Teorie nutná ke zvládnutí cvičení 1

1. Pravděpodobnost náhodného jevu

1.1. Pravděpodobnostní prostor

Definice (Pravděpodobnostní prostor)**.** Trojice $(\Omega, \mathcal{F}, \mathbb{P})$, kde $\Omega \neq \emptyset$ ... množina všech možných výsledků náhodného pokusu, $\mathcal{F} \subset 2^{\Omega}$ ... systém podmnožin Ω a platí

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(A) \geq 0 \quad \forall A \in \mathcal{F}$
- $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad \forall A_i \in \mathcal{F}$ po dvou disjunktní

Terminologie
$\omega \in \Omega$ ... elementární jev
$A \subset \Omega$ ... náhodný jev

Definice (Speciální příklad: Klasický pravděpodobnostní prostor)**.** Množina Ω je neprázdná a obsahuje konečný počet prvků, tj. $\Omega = \{\omega_1, \ldots, \omega_n\}$. Dále předpokládáme $\mathcal{F} = 2^{\Omega}$ a všechny elementární jevy jsou stejně pravděpodobné ($\mathbb{P}(\omega_1) = ... = \mathbb{P}(\omega_n) = \frac{1}{n}$). Pravděpodobnost náhodného jevu $A \subset \Omega$ odpovídá

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n}.$$

Příklad (Hod kostknou)
náhodný pokus: hod kostkou
$\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^{\Omega} = \{\{\emptyset\}, \{1\}, \{2\}, ..., \{1, 2\}, ..., \{1, 2, 3\}, ..., \Omega\}$
příklad elementárního jevu: padla šestka ... $\omega = \{6\}$
příklad náhodného jevu: padlo sudé číslo ... $A = \{2, 4, 6\}$

Příklad (Dva hody jednou mincí) V tomto příkladě záleží na pořadí. Označme jevy H = padla hlava, O = padl orel.
náhodný pokus: dva hody jednou mincí
$\Omega = \{HH, HO, OH, OO\}$,
$\mathcal{F} = 2^{\Omega} = \{\{\emptyset\}, \{HH\}, \{HO\}, ..., \{HH, HO\}, ..., \{HH, HO, OH\}, ..., \Omega\}$
příklad elementárního jevu: padla hlava a pak orel ... $\omega = \{HO\}$
příklad náhodného jevu: padlo dvakrát to samé ... $A = \{HH, OO\}$

Příklad Jak by se dal zapsat náhodný pokus jednoho hodu dvěma mincemi?

Užitečné vztahy pro výpočet pravděpodobnosti

- $0 \leq \mathbb{P}(A) \leq 1, \quad \forall A \in \mathcal{F}$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \quad A^c = \Omega \setminus A$
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$
- $\mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c)$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- Princip inkluze a exkluze (PIE):

$$\mathbb{P}(A_1 \cup ... \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) ... + (-1)^{n+1} \mathbb{P}(A_1 \cap ... \cap A_n)$$

Připomenutí definic z kombinatoriky
Vybíráme prvky z množiny $\{1, \ldots, n\}$ a zajímá nás, kolik existuje uspořádání různého typu.

1. **Permutace**: uspořádaná n-tice, která má stejný počet prvků jako množina, ze které vybíráme
   - počet všech permutací bez opakování prvků: $n!$
   - počet všech permutací s opakováním prvků (i-tý prvek se opakuje k_i-krát): $\frac{k_1+\ldots+k_n}{k_1!\cdots k_n!}$
2. **Variace**: uspořádaná k-tice, která může mít jiný počet prvků než množina, ze které vybíráme
   - počet všech variací bez opakování prvků: $\frac{n!}{(n-k)!}$
   - počet všech variací s opakováním prvků (i-tý prvek se může opakovat nejvýše k-krát): n^k
3. **Kombinace**: neuspořádaná k-tice, která může mít jiný počet prvků než množina, ze které vybíráme
   - počet všech kombinací bez opakování prvků: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{\text{variace}}{\text{permutace}}$
   - počet všech kombinací s opakováním prvků (i-tý prvek se může opakovat nejvýše k-krát): $\binom{n+k-1}{k}$

I.2. Podmíněná pravděpodobnost

Definice (Podmíněná pravděpodobnost)**.** Podmíněná pravděpodobnost jevu A za podmínky jevu B je definována jako

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

Poznámky

1. Často známe $\mathbb{P}(A|B)$ a pomocí toho dopočítáme $\mathbb{P}(A \cap B)$ nebo $\mathbb{P}(B)$ ze vztahu
$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B).$$
2. Pro pevné B definujeme zobrazení $P_B(A) = \mathbb{P}(A|B)$. Toto zobrazení splňuje definici pravděpodobnosti. POZOR: Pro zobrazení $P_B(A) = \mathbb{P}(B|A)$ toto tvrzení neplatí.
3. $\mathbb{P}(B|B) = 1$, $\mathbb{P}(B^c|B) = 0$

Věta (Věta o úplné pravděpodobnosti)**.** Buďte $A, B_1, B_2, \ldots \in \mathcal{F}$, $B_i \cap B_j = \emptyset \,\forall i \neq j$, $\cup_i B_i = \Omega$, $\mathbb{P}(B_i) > 0 \,\forall i$. Pak

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i).$$

Věta (Bayesova věta)**.** Buďte $A, B_1, B_2, \ldots \in \mathcal{F}$, $B_i \cap B_j = \emptyset \,\forall i \neq j$, $\cup_i B_i = \Omega$, $\mathbb{P}(B_i) > 0 \,\forall i$, $\mathbb{P}(A) > 0$. Pak

$$\mathbb{P}(B_i|A) \overset{def}{=} \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} \overset{VoUP}{=} \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}.$$

Věta (Věta o násobení pravděpodobnosti - o postupném podmiňování)**.** Buďte $B_1, \ldots, B_n \in \mathcal{F}, \mathbb{P}(\cap_i B_i) > 0$. Pak

$$\mathbb{P}(\cap_{i=i}^n B_i) = \mathbb{P}(B_1) \cdot \mathbb{P}(B_2|B_1) \cdot \mathbb{P}(B_3|B_1 \cap B_2) \cdot \ldots \cdot \mathbb{P}(B_n| \cap_{i=1}^{n-1} B_i).$$

Teorie nutná ke zvládnutí cvičení 2

I.3. Nezávislost náhodných jevů

Definice. (Nezávislé jevy) Jevy $A_1, \dots, A_n \in \mathcal{F}$ jsou **nezávislé**, pokud pro každé $r \leq n$ a každou $\{i_1, \dots, i_r\} \subset \{1, \dots, n\}$ platí

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) = \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_r}).$$

Speciálně dva jevy $A, B \in \mathcal{F}$ jsou nezávislé, pokud

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

Definice. (Po dvou nezávislé jevy) Jevy $A_1, \dots, A_n \in \mathcal{F}$ jsou **po dvou nezávislé**, pokud pro každé $i, j \in \{1, \dots, n\}, i \neq j$ platí

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j).$$

Poznámka (Vztah nezávislosti a podmíněné pravděpodobnosti) Jevy $A, B \in \mathcal{F}$ jsou nezávislé právě když $\mathbb{P}(A|B) = \mathbb{P}(A)$, má-li tato podmíněná pravděpodobnost smysl.

II. Náhodné veličiny

II.1. Rozdělení náhodné veličiny

Definice (Náhodná veličina)**.** (Reálná) náhodná věličina X je měřitelné zobrazení (funkce) z prostoru $(\Omega, \mathcal{F})$ do prostoru $(\mathbb{R}, \mathcal{B})$, kde $\mathcal{B}$ značí Borelovskou σ-algebru. Prvkům $\omega \in \Omega$ tedy přiřazuje reálné číslo $X(\omega) \in \mathbb{R}$.

Definice (Rozdělení náhodné veličiny)**.** Buď $X : \Omega \to \mathbb{R}$ náhodná veličina. Rozdělení náhodné veličiny X je pravděpodobnostní míra P_X na $\mathbb{R}$ definovaná předpisem

$$P_X(B) = \mathbb{P}(X \in B) \quad B \in \mathcal{B}$$

Definice (Distribuční funkce)**.** Distribuční funkcí náhodné veličiny rozumíme funkci $F_X : \mathbb{R} \to [0, 1]$ definovanou předpisem

$$F_X(x) := \mathbb{P}(X \leq x) = P_X((-\infty, x]), \quad x \in \mathbb{R}.$$

Poznámka Rozdělení náhodné veličiny X je jednoznačně určeno jeho distribuční funkcí F_X.

Poznámka (Vlastnosti distribuční funkce) Pro distribuční funkci F_X vždy platí

- je neklesající
- je zprava spojitá
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$, pro libovolné $a < b$

Definice (Diskrétní náhodná veličina)**.** Diskrétní náhodná veličina nabývá nejvýše spočetně mnoha hodnot $x_1, x_2, ... \in \mathbb{R}$.

Definice (Spojitá náhodná veličina)**.** Spojitá náhodná veličina nabývá nespočetně mnoha hodnot z nějakého podintervalu $\mathbb{R}$.

Teorie nutná ke zvládnutí cvičení 3

II.2. Diskrétní a spojité náhodné veličiny

Definice (Diskrétní náhodná veličina)**.** Náhodná veličina je **diskrétní**, pokud nabývá nejvýše spočetně mnoha hodnot (např. $x_1, x_2, ... \in \mathbb{R}$).

Definice (Spojitá náhodná veličina)**.** Náhodná veličina je **spojitá**, pokud nabývá nespočetně mnoha hodnot (např. interval $[a, b] \subset \mathbb{R}$).

Poznámka (Rozdělení)

(a) Rozdělení diskrétní náhodné veličiny X je jednoznačně určeno pravděpodobnostmi

$$p_k := \mathbb{P}(X = x_k), k \in \mathbb{N}.$$

Vždy platí

$$\sum_{k \in \mathbb{N}} p_k = 1.$$

(b) Rozdělení spojité náhodné veličiny X je jednoznačně určeno **hustotou** $f_X : \mathbb{R} \to \mathbb{R}_+$, pro níž

$$\mathbb{P}(X \in B) = \int_B f_X(x)\,\mathrm{d}x \quad \forall B \in \mathcal{B}.$$

Tedy např. pro $B = [a, b], a < b$, je $\mathbb{P}(a < X < b) = \int_a^b f_X(x)\,\mathrm{d}x$, pro $B = \{a\}, a \in \mathbb{R}$, je $\mathbb{P}(X = a) = \int_a^a f_X(x)\,\mathrm{d}x = 0$. Vždy platí

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1.$$

Poznámka (Distribuční funkce)

(a) Distribuční funkce diskrétní náhodné veličiny nabývající hodnot $x_k, k \in \mathbb{N}$, je skokovitá, po částech konstantní se skoky v x_k o velikosti $p_k = \mathbb{P}(X = x_k)$.

(b) Distribuční funkce spojité náhodné veličiny je spojitá a platí

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t, \quad f_X(x) = F_X'(x), \text{ pokud derivace v tomto bodě existuje.}$$

Pro libovolné $a < b$ pak

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a < X \le b) = \mathbb{P}(a \le X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,\mathrm{d}x.$$

Opakování (Základní vzorce pro integrály a derivace)

Tvar základních primitivních funkcí $F \overset{c}{=} \int f(x)\,\mathrm{d}x$:

- $\int x^a\,\mathrm{d}x \overset{c}{=} \frac{x^{a+1}}{a+1}, \quad a \neq -1$
- $\int \frac{1}{x}\,\mathrm{d}x \overset{c}{=} \ln|x|, \quad x \neq 0$
- $\int \mathrm{e}^x\,\mathrm{d}x \overset{c}{=} \mathrm{e}^x$
- $\int \sin(x)\,\mathrm{d}x \overset{c}{=} -\cos(x)$
- $\int \cos(x)\,\mathrm{d}x \overset{c}{=} \sin(x)$

Základní vzorce pro derivace:

- $(x^n)' = nx^{n-1} \quad x \in \mathbb{R}, n \in \mathbb{N}$
- $(x^a)' = ax^{a-1} \quad x > 0, a \in \mathbb{R}$
- $(\mathrm{e}^x)' = \mathrm{e}^x$
- $(\sin(x))' = \cos(x)$
- $(\cos(x))' = -\sin(x)$

Určitý integrál pak počítáme jako $\int_a^b f(x)\,\mathrm{d}x = \lim_{x \to b-} F(x) - \lim_{x \to a+} F(x)$.
Linearita neurčitého integrálu: $\int af(x) + bg(x)\,\mathrm{d}t \overset{c}{=} aF + bG$
Věta o substituci: $\int f(\Phi(t)) \cdot \Phi'(t)\,\mathrm{d}t \overset{c}{=} F(\Phi(t))$
Derivace složené funkce: $[f(g(x))]' = f'(g(x)) \cdot g'(x)$

Teorie nutná ke zvládnutí cvičení 4

II.3. Momenty náhodné veličiny

Definice (Střední hodnota)**.** Buď X náhodná veličina.

1. Je-li X **diskrétní** náhodná veličina, pak její střední hodnotu počítáme jako
$$\mathbb{E}\, X = \sum_k x_k p_k, \quad \text{(existuje-li)},$$
kde $p_k = \mathbb{P}(X = x_k)$ a sčítáme přes všechna k, pro která X nabývá nějaké hodnoty $x_k \in \mathbb{R}$.

2. Je-li X **spojitá** náhodná veličina, pak její střední hodnotu počítáme jako
$$\mathbb{E}\, X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x, \quad \text{(existuje-li)},$$
kde f je hustota náhodné veličiny X.

Definice (Obecný moment)**.** Buď X náhodná veličina a $h : \mathbb{R} \to \mathbb{R}$.

1. Je-li X **diskrétní** náhodná veličina, pak střední hodnota $h(X)$ je
$$\mathbb{E}\, h(X) = \sum_k h(x_k) p_k, \quad \text{(existuje-li)}.$$

2. Je-li X **spojitá** náhodná veličina, pak střední hodnota $h(X)$ je
$$\mathbb{E}\, h(X) = \int_{-\infty}^{\infty} h(x) f(x)\,\mathrm{d}x, \quad \text{(existuje-li)}.$$

Definice (Rozptyl)**.** Buď X náhodná veličina, rozptyl X spočítáme jako
$$\mathrm{var} X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Poznámka.

1. Střední hodnota nám říká, jakou hodnotu bychom očekávali, že bude mít náhodná veličina X. Rozptyl je míra variability a říká nám, jak moc se náhodná veličina může lišit od očekávané hodnoty.

2. Z první rovnosti v definici rozptylu vidíme, že při výpočtu sčítáme a nebo integrujeme nezáporná čísla (druhou mocninu $X - \mathbb{E}X$). A tedy...

Rozptyl nikdy nemůže vyjít záporný!

Vlastnosti střední hodnoty.

1. Pro každé $a, b \in \mathbb{R}$ a náhodnou veličinu X platí: $\mathbb{E}(a + bX) = a + b\mathbb{E}X$.

2. Pro každé dvě náhodné veličiny X, Y platí: $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$.

Vlastnosti rozptylu.

1. Pro každé $a, b \in \mathbb{R}$ a náhodnou veličinu X platí: $\mathrm{var}(a + bX) = b^2 \mathrm{var} X$.

2. Obecně **neplatí**, že $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y$, pouze pokud X a Y jsou nezávislé (bude vysvětleno příště).

Připomenutí - integrace per partes.
$$\int u \cdot v' = u \cdot v - \int u' v$$

Teorie nutná ke zvládnutí cvičení 5

II.4. Náhodné vektory

Pod pojmem náhodný vektor v rámci tohoto cvičení chápeme dvojici náhodných veličin (X, Y). Navíc se zde omezíme jen na případy, kdy jsou obě náhodné veličiny diskrétní a nebo jsou obě spojité.

Definice. (Náhodný vektor)

- Diskrétní náhodný vektor je dvojice (X, Y) nabývající nejvýše spočetně mnoha hodnot (x_i, y_j).
- Spojitý náhodný vektor je dvojice (X, Y) nabývající nekonečně mnoha hodnot.

Definice. (Sdružené rozdělení)

- Sdružené rozdělení diskrétního náhodného vektoru (X, Y) je určeno pravděpodobnostmi
$$p_{i,j} = \mathbb{P}(X = x_i, Y = y_j)$$
a platí $\sum_i \sum_j p_{i,j} = 1$.
- Sdružené rozdělení spojitého náhodného vektoru (X, Y) je určeno sdruženou hustotou $f : \mathbb{R}^2 \to \mathbb{R}_+$, pro kterou platí
$$\mathbb{P}((X, Y) \in B) = \iint\limits_B f(x, y) \, \mathrm{d}x \, \mathrm{d}y, \quad B \subset \mathbb{R}^2$$
a platí $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1$.

Definice. (Marginální rozdělení)

- Marginální rozdělení diskrétního náhodného vektoru (X, Y) nabývající hodnot $(x_i, y_j); i, j \in \mathbb{N}$ jsou určena marginálními pravděpodobnostmi
$$p_i^X = \mathbb{P}(X = x_i) = \sum_j \mathbb{P}(X = x_i, Y = y_j),$$
$$p_j^Y = \mathbb{P}(Y = y_j) = \sum_i \mathbb{P}(X = x_i, Y = y_j).$$
- Marginální rozdělení spojitého náhodného vektoru (X, Y) jsou určena marginálními hustotami
$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}y,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x.$$

Definice. (Nezávislost náhodných veličin)

- Dvě diskrétní náhodné veličiny X, Y jsou nezávislé, pokud sdružené pravděpodobnosti vektoru (X, Y) spočítáme jako součin marginálních pravděpodobností X a Y, tj.
$$\mathbb{P}(X = x_i, Y = y_i) = \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_i)$$
pro každé (i, j), pro které (X, Y) nabývá nějakou hodnotu (x_i, y_j).
- Dvě spojité náhodné veličiny X, Y jsou nezávislé, pokud sdruženou hustotu vektoru (X, Y) spočítáme jako součin sdružených hustot X a Y, tj.
$$f(x, y) = f_X(x) \cdot f_Y(y)$$
pro skoro všechny $(x, y) \in \mathbb{R}^2$.

Teorie nutná ke zvládnutí cvičení 6

II.5. Kovariance a korelace

Definice (Rozptyl)**.** Pro náhodnou veličinu X definujeme **rozptyl** jako

$$\mathrm{var}X = \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X).$$

Definice (Kovariance)**.** Pro dvě náhodné veličiny X, Y s $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$ definujeme **kovarianci** jako

$$\mathrm{cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Definice (Korelace)**.** Pro dvě náhodné veličiny X, Y s $0 < \mathrm{var}X, \mathrm{var}Y < \infty$ definujeme **korelaci** jako

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}X} \cdot \sqrt{\mathrm{var}Y}}.$$

Definice (Varianční matice)**.** Pro dvě náhodné veličiny X, Y s $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$ definujeme **varianční matici** jako

$$\mathrm{Var}(X, Y) = \begin{pmatrix} \mathrm{var}X & \mathrm{cov}(X, Y) \\ \mathrm{cov}(Y, X) & \mathrm{var}Y \end{pmatrix}.$$

Poznámka.

(a) $\mathrm{var}X \geq 0, \mathrm{cov}(X, Y) \in \mathbb{R}, \mathrm{corr}(X, Y) \in [-1, 1]$

(b) X, Y nezávislé, pak $\mathrm{cov}(X, Y) = \mathrm{corr}(X, Y) = 0$

(c) (Linearita rozptylu) Pro X, Y náhodné veličiny a $a, b \in \mathbb{R}$ platí

$$\mathrm{var}(aX + bY) = a^2\mathrm{var}X + b^2\mathrm{var}Y + 2a\,b\,\mathrm{cov}(X, Y)$$

Speciálně pro X, Y nezávislé je

$$\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y.$$

Teorie nutná ke zvládnutí cvičení 7

III.1. Náhodný výběr

Definice (Náhodný výběr)**.** Náhodný výběr je posloupnost náhodných veličin $X_1, \ldots, X_n$, které jsou nezávislé a mají všechny stejné rozdělení.

Definice (Bodový odhad)**.** Buď $X_1, \ldots, X_n$ náhodný výběr z rozdělení daného distribuční funkcí F a $g_n : \mathbb{R}^n \to \mathbb{R}$ funkce, jejíž předpis nezávisí na F. Bodovým odhadem rozumíme náhodnou veličinu

$$T_n := g_n(X_1, \ldots, X_n).$$

Příklady.

(a) $g_n(X_1, \ldots, X_n) = \min_{i=1,\ldots,n} X_i$

(b) $g_n(X_1, \ldots, X_n) = \max_{i=1,\ldots,n} X_i$

(c) výběrový průměr: $g_n(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^n X_i =: \bar{X}_n$

(d) výběrový rozptyl: $g_n(X_1, \ldots, X_n) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 =: S_n^2$

(e) empirická distribuční funkce: $g_n(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i \le x\} =: \hat{F}_n(x)$

Teorie nutná ke zvládnutí cvičení 8

III.1. Náhodný výběr

Zkoumané vlastnosti bodových odhadů Nechť bodový odhad T_n je odhadem parametru θ rozdělení daného distribuční funkcí F (např. $\mathbb{E}X, \mathbb{P}(X \leq b), ...$). Parametr θ může nabývat hodnot v nějaké množině $\Theta \subset \mathbb{R}$. Pro posouzení kvality odhadu T_n vyšetřujeme následující dvě vlastnosti.

1. **Nestrannost** T_n je nestranným odhadem θ, pokud

$$\mathbb{E}T_n = \mathbb{E}_\theta T_n = \theta, \quad \forall \theta \in \Theta.$$

2. **Konzistence** T_n je konzistentním odhadem θ, pokud pro všechna $\varepsilon > 0$ je

$$\mathbb{P}(|T_n - \theta| > \varepsilon) \xrightarrow{n\to\infty} 0, \quad \forall \theta \in \Theta.$$

Věta (Čebyševova nerovnost)**.** Buď X náhodná veličina s $\mathbb{E}|X| < \infty$. Pak

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathrm{var}X}{\varepsilon^2}.$$

Odhad momentovou metodou Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$ (např. $Alt(\theta), Exp(\theta), Hypergeom(100, \theta, 5)$). Předpokládejme, že jsme schopni najít spojitou funkci $g : \mathbb{R} \to \mathbb{R}$ tak, že

$$\mathbb{E}X = g(\theta).$$

Odhad momentovou metodou je odhad $\hat{\theta}_n$ splňující rovnost

$$\bar{X}_n = g(\hat{\theta}_n).$$

Poznámka Pokud $\mathbb{E}X$ nezávisí na θ, využijeme některý z odhadů momentů vyšších řádů, tj.

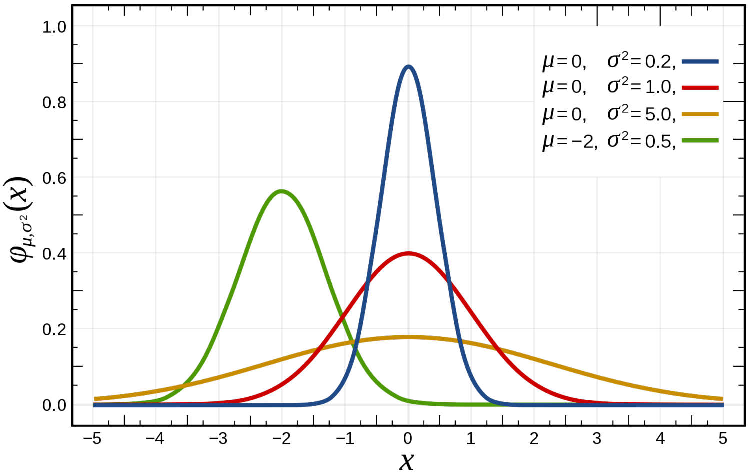
$$\frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k \geq 2$$

Teorie nutná ke zvládnutí cvičení 9

III.2. Centrální limitní věta

Definice (Normální rozdělení)**.** Normální rozdělení $N(\mu, \sigma^2)$ má hustotu

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$



Věta (Centrální limitní věta)**.** Máme posloupnost $X_1, X_2, \ldots$ nezávislých, stejně rozdělených náhodných veličin, pro které je $0 < \text{var}X_1 < \infty$. Pak

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1}{\sqrt{n}\sqrt{\text{var}X_1}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x), \quad x \in \mathbb{R},$$

kde $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-x^2/2\right\} \mathrm{d}x$ je distribuční funkce normovaného normálního rozdělení $N(0,1)$.

Tabulka hodnot Φ a Φ^{-1}
Hodnoty distribuční funkce $\Phi(x)$ lze počítat jen numericky a lze je najít v následující tabulce pro $x \geq 0$. Ostatní hodnoty dopočítáme ze vztahu $\Phi(-x) = 1 - \Phi(x)$.

x	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\Phi(x)$	0.500	0.540	0.579	0.618	0.655	0.691	0.726	0.758	0.788	0.816	0.841
x		1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
$\Phi(x)$		0.864	0.885	0.903	0.919	0.933	0.945	0.955	0.964	0.971	0.977
x		2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
$\Phi(x)$		0.982	0.986	0.989	0.992	0.994	0.995	0.997	0.997	0.998	0.999

Někdy se nám budou hodit hodnoty inverzní funkce, tzv. kvantily $N(0,1)$ definované jako

$$q_\alpha = \Phi^{-1}(\alpha), \quad \alpha \in [0,1].$$

Ty najdeme v tabulce pro $\alpha \geq 0.5$. Ostatní hodnoty dopočítáme ze vztahu $q_\alpha = -q_{1-\alpha}$

α	0.5	0.8	0.9	0.95	0.975	0.99	0.995
$\Phi^{-1}(\alpha) = q_\alpha$	0	0.842	1.282	1.645	1.960	2.326	2.576

Teorie nutná ke zvládnutí cvičení 10

Součty náhodných veličin

Jsou-li $X_1, ..., X_n$ náhodné veličiny a $a_1, ..., a_n$ reálná čísla, pak (za předpokladu, že obě strany existují)

$$\mathbb{E}\sum_{i=1}^{n} a_i X_i = \sum_{i=1}^{n} a_i \mathbb{E} X_i$$

$$\mathrm{var}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} a_i^2 \mathrm{var} X_i + \sum\sum\nolimits_{i\neq j} a_i a_j \mathrm{cov}(X_i, X_j)$$

Věta (Centrální limitní věta)**.** Máme posloupnost $X_1, X_2, ...$ nezávislých, stejně rozdělených náhodných veličin, pro které je $0 < \mathrm{var} X_1 < \infty$. Pak

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - n\mathbb{E}X_1}{\sqrt{n}\sqrt{\mathrm{var}X_1}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x), \quad x \in \mathbb{R}.$$

Věta (Slabý zákon velkých čísel)**.** Máme posloupnost $X_1, X_2, ...$ nezávislých, stejně rozdělených náhodných veličin s konečnou střední hodnotou $\mathbb{E}X_1 = \mu$ a konečným rozptylem. Pak

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

Věta (Věta o spojité transformaci)**.** Jestliže pro posloupnost náhodných veličin $\{Y_n\}_{n\geq 1}$ a $a \in \mathbb{R}$ platí $Y_n \xrightarrow{p} a$ a g je spojitá funkce, pak

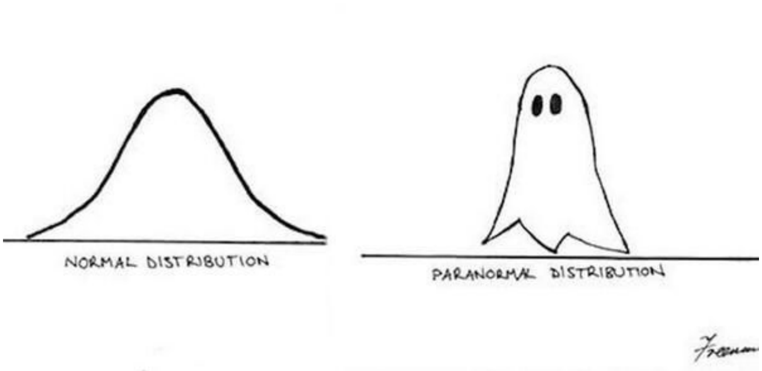
$$g(Y_n) \xrightarrow{p} g(a).$$

Přehled momentů základních rozdělení

Název	Hodnoty	Rozdělení	Střední hodnota	Rozptyl
Alternativní $Alt(p)$	$\{0,1\}$	$p_1 = p, p_0 = 1-p$	p	$p(1-p)$
Binomické $Bi(n,p)$	$\{0,\dots,n\}$	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Poissonovo $Poiss(\lambda)$	$\{0,1,2,\dots\}$	$p_k = \frac{\lambda^k}{k!} e^{\lambda}$	λ	λ
Hypergeometrické $Hg(N,K,n)$	$\{\max(0, K-N+n), ..., \min(n,K)\}$	$p_k = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n\frac{K}{N}$	$n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$
Rovnoměrné $R(a,b)$	$[a,b]$	$f_X(x) = \frac{1}{b-a}$, $x \in [a,b]$ $= 0$ jinak	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponenciální $Exp(\lambda)$	$[0,\infty)$	$f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$ $= 0$ jinak	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normální $N(\mu,\sigma^2)$	$\mathbb{R}$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$	μ	σ^2

Teorie nutná ke zvládnutí cvičení 12

Opakování



Věta (Centrální limitní věta)**.** Máme posloupnost $X_1, X_2, ...$ nezávislých, stejně rozdělených náhodných veličin, pro které je $0 < \text{var}X_1 < \infty$. Pak

$$\mathbb{P}\left(\frac{\sum_{i=1}^n X_i - n\mathbb{E}X_1}{\sqrt{n}\sqrt{\text{var}X_1}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x), \quad x \in \mathbb{R}.$$

Věta (Sluckého věta)**.** Nechť X_n, Y_n, Z_n jsou posloupnosti náhodných veličin takových, že $X_n \xrightarrow{D} N(0,1)$, $Y_n \xrightarrow{P} c$ a $Z_n \xrightarrow{P} d$, kde $c, d \in \mathbb{R}$ jsou reálné konstanty. Pak

$$Y_n \cdot X_n + Z_n \xrightarrow{D} N(d, c^2).$$

III.3. Intervalové odhady

Definice (Intervalový odhad)**.** Buď $X_1, ..., X_n$ náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in \Theta \subset \mathbb{R}$. Zvolme hladinu $\alpha \in (0,1)$. Intervalový odhad parametru θ na hladině α je interval $[D, H]$, jehož meze tvoří náhodné veličiny $D = D(X_1, ..., X_n), H = H(X_1, ..., X_n)$ splňující

$$\mathbb{P}(\theta \in [D, H]) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$$

Poznámka. Někdy je obtížné nalézt přesný intervalový odhad a spokojíme se s asymptotickým intervalovým odhadem (založeným na CLV), který uvedenou podmínku splňuje pro $n \to \infty$ (viz cvičení 1).

IV.1. Podmíněné rozdělení

Definice (Podmíněné rozdělení)**.** Nechť (X, Y) je náhodný vektor s diskrétním rozdělením s hodnotami v $\mathbb{N}^2$ a nechť $n \in \mathbb{N}$ je takové, že $\mathbb{P}(Y = n) > 0$. Pak definujeme podmíněné rozdělení X za podmínky $Y = n$ jako

$$\mathbb{P}(X = k | Y = n) = \frac{\mathbb{P}(X = k, Y = n)}{\mathbb{P}(Y = n)}.$$

Definice (Podmíněná střední hodnota)**.** Je-li $\mathbb{E}X < \infty$, pak definujeme podmíněnou střední hodnotu X za podmínky $Y = n$ jako

$$\mathbb{E}[X | Y = n] = \sum_{k \in \mathbb{N}} k\mathbb{P}(X = k | Y = n).$$

Tabulka základních rozdělení

1. Příklady diskrétních rozdělení

Název	Značení	Hodnoty	Rozdělení	Interpretace
Alternarivní	$X \sim Alt(p)$ $p \in (0,1)$	$\{0,1\}$	$p_1 = p, p_0 = 1-p$	Výsledek jednoho pokusu, který dopadne úspěchem s pravděpodobností p
Binomické	$X \sim Bi(n,p)$ $n \in \mathbb{N}, p \in (0,1)$	$\{0, \cdots, n\}$	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$	Výsledek n nezávislých pokusů, které dopadnou úspěchem s pravděpodobností p, $Bi(1,p) = Alt(p)$
Geometrické	$X \sim Ge(p)$ $p \in (0,1)$	$\{0, \cdots, \infty\}$	$p_k = p(1-p)^k$	Počet neúspěchů před prvním úspěchem, kde úspěch nastane s pravděpodobností p
Poissonovo	$X \sim Poiss(\lambda)$ $\lambda > 0$	$\{0, \cdots, \infty\}$	$p_k = \frac{\lambda^k}{k!} \mathrm{e}^{\lambda}$	Počet výskytu nějaké události v daném časovém intervalu, kde události nastávají nezávisle na sobě, λ určuje intenzitu výskytů
Hyper-geometrické	$X \sim Hg(N,K,n)$ $N \in \mathbb{N}$, $n, K \in \{1, ..., N\}$	$\{\max(0, K-N+n),$ $..., \min(n,K)\}$	$p_k = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	Máme množinu N prvků, z nichž K má zkoumanou vlastnost. Vybíráme náhodně bez vracení n z nich a zkoumáme, kolik vybraných mělo danou vlastnost

2. Příklady spojitých rozdělení

Název	Značení	Hodnoty	Rozdělení	Interpretace
Rovnoměrné	$X \sim R(a,b)$ $a < b, a, b \in \mathbb{R}$	$[a,b]$	$f_X(x) = \frac{1}{b-a}$, $x \in [a,b]$ $= 0$ jinak	Generátor náhodných čísel z intervalu $[a,b]$
Exponenciální	$X \sim Exp(\lambda)$ $\lambda > 0$	$[0,\infty)$	$f_X(x) = \lambda \mathrm{e}^{-\lambda x}$, $x \geq 0$ $= 0$ jinak	Doba do výskytu události (např. poruchy), λ určuje intenzitu výskytu událostí
Normální (Gaussovo)	$X \sim N(\mu, \sigma^2)$ $\mu \in \mathbb{R}, \sigma^2 \geq 0$	$\mathbb{R}$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$	Dobrá aproximace pro neznámé rozdělení (např. rozdělení IQ v populaci, náhodná odchylka měření, apod.). Parametr μ je očekávaná hodnota, parametr σ^2 určuje variabilitu. Rozdělení $N(0,1)$ se nazývá **standardní normální**.