kvantitativně pomocí jejich distribučních funkcí:

$$P[X \leq x] = P(\{\omega \mid X(\omega) \leq x\}) .$$

4.2. Definice. Předpokládejme, že X je náhodná veličina na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. *Distribuční funkce*, F_X, náhodné veličiny X je funkce

$$F_X(x) = P[X \leq x], \qquad x \in \mathbb{R} .$$

4.3. Příklad. X ... počet ok při hodu kostkou
X nabývá šesti hodnot, 1,2,3,4,5,6; distribuční funkce je po částech spojitá funkce.

4.4. Příklad. X ... poloha ručičky hodinek při náhodném zastavení:

$$F(x) = \begin{cases} \frac{x}{2\pi} & x \in < 0, 2\pi > \\ 0 & x \leq 0 \\ 1 & x \geq 2\pi \,. \end{cases}$$

4.5. Příklad. Vlak projíždí přejezdem jedenkrát za hodinu, Závory jsou staženy na dvanáct minut. Náhodná veličina X je doba čekání.

$$P[X = 0] = \frac{60 - 12}{60} = \frac{48}{60} = 0,8\,.$$

Pro $x \in (0, 12 >$ máme:

$$P[X \leq x] = P[X = 0] + P[0 < X \leq x] = 0,8 + \frac{x}{60}\,.$$

Tedy

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0,8 & x = 0 \\ 0,8 + \frac{x}{60} & x \in (0, 12 > \\ 1 & x \geq 12 \end{cases}$$

4.6. Věta. *Distribuční funkce F_X náhodné veličiny X splňuje následující podmínky:*

(i) $0 \leq F_X \leq 1$

(ii) F_X *je neklesající*

(iii) F_X *je zprava spojitá*

(iv) $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.

Dukaz: (ii) Je-li $x \leq y$, pak $[X \leq x] \subset [X \leq y]$, a tedy

$$F_X(x) \leq F_X(y).$$

(iii) Volme (δ_n) klesající posloupnost kladných čísel s nulovou limitou a $a \in \mathbb{R}$. Uvažujme množiny

$$A_n = [X \leq a + \delta_n].$$

Platí

$$\bigcap_{n=1}^{\infty} A_n = [X \leq a].$$

Dle spojitosti pravděpodobnosti Věta 2.16 platí

$$P[X \leq a] = \lim_{n\to\infty} P(A_n).$$

Jinými slovy

$$F_X(a) = \lim_{n\to\infty} F_X(a + \delta_n),$$

a proto

$$F_X(a) = \lim_{x\to a+} F_X(x).$$

(iv)

$$A_n = [X \leq -n]\,, \qquad n = 1, 2 \dots$$

Pak A_n je klesající posloupnost množin s prázdným průnikem. Dle Věty 2.16 máme

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} P(A_n) = P(\emptyset) = 0\,.$$

F_X je neklesající, a proto musí mít v $-\infty$ limitu nula. (Druhá limita podobně.)

V distribuční funkci F_X jsou všechny relevantní informace o náhodné veličině X:

- $$P[X > a] = 1 - P[X \leq a] = 1 - F_X(a)\,.$$
- $$\begin{aligned} P[X \in (a, b>] &= P[(X \leq b) \wedge (X \leq a)^c] = \\ &= P[X \leq b] - P[X \leq a] = F_X(b) - F_X(a)\,. \end{aligned}$$
- $$P[X < a] = \lim_{n\to\infty} P\left[X \leq a - \frac{1}{n}\right] = \lim_{n\to\infty} F_X\left(a - \frac{1}{n}\right) = \lim_{x\to a-} F_X(x)\,.$$
- $$P[X = a] + P[X < a] = P[X \leq a]\,.$$

Odtud

$$P[X = a] = F_X(a) - \lim_{x\to a-} F_X(x)\,.$$

(Velikost případného skoku.)

4.7. Věta. *Ke každé zprava spojité, neklesající funkci $F(x)$ na $\mathbb{R}$, s limitami $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, existuje náhodná veličina X tak, že*

$$F_X = F\,.$$

Důkaz je mimo naše možnosti – teorie míry.

distribuční funkce (náhodná rozdělení) = náhodné veličiny.

Typy náhodných veličin:

- diskrétní rozdělení
- spojité rozdělení
- smíšené rozdělení

4.8. Definice. Náhodná veličina X se nazývá diskrétní, jestliže existuje konečná nebo nekonečná posloupnost (x_n) taková, že

$$\sum_n P[X = x_n] = 1\,.$$

Daná tabulkou resp. **pravděpodobnostní funkcí**:

x_1	x_2	x_3	$\cdots$
p_1	p_2	p_3	$\cdots$

$$P[X = x_i] = p_i\,, \qquad (x_n)_n \ldots \text{ uzly}$$

4.9. Příklad. X … počet ok při hodu hrací kostkou

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

4.10. Příklad. X … doba kdy poprvé padne líc při sérii hodů symetrickou mincí.

1	2	3	$\cdots$
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{3}$	$\cdots$

$$P[X = n] = \frac{1}{2^n}\,.$$

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

4.11. Tvrzení. *Má-li náhodná veličina X diskrétní rozdělení s pravděpodobnostní funkcí $(x_n, p_n)_n$, pak pro $M \subset \mathbb{R}$ platí*

$$P[X \in M] = \sum_{\{n \mid x_n \in M\}} p_n \,.$$

4.12. Příklad. Jaká je pravděpodobnost, že při házení symetrickou mincí padne líc poprvé po sudém počtu hodů ?

X z Příkladu 4.10.

$$P[X = \text{sudé}] = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \cdots = \frac{1}{4}\frac{1}{1-\frac{1}{4}} = \frac{1}{3}\,.$$

4.13. Definice. Náhodná veličina X se nazývá *spojitá*, jestliže existuje nezáporná funkce f taková, že

$$F_X(x) = \int\limits_{-\infty}^{x} f(t)\,dt\,.$$

Funkce f se přitom nazývá hustotou náhodné veličiny X.

- Funkce f je hustotou náhodné veličiny právě tehdy když je nezáporná a $\int_{-\infty}^{\infty} f(t)\,dt = 1$.
- Je-li x bod spojitosti hustoty f, pak $f(x) = F_X(x)'$.

Je-li f hustota náhodné veličiny X, pak

$$P[a \leq X \leq b] = \int\limits_{a}^{b} f(x)\,dx\,.$$

Hustota dává preference hodnotám.

4.14. Příklad. $f(x) = \begin{cases} 1 & x \in < 0, 1 > \\ 0 & \text{jinak} \end{cases}$

(Rovnoměrné rozdělení)
Pro $x \in < 0, 1 >$

$$F_X(x) = \int_0^x dt = x\,.$$

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & x \in < 0, 1 > \\ 1 & x \geq 1 \end{cases}$$

4.15. Příklad. $f(x) = \begin{cases} \frac{x+1}{2} & x \in < -1, 1 > \\ 0 & \text{jinak} \end{cases}$

Pro distribuční funkci $F(x)$ na intervalu $< -1, 1 >$ platí

$$F_X(x) = \int_{-1}^x \frac{t+1}{2}\,dt = \left[\frac{t^2}{4} + \frac{t}{2}\right]_{-1}^x =$$

$$= \frac{x^2}{4} + \frac{x}{2} + \frac{1}{4} = \frac{(x+1)^2}{4}\,.$$

Pravděpodobnost roste s kvadratickou rychlostí. Pro distribuční funkci máme

$$F_X(x) = \begin{cases} 0 & x < -1 \\ \frac{(x+1)^2}{4} & x \in < -1, 1 > \\ 1 & x \geq 1 \end{cases}$$

4.16. Příklad. $$f(x) = \begin{cases} x+1 & x \in < -1, 0 > \\ 1-x & x \in < 0, 1 > \\ 0 & \text{jinak} \end{cases}$$

Pro $x \in < -1, 0 >$

$$F_X(x) = \int_{-1}^{x} (t+1)\, dt = \left[\frac{t^2}{2} + t\right]_{-1}^{x} = \frac{x^2}{2} + x + \frac{1}{2}\,.$$

Pro $x \in < 0, 1 >$

$$F_X(x) = \frac{1}{2} + \int_{0}^{x} (1-t)\, dt = \frac{1}{2} + \left[t - \frac{t^2}{2}\right]_{0}^{x} = -\frac{x^2}{2} + x + \frac{1}{2}\,.$$

4.17. Příklad. (Semicircular law)

$$f(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & x \in < -2, 2 > \\ 0 & \text{jinak} \end{cases}$$

$$\int \sqrt{4-x^2}\, dx = \frac{1}{2}(x\sqrt{4-x^2} + 4 \arcsin \frac{x}{2}) + c\,.$$

Pro $x \in < -2, 2 >$ tedy máme:

$$F_X(x) = \frac{1}{4\pi} x\sqrt{4-x^2} + \frac{1}{\pi} \arcsin \frac{x}{2} + \frac{1}{2}\,.$$

4.18. Příklad. $$f(x) = \frac{1}{2}e^{-|x|}\,.$$

Pro $x < 0$

$$F_X(x) = \frac{1}{2}\int\limits_{-\infty}^{x} e^t\,dt = \frac{e^x}{2}\,.$$

Pro $x \geq 0$

$$F_X(x) = \frac{1}{2} + \frac{1}{2}\int\limits_{0}^{x} e^{-t}\,dt = \frac{1}{2} - \frac{1}{2}e^{-x} + \frac{1}{2} = 1 - \frac{1}{2}e^{-x}\,.$$

4.19. Příklad. Střílíme na terč o poloměru r. Výhra je dána vzdáleností zásahu d od středu terče vzorcem

$$X = 10(r - d)\,.$$

Nalezněte hustotu veličiny X.

Řešení:
Pro $0 \le x \le 10r$

$$P[X \le x] = P[10r - 10d \le x] = P\left[d \ge r - \frac{x}{10}\right] =$$

$$= 1 - P\left[d \le r - \frac{x}{10}\right] = 1 - \frac{\pi (r - \frac{x}{10})^2}{\pi\, r^2} = 1 - \frac{(r - \frac{x}{10})^2}{r^2}\,.$$

Hustota pro $x \in (0, 10r >$:

$$f(x) = F_X'(x) = \frac{-2(r - \frac{x}{10})}{r^2} \cdot \frac{-1}{10} = \frac{1}{5r^2}\left(r - \frac{x}{10}\right).$$

Hustota je nulová mimo interval $< 0, 10r >$.

Důležité je reprezentovat náhodnou veličinu číselnými charakteristikami. Jednou z nich je střední hodnota.

Motivace: Ve škole je N žáků z toho

n_1 má prospěch $x_1 = 1$
n_2 má prospěch $x_2 = 2$
n_3 má prospěch $x_3 = 3$
n_4 má prospěch $x_3 = 4$
n_5 má prospěch $x_5 = 5$

$p_i = \frac{n_i}{N}$ ... relativní četnost.

$$
\begin{aligned}
&\text{Průměrný prospěch } = \\
&\quad = \frac{n_1x_1 + n_2x_2 + n_3x_3 + n_4x_4 + n_5x_5}{N} = \\
&\quad = \frac{n_1}{N}x_1 + \frac{n_2}{N}x_2 + \frac{n_3}{N}x_3 + \frac{n_4}{N}x_4 + \frac{n_5}{N}x_5 = \\
&\qquad\qquad = x_1p_1 + x_2p_2 + x_3p_3 + x_4p_4 + x_5p_5 \,.
\end{aligned}
$$

4.20. Definice. Nechť X je diskrétní náhodná veličina, která nabývá hodnot $x_1, x_2, \ldots$ s pravděpodobnostmi

$$P[X = x_i] = p_i, \qquad \sum_i p_i = 1\,.$$

Předpokládejme, že

$$\sum_i |x_i|\, p_i < \infty\,.$$

Střední hodnota, EX, veličiny X je definovaná vztahem

$$\boxed{EX = \sum_i x_i\, p_i\,.}$$

4.21. Definice. Nechť X je náhodná veličina s hustotou $f(x)$ taková, že

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty\,.$$

Střední hodnota, EX, veličiny X je definována vztahem:

$$\boxed{EX = \int_{-\infty}^{\infty} x\, f(x)\, dx\,.}$$

4.22. Příklad. Konstantní náhodná veličina

$$P[X = c] = 1$$

$$EX = c \cdot 1 = c\,.$$

4.23. Příklad. Počet ok při hodu hrací kostkou

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

4.24. Příklad. Alternativní rozdělení $\begin{array}{c|c} 0 & 1 \\ \hline 1-p & \mathsf{p} \end{array}$

$$EX = 0 \cdot (1 - p) + 1 \cdot p = p\,.$$

V tomto případě splývá střední hodnota s pravděpodobností p.

4.25. Příklad. Házíme symetrickou mincí. X je počet hodů než padne první líc.

$$P[X = n] = \frac{1}{2^n}\,, \qquad n = 1, 2, \dots$$

$$EX = \sum_{n=1}^{\infty} \frac{n}{2^n}\,.$$

Při výpočtu si pomůžeme teorií mocninných řad.

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}\,, \qquad \text{pro } |x| < 1\,.$$

Derivace člen po členu:

$$\sum_{n=1}^{\infty} n\, x^{n-1} = \left(\frac{1}{1-x}\right)' = \frac{1}{(1-x)^2}\,.$$

Pronásobením x

$$\sum_{n=1}^{\infty} n\, x^n = \frac{x}{(1-x)^2}$$

Dosazením $x = \frac{1}{2}$

$$EX = \sum_{n=1}^{\infty} \frac{n}{2^n} = \frac{2}{1} = 2\,.$$

Na každá náhodná veličina má definovánu střední hodnotu.
Např. diskrétní veličina s pravděpodobnostní funkcí

$$P[X = 2^n] = \frac{1}{2^n}\,, \qquad n = 1, 2 \dots$$

nemá střední hodnotu neboť

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

4.26. Příklad. Hustota

$$f(x) = \begin{cases} \frac{x+1}{2} & x \in < -1, 1 > \\ 0 & \text{jinak} \end{cases}$$

$$EX = \int_{-1}^{1} x \frac{x+1}{2}\, dx = \left[\frac{x^3}{6} + \frac{x^2}{4}\right]_{-1}^{1} = \frac{1}{3}\,.$$

4.27. Příklad. Semicircular law

$$EX = \frac{1}{2\pi} \int_{-2}^{2} x \sqrt{1 - x^2} \, dx = 0 \, .$$

4.28. Věta. *Jsou-li* X_1 *a* X_2 *náhodné veličiny na pravděpodobnostním prostoru* $(\Omega, \mathcal{A}, P)$*, pak*

(i) $E(X_1 + X_2) = EX_1 + EX_2,$

(ii) $E(\alpha X_1) = \alpha E(X_1) \, , \qquad \alpha \in \mathbb{R} \, .$

(iii) $E(X_1) \geq 0$ *je-li* $X_1 \geq 0$.

Podrobnější popis rozložení hodnot kolem střední hodnoty poskytuje rozptyl

4.29. Definice. Rozptyl (variance) náhodné veličiny X, pro kterou existuje EX a EX^2 je definován

$$varX = E[(X - EX)^2] .$$

Značení: $var(X)$, $D(X)$,

Směrodatná odchylka: $\sqrt{var(X)}$.

$$E[(X - EX)^2] = E[X^2 - 2\,X \cdot (EX) + (EX)^2] = \\ = E(X^2) - (EX)^2 .$$

Způsob výpočtu:

$$EX^2 = \sum_i x_i^2\, p_i \ \ldots \text{ diskrétní veličina}$$

$$EX^2 = \int_{-\infty}^{\infty} x^2\, f(x)\, dx \ldots \text{ spojitá veličina}$$

4.30. Příklad. Alternativní rozdělení X

0	1
$1-p$	p

X^2 :

0	1
$1-p$	p

$$var(X) = EX^2 - (EX)^2 = p - p^2 = p(1-p)\,.$$

$$var(X) \leq \frac{1}{4}$$

Nejvyšší možná hodnota rozptylu je pro $p = \frac{1}{2}$ a to $\frac{1}{4}$.

4.31. Definice. Nechť $F_X(x)$ je distribuční funkce náhodné veličiny X. Předpokládejme, že pro každé $\alpha \in (0,1)$ existuje právě jedno β tak, že $F(\beta) = \alpha$.

α-*kvantil* náhodné veličiny X je číslo, pro které platí

$$F_X(x_\alpha) = \alpha\,.$$

$$\alpha = P[X \leq x_\alpha]\,.$$

Je-li F_X prostá, pak $x_\alpha = F_X^{-1}(\alpha)\,.$
F_X^{-1} se v tomto případě nazývá *kvantilová funkce*.

Významné kvantily:

$x_{0,5}\ldots$	medián
$x_{0,75}\ldots$	horní kvartil
$x_{0,25}\ldots$	dolní kvartil
$x_{0,9}\ldots$	horní decil
$x_{0,1}\ldots$	dolní decil
$x_{0,99}\ldots$	horní percentil

Statistické tabulky: Průměrný čistý plat v ČR na osobu v domácnosti v roce 2003 byl 8175 KČ.
Dolní decil $x_{0,1}$=4524 KČ.

4.32. Příklad. X je spojité rozdělení s hustotou
$f(x) = \frac{x+1}{2}$, $x \in < -1, 1 >$ a nula jinak.
Viz Příklad 4.15.

$$F(x) = \frac{1}{4}\,(x+1)^2\,, \qquad x \in < -1, 1 > \ .$$

$\alpha \in (0,1)$

$$\begin{aligned} F(x) &= \frac{1}{4}\,(x+1)^2 = \alpha \\ |x+1| &= \sqrt{4\alpha} \\ x_\alpha &= 2\sqrt{\alpha} - 1 \end{aligned}$$

Např. medián

$$x_{0,5} = 2\sqrt{1/2} - 1\,.$$

4.33. Příklad. Doba rozpadu radioaktivního atomu je náhodná veličina s hustotou

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & x \geq 0 \\ 0 & x < 0\,, \end{cases}$$

kde $\lambda > 0$. Určete poločas rozpadu.

Řešení:

Poločas rozpadu je medián. Pro distribuční funkci máme

$$F(x) = \int\limits_0^x \lambda e^{-\lambda t}\, dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}\,.$$

$$\begin{aligned} 1 - e^{-\lambda x} &= 0,5 \\ e^{-\lambda x} &= 0,5 \\ -\lambda x &= -\ln 2 \\ x &= \frac{\ln 2}{\lambda} \end{aligned}$$

Poločas rozpadu je medián

$$x_{0,5} = \frac{\ln 2}{\lambda}\,.$$

4.34. Tvrzení. *Platí-li pro náhodné veličiny X a Y vztah*

$$Y = aX + b\,, \qquad a, b \in \mathbb{R}, a > 0\,,$$

pak

$$y_\alpha = ax_\alpha + b$$

pro všechna $\alpha \in (0, 1)$.

Důkaz:

$$\alpha = P[X \leq x_\alpha] = P[aX + b \leq ax_\alpha + b] = P[Y \leq \underbrace{ax_\alpha + b}_{y_\alpha}]\,.$$

5 Důležitá rozdělení

Diskrétní rozdělení

Alternativní rozdělení $A(p)$, $0 < p < 1$.

0	1
$1-p$	p

$$EX = p$$

$$varX = p(1-p)\,.$$

Binomické rozdělení $Bi(n,p)$, $n \in \mathbb{N}$, $0 < p < 1$.

0	1	2	...	n
p_0	p_1	p_2	$\ldots$	p_n

$$P[X = k] = p_k = \binom{n}{k} p^k\,(1-p)^{n-k} \qquad k = 0, 1, \ldots, n\,.$$

X = počet zdarů v sérii n pokusů Bernoulliova schématu.

- počet líců v sérii n hodů
- počet osob volící daný politický subjekt z n dotázaných
- počet osob sledujících daný TV pořad z n sledovaných
- počet částic v náhodně zvolené přihrádce z n přihrádek
- počet vadných sučástek z n náhodně vybraných součástek

$$X_i = \begin{cases} 1 & \text{nastal-li zdar v } i - \text{tém pokusu} \\ 0 & \text{jinak} \end{cases}$$

X_i má rozdělení $A(p)$.

$$\boxed{X = X_1 + X_2 + \cdots + X_n\,.}$$

5.1. Příklad. S.Pepys (1693), náruživý hráč v kostky. Co je pravděodobnější, že šesti kostkami hodíme alespoň jednu šestku (jev A), nebo že dvanácti kostkami hodíme alespoň dvě šestky (jev B)? Vyřešil Newton.

počet šestek při hodu šesti kostkami ... $Bi(6, 1/6)$.
$(p = \frac{1}{6})$

$$P(A) = \sum_{k=1}^{6} \binom{6}{k} p^k (1-p)^{n-k} =$$

$$1 - \binom{6}{0} p^0 (1-p)^6 = 0,6651\,.$$

počet šestek při hodu dvanácti kostkami ... $Bi(12, 1/6)$.
$(p = \frac{1}{6})$

$$P(B) = \sum_{k=2}^{12} \binom{12}{6} p^k (1-p)^{n-k} =$$

$$1 - \binom{12}{0} p^0 (1-p)^{12} - \binom{12}{1} p (1-p)^{11} = 0,6189\,.$$

Domácí cvičení: Pravděpodovbnost, že 18 kostkami hodíme alespoň 3 šestky je 0,5973.

5.2. Příklad. Náhodná procházka
Částice se pohybuje po ose x. Začíná v bodě 0. V daném stádiu se rozhodne s pravděpodobností 1/2 jít doprava a s pravděpodobností 1/2 doleva. S_n je poloha částice v čase n. Jaké je rozdělení S_n?

Bernoulliovo schéma, n pokusů; $p = \frac{1}{2}$.
1.. jdeme doprava, -1... jdeme doleva.

Je-li k jedniček a $n-k$ -jedniček, pak je poloha

$$k - (n-k) = 2k - n \quad k = 0, \dots n\,.$$

$$P[S_n = 2k - n] = \binom{n}{k} 2^{-n}\,.$$

(Dá se ukázat, že částice s pravděpodobnosí jedna navštíví každý bod. Totéž pro dvě dimenze, ne však pro tři.)

5.3. Příklad.
Maxwellovo-Boltzmanovo schéma

Máme n částic a r příhrádek. Každá částice si vybírá nějakou příhrádku. Všechny možnosti mají stejnou šanci. Jaké je rozložení počtu částic v pevně zvolené příhrádce?

Bernoulliovo schéma: vybraná částice si zvolí danou přihrádku, celkem n pokusů (máme n částic). Šance zdaru je $\frac{1}{r}$.

Náhodná veličina má rozdělení $Bi(n, \frac{1}{r})$.

Konkrétní situace:

5.4. Příklad. Máme $n = 500$ osob a $r = 365$ příhrádek (narozeniny). Počet osob mající narozeniny dne 18.7. (jako přednášející) se řídí $Bi(500, \frac{1}{365})$.

Tabulka numerických hodnot:

počet	0	1	2	3	4	5	6
pravděpodobnost	0,2537	0,3484	0,2388	0,1089	0,0372	0,0101	0,0023

Pravděpodobnost, že tři osoby mají narozeniny 18.7. je 0,1089 .
další modely: osoby obsazující vagóny, výsledky hodu kostkou padající do 6 možností, . . .

Charakteristiky $Bi(n,k)$.

$$X = X_1 + X_2 + \cdots + X_n$$

$X_i \dots A(p)$, $EX_i = p$.

$$EX = EX_1 + EX_2 + \cdots + EX_n = np\,.$$

$$E(X^2) = E(X_1{+}X_2{+}\cdots{+}X_n)^2 = EX_1^2{+}EX_2^2\cdots{+}EX_n^2{+}$$
$$+\, 2E(X_1X_2) + 2E(X_1X_3) + \cdots\,.$$

Je-li $i \neq j$ máme pro $X_i\,X_j$ rozdělení

0	1
$1-p^2$	p^2

Tedy $E(X_iX_j) = p^2$, což znamená, že

$$EX^2 = np + 2\cdot\binom{n}{2}p^2\,.$$

Konečně,

$$var(X) = EX^2{-}(EX)^2 = np{+}2{\cdot}\binom{n}{2}p^2{-}n^2p^2 =$$

$$np{+}2\frac{n(n-1)}{2}p^2{-}n^2p^2 = np{-}np^2 = np(1{-}p)\,.$$

Má-li X rozdělení $Bi(n,p)$, pak

$$\boxed{EX = np \qquad varX = np(1-p)}$$

Průměrný počet šestek při sérii n hodů hrací kostkou je $\frac{n}{6}$.

Průměrný počet částic v jedné přihrádce u 500 částic náhodně rozptýlených v 365 příhrádkách je $\frac{500}{365} = 1,369863014.$

...

Co se děje s binomickým rozdělením, jestliže se nemění střední hodnota, ale počet pokusů jde do nekonečna?

5.5. Věta. Poissonova věta
Předpokládejme, že $(X_n)_{n=1}^{\infty}$ *je posloupnost náhodných veličin majících rozdělení* $Bi(n, \frac{\lambda}{n})$*, kde* $\lambda > 0$*. (Tj.* $EX_n = \lambda$ *pro všechna* n*.) Pak*

$$\lim_{n\to\infty} P[X_n = k] = \frac{\lambda^k}{k!}\, e^{-\lambda}\,.$$

Důkaz:

$$p_n = \frac{\lambda}{n}\,.$$

$$P[X_n = k] = \binom{n}{k} p_n^k (1-p_n)^{n-k} =$$
$$= \frac{1}{k!}\, \frac{np_n \cdot (n-1)p_n \cdots (n-k+1)p_n}{(1-p_n)^k} \left(1 - \frac{\lambda}{n}\right)^n .$$

$$\lim_{n\to\infty} \frac{np_n}{1-p_n} = \lim_{n\to\infty} \frac{\lambda}{1-\frac{\lambda}{n}} = \lambda\,.$$

$$\lim_{n\to\infty} \frac{(n-1)p_n}{1-p_n} = \lim_{n\to\infty} \frac{\lambda - \frac{\lambda}{n}}{1-\frac{\lambda}{n}} = \lambda\,.$$

Tedy

$$\lim_{n\to\infty} P[X_n = k] = \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^k}{k!} e^{-\lambda}\,.$$

Aproximujeme pro n velké a p_n malé

$$P[X = k] \sim \frac{\lambda^k}{k!} e^{-\lambda}\,.$$

5.6. Příklad. Stroj produkuje 1% zmetků. Jaká je pravděpodobnost, že z 200 náhodně vybraných výrobků neni žádný zmetek?

$$X \sim Bi(200, \frac{1}{100})$$

$$P[X = 0] = 0,99^{200} = 0,1340\,.$$

Aproximace pomocí Poissonovy věty:

$$\lambda = 200 \cdot \frac{1}{100} = 2$$

$$P[X = 0] \doteq e^{-2} = 0,1353.$$

n částic se náhodně rozděluje do r příhrádek, přičemž $n, r \to \infty$ při konstantním poměru $\lambda = \frac{n}{r}$. Počet částic v pevně zvolené příhrádce se asymtoticky řídí Poissonovým zákonem s parametrem λ.

5.7. Příklad. $X \sim Bi(500, 365) \ldots$ viz Příklad 5.4.

počet	0	1	2	3	4	5	6
binomický zákon	0,2537	0,3484	0,2388	0,1089	0,0372	0,0101	0,0023
Poissonův zákon	0,2541	0,3481	0,2385	0,1089	0,0372	0,0102	0,0023

Poissonovo rozdělení

Náhodná veličina X, která nabývá hodnot $0, 1, \dots$ s pravděpodobnostmi

$$\boxed{P[X_n = n] = \frac{\lambda^n}{n!}\, e^{-\lambda}\,, n = 0, 1, \dots}$$

$$Po(\lambda)$$

$$P[X = 0] = e^{-\lambda} \qquad P[X > 0] = 1 - e^{-\lambda}\,.$$

5.8. Příklad. Lahve se vyrábějí ze skloviny obsahující kazy, keré jsou rozděly nepravidelně tak, že v každém metrickém centu skloviny je průměrně x kazů. Láhev váží 1 kg a je vadná obsahuje-li jeden či více kazů. Stanovte procento vadných lahví.

Řešení Z M metrických centů se vyrobí $100M$ lahví, které budou obsahovat přibližně xM kazů. Pro počet kazů v jedné lahvi tedy máme rozdělení

$$Bi(xM, \frac{1}{100M})\,.$$

$$EX = \lambda = xM\frac{1}{100M} = \frac{x}{100}\,.$$

Pro $M \to \infty$ máme rozdělení

$$Po\left(\frac{x}{100}\right)$$

Pravděpodobnost, že láhev bude bez kazu je

$$1 - e^{-\frac{x}{100}} \; .$$

Je-li například $x = 30$, pak procento vadných lahví bude $1 - e^{-0,3} = 0,2592$.

Při velkém počtu kazů je výhodnější vyrábět menší lahve. Je-li např. váha lahve $0,25 kg$ je procento zmetků $7,22\%$.

Charakteristiky $P(\lambda)$:

$$EX = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda \left(\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \right) e^{-\lambda} = \\ = \lambda e^{\lambda} e^{-\lambda} = \lambda \,.$$

$$EX^2 = \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=2}^{\infty} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} + \overbrace{\sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda}}^{=\lambda} = \\ = e^{-\lambda} \sum_{n=2}^{\infty} \frac{\lambda^n}{(n-2)!} + \lambda = e^{-\lambda} \lambda^2 \sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!} + \lambda = \\ = e^{-\lambda} \lambda^2 \, e^{\lambda} + \lambda = \lambda^2 + \lambda \,.$$

Tedy

$$var(X) = EX^2 - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \,.$$

$$\boxed{\begin{aligned} E(X) &= \lambda \\ var(X) &= \lambda \end{aligned}}$$

Příklady Poissonova rozdělení: (homogenní chaos v prostoru nebo čase)

- počet volání na telofonní ústřednu za jednotku času
- počet atomů radioaktivní látky rozpadlých za jednotku času
- počet hvězd v daném objemu galaxie
- počet létavic meteorického roje za jednotku času
- počet střel zasahující danou oblast
- počet defektů kola (bad luck) za jednotku času
- počet zákazníků za jednotku času

Geometrické rozdělení $Ge(p)$, $0 < p < 1$.

X je počet zdarů v Bernoulliově schématu před prvním nezdarem.

$$P[X = 0] = 1 - p\,.$$

$$P[X = 1] = (1 - p)p\,.$$

$$P[X = 2] = (1 - p)p^2\,.$$

....

$$P[X = k] = (1 - p)p^k \qquad k = 0, 1, \ldots\,.$$

5.9. Příklad. Dva hráči se střídají a házejí hrací kostkou. Vyhrává ten komu padne šestka. Jaká je pravděpodobnost výhry u jednotlivých hráčů?

$$X \ldots \text{geometrické rozdělení s } p = \frac{5}{6}\,.$$

A ... vyhrává hráč, který začíná

$$P(A) = P[X = \text{sudé}] = \sum_{k=0}^{\infty}(1-p)p^{2k} = (1-p)\sum_{k=0}^{\infty}p^{2k} =$$

$$= (1-p)\frac{1}{1-p^2} = \frac{1}{1+p} =$$

$$= \frac{1}{1+\frac{5}{6}} = \frac{6}{11} = 0,54545455\,.$$

Střední hodnota geometrického rozdělení:

$$EX = \sum_{n=0}^{\infty} n\,(1-p)\,p^n = (1-p)\sum_{n=0}^{\infty} n\,p^n =$$
$$= (1{-}p)\,p\,\frac{d}{dp}\left(\sum_{n=0}^{\infty} p^n\right) = (1{-}p)p\frac{d}{dp}\left(\frac{1}{1-p}\right) =$$
$$= (1-p)p\frac{1}{(1-p)^2} = \frac{p}{1-p}\,.$$

5.10. Příklad. Žák umí 90% látky. Kolik přežije průměrně otázek?

$Ge(p), p = 0,9$

$$EX = \frac{0,9}{1-0,9} = 9\,.$$

Rozptyl

$$\frac{d^2}{dp^2}\sum_{n=0}^{\infty} p^n = \sum_{n=0}^{\infty} n(n-1)p^{n-2}\,.$$

$$E(X^2) - EX = \sum_{n=0}^{\infty} n(n-1)(1-p)p^n =$$
$$= (1-p)p^2 \frac{d^2}{dp^2}\sum_{n=0}^{\infty} p^n = (1-p)p^2\left(\frac{1}{1-p}\right)'' =$$
$$= (1-p)p^2\frac{2}{(1-p)^3} = \frac{2p^2}{(1-p)^2}\,.$$

$$E(X^2) - (EX)^2 = \frac{2p^2}{(1-p)^2} + \frac{p}{1-p} - \frac{p^2}{(1-p)^2} =$$
$$\frac{p^2}{(1-p)^2} + \frac{p}{1-p} = \frac{p^2 + p(1-p)}{(1-p)^2} =$$
$$= \frac{p}{(1-p)^2}\,.$$

$$\boxed{\begin{aligned} EX &= \frac{p}{1-p} \\ varX &= \frac{p}{(1-p)^2}\,. \end{aligned}}$$

Rovnoměrné rozdělení na $< a, b >$
$R < a, b >$.

Hustota:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in < a, b > \\ 0 & \text{jinak} \end{cases}$$

Distribuční funkce:

$$F(x) = \begin{cases} \frac{x-a}{b-a} & x \in < a, b > \\ 0 & x < a \\ 1 & x > b \end{cases}$$

$$EX = \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b =$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}\,.$$

$$EX^2 = \frac{1}{b-a}\int_a^b x^2\, dx = \frac{1}{b-a}\frac{b^3-a^3}{3} =$$

$$= \frac{a^2+ab+b^2}{3}\,.$$

$$var(X) = E(X^2) - (EX)^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} =$$

$$= \frac{a^2-2ab+b^2}{12} = \frac{1}{12}(b-a)^2\,.$$

$$\boxed{\begin{aligned} EX &= \frac{a+b}{2} \\ varX &= \frac{1}{12}(b-a)^2\,. \end{aligned}}$$

Normální rozdělení (Gaussovo rozdělení)

$N(\mu, \sigma^2)$.
$\mu \in \mathbb{R}, \sigma^2 > 0$.

hustota:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

Vychází z Laplaceova integrálu:

$$\int\limits_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}\,.$$

standardní, normované normální rozdělení: $N(0, 1)$.

značení:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

počítá se numericky, tabelována.

Souvislost mezi normálními rozděleními různých parametrů.

• Má-li Y rozdělení $N(0,1) \Longrightarrow X = \mu + \sigma Y$ má rozdělení $N(\mu, \sigma^2)$.

Odvození:

$$
\begin{aligned}
F_X(x) = P[X \leq x] = P[\mu + \sigma Y \leq x] = \\
= P\left[Y \leq \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dx} F_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\frac{x-\mu}{\sigma})^2}{2}} \cdot \frac{1}{\sigma} = \\
= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .
\end{aligned}
$$

• Má-li X rozdělení $N(\mu, \sigma)$, pak $Y = \frac{X-\mu}{\sigma}$ má rozdělení $N(0,1)$.

5.11. Příklad. S jakou pravěpodobností má veličina X s rozdělením $N(1,4)$ hodnotu v intervalu $< 3,5 >$?

$$P[3 \leq X \leq 5] = \Phi\left(\frac{5-1}{2}\right) - \Phi\left(\frac{3-1}{2}\right) =$$
$$= \Phi(2) - \Phi(1) = 0,97250 - 0,841345 = 0,131155.$$

5.12. Tvrzení. *Vzhledem k tomu, že hustota standardního normálního rozdělení je sudá funkce, platí*

$$\Phi(x) = 1 - \Phi(-x) \qquad x \in \mathbb{R}\,.$$

5.13. Příklad. Spočtěte pravděpodobnost, že veličina X s rozdělením $N(0, \sigma^2)$ má hodnotu v intervalu $< -a, a >$, kde $a > 0$.

Řešení:

$$P[-a \leq X \leq a] = P[-\frac{a}{\sigma} \leq \frac{X}{\sigma} \leq \frac{a}{\sigma}] =$$

$$= \Phi\left(\frac{a}{\sigma}\right) - \Phi\left(-\frac{a}{\sigma}\right) = \Phi\left(\frac{a}{\sigma}\right) - \left(1 - \Phi\left(\frac{a}{\sigma}\right)\right) =$$

$$2\Phi\left(\frac{a}{\sigma}\right) - 1\,.$$

Střední hodnota a rozptyl normálního rozdělení:

Pro Y s rozdělením $N(0,1)$ platí

$$EY = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x e^{-\frac{x^2}{2}}\, dx = 0\,.$$

$$1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx\,.$$

Použijeme metodu per partes pro

$$\begin{array}{rl} u' = 1 & v = e^{-\frac{x^2}{2}} \\ u = x & v' = -xe^{-\frac{x^2}{2}} \end{array}$$

a dostaneme

$$1 = \frac{1}{\sqrt{2\pi}} \underbrace{\left[xe^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty}}_{=0} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\, dx\,.$$

Odtud plyne, že

$$var(Y) = E(Y^2) = 1\,.$$

Obecně: X má rozdělení $N(\mu, \sigma^2)$

$$X = \sigma Y + \mu\,,$$

a proto

$$EX = \sigma EY + \mu = \mu\,.$$

$$X^2 = \sigma^2 Y^2 + 2\mu\sigma Y + \mu^2$$

$$EX^2 = \sigma^2 + 0 + \mu^2 \,.$$

$$varX = E(X^2) - (EX)^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2 \,.$$

Závěr:

$$\boxed{\begin{aligned} EX &= \mu \\ var(X) &= \sigma^2 \,. \end{aligned}}$$

kvantilová funkce a kvantily

u_α ... α-kvantil $N(0,1)$.

kvantilová funke je inverzní funkce Φ^{-1}.

Některé numerické hodnoty:

$u_{0,5} = 0$,
$u_{0,95} = 1,644$,
$u_{0,975} = 1,95996$
$u_{0,999} = 3,09023$

Pro $\alpha \to 1$ jde $u_\alpha \to \infty$.

5.14. Tvrzení.

(i) $u_{1-\alpha} = -u_\alpha$ *pro všechna* $\alpha \in (0,1)$.

(ii) *Pro* α*-kvantil* x_α *rozdělení* $N(\mu, \sigma^2)$ *platí*

$$x_\alpha = \mu + \sigma\, u_\alpha\,.$$

5.15. Příklad. Určete interval $< -a, a >$ tak, aby náhodná veličina Y s rozdělením $N(0,1)$ měla v tomto intervalu hodnotu s pravděpodobností 0,95.

$$a = u_{0,975} \doteq 1,96\,.$$

Pravidlo 3σ
Máme rozdělení X typu $N(\mu, \sigma^2)$. Určeme

$$P[|X-\mu| \le 3\sigma]\,.$$

Řešení:

$$P\left[\frac{|X-\mu|}{\sigma} \le 3\right] = \Phi(3) - \Phi(-3) =$$
$$= 2\Phi(3) - 1 = 2\cdot 0,99865 - 1 = 0,99730\,.$$

Po třech σ zbývají asi tři promile případů.

5.16. Příklad. Pro oděvní továrnu je neziskové vyrábět šaty pro velmi malé a velmi velké muže. Záměr je nevyrábět pro 7,5% největších a 7,5% nejmenších mužů. Ví se, že výška mužů (v palcích) má rozdělení $N(69, 2, 8^2)$. Nalezněte největší a nejmenší výšku pro kterou vyrábět.

Řešení

$$u_{0,925} = 1,43953\,.$$

$$\begin{aligned} x_{0,925} &= 69 + 2,8 \cdot 1,43953 = 73,03068 \\ x_{0,075} &= 69 - 2,8 \cdot 1,43953 = 64,96932 \end{aligned}$$

5.17. Příklad. Výsledky přijímacích zkoušek se řídí normálním rozdělením s rozptylem 100. Je přijato 30% uchazečů. Hranice pro přijetí je 85 bodů. Jaký je průměrný výsledek u zkoušky?

Řešení:

$$N(\mu, \sigma^2)$$

$$x_{0,7} = \mu + 10 \cdot u_{0.7}$$

$$\mu = 85 - 10 \cdot u_{0,7} = 85 - 10 \cdot 0,52440 \doteq 79,8\,.$$

5.18. Příklad. Máme rozdělení $N(\mu, 0, 5)$. Jak zvolit střední hodnotu, aby

$$P[X \geq 2] = 0,01\,.$$

Řešení:

$$P\left[\frac{X-\mu}{\sqrt{0,5}} \geq \frac{2-\mu}{\sqrt{0,5}}\right] = 0,01$$

$$\frac{2-\mu}{\sqrt{0,5}} = u_{0,99}$$

$$\mu = 2 - \sqrt{0,5} \cdot u_{0,99} \doteq 0,355023643$$

Exponenciální rozdělení

$$Exp(\lambda) \qquad \lambda > 0$$

hustota:

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & x \geq 0 \\ 0 & x < 0\,. \end{cases}$$

distribuční funkce:
$x \geq 0$

$$F(x) = \int_0^x \lambda\, e^{-\lambda x}\, dx = 1 - e^{-\lambda\, x}\,.$$

funkce přežití:

$$P[X \geq x] = e^{-\lambda\, x}$$

Exponenciální rozdělení popisuje čas do první "poruchy" u systému "bez paměti"

Odvození:

Hledáme funkci přežití

$$R(t) = P[X \geq t]$$

tak, aby byly splněny následující předpoklady:

(i) $R(0) = 1$

(ii) $P[X \geq t + h | X \geq t] = P[X \geq h]$ pro všechna $x, h \geq 0$.

(iii) R je diferencovatelná klesající funkce

Z toho plyne:

$$P[X \geq t + h] = P[X \geq t] \cdot P[X \geq h] .$$

$$R(t + h) = R(t)R(h)$$

$$\begin{aligned} \frac{R(t+h) - R(t)}{h} = \frac{R(t)R(h) - R(t)}{h} &= \\ &= R(t) \frac{R(h) - 1}{h} \end{aligned}$$

Limitním přechodem $h \to 0+$ dostaneme

$$R'(t) = R(t) \cdot R'(0)$$

$$R(0) = 1$$

Diferenciální rovnice s počáteční podmínkou.

Označme

$R'(0) = -\lambda$ ($\lambda > 0$).

Řešení (jediné):

$$\boxed{R(t) = e^{-\lambda t}.}$$

$R(t)$ tedy vede na exponenciální rozdělení.

Střední hodnotu a rozptyl získáme integrací (per partes)

$$\boxed{\begin{aligned} E(X) &= \frac{1}{\lambda} \\ var(X) &= \frac{1}{\lambda^2} \end{aligned}}$$

λ ... "intenzita poruch"

Příklady exponenciálního rozdělení:

- doba rozpadu atomu
- doba do registrace zákazníka
- doba do příletu létavice v meteorickém roji

5.19. Příklad. Na přílet meteoritu se průměrně čaká deset minut. Jaká je pravděpodobnost, že budeme na "padající hvězdu" čekat dvě minuty?

Řešení:

$$\frac{1}{\lambda} = 10 \qquad \lambda = 0,1$$

$$F(2) = 1 - e^{-2\cdot 0,1} = 1 - e^{-0,2} \doteq 0,18127\,.$$

6 Transformace náhodných veličin

Nutnost přepočítat distrubuční funkci. Například máme měření rychlosti a chceme ho přepočítat na energii.

Obecná úloha: X je náhodná veličina, $h : \mathbb{R} \mapsto \mathbb{R}$

$$Y = h(X)$$

• Diskrétní náhodná veličina se vždy zobrazí na diskrétní

6.1. Příklad. Diskrétní rozdělení X s pravděpodobnostní funkcí

-1	0	1
0,3	0,2	0,5

$Y = X^2 \dots$

1	0	1
0,3	0,2	0,5

$Y \dots$

0	1
0,2	0,8

Obecně stanovíme transformaci pomocí distribuční funkce:

$$Y = h(X)$$
$$F_Y(y) = P[h(X) \leq y]$$

6.2. Příklad. Rychlost molekul plynu má rozdělení $N(0,1)$. Molekula má hmotnost m. Nalezněte distribuční funkci a hustotu energie částice.

$$X \sim N(0,1)$$
$$Y = \frac{1}{2} m X^2$$

Nerovnice $\frac{1}{2} m X^2 \leq y$ má řešení pouze pro $y \geq 0$.

$y \geq 0$

$$F_Y(y) = P[\frac{1}{2} m X^2 \leq y] = P[X \in < -\sqrt{\frac{2}{m} y}, \sqrt{\frac{2}{m} y} >] =$$
$$= \Phi\left(\sqrt{\frac{2}{m} y}\right) - \Phi\left(-\sqrt{\frac{2}{m} y}\right) = 2\,\Phi\left(\sqrt{\frac{2}{m} y}\right) - 1\,.$$

Hustota je pro $y > 0$ derivací distribuční funkce:

$$g(y) = 2\frac{1}{\sqrt{2\pi}}e^{-\frac{\frac{2}{m}y}{2}} \cdot \sqrt{\frac{2}{m}} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{m\pi y}}e^{-\frac{y}{m}}$$

Pro $y \leq 0$ je $g(y) = 0$.

Důležitý je případ lineární transformace.

$$Y = aX + b, a \neq 0$$

$a > 0$

$$F_Y(y) = P[aX + b \leq y] = P\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right).$$

$a < 0$

$$F_Y(y) = P[aX + b \leq y] = P\left[X \geq \frac{y-b}{a}\right] = 1 - F_X\left(\frac{y-b}{a}-\right).$$

Užitečné je aplikovat toto pravidlo na spojité rozdělení

6.3. Tvrzení. *Je-li X spojitá náhodná veličina s hustotou $f(x)$, pak náhodná veličina*

$$Y = aX + b\,, \qquad a \neq 0$$

je spojitá a má hustotu

$$\boxed{g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right).}$$

Důkaz:

$a > 0$

$$F_Y(y) = F\left(\frac{y-b}{a}\right)$$

Derivací podle y:

$$g(y) = \frac{1}{a} f\left(\frac{y-b}{a}\right).$$

$a < 0$

$$F_Y(y) = 1 - F\left(\frac{y-b}{a}\right)$$

Derivací podle y:

$$g(y) = -\frac{1}{a} f\left(\frac{y-b}{a}\right).$$

6.4. Příklad. X má rovnoměrné rozdělení na intervalu $< 0, 3 >$. Určete hustotu

$$Y = 2X + 1$$

$$g(y) = \frac{1}{2} f\left(\frac{y-1}{2}\right).$$

$$\frac{y-1}{2} \in < 0, 3 > \iff y \in < 1, 7 > .$$

$$g(y) = \begin{cases} 1/6 & y \in < 1, 7 > \\ 0 & \text{jinak} \end{cases}$$

Y má rovnoměrné rozdělení na intervalu $< 1, 7 >$.

6.5. Příklad. $Y = -X$, kde X má rozdělení $N(0,1)$.

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-y)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Y má také rozdělení $N(0,1)$.

6.6. Příklad. X má rozdělení $Exp(1)$. Určete rozdělení $-X$.

$$g(y) = f(-y)$$

$$g(y) = \begin{cases} e^y & y \leq 0 \\ 0 & y > 0\,. \end{cases}$$

Nelineární transformace náhodné veličiny

Předpoklady: X má hustotu $f(x)$ soustředěnou na intervalu I a h je rostoucí diferencovatelná funkce definovaná na I, jejíž obor hodnot je interval J.

$$Y = h(X)$$

Pro $y \notin J$ bude hustota nulová.

Pro $y \in J$

$$F_Y(y) = \int_{-\infty}^{h^{-1}(y)} f(x)\,dx$$

Substituce: $t = h(x)$, $x = h^{-1}(t)$, $dt = h'(x)\,dx$.

$$F_Y(y) = \int_{-\infty}^{y} f(h^{-1}(t))\,\frac{dt}{h'(h^{-1}(t))}$$

Podobně lze postupovat v případě, kdy h je klesající, nebo je možno použít $-(-h)$.

Závěr:

$$\boxed{g(y) = \frac{f(h^{-1}(y))}{|h'(h^{-1}(y))|}\,.}$$

pro $y \in J$ a nula jinak.

6.7. Příklad. Logaritmicko-normální rozdělení

$$Y = e^X\,,$$

kde X má rozdělení $N(\mu, \sigma^2)$.

$$h(x) = e^x\,, \quad h^{-1}(x) = \ln x\,, \quad J = h(\mathbb{R}) = (0, \infty)\,.$$

$$g(y) = \frac{f(\ln y)}{e^{\ln y}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \frac{1}{y}\,,$$

pro $y > 0$; a nula jinak.

K výpočtu střední hodnoty transformované veličiny nepotřebujeme znát rozdělení transformace:

6.8. Věta. *Předpokládejme, že X je náhodná veličina a $Y = h(X)$. Pak*

(i) $$\boxed{E(Y) = \sum_{i \in I} h(x_i) p_i \,,}$$

je-li X diskrétní náhodná veličina s pravděpodobnostní funkcí $(x_i, p_i)_{i \in I}$.

(ii) $$\boxed{E(Y) = \int_{-\infty}^{\infty} f(x) h(x)\, dx}$$

je-li X spojitá náhodná veličina s hustotou $f(x)$.

Důkaz:

(i) $h(Y)$ má (včetně násobnosti) pravděpodobnostní funkci $(h(x_i), p_i)_{i \in I}$, a proto

$$E(Y) = \sum_{i \in I} h(x_i) p_i \,.$$

(ii) je spojitou verzí.

6.9. Příklad. Určete střední hodnotu třetí mocniny rozdělení $Exp(1)$.

$$E(Y) = \int\limits_0^\infty x^3 e^{-x}\, dx = 6\,.$$

7 Náhodné vektory

Značení: $\mathbf{x} = (x_1, x_2, \ldots x_n) \in \mathbb{C}^n$

7.1. Definice. Nechť $(\Omega, \mathcal{A}, P)$ je pravděpodobnostní prostor. Zobrazení $\mathbf{X} : \Omega \to \mathbb{R}^n$

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots , X_n(\omega)) \,,$$

kde $X_1, \ldots , X_n$ jsou náhodné veličiny na Ω, se nazývá *náhodný vektor*. Veličiny $X_1, \ldots , X_n$ se nazývají *marginální rozdělení* vektoru X.

Nestačí znát marginální distribuční funkce, ale distribuční funkci sdruženou.

Značení:
$[\mathbf{X} \leq \mathbf{x}] = [X_1 \leq x_1 \wedge X_2 \leq x_2 \wedge \cdots \wedge X_n \leq x_n]$.

7.2. Definice. Je-li $\mathbf{X} : \Omega \to \mathbb{R}^n$ náhodný vektor, pak distribuční funkci $F_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}$ definujeme jako

$$F_{\mathbf{X}}(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] \,.$$

7.3. Příklad. X má distribuční funkci F. Určete distribuční funkci náhodného vektoru

$$\mathbf{X} = (X, X)$$

$$F_{\mathbf{X}}(x, y) = P[X \leq x \wedge X \leq y] = \\ = P[X \leq \min(x, y)] = F(\min(x, y))$$

Pomocí distribuční funkce spočítáme všechny relevantní pravděpodobnosti:

např.

$$\mathbf{X} = (X, Y)$$

$$P[(a < X \leq b) \wedge (c < Y \leq d)] = \\ = F_{\mathbf{X}}(b, d) - F_{\mathbf{X}}(a, d) - F_{\mathbf{X}}(b, c) + F_{\mathbf{X}}(a, c)\,.$$

Základní vlastnosti vícerozměrné distribuční funkce:

7.4. Věta. *Je-li $F(x_1,\dots,x_n)$ sdružená distribuční funkce náhodných veličin $X_1,\dots,X_n$, pak*

(i) $\lim_{x_1\to\infty,x_2\to\infty,\dots,x_n\to\infty} F(x_1,\dots,x_n)=1$

(ii) $$\lim_{x_1\to-\infty,x_2\to-\infty,\dots,x_n\to-\infty} F(x_1,\dots,x_n)=0$$

(iii) *F je zprava spojitá a neklesající v každé proměnné.*

Tyto vlastnosti nestačí k tomu, aby F byla distribuční funkcí. Musí splňovat složitejší podmínku pro hodnoty ve vrcholech vícerozměrných intervalů.

Jak spočítat marginální rozdělení vektoru

$$\mathbf{X}=(X_1,\dots,X_n)?$$

$$F_{X_1}(x)=\lim_{x_2\to\infty,x_3\to\infty,\dots,x_n\to\infty} F_{\mathbf{X}}(x,x_2,x_3,\dots,x_n)\,.$$

7.5. Příklad. $\mathbf{X}$ má rovnoměrné rozdělení na jednotkovém kruhu se středem v počátku. Nalezněte marginální rozdělení.

$$\mathbf{X} = (X, Y)$$

Pro $-1 < x < 1$.

$$F_X(x) = \frac{2}{\pi}\int_{-1}^{x} \sqrt{1-t^2}\,dt = \frac{2}{\pi}\left[\frac{1}{2}t\sqrt{1-t^2}+\frac{1}{2}arcsin\,t\right]_{-1}^{x} =$$

$$= \frac{x\sqrt{1-x^2}}{\pi} + \frac{arcsin\,x}{\pi} + \frac{1}{2}\,.$$

7.6. Definice. Náhodný vektor se nazývá diskrétní, jestliže všechny jeho složky mají diskrétní rozdělení.

Diskrétní vektor je dán pravděpodobnostní funkcí:

$$(\mathbf{x}_1, p_1);\ (\mathbf{x}_2, p_2);\ (\mathbf{x}_3, p_3), \dots$$

$$p_i > 0\,, \qquad \sum_{i=1}^{n} p_i = 1\,.$$

7.7. Příklad. Pravděpodobnost soustředěná ve vrcholech čtverce:

$$P[\mathbf{X}=(0,0)]=1/8, \qquad P[\mathbf{X}=(0,1)]=1/4,$$
$$P[\mathbf{X}=(1,0)]=1/8, \qquad P[\mathbf{X}=(1,1)]=1/2$$

tabulka:

Y/X	0	1
0	1/8	1/4
1	1/8	1/2

Diskrétní rozdělení pro X

$$P[X=0]=1/8+1/4=3/8$$
$$P[X=1]=1/8+1/2=5/8$$

$X \sim A(5/8)$.
Podobně $Y \sim A(3/4)$.

• Distribuční funkce diskrétního vektoru je po částech konstantní.

• Má-li $\mathbf{X}$ diskrétní rozdělení je

$$P[\mathbf{X} \in A] = \sum_{\{i \mid \mathbf{x}_i \in A\}} p_i \,.$$

• Marginální rozdělení diskrétního rozdělení má pravděpodobnostní funkci

$$P[X_1 = a] = \sum_{\{i \mid \mathbf{x}_i \in A\}} p_i \,,$$

kde $A = \{\mathbf{x}_i \mid (\mathbf{x}_i)_1 = a\}$.

7.8. Definice. Nechť $f(x_1, \dots, x_n) \geq 0$ je funkce definovaná na $\mathbb{R}^n$ s

$$\int\limits_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x} = 1\,.$$

Náhodný vektor $\mathbf{X} = (X_1, \dots, X_n)$ má *spojité rozdělení s hustotou* f, jestliže

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int\limits_{-\infty}^{x_1} \int\limits_{-\infty}^{x_2} \cdots \int\limits_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n\,.$$

•

$$f(x_1, \dots, x_n) = \frac{\partial^n F_{\mathbf{X}}(x_1, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

v bodech spojitosti funkce f

•

$$P[\mathbf{X} \in A] = \int\limits_A f(\mathbf{x})\, d\mathbf{x}\,.$$

Například: $\mathbf{X} = (X, Y)$

$$P[X \in (a, b > \wedge Y \in (c, d >] = \int\limits_a^b \int\limits_c^d f(x, y)\, dx dy\,.$$

7.9. Příklad. Rovnoměrné rozdělení na jednotkové kouli se středem v počátku má hustotu

$$f(x,y,z) = \begin{cases} \frac{1}{4/3\pi} & \text{pro } x^2+y^2+z^2 \le 1 \\ 0 & \text{jinak} \end{cases}$$

7.10. Příklad. Dvourozměrná distribuční funkce

$$F(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}) & \text{je-li } x,y > 0 \\ 0 & \text{jinak} \end{cases}$$

Pro $x, y > 0$

$$f(x,y) = \frac{\partial^2}{\partial x \partial y}(1-e^{-x})(1-e^{-y}) =$$
$$= e^{-x}e^{-y}\,.$$

(Jinak je $f(x,y)$ nulová.) Je správně neboť

$$f(x,y) = \int\limits_0^\infty \int\limits_0^\infty e^{-x}\,e^{-y} = 1\,.$$

Marginální rozdělení:

$\mathbf{X} = (X_1, \dots, X_n)$ má hustotu $f(x_1, \dots, x_n)$.
Pak X_1 má distribuční funkci:

$$F_{X_1}(x) = \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(n-1)\times} f(t_1, \dots, t_n)\, dt_1 \cdots dt_n =$$

$$= \int_{-\infty}^{x} \Bigg(\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(n-1)\times} f(t_1, \dots, t_n)\, dt_2 \cdots dt_n \Bigg) dt_1$$

Hustota tedy bude

$$f_{X_1}(x) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(n-1)\times} f(x, t_2, \cdots t_n)\, dt_2 \cdots dt_n \,.$$

7.11. Příklad. Dvourozměrné Gaussovo rozdělení
hustota:

$$f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\,.$$

$$X, Y \sim N(0,1)$$

$$F_{\mathbf{X}}(x,y) = \Phi(x)\Phi(y)\,.$$

Např.
$P[X \leq 3 \wedge Y \leq 5] = \Phi(3)\Phi(5) = 0,9986498158.$

Jaká je pravděpodobnost, že $(X,Y) \in A$, kde A je mezikruží

$$A = \{(x,y) \mid 4 \leq x^2 + y^2 \leq 9\}?$$

$$P = \frac{1}{2\pi} \int\int_A e^{-\frac{x^2+y^2}{2}}\, dxdy$$

Substituce:
$x = \varrho \sin\varphi, y = \varrho \cos\varphi$,
Jakobián: $dxdy = \varrho\, d\varrho\, d\varphi$.

$$P = \frac{1}{2\pi} \int_0^{2\pi} \int_2^3 e^{-\varrho^2/2} \varrho \, d\varrho \, d\varphi =$$

$$= \frac{1}{2\pi} \, 2\pi [-e^{-\varrho^2/2}]_2^3 = e^{-2} - e^{-9/2} = 0,12422 \, .$$

Nezávislost náhodných veličin

7.12. Definice. Náhodné veličiny $X_1, \dots, X_n$ na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$ jsou *nezávislé*, jestliže sdružená distribuční funkce vektoru

$$\mathbf{X} = (X_1, \dots, X_n)$$

je součinem marginálních distribučních funkcí.

7.13. Tvrzení. *Náhodné veličiny* $X_1, \dots, X_n$ *jsou nezávislé právě tehdy když*

$$P[(X_1 \in I_1) \wedge (X_2 \in I_2) \wedge \cdots \wedge (X_n \in I_n)] = \\ = P[X_1 \in I_1] \cdot P[X_2 \in I_2] \cdots P[X_n \in I_n]$$

pro všechny možné výběry intervalů $I_1, \dots, I_n \subset \mathbb{R}$.

Zdůvodnění pro $\mathbf{X} = (X, Y)$, kde X, Y jsou nezávislé :

$$P[X \in (a, b > \wedge Y \in (c, d >] = \\ = F_{\mathbf{X}}(b, d) - F_{\mathbf{X}}(a, d) - F_{\mathbf{X}}(b, c) + F_{\mathbf{X}}(a, c) = \\ = F_X(b)F_y(d) - F_X(a)F_Y(d) - F_X(b)F_Y(c) + F_X(a)F_Y(c) = \\ = (F_X(b) - F_X(a))(F_Y(d) - F_Y(c))\,.$$

7.14. Příklad. $\mathbf{X} = (X, Y)$ je rovnoměrné rozdělení na čtverci $< 0, 1 > \times < 0, 1 >$. Pak X a Y jsou nezávislé:
$0 \leq x, y \leq 1$

$$F_{\mathbf{X}}(x, y) = xy = F_X(x) \cdot F_Y(y)$$

7.15. Tvrzení. *Diskrétní náhodné veličiny* $X_1, \ldots, X_n$ *jsou nezávislé právě tehdy když všechny jevy* $[X_1 = x_1]$, $[X_2 = x_2], \ldots, [X_n = x_n]$, *kde* $x_1, \ldots, x_n \in \mathbb{R}$, *jsou nezávislé.*

7.16. Příklad. Pravděpodobnost soustředěná ve vrcholech čtverce:

$$P[\mathbf{X} = (0,0)] = 1/8, \qquad P[\mathbf{X} = (0,1)] = 1/4,$$
$$P[\mathbf{X} = (1,0)] = 1/8, \qquad P[\mathbf{X} = (1,1)] = 1/2$$

tabulka:

Y/X	0	1
0	1/8	1/4
1	1/8	1/2

X a Y nejsou nezávislé, protože

$$P[X = 0 \wedge Y = 0] = 1/8 \neq P[X = 0]P[Y = 0] = 3/8 \cdot 1/4\,.$$

$$\mathbf{X} = (X_1, \ldots, X_n)$$

náhodné veličiny v Bernoulliově schématu. Tyto veličiny jsou nezávislé.

Binomické rozdělení $Bi(n,p)$ je součtem n nezávislých alternativních rozdělení $A(p)$.

7.17. Tvrzení. *Spojité vícerozměrné rozdělení je rozdělení nezávislých veličin právě tehdy když sdružená hustota je součinem hustot marginálních.*

Důvod: hustota je derivací distribuční funkce.

7.18. Příklad. Je-li (X, Y) rovnoměrné rozdělení na jednotkovém kruhu, pak X a Y nejsou nezávislé, protože součin marginálních hustot je nenulový v každém bodě čtverce
$< 0, 1 > \times < 0, 1 >$.

7.19. Příklad. X a Y jsou nezávislé náhodné veličiny s rozděleními $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$. Jaká je hustota součinu?

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}}$$

Gaussovská plocha.

---

charakteristiky nezávislých veličin:

7.20. Věta. *Jsou-li* $X_1, \ldots, X_n$ *nezávislé náhodné veličiny na pravděpodobnostním prostoru* $(\Omega, \mathcal{A}, P)$, *pak*

$$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n)\,.$$

Důkaz: pro diskrétní rozdělení

X má pravděpodobnostní funkci $(\mathbf{x}_i, p_i)_{i \in I}$

Y má pravděpodobnostní funkci $(\mathbf{y}_j, p_j)_{j \in J}$

$$\begin{aligned} E(XY) = \sum_{i \in I, j \in J} x_i\, y_j P[X = x_i \wedge Y = y_j] = \\ = \sum_{i \in I, j \in J} x_i\, y_j P[X = x_i] P[Y = y_j] = \\ = \left(\sum_{i \in I} x_i P[Y = x_i]\right) \cdot \left(\sum_{j \in J} y_j P[Y = y_j]\right) = \\ = E(X)E(Y) \end{aligned}$$

7.21. Věta. *Jsou-li $X_1, \dots, X_n$ nezávislé náhodné veličiny na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$, pak*

$$\boxed{var(X_1 + X_2 + \dots + X_n) = var(X_1) + var(X_2) + \dots + var(X_n)\,.}$$

Důkaz pro $n = 2$
Víme, že $E(X_1X_2) = E(X_1)E(X_2)$.

$$\begin{aligned} var(X_1+X_2) &= E[(X_1+X_2)^2] - (E(X_1)+E(X_2))^2 = \\ &= E(X_1^2) + E(X_2^2) + 2E(X_1X_2) - \\ &- (E(X_1))^2 - (E(X_2))^2 - 2E(X_1)E(X_2) = \\ &E(X_1)^2 - (E(X_1))^2 + E(X_2)^2 - (E(X_2))^2 = \\ &= var(X_1) + var(X_2)\,. \end{aligned}$$

7.22. Příklad. Pro X s rozdělením $Bi(n,p)$ a Y s rozdělením $A(p)$ platí

$$var(X) = n\,var(Y) = np(1-p)\,.$$

Funkce nezávislých náhodných veličin

7.23. Příklad. Doba kdy lano vydrží zátěž se řídí exponeniálním rozdělením $Exp(1)$. Dvě lana zapojíme a) paralelně b) sériově. Určete rozdělení doby po kterou systém lan vydrží v obou případech a stanovte střední hodnotu těchto náhodných veličin.

X a Y jsou nezávislé náhodné veličiny s rozdělením $Exp(1)$.

a) $Z_{par} = \max(X,Y)$

b) $Z_{serie} = \min(X,Y)\,.$

a) paralelně

Distribuční funkce je nulová pro $t < 0$. Pro $t > 0$ je

$$F_{Z_{par}} = (1 - e^{-t})(1 - e^{-t}) = (1 - e^{-t})^2$$

Hustota pro $t > 0$:

$$f(t) = 2(1 - e^{-t})e^{-t}$$

Střední hodnota

$$EZ_{par} = 2\int_0^\infty te^{-t}\,dt - 2\int_0^\infty te^{-2t}\,dt = 1,5\,.$$

a) sériově

Distribuční funkce je nulová pro $t < 0$. Pro $t > 0$ je

$$F_{Z_{par}} = 1 - e^{-t}e^{-t} = 1 - e^{-2t}$$

Hustota pro $t < 0$:

$$g(t) = 2e^{-2t}$$

Střední hodnota

$$EZ_{serie} = 2\int_0^\infty e^{-2t}\,dt = 0,5\,.$$

Rozdělení součtu spojitých nezávislých náhodných veličin

$$Z = X + Y\,,$$

kde X a Y jsou nezávislé, X má hustotu $f(x)$ a Y má hustotu $g(x)$.
Úlohou je stanovit hustotu náhodné veličiny Z.

Sdružená hustota vektoru (X, Y) je funkce $h(x, y) = f(x)g(y)$.

$$P[Z \le z] = P[X+Y \le z] = \iint\limits_{\{(x,y)\,|\,x+y\le z\}} f(x)g(y)\,dxdy =$$
$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} f(x)g(y)\,dy\right)dx\,.$$

Substituce ve vnitřním integrálu
$y = v - x$, $dv = dy$

$$= \int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{z} f(x)\, g(v-x)\, dv \right) dx =$$

$$= \int\limits_{-\infty}^{z} \underbrace{\left(\int\limits_{-\infty}^{\infty} f(x) g(v-x)\, dx \right)}_{hustota\, h(v)} dv$$

Závěr: Hustota součtu $X + Y$ je funkce

$$\boxed{h(y) = \int\limits_{-\infty}^{\infty} f(x) g(y-x)\, dx\,.}$$

Terminologie : funkce h je konvolutivní součin funkcí f a g.

7.24. Příklad. Čas do první poruchy daného zařízení se řídí exponenciálním zákonem $Exp(\lambda)$. Náhodná veličina Z je čas do druhé poruchy. Určete její hustotu.

$$Z = X + Y\,,$$

kde X a Y jsou nezávislé s rozdělením $Exp(\lambda)$. Hustota $h(y)$ je

$$h(y) = \int_{-\infty}^{\infty} f(x)g(y-x)\,dx = \int_0^y \lambda e^{-\lambda x}\lambda e^{-\lambda(y-x)}\,dx =$$

$$= \lambda^2 e^{-\lambda y}\int_0^y dx = \lambda^2 y e^{-\lambda y}$$

pro $y \geq 0$ a nula jinak.

7.25. Příklad. Souprava metra přijíždí kdykoliv během jedné minuty. Dvakrát přestupujeme. Jaká je hustota čekací doby? Která hodnota je nejvíc preferována?

$$Z = X + Y\,,$$

kde X a Y jsou nezávislé náhodné veličiny s rozdělením $R < 0, 1 >$.

$$h(y) = \int\limits_{-\infty}^{\infty} f(x)g(y-x)\,dx\,,$$

kde f a g jsou chrakteristické funkce intervalu $< 0, 1 >$.
Tento integrál je délkou průniku intervalu $< 0, 1 >$ s intervalem $< y - 1, y >$. Tedy

$$h(y) = \begin{cases} 0 & y \leq 0 \\ y & 0 \leq y \leq 1 \\ 2 - y & 1 \leq y \leq 2 \\ 0 & y \geq 2 \end{cases}$$

Trojúhelníkové rozdělení, preferována je čekací doba 1.

8 Kovariance a korelace náhodných vektorů

Je-li $\mathbf{X}$ náhodný vektor, je vektor středních hodnot vektor

$$E\mathbf{X} = (EX_1, EX_2, \ldots EX_n)\,.$$

Nepopisuje interakci mezi náhodnými veličinami.

8.1. Definice. Nechť X a Y jsou náhodné veličiny na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$, které mají rozptyl. *Kovariance* náhodných veličin X a Y je definována

$$cov(X, Y) = E[(X - EX)(Y - EY)]\,.$$

- $cov(X,X) = var(X)$
- $cov(X,Y) = cov(Y,X)$
- $cov(X,Y) = E(XY) - E(X)E(Y)$:

$$E[(X-EX)(Y-EY)] = E(XY)-2EX\cdot EY+EX\cdot EY = \\ = E(XY) - E(X)E(Y)$$

Pro výpočet potřebujeme:

1. Diskrétní vektor (X,Y) s pravděpodobnostní funkcí $((x_i,y_i);p_i)_{i\in I}$.

$$EX = \sum_{i\in I} x_i\, p_i$$

$$EY = \sum_{i\in I} y_i\, p_i$$

$$E(XY) = \sum_{i\in I} x_i y_i\, p_i$$

---

1. Spojitý vektor (X, Y) se sdruženou hustotou $f(x, y)$.

$$EX = \int_{\mathbb{R}^2} x f(x,y)\,dxdy$$

$$EY = \int_{\mathbb{R}^2} y f(x,y)\,dxdy$$

$$E(XY) = \int_{\mathbb{R}^2} xy f(x,y)\,dxdy$$

8.2. Příklad. X má rovnoměrné rozdělení na intervalu $< 0, 1 >$. Určete $cov(X, X^2)$.

$$cov(X, X^2) = EX^3 - EX \cdot EX^2$$

$$EX^3 \;=\; \int_0^1 x^3\,dx = \frac{1}{4}$$

$$EX^2 \;=\; \int_0^1 x^2\,dx = \frac{1}{3} \qquad EX = \frac{1}{2}\,.$$

$$cov(X, X^2) = \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}\,.$$

8.3. Příklad. $\mathbf{X} = (X, Y)$ má rovnoměrné rozdělení na jednotkovém kruhu. Stanovte $cov(X, Y)$.

$$\begin{aligned} EX &= \iint_K x\,dxdy = 0 \\ E(XY) &= \iint_K xy\,dxdy = 0\,. \end{aligned}$$

$$cov(X, Y) = 0\,.$$

8.4. Věta. *Jsou-li X a Y nezávislé náhodné veličiny je $cov(X,Y) = 0$.*

Důkaz:

Jsou-li X a Y nezávislé, pak $E(XY) = EX \cdot EY$.

Veličiny X, Y souřadnice rovnoměrného rozdělení na jednotkovém kruhu mají nulovou kovarianci, a přesto nejsou nezávislé.

Normování náhodné veličiny X : $\frac{X-EX}{\sqrt{varX}}$ má nulovou střední hodnotu a jednotkový rozptyl.

$$cov\left(\frac{X-EX}{\sqrt{varX}}, \frac{Y-EY}{\sqrt{varY}}\right) = \frac{cov(X,Y)}{\sqrt{varX}\sqrt{varY}}\,.$$

8.5. Definice. Předpokládejme, že X a Y jsou náhodné veličiny s nenulovým rozptylem. *Korelace* $\varrho(X,Y)$ je definována vztahem

$$\varrho(X,Y) = \frac{cov(X,Y)}{\sqrt{varX}\sqrt{varY}}\,.$$

8.6. Příklad. (X,Y) má diskrétní rozdělení s pravděpodobnostní funkcí

$$((x_1,y_1),1/3);((x_2,y_2),1/3);((x_3,y_3),1/3);$$

Předpokládejme dále, že

$$x_1+x_2+x_3=0 \qquad y_1+y_2+y_3=0\,.$$

$$cov(X,Y)=\frac{\frac{1}{3}x_1y_1+\frac{1}{3}x_2y_2+\frac{1}{3}x_3y_3}{\sqrt{\frac{1}{3}(x_1^2+x_2^2+x_3^2)}\sqrt{\frac{1}{3}(y_1^2+y_2^2+y_3^2)}}=$$
$$=\frac{x_1y_1+x_2y_2+x_3y_3}{\sqrt{x_1^2+x_2^2+x_3^2}\sqrt{y_1^2+y_2^2+y_3^2}}=\cos\varphi\,,$$

kde φ je úhel mezi vektory
$\mathbf{x}=(x_1,x_2,x_3)$ a $\mathbf{y}=(y_1,y_2,y_3)$.
Pokud je tedy $cov(X,Y)=1$ je $\mathbf{y}$ kladným násobkem $\mathbf{x}$, pokud je $cov(X,Y)=-1$ je $\mathbf{y}$ záporným násobkem $\mathbf{x}$

8.7. Věta. *Pro korelaci $\varrho(X, Y)$ náhodných veličin X a Y platí*

$$|\varrho(X, Y)| \leq 1 \,.$$

Přitom $\varrho(X, Y) = 1$ právě tehdy když hodnoty X a Y leží s pravděpodobností 1 na jedné přímce s kladnou směrnicí.
$\varrho(X, Y) = -1$ právě tehdy když hodnoty X a Y leží s pravděpodobností 1 na jedné přímce se zápornou směrnicí.

Důkaz:

Pro všechna $t \in \mathbb{R}$ máme

$$0 \leq var\left(\frac{X - EX}{\sqrt{varX}} + t\,\frac{Y - EY}{\sqrt{varY}}\right) = \\ = 1 + 2t\varrho(X, Y) + t^2 \,. \quad (2)$$

Diskuse kvadratického výrazu:

$$4\varrho^2(X, Y) - 4 \leq 0$$

a tedy

$$|\varrho(X, Y)| \leq 1 \,.$$

Předpokládejme, že $cov(X,Y) = 1$. Pak v nerovnosti (2) nastane rovnost a to může nastat právě když $t = -1$. Dosazením $t = -1$ pak vede k

$$\frac{X - EX}{\sqrt{varX}} - \frac{Y - EY}{\sqrt{varY}} = konst\,.$$

Tedy

$$Y = \frac{\sqrt{varY}}{\sqrt{varX}} \cdot X + \text{konstanta}\,.$$

$\varrho(X,Y)$ je míra lineární závislosti X a Y.

9 Asymptotické vlastnosti náhodných veličin

Větou asymptotického typu byla již věta Poissonova: binomické rozdělení s počtem pokusů jdoucím k nekonečnu a střední hodnotou jdoucí k λ se blíží Poissonově rozdělení $Po(\lambda)$.

Základní otázka: Co se děje s $Bi(n,p)$, jestliže $n \to \infty$ a p se nemění?

Orientační odhady poskytuje Čebyševova nerovnost.

9.1. Věta. Čebyševova nerovnost
Nehť X je náhodná veličina s $E(X^2) < \infty$. Potom platí

$$\boxed{P[|X| \geq \varepsilon] \leq \frac{E(X^2)}{\varepsilon^2}\,.}$$

Speciálně, má-li X rozptyl, pak

$$\boxed{P[\,|X - EX| \geq \varepsilon] \leq \frac{var(X)}{\varepsilon^2}\,.}$$

Důkaz pro spojité rozdělení s hustotou $f(x)$.

$$P[\,|X| \geq \varepsilon] = \int\limits_{\{x \mid |x| \geq \varepsilon\}} f(x)\,dx$$

$$E(X^2) = \int\limits_{-\infty}^{\infty} x^2\,f(x)\,dx = \int\limits_{\{x \mid |x| \geq \varepsilon\}} x^2\,f(x)\,dx + \int\limits_{-\varepsilon}^{\varepsilon} x^2\,f(x)\,dx \geq$$

$$\geq \int\limits_{\{x \mid |x| \geq \varepsilon\}} \varepsilon^2 f(x)\,dx = \varepsilon^2 P[\,|X| \geq \varepsilon]\,.$$

9.2. Důsledek. *Je-li X_n náhodná veličina s rozdělením $Bi(n,p)$, pak pro každé $\varepsilon > 0$ platí*

$$P\left[\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right] \to 0 \qquad \textit{pro } n \to \infty\,.$$

Důkaz:

$$E\left(\frac{X_n}{n}\right) = \frac{np}{n} = p$$

$$var\left(\frac{X_n}{n}\right) = \frac{1}{n^2}np(1-p) \leq \frac{1}{4n}$$

Dle Čebyševovy nerovnosti

$$P\left[\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right] \leq \frac{var(\frac{X_n}{n})}{\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}\,.$$

9.3. Příklad. Kolika respondentů je třeba se zeptat, abychom odhadli volební preference p s tolerancí 1% s pravděpodobností alespoň 90% ?

n ... počet osob
$X_n \ldots Bi(n,p)$ (počet osob volících danou stranu)
$p \approx \frac{X_n}{n}$ (aproximace)

Chceme:

$$P\left[\left|\frac{X_n}{n} - p\right| < 0,01\right] \geq 0,9\,.$$

Nebo-li

$$\underbrace{P\left[\left|\frac{X_n}{n} - p\right| \geq 0,01\right]}_{\leq \frac{1}{4n(0,01)^2}} \leq 0,1$$

$$\frac{1}{4n(0,01)^2} \le 0,1$$

$$n \ge \frac{10^5}{4} = 25000.$$

Je pesimistický, nicméně jistý horní odhad.

9.4. Věta. Centrální limitní věta
Nehť $X_1, X_2, X_3, \dots$ je posloupnost nezávislých náhodných veličin, pro které platí

$$EX_i = \mu\,, \quad varX_i = \sigma^2\,, \quad E|X_i|^3 < \infty\,.$$

pro všechna $i \in \mathbb{N}$. Pro veličinu S_n (normovaný součet),

$$S_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}\,,$$

platí

$$\boxed{\lim_{n\to\infty} P[\,S_n \le x] = \Phi(x)}\,,$$

pro všechna $x \in \mathbb{R}$.

Tedy $X_1 + X_2 + \cdots + X_n$ má přibližně rozdělení $N(n\mu, n\sigma^2)$.

9.5. Věta. Moaivrova-Laplaceova věta
Pro náhodnou veličinu X_n s rozdělením $Bi(n,p)$ platí

$$\lim_{n\to\infty} P\left[\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right] = \Phi(x)$$

pro každé $x \in \mathbb{R}$.

Důkaz:

Binomické rozdělení je součtem nezávislých alternativních rozdělení.

$Bi(n,p) \approx N(np, np(1-p))$
Empiriké pravidlo pro velikost n :
$np \geq 5, n(1-p) \geq 5.$

9.6. Příklad. 250 krát hodíme symetrickou mincí. Určete pravděpodobnost, že rub padne 100 až 150 krát.

$$\frac{X_{250} - 125}{\sqrt{250 \cdot 0,25}} \approx N(0,1)$$

$$\begin{aligned} P[100 \leq X_{250} \leq 150] = \\ = P\left[\frac{-25}{\sqrt{250 \cdot 0,25}} \leq \frac{X_{250} - 125}{\sqrt{250 \cdot 0,25}} \leq \frac{25}{\sqrt{250 \cdot 0,25}}\right] = \\ = 2\Phi\left(\frac{25 \cdot 2}{\sqrt{250}}\right) - 1 = 2\Phi(\sqrt{10}) - 1 \doteq 2\Phi(3,16) - 1 \doteq \\ = 0,99842\,. \end{aligned}$$

Přesně: $\sum_{k=125}^{150} \binom{250}{k} \frac{1}{2^{250}}$

9.7. Příklad. Kolika respondentů je třeba se zeptat, abychom odhadli volební preference p s tolerancí 1% a to s pravděpodobností alespoň 90%.

$$P\left[\left|\frac{X_n}{n} - p\right| \leq 0,01\right] \geq 0,9$$

$$P[\,|X_n - np| \le 0,01n] = P\left[\frac{|X_n - np|}{\sqrt{np(1-p)}} \le \frac{0,01n}{\sqrt{np(1-p)}}\right] \doteq$$

$$\doteq 2\Phi\left(\frac{0,01\cdot\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \ge 2\Phi\left(\frac{0,01\cdot\sqrt{n}}{1/2}\right) - 1 =$$

$$= 2\Phi(0,02\sqrt{n}) - 1\,.$$

Volíme n tak aby

$$2\Phi(0,02\sqrt{n}) - 1 \ge 0,9$$

$$\Phi(0,02\sqrt{n}) \ge 0,95$$

$$0,02\sqrt{n} \ge u_{0,95}$$

$$\sqrt{n} \ge \frac{u_{0,95}}{0,02}$$

$$n \ge \left(\frac{u_{0,95}}{0,02}\right)^2 = \left(\frac{1,644}{0,02}\right)^2 \doteq 6763,8586$$

Pro srovnání při 5% toleranci potřebujeme 270 respondentů.

10 Statistika

Základní výběrové statistiky

10.1. Definice. Náhodný výběr z rozdělení s distribuční funkcí $F(x)$ je vektor

$$\mathbf{X} = (X_1, \dots, X_n)\,,$$

kde $X_1, \dots, X_n$ jsou nezávislé náhodné veličiny mající stejnou distribuční funkci $F(x)$.

Základní výběrové statistiky:

Výběrový průměr

$$\overline{X_n} = \frac{1}{n}(X_1 + X_2 + \cdots X_n)$$

Výběrový rozptyl

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$$

Výběrová směrodatná odchylka

$$S_n$$

Výběrové maximum

$$\max(X_1, \dots, X_n)$$

Výběrové minimum

$$\min(X_1, \dots, X_n)$$

Výběrové rozpětí

$$\max(X_1, \dots, X_n) - \min(X_1, \dots, X_n)$$

Výběrový průměr

$EX = \mu, var(X) = \sigma^2$

$$E = E(\frac{1}{n}(X_1 + \cdots + X_n)) = \frac{n\mu}{n} = \mu\,.$$

$$var(\overline{X_n}) = var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}\,.$$

$$\boxed{\begin{aligned} E(\overline{X_n}) &= \mu \\ var(\overline{X_n}) &= \frac{\sigma^2}{n} \end{aligned}}$$

Důležité je, že $var(X_n) \to 0$ pro $n \to \infty$.

Čebyševova nerovnost:

$$P[\,|\overline{X_n} - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}\,.$$

Výběrový rozptyl

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2\,.$$

pomocný výpočet:

$$\sum_{i=1}^{n}(X_i - \overline{X_n})^2 = \sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X_n} + \overline{X_n}^2) =$$
$$= \sum_{i=1}^{n} X_i^2 - 2\overline{X_n}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}\overline{X_n}^2 =$$
$$= \sum_{i=1}^{n} X_i^2 - 2n\overline{X_n}^2 + n\overline{X_n}^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X_n}^2\,.$$

Výpočet střední hodnoty:

$$(n-1)E(S_n^2) = E\left[\sum_{i=1}^{n} X_i^2 - n\overline{X_n}^2\right] = \sum_{i=1}^{n} E(X_i^2) - nE(\overline{X_n}^2) =$$

$$= \sum_{i=1}^{n}(var(X_i)+E^2(X_i)) - n(var(\overline{X_n})+E^2(\overline{X_n})) =$$

$$= n\sigma^2 + n\mu^2 - n\frac{\sigma^2}{n} - n\mu^2 = (n-1)\sigma^2$$

$$\boxed{E(S_n^2) = \sigma^2\,.}$$

Výběrové statistiky odvozené od normálního rozdělení.

10.2. Tvrzení. *Výběrový průměr z rozdělení* $N(\mu, \sigma^2)$ *má rozdělení* $N(\mu, \frac{\sigma^2}{n})$.

Důkaz je založen na tom, že součet nezávislých normálních rozdělení je normální rozdělení, což se dá dokázt pomocí konvoluce gaussovských funkcí.

Pro popis S_n^2 potřebujeme zavést nové rozdělení

10.3. Definice. Nechť $X_1, \dots, X_n$ jsou nezávislé veličiny s normovaným normálním rozdělením. Rozdělení χ^2 o n stupních volnosti je rozdělení náhodné veličiny

$$Y = X_1^2 + X_2^2 + \dots + X_n^2$$

Značení : χ_n^2
kvantily: $\chi_{n,\alpha}^2$ $\chi_n^2(\alpha)$

Má-li X rozdělení χ_n^2, pak

$$\boxed{EX = n \qquad var(X) = 2n\,.}$$

10.4. Příklad. X, Y, Z jsou nezávislé náhodné veličiny s rozdělením $N(0,1)$. Stanovte r tak, aby

$$P[\sqrt{X^2 + Y^2 + Z^2} \le r] = 0,995$$

(Legenda: složky rychlosti molekul plynu)

$$r = \sqrt{\chi_3^2(0,995)} = \sqrt{12,838} = 3,58$$

Základní věta klasické statistiky

10.5. Věta. *Nechť* $\mathbf{X} = (X_1, \dots, X_n)$ *je náhodný výběr z rozdělení* $N(\mu, \sigma^2)$. *Pak*

(i) $\overline{X_n}$ *a* S_n^2 *jsou nezávislé náhodné veličiny*

(ii) $\frac{(n-1)S_n^2}{\sigma^2}$ *má rozdělení* χ^2_{n-1}.

Statistická varianta normování veličiny:

$$\frac{\overline{X_n} - \mu}{S_n/\sqrt{n}}$$

10.6. Definice. Nechť $\mathbf{X} = (X_1, \dots, X_n)$ je náhodný výběr z rozdělení $N(\mu, \sigma^2)$. *Studentovo rozděleni s* $n-1$ *stupni volnosti* je rozdělení náhodné veličiny

$$\frac{\overline{X_n} - \mu}{S_n/\sqrt{n}}$$

Značení: t_{n-1}, kvantily t_{n-1}, α; $t_{n-1}(\alpha)$.

W. Gosset – pseudonym Student, Studentovo rozdělení t_n má sudou hustotu, $f_n(x)$, která pro $n \to \infty$ konverguje bodově k $\Phi(x)$. (Aproximuje se obvykle pro $n \geq 31$)

Intervalové odhady

Lokalizace neznámého parametru pomocí dat

10.7. Definice. Nechť $\mathbf{X} = (X_1, \dots, X_n)$ je náhodný výběr z rozdělení s neznámým parametrem θ a $\alpha \in (0, 1)$.

(i) Interval $< T_D(\mathbf{X}), T_H(\mathbf{X}) >$ se nazývá $100(1-\alpha)\%$ oboustranným intervalem spolehlivosti parametru θ jestliže

$$P[\, T_D(\mathbf{X}) \leq \theta \leq T_H(\mathbf{X})\,] = 1 - \alpha$$

(ii) Interval $< T_D(\mathbf{X}), \infty)$ se nazývá dolním $100(1-\alpha)\%$ intervalem spolehlivosti parametru θ, jestliže

$$P[\theta \geq T_D(\mathbf{X})] = 1 - \alpha$$

(iii) Interval $(-\infty, T_H(\mathbf{X}) >$ se nazývá horním $100(1-\alpha)\%$ intervalem spolehlivosti parametru θ, jestliže

$$P[\theta \leq T_H(\mathbf{X})] = 1 - \alpha$$

Pro odvození intervalových odhadů pro parametry μ a σ^2 používáme statistiky

$$Z = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$T = \frac{\overline{X_n} - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$$

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

10.8. Věta. *Nechť* $\mathbf{X} = (X_1, \dots, X_n)$ *je náhodný výběr z rozdělení* $N(\mu, \sigma^2)$. *Pak*

(i) $$P[\,\overline{X_n} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X_n} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\,] = 1 - \alpha$$

(ii) $$P[\,\overline{X_n} - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \le \mu \le \overline{X_n} + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}\,] = 1 - \alpha$$

Důkaz: (i) $Z = \frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$P[u_{\alpha/2} \le Z \le u_{1-\alpha/2}] = 1 - \alpha\,.$$

$$\frac{\sigma}{\sqrt{n}}u_{\alpha/2} \le \overline{X_n} - \mu \le \frac{\sigma}{\sqrt{n}}u_{1-\alpha/2}$$

(ii) Totéž s využitím statistiky $T = \frac{\overline{X_n}-\mu}{S_n/\sqrt{n}} \sim t_{n-1}$.

10.9. Příklad. Je měřena výška 16 rostlin. Průměr naměřených hodnot je 72,5 cm, výběrová směrodatná odchylka je 4,5 cm. Nalezněte 90% interval spolehlivosti pro střední výšku.

$1-\alpha = 0,9; \alpha = 0,1;\ 1-\alpha/2 = 0,95$

$t_{15}(0,95) = 1,75$

$r = 1,75\frac{4,5}{\sqrt{16}} = 1,97$

$$(70,53; 74,47)$$

10.10. Příklad. $X \sim N(\mu, \sigma^2)$. Při padesáti měření byla získána směrodatná odchylka

$$S_{50} = \sqrt{2,192}$$

Určete horní 95% interval spolehlivosti pro σ^2.

$$\frac{(n-1)\sigma^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$P\left[\frac{(n-1)S_n^2}{\sigma^2} \geq \chi^2_{n-1,\alpha}\right] = 1-\alpha$$

$$P\left[\sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha}}\right] = 1-\alpha$$

$$\chi^2_{49}(0,05) = 33,930$$

$$\frac{49\cdot 2,192}{33,930} = 3,166\,.$$

Interval: (0, 3,166).

Přibližné intervalové odhady: $\overline{X_n}$ má podle Centrální limitní věty pro velké n přibližně normální rozdělení:

$$\frac{X_1+X_2+\cdots+X_n-n\mu}{\sqrt{n}\sigma} = \frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}$$

má přibližně rozdělení $N(0,1)$. Dá se použít intervalový odhad pro normální rozdělení:

$$\boxed{\overline{X_n} - u_{1-\alpha/2}\frac{S_n}{\sqrt{n}} \leq E(X) \leq \overline{X_n} + u_{1-\alpha/2}\frac{S_n}{\sqrt{n}}}$$

s pravděpodobností $1-\alpha$.

U alternativního rozdělení $A(p)$ se navíc aproximuje σ^2 hodnotou

$$\overline{X_n}(1-\overline{X_n})$$

$$\boxed{\overline{X_n} - u_{1-\alpha/2}\frac{\sqrt{\overline{X_n}(1-\overline{X_n})}}{\sqrt{n}} \le p \le \overline{X_n} + u_{1-\alpha/2}\frac{\sqrt{\overline{X_n}(1-\overline{X_n})}}{\sqrt{n}}}$$

s pravděpodobností $1-\alpha$.

10.11. Příklad. Při průzkumu se zjistilo, že z 1500 oslovených osob poslouchá rádiovou stanici kvůli hudbě 63%. Určete 98% interval spolehlivosti pro skutečné procento takovýchto posluchačů.

$$\overline{X}_{1500} = 0,63$$

$$r = u_{0,99} \cdot \sqrt{\frac{0,63 \cdot 0,37}{1500}} = 0,029$$

interval

$$(60,1\%; 65,9\%)$$

11 Testování statistických hypotéz

statistické rozhodování při neúplné informaci, vždy s jistým rizikem.

H_0 ... nulová hypotéza:

$k-$tice parametrů $(\theta_1, \ldots, \theta_k) \in A \subset \mathbb{R}^n$

H_1 ... alternatvní hypotéza:

$k-$tice parametrů $(\theta_1, \ldots, \theta_k) \in B \subset \mathbb{R}^n$

$$A \cap B = \emptyset$$

kritický obor:

$W \subset \mathbb{R}^n$

naměřená data: $\mathbf{X} = (X_1, \ldots, X_n)$.

H_0 zamítneme ve prospěch H_1 pokud $\mathbf{X} \in W$, volí se tak, aby

$$P[\mathbf{X} \in W | H_0] \le \alpha$$

α ... hladina významnosti

standardy: $\alpha = 0,5; 0,01$

Chyba prvního druhu: H_0 platí a test ji zamítá

Chyba druhého druhu: H_0 neplatí a test ji nezamítá

11.1. Příklad. V roce 1951 byla naměřena průměrná výška 10 letých chlapců v ČSSR 136,1 cm. V roce 1961 se u 15 náhodně vybraných 10 letých chlapců zjistila průměrná výška

$$\overline{X_{15}} = 139,133cm$$

V obou případech je $\sigma = 6,4cm$

Vzrostla výška?

Výška v r. 1961, $N(\mu, 6,4^2)$

$$H_0 : \mu = 136,1 = \mu_0$$

$$H_1 : \mu > 136,1$$

Hypotézu H_0 zamítneme pokud $\overline{X_{15}}$ bude příliš velký, tj. pokud $\overline{X_{15}} > c$, kde c je vhodně zvolená konstanta.

Platí-li H_0, pak $\overline{X_{15}}$ má rozdělení $N(\mu_0, \frac{6{,}4^2}{15})$. Tedy

$$Z = \frac{\overline{X_{15}} - \mu_0}{\sigma/\sqrt{15}} \sim N(0,1)\,.$$

Zamítáme na hladině α jestliže $Z > u_{1-\alpha}$

Volme $\alpha = 0,05$. $u_{0,95} \doteq 1,64$

$$Z = \frac{139,133 - 136,1}{6,4/\sqrt{15}} \doteq 1,835$$

Protože $Z > u_{0,95}$ zamítáme hypotézu H_0 na hladině významnosti 5%.

Kritický obor: $\{\mathbf{x} \in \mathbb{R}^{15} \mid \frac{1}{15}(x_1 + \cdots + x_{15}) \geq c\}$

$$c = 136,1 + u_{0,95} \cdot 6,4\sqrt{15} = 136,1 + 1,838 = 137,938$$

Lze volit mnoho kritických oborů. Existuje matematický výsledek (Neymannovo-Pearsonovo lemma) říkající, že tento test má největší sílu.
Síla testu: pravděpodobnost s jakou zamítneme nulovou hypotézu když platí hypotéza alternativní.

Vztah mezi chybou prvního a druhého druhu α a β. Pokud H_0 neplatí, pak $\alpha \to 0$ implikuje $\beta \to 1$.

Testy o střední hodnotě normálního rozdělení

$$\mathbf{X} = (X_1, \ldots, X_n)$$

… náhodný výběr z rozdělení $N(\mu, \sigma^2)$.

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$...oboustranný test

$H_2 : \mu > \mu_0$...jednostranný test

$H_3 : \mu < \mu_0$...jednostranný test

Z-test

známe σ

$$Z = \frac{\overline{X_n} - \mu_0}{\sigma/\sqrt{n}}$$

má za podmínky H_0 rozdělení $N(0,1)$.

H_0 zamítneme na hladině významnosti α

a) proti H_1: pokud $|Z| > u_{1-\alpha/2}$

b) proti H_2: pokud $Z > u_{1-\alpha}$

a) proti H_3: pokud $Z < u_\alpha = -u_{1-\alpha}$

---

t-test

neznáme σ

$$T = \frac{\overline{X_n} - \mu_0}{S_n/\sqrt{n}}$$

má za podmínky H_0 rozdělení t_{n-1}.

H_0 zamítneme na hladině významnosti α

a) proti H_1: pokud $|T| > t_{n-1}(1-\alpha/2)$

b) proti H_2: pokud $T > t_{n-1}(1-\alpha)$

a) proti H_3: pokud $T < t_{n-1}(\alpha) = -t_{n-1}(1-\alpha)$

Pro velké n aproximujeme t_{n-1} rozdělením $N(0,1)$.

11.2. Příklad. Automat plní krabice práškem. Norma je 2kg. Náhodně bylo vybráno 6 krabic. Zjistily se následující odchylky od normy v dkg.

$$-5, 1, -1, -8, 7, -6$$

Testujeme správnost funkce automatu.

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0\,.$$

$n = 6$, $\overline{X_6} = -2$, $S_6^2 = 30,4$

$$T = \frac{-2 - 0}{\sqrt{30,4}/\sqrt{6}} \doteq -0,889\,.$$

$$t_5(0,975) = 2,5706$$

$$|T| < 2,5706$$

Závěr: H_0 nezamítáme.

Párový t-test: Sledujeme související veličiny, předpokládáme, že rozdíl má normální rozdělení.

11.3. Příklad.

váha osob před dietou	82	70	91
váha osob po dietě	81	69,5	89

Má dieta efekt ?

Rozdíly před a po:

1— 0,5— 2

$$H_0 : \mu = 0$$

$$H_1 : \mu > 0$$

$$\overline{X_3} = 1,167$$

$$S_3^2 = 0,583$$

$$T = \frac{1,167}{S_3/\sqrt{3}} = 2,647\,.$$

$$t_2(0,95) = 2,92$$

Závěr: nezamítáme nulovou hypotézu, pokles váhy neni průkazný. (I když se zdá být ztráta váhy velká, máme málo pokusných osob).

Asymtotický test proporce

$A(p)$ ... alternativní rozdělení

Centrální limitní věta: $\overline{X_n}$ se aproximuje rozdělením $N(p, \frac{p(1-p)}{n})$.

$$H_0 : p \leq p_0$$

$$H_1 : p > p_0$$

H_0 zamítáme, jestliže $\overline{X_n} > c$. Volíme vhodně c tak, aby

$$P[\overline{X_n} > c | H_0] \leq \alpha$$

$$P\left[\overline{X_n} > c | H_0\right] \doteq 1 - \Phi\left(\frac{c-p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}}\right) = \\ = 1 - \Phi\left(\frac{\sqrt{n}(c-p)}{\sqrt{p(1-p)}}\right)$$

$$\sqrt{p(1-p)} \leq 1/2 \qquad c - p \geq c - p_0$$

implikuje

$$\frac{\sqrt{n}(c-p)}{\sqrt{p(1-p)}} \geq 2\sqrt{n}(c - p_0)$$

Tedy

$$1-\Phi\left(\frac{\sqrt{n}(c-p)}{\sqrt{p(1-p)}}\right) \leq 1-\Phi(2\sqrt{n}(c-p_0))$$

Podmmínce vyhovíme, jestliže

$$1-\Phi(2\sqrt{n}(c-p_0)) \leq \alpha$$

$$\Phi(2\sqrt{n}(c-p_0)) \geq 1-\alpha$$

$$2\sqrt{n}(c-p_0) \geq u_{1-\alpha}$$

$$c \geq p_0 + \frac{u_{1-\alpha}}{2\sqrt{n}}$$

Závěr: H_0 zamítáme ve prospěch H_1 jestliže

$$\boxed{\overline{X_n} \geq p_0 + \frac{u_{1-\alpha}}{2\sqrt{n}}}$$

11.4. Příklad. Průzkum zahrnuje 1600 osob. Kolik procent z tohoto vzorku má daná koalice získat hlasů, abychom na hladině významnosti 1% potvrdili hypotézu, že koalice vyhraje volby.

$$\overline{X_{1600}} \geq 0,5 + \frac{u_{0,99}}{80} \doteq 0,5 + \frac{2,326}{80} = 0,529\,.$$

Musíme tedy získat $0,529 \cdot 1600 \doteq 847$ hlasů.
Musíme vždy získat o asi $2,9\%$ více než je daná mez.

Test rozptylu normálního rozdělení

$\mathbf{X} = (X_1, \ldots, X_n) \ldots$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 > \sigma_0^2$$

Využíváme statistiku

$$S = \frac{(n-1)S_n^2}{\sigma_0^2}$$

která má za předpokladu H_0 rozdělení χ^2_{n-1}.

Hypotézu H_0 zamítáme ve prospěch H_1 při hladině významnosti α jestliže

$$\frac{(n-1)S_n^2}{\sigma_0^2} > \chi^2_{n-1}(1-\alpha)\,.$$

Testy dobré shody

$(\Omega, \mathcal{A}, P)$ ... pravděpodobnostní prostor
$A_1, \dots, A_k$ ... úplný systém jevů

$$P(A_i) = p_i \qquad \sum_{i=1}^{k} p_i = 1\,.$$

A_1	A_2	$\cdots$	A_k
p_1	p_2	$\cdots$	p_k

Odpovídá rozdělení dat do disjunktních tříd. p_i jsou apriorní pravděpodobnosti. Testujeme jejich shodu s empirickými daty.

$$H_0 : P(A_i) = p_i$$

$(k = 2)$ máme vpodstatě alternativní rozdělení.

Učiníme sérii n pokusů, v každém indikujeme jeden z jevů $A_1, \dots, A_k$. Počítáme kolikrát který jev nastane. Při $k = 2$ tak máme binomické rozdělení.

Značení:
n ... počet nezávislých pokusů

O_i ... počet výskytů jevu A_i v sérii n pokusů.

Náhodná veličina – empirická četnost.

$n\,p_i$ ... teoretická četnost

($k = 2$ pak $O_i \sim Bi(n, p_i)$)

Testujeme shodu $np_i \approx O_i$

11.5. Věta. *Náhodná veličina*

$$\boxed{\sum_{i=1}^{k} \frac{(O_i - np_i)^2}{np_i}}$$

má při $n \to \infty$ *přibližně rozdělení* χ^2_{k-1}.

Důkaz je založen na centrální limitní větě

Hypotézu H_0 o apriorních pravděpodobnostech zamítáme na hladině významnosti α, pokud

$$\sum_{i=1}^{k} \frac{(O_i - np_i)^2}{np_i} > \chi^2_{k-1}(1-\alpha)$$

Tento test se nazývá χ^2 test, nebo též Pearsonův test. Tento test je asymptotický, doporučuje se ho použít pro n tak velké, že

$$np_i > 5 \text{ pro všechna } i = 1, \dots, k\,.$$

11.6. Příklad. Testujeme zda hrací kostka neni falešná. Provedeno 120 hodů. Výsledky jsou

hodnota	1	2	3	4	5	6
četnost	15	16	25	31	15	18

$n = 120$, $p_i = 1/6$, $np_i = 120 \cdot \frac{1}{6} = 20$.

$$\sum_{i=1}^{6} \frac{(O_i - np_i)^2}{np_i} =$$
$$= \frac{5^2}{20} + \frac{4^2}{20} + \frac{5^2}{20} + \frac{11^2}{20} + \frac{5^2}{20} + \frac{2^2}{20} = 10,8\,.$$

$\chi^2_5(0,95) = 11,07\,.$

Závěr: nelze zamítnout hypotézu, že kostka je falešná.

Často se používá na test typu rozdělení:

11.7. Příklad. Testujeme hypotézu, že data pocházejí z rozdělení $N(0,1)$. Provedeno je 1000 měření, data jsou rozdělena do tří skupin

$u_{0,8} = 0,84162$

250 hodnot v $(-\infty, -u_{0,8})$

550 hodnot v $(-u_{0,8}, u_{0,8})$

200 hodnot v $(u_{0,8}, \infty)$

$H_0 : p_1 = 0,2; \quad p_2 = 0,6; \quad p_3 = 0,2$

$np_1 = 200$

$np_2 = 600$

$np_3 = 200$

$$\sum_{i=1}^{3} \frac{(O_i - np_i)^2}{np_i} = \frac{50^2}{200} + \frac{50^2}{600} + \frac{0}{200} = 16,667\,.$$

$\chi^2_2(0,95) = 5,9915$

Závěr: Hypotézu o rozdělení $N(0,1)$ zamítáme.

Testy shody při neznámých parametrech

Testujeme zda data pocházejí z rozdělení s neznámými parametry.

Je-li l počet odhadnutých parametrů, pak

$$\sum_{i=1}^{k} \frac{(O_i - np_i)^2}{np_i}$$

má asymptoticky rozdělení χ^2_{k-1-l}.

Postup:
1. Neznámé parametry odhadneme z dat.
2. Pomocí nich spočítáme apriorní pravděpodobnosti
3. Hypotézu o daném rozdělení zamítneme, jestliže

$$\sum_{i=1}^{k} \frac{(O_i - np_i)^2}{np_i} > \chi^2_{k-1-l}(1-\alpha)$$

11.8. Příklad. Za války byl Londýn bombardován střelami V1 a V2. Hypotéza je, že střely dopadaly náhodně (bez zaměřování) do jednotlivých lokalit.
Celkem na Londýn dopadlo 537 raket.
Území města bylo rozděleno na $24^2 = 576$ čtverců stejné velikosti.

Empirická data:

počet zásahů:	0	1	2	3	4 a více
počet oblastí:	229	211	93	35	8

X: počet zásahů v náhodně vybrané oblasti:

$$Bi\left(\text{počet střel}\ , \frac{1}{\text{počet oblastí}}\right) \approx Po\left(\text{počet střel}\ \cdot \frac{1}{\text{počet oblastí}}\right).$$

1. Odhadneme:

$$\lambda \doteq \frac{537}{576} \left(= \frac{\text{počet střel}}{\text{počet oblastí}}\right).$$

2. Platí-li H_0 je

$$P[X = i] = \frac{\lambda^i e^{-\lambda}}{i!}.$$

počet pokusů = počet oblastí

$$n = 576$$

$$nP[X = 0] = 576\, e^{-\frac{537}{576}} \doteq 226,7\,.$$

$$nP[X = 1] = 576 \cdot \frac{537}{576} e^{-\frac{537}{576}} \doteq 211,4\,.$$

$$nP[X = 2] \doteq 98,5\,.$$

$$nP[X = 3] \doteq 30,6\,.$$

$$nP[X \geq 4] \doteq 8,7\,.$$

počet zásahů:	0	1	2	3	4 a více
skutečnost:	229	211	93	35	8
teorie:	226,7	211,4	98,5	30,6	8,7

Numerický výpočet:

$$\sum_{i=1}^{5} \frac{(0_i - np_i)^2}{np_i} = 1,569$$

$$\chi^2_{5-1-1}(0,95) = 7,81\,.$$

Závěr: Na hladině významnosti 5% potvrzujeme hypotézu H_0.