# Pravděpodobnost a matematická statistika

## NMSA202

Michal Pešta

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta
Univerzita Karlova
Praha

LS 24/25

# Předehra

- "*Lasciate ogni speranza o voi ch'entrate.*"
  (Zanechte vší naděje, kdo vstupujete.)

  [Dante Alighieri; Božská komedie, Peklo]

- "*Und die Pforte ist enge, und der Weg ist schmal, der zum Leben führet; und wenig ist ihrer, die ihn finden.*"
  (Těsná brána a úzká cesta však vede k životu a málokdo ji nachází.)

  [Matouš 7:14; Lutherova Bible]

# Těsná brána a úzká cesta

- Potřeba stochastického myšlení: 95% & 95% = ?

- Tři běžné záležitosti mimo naší dlouhé cesty:
  - Derivace měla být definována jako log-derivace, ale . . . fyzika
  - Centrální limitní věta je pouze svatý předgrál, ale všude nediferencovatelná spojitá funkce, ke které všechno konverguje, je svatý grál
  - Každý používá neuronové sítě, ale pouze hrstka jim rozumí

# Úvodní kurz

$\neq$ Základní kurz

$>$ Přípravný kurz

$<$ Komplexní kurz

# Agenda

Úvod a motivace

1. Náhodné jevy
2. Náhodné veličiny
3. Střední hodnota
4. Stochastické nerovnosti
5. Stochastické konvergence
6. Statistické učení
7. Statistické funkcionály
8. Parametrická inference
9. Testování hypotéz
10. Literatura

# Agenda

## Úvod a motivace

0.1 Literatura
0.2 Struktura
0.3 Data science
0.4 Data mining
0.5 Machine learning

# Literatura

- Hlavní zdroj [Wasserman, 2013]
- Česká částečná alternativa [Dupač and Hušková, 2013]
- Dodatečný a rozšiřující materiál
  [Casella and Berger, 2001, Chung, 2001, Resnick, 2013, Rosenthal, 2006, Ross, 2020]

# Bloky zvýrazněného textu

Některé důležité texty budou zvýrazněny, protože jsou důležité.

**Definice 0.1 (Název definice)**

Definice jsou červeně.

**Příklad 0.2 (Název příkladu)**

Příklady jsou zeleně.

**Věta 0.3 (Název věty)**

*Věty jsou kurzívou a modře.*

# Data science (datová věda)

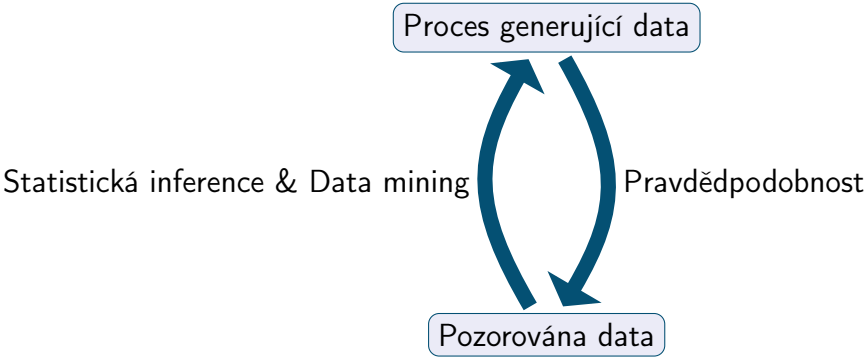
= Matematická statistika + Algoritmické výpočty

- Matematická statistika = odůvodněná statistika

! ne pouze “Jako?”, ale také “Proč?”

# Data mining (dolování dat?)

+ Statistická inference = Stochastické modelování



# Machine learning (strojové učení)

$\subsetneq$ Statistical learning (statistické učení)

| **Matematická statistika** | **Strojové učení** |
|---|---|
| odhadování | učení |
| klasifikace | učení s učitelem |
| shlukování | uční bez učitele |
| data | trénovací soubor |
| kovariáty | znaky |
| klasifikátor | hypotéza |
| hypotéza | — |
| interval spolehlivosti | — |

# Agenda


## 1. Náhodné jevy

1.1 Měřitelný prostor
1.2 Pravděpodobnostní prostor
1.3 Nezávislé události
1.4 Podmíněná pravděpodobnost
1.5 Bayesova věta

# Měřitelný prostor

Výběrový prostor $\Omega$ je množina výsledků experimentu. Prvky $\omega \in \Omega$ se nazývají elementární jevy, výsledky nebo realizace. Podmnožiny $\Omega$ se nazývají (náhodné) jevy/události.

**Příklad 1.1 (Dvojnásobný hod mincí)**

Pokud hodíme mincí dvakrát, pak $\Omega = \{PP, PO, OP, OO\}$. Událost, že první hod je panna, je $A = \{PP, PO\}$.

**Definice 1.2 (Měřitelný prostor)**

Nechť $\Omega \neq \emptyset$ je nějaká množina a nechť $2^\Omega$ představuje její potenční množinu. Podmnožina $\mathcal{A} \subseteq 2^\Omega$ se nazývá *$\sigma$-algebra* nebo *$\sigma$-pole*, pokud splňuje:

(i) $\emptyset \in \mathcal{A}$;

(ii) pokud $A \in \mathcal{A}$, pak $A^\complement \in \mathcal{A}$;

(iii) pokud $A_1, A_2, \ldots \in \mathcal{A}$, pak $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Dvojice $(\Omega, \mathcal{A})$ se nazývá *měřitelný prostor*.

# Pravděpodobnostní prostor

Každé události $A \in \mathcal{A}$ přiřadíme reálné číslo $\mathbb{P}(A)$, které nazýváme pravděpodobnost události $A$.

**Příklad 1.3 (Dva hody mincí)**

Nechť $H_1$ je událost, že při prvním hodu padne panna, a nechť $H_2$ je událost, že při druhém hodu padne panna. Pokud jsou všechny výsledky stejně pravděpodobné, pak pravděpodobnost, že padne alespoň jedna panna (tj. $H_1 \cup H_2$), je 3/4.

**Definice 1.4 (Pravděpodobnostní prostor)**

Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. Zobrazení (množinová funkce) $\mathbb{P} : \mathcal{A} \to [0, 1]$ je *pravděpodobnostní míra*, pokud splňuje:

(i) $\mathbb{P}(\Omega) = 1$;

(ii) pokud $A_1, A_2, \ldots \in \mathcal{A}$ jsou (dvojice) disjunktní, pak $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

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

# Základní vlastnosti pravděpodobnosti

Definice 1.4 jasně implikuje: $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$, $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

**Lemma 1.5 (Pravděpodobnost sjednocení)**

*Pro libovolné události $A, B \in \mathcal{A}$ platí:*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

**Věta 1.6 (Spojitost pravděpodobností)**

*Buď $A_n \uparrow A$ nebo $A_n \downarrow A$, pak*

$$\mathbb{P}(A_n) \to \mathbb{P}(A).$$

# Konečné výběrové prostory

Předpokládejme, že výběrový prostor $\Omega = \{\omega_1, \ldots, \omega_n\}$ je konečný.

**Příklad 1.7 (Hod dvěma kostkami)**

$\Omega$ má 36 prvků: $\Omega = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$. Pokud jsou všechny výsledky stejně pravděpodobné, pak $\mathbb{P}(A) = |A|/36$, kde $|A|$ označuje počet prvků v $A$. Pravděpodobnost, že součet na kostkách je 11, je 2/36, protože dva výsledky odpovídají této události.

Pokud je $\Omega$ konečné a pokud jsou všechny výsledky stejně pravděpodobné, pak

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

což nazýváme rovnoměrné rozdělení pravděpodobnosti.

# Nezávislost I

**Definice 1.8 (Nezávislé události)**

Dvě události $A, B \in \mathcal{A}$ jsou nezávislé, pokud

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

a zapisujeme $A \perp\!\!\!\perp B$. Množina událostí $\{A_i \in \mathcal{A} : i \in I\}$ je nezávislá, pokud

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

pro každou konečnou podmnožinu $J \subseteq I$. Pokud $A$ a $B$ nejsou nezávislé, zapisujeme $A \not\perp\!\!\!\perp B$.

- Nezávislost může být někdy předpokládána (věřena) nebo někdy odvozena (dokázána)
- Disjunktní události s kladnou pravděpodobností nejsou nezávislé

# Nezávislé experimenty

**Příklad 1.9 (Hod spravedlivou mincí 10krát)**

Nechť $A =$ "alespoň jedna panna". Nechť $T_j$ je událost, že při $j$-tém hodu padne orel. Pak

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^{\complement}) \\
&= 1 - \mathbb{P}(\text{samé orly}) \\
&= 1 - \mathbb{P}(T_1 \cap \ldots \cap T_{10}) && \text{používáme nezávislost} \\
&= 1 - \mathbb{P}(T_1) \ldots \mathbb{P}(T_{10}) \\
&= 1 - (1/2)^{10} \approx .999.
\end{aligned}$$

# Podmíněná pravděpodobnost

**Definice 1.10 (Podmíněná pravděpodobnost)**

Pokud $\mathbb{P}(B) > 0$, pak podmíněná pravděpodobnost $A$ za podmínky $B$ je

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

- Můžete si představit $\mathbb{P}(A|B)$ jako podíl případů, kdy nastane $A$, mezi těmi, kdy nastane $B$. Pro pevné $B$ takové, že $\mathbb{P}(B) > 0$, je $\mathbb{P}(\cdot|B)$ pravděpodobnost.
- Obecně $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$.
- Události $A$ a $B$ jsou nezávislé tehdy a jen tehdy, když $\mathbb{P}(A|B) = \mathbb{P}(A)$, za předpokladu, že $\mathbb{P}(B) > 0$.
- $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$, za předpokladu, že $\mathbb{P}(A)\mathbb{P}(B) > 0$.

# Citlivost a specificita I

**Příklad 1.11 (Lékařský test na nemoc $D$ s výsledky + a −)**

Pravděpodobnosti jsou:

| | $D$ | $D^{\complement}$ |
|---|---|---|
| + | .009 | .099 |
| − | .001 | .891 |

Podle definice podmíněné pravděpodobnosti:

$$\mathbb{P}(+|D) = \frac{\mathbb{P}(+ \cap D)}{\mathbb{P}(D)} = \frac{.009}{.009 + .001} = .9, \ \mathbb{P}(-|D^{\complement}) = \frac{\mathbb{P}(- \cap D^{\complement})}{\mathbb{P}(D^{\complement})} = \frac{.891}{.891 + .099} \approx .9.$$

Zdá se, že test je poměrně přesný. Nemocní lidé vykazují pozitivní výsledek v 90 % případů a zdraví lidé negativní výsledek přibližně v 90 % případů.

# Citlivost a specificita II

**Příklad 1.12 (Lékařský test na nemoc $D$ s výsledky + a − (pokračování))**

Předpokládejme, že jdete na test a výsledek je pozitivní. Jaká je pravděpodobnost, že máte nemoc? Většina lidí odpoví .90. Správná odpověď je:

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(D \cap +)}{\mathbb{P}(+)} = \frac{.009}{.009 + .099} \approx .08.$$

Nevěřte své intuici.

Nicméně ani *černé skříňce* nelze slepě důvěřovat: Náhodně vybraná osoba je posuzována (nikoli symptomatická).

# Zákon úplné pravděpodobnosti a Bayesova věta

**Věta 1.13 (Zákon úplné pravděpodobnosti)**

*Nechť $A_1, A_2, \ldots$ je disjunktní spočetný rozklad $\Omega$ takový, že $\mathbb{P}(A_i) > 0$ pro každé $i \in \mathbb{N}$. Pak pro libovolnou událost $B$ platí:*

$$\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

$\mathbb{P}(A_i)$ je apriorní pravděpodobnost $A_i$ a $\mathbb{P}(A_i|B)$ je aposteriorní pravděpodobnost $A_i$.

**Věta 1.14 (Bayesova věta)**

*Nechť $A_1, A_2, \ldots$ je disjunktní spočetný rozklad $\Omega$ takový, že $\mathbb{P}(A_i) > 0$ pro každé $i \in \mathbb{N}$. Pokud $\mathbb{P}(B) > 0$, pak pro každé $i$ platí:*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

# Filtrování emailů

**Příklad 1.15 (Tři kategorie emailů)**

$A_1$=„spam“, $A_2$=„nízká priorita“ a $A_3$=„vysoká priorita“. Na základě předchozí zkušenosti zjistíme, že $\mathbb{P}(A_1) = .7$, $\mathbb{P}(A_2) = .2$ a $\mathbb{P}(A_3) = .1$. Samozřejmě $.7 + .2 + .1 = 1$. Nechť $B$ je událost, že email obsahuje slovo „zdarma“. Na základě předchozí zkušenosti víme, že $\mathbb{P}(B|A_1) = .9$, $\mathbb{P}(B|A_2) = .01$, $\mathbb{P}(B|A_3) = .01$. (Poznámka: $.9 + .01 + .01 \neq 1$.) Přijde vám email obsahující slovo „zdarma“. Jaká je pravděpodobnost, že jde o spam? Bayesova věta říká:

$$\mathbb{P}(A_1|B) = \frac{.9 \times .7}{.9 \times .7 + .01 \times .2 + .01 \times .1} = .995.$$

# Agenda

## 2. Náhodné veličiny


2.1 Měřitelná zobrazení
2.2 Rozdělení náhodné veličiny
2.3 Hustota
2.4 Distribuční funkce
2.5 Diskrétní a spojitá rozdělení
2.6 Kvantil
2.7 Diskrétní náhodné veličiny
2.8 Spojité náhodné veličiny
2.9 Náhodné vektory
2.10 Marginální rozdělení
2.11 Nezávislé náhodné veličiny
2.12 Podmíněné rozdělení
2.13 Mnohorozměrná rozdělení
2.14 Transformace náhodných veličin
2.15 Transformace náhodných vektorů

# Měřitelná zobrazení

**Definice 2.1 (Náhodná veličina)**

Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. Náhodná veličina je měřitelné zobrazení, které přiřazuje každému výsledku $\omega$ reálné číslo $X(\omega)$. To znamená, že

$$X : \Omega \to \mathbb{R} \quad \& \quad \{\omega \in \Omega : \ X(\omega) \leq x\} \in \mathcal{A}, \forall x \in \mathbb{R}.$$

- $[X \in A] \equiv X^{-1}(A) := \{\omega \in \Omega : \ X(\omega) \in A\}$
- $\mathcal{B}(\mathbb{R})$ je borelovská $\sigma$-algebra na $\mathbb{R}$
- $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad \& \quad \forall B \in \mathcal{B}(\mathbb{R}) : \{\omega \in \Omega : \ X(\omega) \in B\} \in \mathcal{A}$

**Příklad 2.2 (Hod mincí desetkrát)**

Nechť $X(\omega)$ je počet líců v posloupnosti $\omega$. Pokud $\omega = HHTHHTHHTT$, pak $X(\omega) = 6$.

# Rozdělení pravděpodobnosti

**Definice 2.3 (Rozdělení náhodné veličiny)**

*Rozdělením náhodné veličiny* $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ nazveme indukovanou pravděpodobnostní míru $\mathbb{P}_X$ na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definovanou jako

$$\mathbb{P}_X(B) := \mathbb{P}[\{\omega \in \Omega : X(\omega) \in B\}], \quad B \in \mathcal{B}(\mathbb{R}).$$

- $\mathbb{P}_X$ je obraz míry $\mathbb{P}$ v zobrazení $X$ a $(\Omega, \mathcal{A}, \mathbb{P})$ se zobrazí na pstní pr. $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$

**Věta 2.4 (O přenosu integrace)**

*Budiž $g$ měřitelná funkce na měřitelném prostoru $(\mathscr{M}, \mathcal{M})$ a $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathscr{M}, \mathcal{M})$. Nechť $\mathbb{P}_X$ je míra na $\mathcal{M}$ indukovaná zobrazením $X$, tj. $\mathbb{P}_X(M) = \mathbb{P}[X^{-1}(M)]$ pro $M \in \mathcal{M}$. Potom, pokud je alespoň jedna strana definována,*

$$\int_\Omega g[X(\omega)] d\mathbb{P}(\omega) = \int_{\mathscr{M}} g(x) d\mathbb{P}_X(x).$$

- Užitečné volby: $(\mathscr{M}, \mathcal{M}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ nebo $(\mathscr{M}, \mathcal{M}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$

# Hustota

- Nepotřebujeme znát celé $(\Omega, \mathcal{A}, \mathbb{P})$ a $X$, ale stačí nám $\mathbb{P}_X$ na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$
- Radon–Nikodýmova věta: $\mu, \nu$ jsou $\sigma$-konečné míry na $(\mathscr{M}, \mathcal{M})$:

$$\nu \ll \mu \quad \Rightarrow \quad \exists!(\text{až na mnu } \mu\text{-míry } 0)\, f \in \mathcal{L}^1(\mu),\ f \geq 0 : \ \nu(M) = \int_M f \mathrm{d}\mu,\ \forall M \in \mathcal{M}$$

- Značíme $f = \mathrm{d}\nu/\mathrm{d}\mu$ a nazýváme Radon–Nikodýmova derivace (nebo hustota)

**Definice 2.5 (Hustota náhodné veličiny vzhledem k míře)**

Buď $X$ náhodná veličina s rozdělením $\mathbb{P}_X$ a $\mathbb{P}_X \ll \mu$ je $\sigma$-konečná míra na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Potom $f_X = \mathrm{d}\mathbb{P}_X/\mathrm{d}\mu$ se nazývá *hustota náhodné veličiny $X$ vzhledem k míře $\mu$*.

**Věta 2.6 (Výpočet psti)**

*Nechť $X$ je náhodná veličina s rozdělením $\mathbb{P}_X$ a $B \in \mathcal{B}(\mathbb{R})$. Pak $\mathbb{P}[X \in B] = \mathbb{P}_X(B)$. Je-li navíc $f_X$ hustota $X$ vzhledem k $\sigma$-konečné míře $\mu$, potom $\mathbb{P}[X \in B] = \int_B f_X \mathrm{d}\mu$.*

# Distribuční funkce

**Definice 2.7 (Distribuční funkce)**

Nechť $(\Omega, \mathcal{A}, \mathbb{P})$ je pravděpodobnostní prostor. Distribuční funkce (CDF) náhodné veličiny $X$ je funkce $F_X : \mathbb{R} \to [0,1]$ definovaná jako

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

- $F_X(x) = \mathbb{P}_X((-\infty, x]), \forall x \in \mathbb{R}$
- Množinový systém $\{(-\infty, x], x \in \mathbb{R}\}$ je uzavřený na průniky, pro $B_n = (-\infty, n]$, $n \in \mathbb{N}$ platí $B_n \uparrow \mathbb{R}$, a $\mathbb{P}_X$ je konečná míra, takže z věty o jednoznačnosti míry plyne, že distribuční funkce jednoznačně určuje rozdělení $\mathbb{P}_X$ náhodné veličiny $X$
- $F_X(x) = F_Y(x), \forall x \in \mathbb{R} \Rightarrow \mathbb{P}[X \in B] = \mathbb{P}[Y \in B], \forall B \in \mathcal{B}(\mathbb{R})$
- Ale různé náhodné veličiny (definované na různých pstních prostorech, nebo na stejném pstním prostoru, ale s různými hodnotami pro stejná $\omega \in \Omega$) mohou mít stejnou distribuční funkci (a tedy stejné rozdělení)

# Hod mincí dvakrát

## Příklad 2.8

Hodíme spravedlivou mincí dvakrát a nechť $X$ je počet líců. Pak $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ a $\mathbb{P}(X = 1) = 1/2$. Distribuční funkce je

$$F_X(x) = \begin{cases} 0, & x < 0; \\ 1/4, & 0 \leq x < 1; \\ 3/4, & 1 \leq x < 2; \\ 1, & x \geq 2. \end{cases}$$

Všimněte si, že funkce je zprava spojitá, neklesající a že je definována pro všechna $x$, přestože náhodná veličina nabývá pouze hodnot 0, 1 a 2.

# Vlastnosti distribuční funkce

**Věta 2.9 (Tři základní vlastnosti)**

*Pokud $F$ je distribuční funkce náhodné veličiny $X$, pak:*

(i) *$F$ je neklesající: $x_1 < x_2$ implikuje $F(x_1) \leq F(x_2)$.*

(ii) *$F$ je normovaná: $\lim_{x \downarrow -\infty} F(x) = 0$ a $\lim_{x \uparrow +\infty} F(x) = 1$.*

(iii) *$F$ je zprava spojitá: $F(x) = F(x^+)$ pro všechna $x$, kde $F(x^+) = \lim_{y \downarrow x+} F(y)$.*

- Důkaz opačného směru – konkrétně, že pokud funkce $F$ zobrazující $\mathbb{R}$ na $[0, 1]$ splňuje (i), (ii) a (iii), pak je distribuční funkcí nějaké náhodné veličiny – vyžaduje hlubší nástroje analýzy.
- Navíc *existuje právě jedna konečná borelovská míra $\mu$* na $\mathbb{R}$ taková, že $F(x) = \mu((-\infty, x])$, $\forall x \in \mathbb{R}$ – Lebesgue–Stieltjesova míra.
- $\mathbb{P}[X \in (a, b]] = \mathbb{P}[X \leq b] - \mathbb{P}[X \leq a] = F_X(b) - F_X(a)$, $a < b$

! obecně: $\mathbb{P}[X \in (a, b)] = F_X(b^-) - F_X(a)$ & $\mathbb{P}[X \in [a, b]] = F_X(b) - F_X(a^-)$

# Rozklad rozdělení pravděpodobnosti

Rozklad borelovské míry na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\mathbb{P}_X = \mu_{aspoj} + \mu_{singn} + \mu_{diska}$$

- $\mu_{aspoj}$ je absolutně spojitá míra vzhledem k Lebesgueově míře $\lambda$

$$\mu_{aspoj} \ll \lambda \quad (\forall B \in \mathcal{B}(\mathbb{R}) : \lambda(B) = 0 \Rightarrow \mu_{aspoj}(B) = 0)$$

- $\mu_{singn}$ je singulární neatomická míra vzhledem k Lebesgueově míře $\lambda$

$$\forall x \in \mathbb{R} : \mu_{singn}(\{x\}) = 0 \quad \& \quad \mu_{singn} \perp \lambda \quad (\exists N \in \mathcal{B}(\mathbb{R}) : \lambda(N) = 0 \,\&\, \mu_{singn}(\mathbb{R} \setminus N) = 0)$$

- $\mu_{diska}$ je diskrétní (atomická) míra vzhledem k Lebesgueově míře $\lambda$

$$\exists \text{ spočetná } S \subset \mathbb{R} : \lambda(S) = 0 \,\&\, \mu_{diska}(\mathbb{R} \setminus S) = 0$$

pak $\mu_{diska}(\cdot) = \sum_{s \in S} a_s \mathbb{1}\{s \in \cdot\}$, kde $\mathbb{1}\{s \in \cdot\} =: \delta_s(\cdot)$ je Diracova míra v atomu $s$

# Diskrétní rozdělení

**Definice 2.10 (Diskrétní náhodné veličiny)**

Náhodnou veličinu $X$ nazveme diskrétní, pokud existuje (konečná nebo spočetná) indexová množina $I \subseteq \mathbb{N}$ a posloupnosti $\{x_i\}_{i \in I} \subset \mathbb{R}$ a $\{p_i\}_{i \in I} \subset (0, 1]$, takové, že

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i} \quad \text{a} \quad \sum_{i \in I} p_i = 1,$$

kde $\delta_u$ značí Diracovu míru v $u \in \mathbb{R}$.

- Diskrétní distribuční funkce odpovídající *rozdělení* $\mathbb{P}_X$ *diskrétní náhodné veličiny* $X$

$$F_X(x) = \sum_{i \in I} p_i \mathbb{1}_{[x_i, \infty)}(x) = \sum_{i \in I} p_i \mathbb{1}\{x_i \leq x\} = \sum_{i \in I}^{x_i \leq x} p_i$$

# Pravděpodobnostní funkce a diskrétní rozdělení

Nechť $X$ je *diskrétní náhodná veličina* z definice výše a $\nu$ *čítací míra* na $\{x_i\}_{i\in I} \subset \mathbb{R}$, pak:

- $\mathbb{P}_X \ll \nu$
- Hustota $f_X$ náhodné veličiny $X$ vzhledem k čítací míře $\nu$ se nazývá také pravděpodobnostní funkce diskrétního rozdělení $X$
- $f_X(x) = \begin{cases} p_i, & x = x_i; \\ 0, & x \notin \{x_i\}_{i\in I}. \end{cases}$

**Příklad 2.11**

Pravděpodobnostní funkce pro Příklad 2.8 je $f_X(x) = \begin{cases} 1/4, & x = 0; \\ 1/2, & x = 1; \\ 1/4, & x = 2; \\ 0, & \text{jinak.} \end{cases}$

# Spojitá rozdělení

**Definice 2.12 (Spojité náhodné veličiny)**

Náhodnou veličinu $X$ nazveme (absolutně) spojitou náhodnou veličinou, pokud $\mathbb{P}_X \ll \lambda$, kde $\lambda$ značí Lebesgueovu míru na $\mathbb{R}$.

- Distribuční funkce $F_X$ absolutně spojité náhodné veličiny má derivaci $F'_X$ $\lambda$-skoro všude a pro hustotu $f_X$ náhodné veličiny $X$ vzhledem k $\lambda$ platí

$$f_X = F'_X \quad \lambda\text{-s.v.}$$

- Absolutně spojitá distribuční funkce odpovídající *rozdělení* $\mathbb{P}_X$ *absolutně spojité veličiny* $X$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t, \quad \forall x \in \mathbb{R}$$

# Důsledky definice distribuční funkce

**Lemma 2.13 (Důsledky definice distribuční funkce)**

*Nechť $F$ je distribuční funkcí náhodné veličiny $X$. Potom:*

1. $\mathbb{P}(X = x) = F(x) - F(x^-)$, *kde* $F(x^-) = \lim_{y \uparrow x -} F(y)$;
2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$;
3. $\mathbb{P}(X > x) = 1 - F(x)$;
4. *Pokud $X$ je spojitá, pak*

$$\begin{aligned} F(b) - F(a) &= \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) \\ &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b). \end{aligned}$$

# Kvantilová funkce

**Definice 2.14 (Kvantil)**

Nechť $X$ je náhodná veličina s distribuční funkcí $F$. *Inverzní distribuční funkce* neboli kvantilová funkce je definována jako

$$F^{-1}(q) = \inf\Big\{x : F(x) > q\Big\}, \quad q \in (0,1).$$

- Pokud je $F$ striktně rostoucí a spojitá, pak $F^{-1}(q)$ je jediné $x \in \mathbb{R}$ takové, že $F(x) = q$.
- Kvantilová funkce je neklesající a zprava spojitá.
- Z $F^{-1}$ lze jednoznačně určit $F$, takže také charakterizuje rozdělení $\mathbb{P}_X$.
- $F^{-1}(1/4)$ je první kvartil, $F^{-1}(1/2)$ je medián (neboli druhý kvartil) a $F^{-1}(3/4)$ je třetí kvartil.
- Dvě náhodné veličiny $X$ a $Y$ jsou stejně rozdělené – zapisujeme $X \overset{d}{=} Y$ – právě tehdy, když $F_X(x) = F_Y(x)$ pro všechna $x$. To však neznamená, že $X = Y$.

# Příklady – Diskrétní náhodné veličiny I

**Příklad 2.15 (Bodové rozdělení – Konstanta skoro jistě)**

$X$ má bodové rozdělení v bodě $a$ právě tehdy, když $\mathbb{P}[X = x] = \mathbb{1}\{x = a\}$, $x \in \mathbb{R}$. Zapisujeme $X \sim \delta_a$ (Diracovo rozdělení v bodě $a$). Potom $F_X(x) = \mathbb{1}\{x \geq a\}$.

**Příklad 2.16 (Diskrétní rovnoměrné rozdělení)**

$X$ má diskrétní rovnoměrné rozdělení na $\{1, \ldots, k\}$ právě tehdy, když

$$f_X(x) = \begin{cases} 1/k, & \text{pro } x = 1, \ldots, k; \\ 0, & \text{jinak.} \end{cases}$$

# Příklady – Diskrétní náhodné veličiny II

## Příklad 2.17 (Bernoulliho rozdělení – Alternativní rozdělení)

$X$ má Bernoulliho (nebo alternativní; 0–1) rozdělení s parametrem $p \in (0,1)$ právě tehdy, když $f_X(x) = p^x(1-p)^{1-x}$ pro $x \in \{0,1\}$. Zapisujeme $X \sim \mathrm{Alt}(p)$ nebo $X \sim \mathrm{Be}(p)$.

- Rozdělení indikátoru náhodného jevu, úspěch vs. neúspěch.

## Příklad 2.18 (Binomické rozdělení)

$X$ má binomické rozdělení s parametry $n \in \mathbb{N}$ a $p \in (0,1)$ právě tehdy, když

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \mathbb{1}\{x \in \{0,\dots,n\}\}.$$

Zapisujeme $X \sim \mathrm{Bi}(n,p)$.

- Rozdělení součtu nezávislých alternativních veličin (počtu úspěchů mezi $n$ pokusy).
- Přesněji, jsou-li $\perp\!\!\!\perp X_i \sim \mathrm{Alt}(p)$, $i = 1,\dots,n$, pak $\sum_{i=1}^n X_i \sim \mathrm{Bi}(n,p)$.
- $\mathrm{Bi}(1,p)$ jest $\mathrm{Alt}(p)$.

# Příklady – Diskrétní náhodné veličiny III

## Příklad 2.19 (Geometrické rozdělení)

$X$ má geometrické rozdělení s parametrem $p \in (0,1)$ ($X \sim \mathrm{Geo}(p)$) právě tehdy, když

$$f_X(x) = p(1-p)^x \text{ pro } x \in \mathbb{N}_0.$$

- Počet neúspěchů před prvním úspěchem v posloupnosti nezávislých pokusů.

## Příklad 2.20 (Negativně binomické rozdělení)

$X$ má negativně binomické rozdělení s parametry $n \in \mathbb{N}$ a $p \in (0,1)$ ($X \sim \mathrm{NB}(n,p)$) právě tehdy, když

$$f_X(x) = \binom{n+x-1}{n-1} p^n (1-p)^x \text{ pro } x \in \mathbb{N}_0.$$

- Rozdělení počtu neúspěchů předcházejících $n$-tému úspěchu v posloupnosti nezávislých pokusů.
- $\mathrm{NB}(1,p)$ jest $\mathrm{Geo}(p)$.

# Příklady – Diskrétní náhodné veličiny IV

**Příklad 2.21 (Poissonovo rozdělení)**

$X$ má Poissonovo rozdělení s parametrem $\lambda > 0$ ($X \sim \mathrm{Po}(\lambda)$) právě tehdy, když

$$f_X(x) = \exp\{-\lambda\}\lambda^x / x! \text{ pro } x \in \mathbb{N}_0.$$

- Pokud $X \sim \mathrm{Po}(\lambda_X)$ a $Y \sim \mathrm{Po}(\lambda_Y)$ jsou nezávislé ($\perp\!\!\!\perp$), potom $X + Y \sim \mathrm{Po}(\lambda_X + \lambda_Y)$ (později snadné).
- Neplatí bez $\perp\!\!\!\perp$ (zkuste kontrapříklad).
- $\Leftarrow$ netriviální (Raikovova věta).
- Jestliže $n \to \infty$ a $np \to \lambda < \infty$, pak $\mathrm{Bi}(n, p)$ konverguje k $\mathrm{Po}(\lambda)$ (v zmyslu konvergence hustot).

# Příklady – Spojité náhodné veličiny I

## Příklad 2.22 (Rovnoměrné rozdělení)

$X$ má rovnoměrné rozdělení na intervalu $[a, b]$ právě tehdy, když $f_X(x) = (b-a)^{-1}\mathbb{1}\{x \in [a, b]\}$. Zapisujeme $X \sim \mathrm{U}(a, b)$ nebo $X \sim \mathrm{R}(a, b)$.

## Příklad 2.23 (Normální nebo Gaussovo rozdělení)

$X$ má normální (nebo Gaussovo) rozdělení s parametry $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ ($X \sim \mathrm{N}(\mu, \sigma^2)$) právě tehdy, když

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

- $X \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow Z = (X-\mu)/\sigma \sim \mathrm{N}(0, 1)$ (standardní normální rozdělení)
- $X \sim \mathrm{N}(\mu_X, \sigma_X^2) \perp\!\!\!\perp Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2) \Rightarrow X + Y \sim \mathrm{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ (později)
- Neplatí bez $\perp\!\!\!\perp$
- $\Leftarrow$ netriviální (Cramérova věta o dekompozici)

# Příklady – Spojité náhodné veličiny II

- Distribuční funkce $\mathrm{N}(0,1)$: $\Phi(x) = \int_{-\infty}^{x} \phi(t)\mathrm{d}t$ ... nelze vyjádřit uzavřeně, kde $\phi(x) := \frac{1}{\sqrt{2\pi}}\exp\{-x^2/2\}$, $x \in \mathbb{R}$ je hustota *standardního/normovaného/standardizovaného normálního rozdělení*

## Příklad 2.24

Předpokládejme, že $X \sim \mathrm{N}(3,5)$. Najděte $\mathbb{P}[X \geq 1]$. Řešení je:

$$\mathbb{P}[X \geq 1] = \mathbb{P}[X > 1] = 1 - \mathbb{P}[X \leq 1] = 1 - \mathbb{P}\left[Z \leq \frac{1-3}{\sqrt{5}}\right] = 1 - \Phi(-0.8944) = 0.81.$$

Nyní najděte $x = F_X^{-1}(0.2)$. Řešení: Z "tabulek" víme, že $\Phi(-0.8416) = 0.2$.

$$0.2 = \mathbb{P}[X \leq x] = \mathbb{P}\left[Z \leq \frac{x-\mu}{\sigma}\right] = \Phi\left[\frac{x-\mu}{\sigma}\right].$$

Proto $-0.8416 = \frac{x-\mu}{\sigma} = \frac{x-3}{\sqrt{5}}$ a tedy $x = 3 - 0.8416\sqrt{5} = 1.1181$.

# Příklady – Spojité náhodné veličiny III

## Příklad 2.25 (Exponenciální rozdělení)

$X$ má exponenciální rozdělení s parametrem $\lambda > 0$ ($X \sim \text{Exp}(\lambda)$) právě tehdy, když

$$f_X(x) = \lambda \exp\{-\lambda x\} \mathbb{1}\{x > 0\}.$$

## Příklad 2.26 (Gamma rozdělení)

$X$ má Gamma rozdělení s parametry $a, p > 0$ právě tehdy, když

$$f_X(x) = \frac{a^p}{\Gamma(p)} x^{p-1} \exp\{-ax\} \mathbb{1}\{x > 0\},$$

kde $\Gamma(p) = \int_0^\infty t^{p-1} \exp\{-t\} \mathrm{d}t$ je Gamma funkce. Zapisujeme $X \sim \text{Gamma}(a, p)$ nebo $X \sim \Gamma(a, p)$.

- $X \sim \text{Exp}(a) \Rightarrow X \sim \text{Gamma}(a, 1)$
- $X \sim \text{Gamma}(a, p_X) \perp\!\!\!\perp Y \sim \text{Gamma}(a, p_Y) \Rightarrow X + Y \sim \text{Gamma}(a, p_X + p_Y)$

# Příklady – Spojité náhodné veličiny IV

## Příklad 2.27 (Beta rozdělení)

$X$ má Beta rozdělení s parametry $\alpha, \beta > 0$ právě tehdy, když

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}\{x \in (0,1)\}.$$

Zapisujeme $X \sim \mathrm{Beta}(\alpha, \beta)$ nebo $X \sim \mathrm{B}(\alpha, \beta)$.

## Příklad 2.28 ($\chi^2$-rozdělení)

$X$ má $\chi^2$-rozdělení s $p$ stupni volnosti ($X \sim \chi^2_p$) právě tehdy, když

$$f_X(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1}\exp\{-x/2\}\mathbb{1}\{x > 0\}.$$

- Pokud $X_1, \ldots, X_p \overset{\perp\!\!\!\perp}{\sim} \mathrm{N}(0,1)$, pak $\sum_{i=1}^p X_i^2 \sim \chi^2_p$.

# Příklady – Spojité náhodné veličiny V

## Příklad 2.29 (Studentovo $t$-rozdělení)

$X$ má Studentovo $t$-rozdělení s $\nu$ stupni volnosti ($X \sim t_\nu$) právě tehdy, když

$$f_X(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \frac{1}{(1+x^2/\nu)^{(\nu+1)/2}}.$$

- $X \sim \mathrm{N}(0,1) \perp\!\!\!\perp Y \sim \chi^2_\nu \Rightarrow \frac{X}{\sqrt{Y/\nu}} \sim t_\nu$
- Standardní normální rozdělení odpovídá $t$-rozdělení s $\nu = \infty$

## Příklad 2.30 (Cauchyho rozdělení)

Cauchyho rozdělení je speciální případ $t$-rozdělení, když $\nu = 1$ ($X \sim$ Cauchy).

- $X \sim \text{Cauchy} \Rightarrow f_X(x) = \frac{1}{\pi(1+x^2)}$

# Mnohorozměrná náhodnost

## Definice 2.31 (Náhodný vektor)

Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. Náhodný vektor je měřitelné zobrazení, které každému výsledku $\omega$ přiřadí reálný *d-rozměrný vektor* $\mathbb{X}(\omega)$. To znamená, že

$$\mathbb{X} : \Omega \to \mathbb{R}^d \quad \& \quad \{\omega \in \Omega : \mathbb{X}(\omega) \leq \mathbb{x}\} \in \mathcal{A}, \forall \mathbb{x} \in \mathbb{R}^d.$$

## Definice 2.32 (Rozdělení náhodného vektoru)

*Rozdělením náhodného vektoru* $\mathbb{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ nazveme indukovanou pravděpodobnostní míru $\mathbb{P}_X$ na $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ definovanou jako

$$\mathbb{P}_{\mathbb{X}}(B) := \mathbb{P}[\{\omega \in \Omega : \mathbb{X}(\omega) \in B\}], \quad B \in \mathcal{B}(\mathbb{R}^d).$$

- $\mathbb{P}_{\mathbb{X}}$ je obraz míry $\mathbb{P}$ v zobrazení $\mathbb{X}$ a $(\Omega, \mathcal{A}, \mathbb{P})$ se zobrazí na pstní prostor $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_{\mathbb{X}})$ – analogie s náhodnými veličinami.
- $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ je náhodný vektor $\iff$ $X_\ell$ je náhodná veličina $\forall \ell \in \{1, \ldots, d\}$.

# Sdružená distribuční funkce

**Definice 2.33 (Sdružená distribuční funkce)**

Nechť $(\Omega, \mathcal{A}, \mathbb{P})$ je pravděpodobnostní prostor. *Sdružená distribuční funkce* náhodného vektoru $\mathbb{X} = [X_1, \ldots, X_d]^\top$ je funkce $F_{\mathbb{X}} : \mathbb{R}^d \to [0, 1]$ definovaná jako

$$F_{\mathbb{X}}(\mathbb{x}) = \mathbb{P}(\mathbb{X} \leq \mathbb{x}) \equiv \mathbb{P}\big( \cap_{\ell=1}^{d} \{X_\ell \leq x_\ell\}\big), \quad \forall \mathbb{x} = [x_1, \ldots, x_d]^\top \in \mathbb{R}^d.$$

**Věta 2.34 (Vlastnosti sdružené distribuční funkce)**

*Pokud je $F$ sdružená distribuční funkce $d$-rozměrného náhodného vektoru $\mathbb{X}$, pak platí:*

(i) *$F$ je po složkách neklesající a zprava spojitá;*

(ii)

$$\lim_{x_\ell \downarrow -\infty} F(\mathbb{x}) = 0 \quad \textit{pro každé } \ell = 1, \ldots, d$$

*a*

$$\lim_{x_\ell \uparrow +\infty \, \forall \ell} F(\mathbb{x}) = 1;$$

# Marginály

**Věta 2.35 (Marginální distribuční funkce)**

*Pokud je $F_{\mathbb{X}}$ je sdružená distribuční funkce náhodného vektoru $\mathbb{X} = [X_1, \ldots, X_d]^\top$, pak*

$$\lim_{x_d \uparrow +\infty} F_{\mathbb{X}}(x_1, \ldots, x_d) = F_{[X_1, \ldots, X_{d-1}]^\top}(x_1, \ldots, x_{d-1}), \quad \forall \mathbb{x} = [x_1, \ldots, x_d]^\top \in \mathbb{R}^d,$$

*kde $F_{[X_1, \ldots, X_{d-1}]^\top}$ je distribuční funkce náhodného (pod)vektoru $[X_1, \ldots, X_{d-1}]^\top$.*

- Limitní přechod můžeme *zopakovat* vícekrát a "marginalizovat" až na jednorozměrný případ (tedy náhodnou veličinu)
- Složky můžeme zapermutovat *permutací* $\pi : \{1, \ldots, d\} \to \{1, \ldots, d\}$ a pak platí:

$$F_{[X_1, \ldots, X_d]^\top}(x_1, \ldots, x_d) = F_{[\pi(X_1), \ldots, \pi(X_d)]^\top}(\pi(x_1), \ldots, \pi(x_d))$$

# Rozdělení podvektorů

**Definice 2.36 (Rozdělení náhodného podvektoru – Marginální rozdělení)**

Nechť $J \subseteq \{1, \dots, d\}$ a $|J| = m$. *Rozdělením náhodného podvektoru* $\mathbf{Y} \equiv \{X_\ell\}_{\ell \in J} : (\Omega, \mathcal{A}) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ nazveme marginálním rozdělením.

- Marginální rozdělení podvektoru $\mathbf{Y}$ je pak indukovaná pravděpodobnostní míra $\mathbb{P}_\mathbf{Y}$ na $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ taková, že

$$\mathbb{P}_\mathbf{Y}(B) := \mathbb{P}[\{\omega \in \Omega : \mathbf{Y}(\omega) \in B \,\&\, \{X_\ell\}_{\ell \in \{1,\dots,d\} \setminus J} \in \mathbb{R}^{d-m}\}], \quad B \in \mathcal{B}(\mathbb{R}^m).$$

- Speciálně když $J = \{1, \dots, m\}$, pak $\mathbb{P}_\mathbf{Y}(B) = \mathbb{P}_\mathbf{X}(B \times \mathbb{R}^{d-m})$.
- Když $|J| = 1$, jedná se o náhodnou veličinu.

# Diskrétní náhodné vektory

**Definice 2.37 (Diskrétní náhodný vektor)**

Náhodný vektor $\mathbb{X} = [X_1, \dots, X_d]^\intercal$ nazveme diskrétní, pokud existuje (konečná nebo spočetná) indexová množina $I \subseteq \mathbb{N}$ a posloupnosti $\{\mathbb{x}_i\}_{i \in I} \subset \mathbb{R}^d$ a $\{p_i\}_{i \in I} \subset (0, 1]$, takové, že

$$\mathbb{P}_{\mathbb{X}} = \sum_{i \in I} p_i \delta_{\mathbb{x}_i} \quad \text{a} \quad \sum_{i \in I} p_i = 1,$$

kde $\delta_{\mathbb{u}}$ značí Diracovu míru v $\mathbb{u} \in \mathbb{R}^d$.

- Diskrétní sdružená distribuční funkce odpovídající *rozdělení* $\mathbb{P}_{\mathbb{X}}$ *diskrétního náhodného vektoru* $\mathbb{X}$

$$F_{\mathbb{X}}(\mathbb{x}) = \sum_{i \in I} p_i \mathbb{1}_{[\mathbb{x}_i, \infty)}(\mathbb{x}) = \sum_{i \in I} p_i \mathbb{1}\{\mathbb{x}_i \leq \mathbb{x}\} = \sum_{i \in I}^{\mathbb{x}_i \leq \mathbb{x}} p_i$$

# Spojité náhodné vektory

**Definice 2.38 (Spojitý náhodný vektor)**

Náhodný vektor $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ nazveme (absolutně) spojitý náhodný vektor, pokud $\mathbb{P}_\mathbb{X} \ll \lambda^d$, kde $\lambda^d$ značí Lebesgueovu míru na $\mathbb{R}^d$.

- Sdružená distribuční funkce $F_\mathbb{X}$ absolutně spojitého náhodného vektoru má derivaci $\frac{\partial^d}{\partial x_1 \ldots \partial x_d} F_\mathbb{X}$ $\lambda^d$-skoro všude a pro sdruženou hustotu $f_\mathbb{X}$ náhodného vektoru $\mathbb{X}$ vzhledem k $\lambda^d$ platí

$$f_\mathbb{X} = \frac{\partial^d}{\partial x_1 \ldots \partial x_d} F_\mathbb{X} \quad \lambda^d\text{-s.v.}$$

- Absolutně spojitá distribuční funkce odpovídající *rozdělení* $\mathbb{P}_\mathbb{X}$ *absolutně spojitého náhodného vektoru* $\mathbb{X}$

$$F_\mathbb{X}(\mathbb{x}) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_d} f_\mathbb{X}(t_1, \ldots, t_d) \mathrm{d}t_d \ldots \mathrm{d}t_1, \quad \forall \mathbb{x} = [x_1, \ldots, x_d]^\intercal \in \mathbb{R}^d$$

# Hustota mnohorozměrného rozdělení

## Věta 2.39 (Hustota vzhledem k součinové referenční míře)

*Nechť* $\mathbb{P}_{\mathbb{X}}$ *je rozdělení náhodného vektoru* $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ *a nechť existují* $\sigma$*-konečné míry* $\mu_\ell$, $\ell \in \{1, \ldots, d\}$ *na* $\mathbb{R}$ *takové, že pro ich součin platí* $\mathbb{P}_{\mathbb{X}} \ll \mu_1 \otimes \ldots \otimes \mu_d$. *Potom* $\mathbb{P}_{X_\ell} \ll \mu_\ell$, $\forall \ell \in \{1, \ldots, d\}$. *Dále pak existují nezáporné měřitelné funkce (hustoty)* $f_{\mathbb{X}} : \mathbb{R}^d \to [0, \infty)$ *a* $f_{X_\ell} : \mathbb{R} \to [0, \infty)$, $\ell \in \{1, \ldots, d\}$ *takové, že:*

$$\mathbb{P}_{\mathbb{X}}\big(\times_{\ell=1}^{d} (-\infty, x_\ell]\big) = F_{\mathbb{X}}(\mathbb{x}) = \int_{\times_{\ell=1}^{d}(-\infty, x_\ell]} f_{\mathbb{X}}(\mathbb{t}) d(\mu_1 \otimes \ldots \otimes \mu_d)(\mathbb{t}), \ \forall \mathbb{x} = [x_1, \ldots, x_d]^\intercal;$$

$$\mathbb{P}_{\mathbb{X}}(B) = \int_B f_{\mathbb{X}}(\mathbb{t}) d(\mu_1 \otimes \ldots \otimes \mu_d)(\mathbb{t}), \ \forall B \in \mathcal{B}(\mathbb{R}^d).$$

*Navíc* $F_{X_\ell}(x_\ell) = \int_{-\infty}^{x_\ell} f_{X_\ell}(t) d\mu_\ell(t)$, $\forall x_\ell \in \mathbb{R}$, $\forall \ell \in \{1, \ldots, d\}$, *kde pro* $\mu_\ell$*-s.v.* $t_\ell \in \mathbb{R}$:

$$f_{X_\ell}(t_\ell) = \int_{\mathbb{R}^{d-1}} f_{\mathbb{X}}(t_1, \ldots, t_d) d(\mu_1 \otimes \ldots \otimes \mu_{\ell-1} \otimes \mu_{\ell+1} \otimes \ldots \otimes \mu_d)(t_1, \ldots, t_{\ell-1}, t_{\ell+1}, \ldots, t_d).$$

# Pravděpodobnostní funkce pro mnohorozměrné

Nechť $\mathbb{X}$ je *diskrétní náhodný vektor* a $\nu$ *čítací míra* na $\{\mathbb{x}_i\}_{i\in I} \subset \mathbb{R}^d$, pak:

- $\mathbb{P}_{\mathbb{X}} \ll \nu$
- Hustota $f_{\mathbb{X}}$ náhodného vektoru $\mathbb{X}$ vzhledem k čítací míře $\nu$ se nazývá také pravděpodobnostní funkce diskrétního mnohorozměrného rozdělení $\mathbb{X}$
- $f_{\mathbb{X}}(\mathbb{x}) = \begin{cases} p_i, & \mathbb{x} = \mathbb{x}_i; \\ 0, & \mathbb{x} \notin \{\mathbb{x}_i\}_{i\in I}. \end{cases}$

## Příklad 2.40

Dvojrozměrné rozdělení náhodného vektoru $[X, Y]^\intercal$:

| | $Y = 0$ | $Y = 1$ | |
|---|---|---|---|
| $X = 0$ | 1/9 | 2/9 | 1/3 |
| $X = 1$ | 2/9 | 4/9 | 2/3 |
| | 1/3 | 2/3 | 1 |

Pak $f_{[X,Y]^\intercal}(1,1) = \mathbb{P}(X = 1, Y = 1) = 4/9$.

# Hustota pro mnohorozměrné rozdělení

**Příklad 2.41**

Nechť dvourozměrný náhodný vektor $[X, Y]^\intercal$ je rovnoměrně rodělen na jednotkovém čtverci. Pak $f_{[X,Y]^\intercal}(x, y) = \mathbb{1}\{(x, y) \in [0, 1]^2\}$. Určete $\mathbb{P}[X < 1/2, Y < 1/2]$.

Událost $B = \{X < 1/2, Y < 1/2\}$ odpovídá podmnožině jednotkového čtverce. Integrace $f_{[X,Y]^\intercal}$ přes tuto podmnožinu v tomto případě odpovídá výpočtu plochy množiny $\{x < 1/2, y < 1/2\}$, která je $1/4$. Takže $\mathbb{P}[X < 1/2, Y < 1/2] = 1/4$.

# Marginální rozdělení – Diskrétní a spojité

**Důsledek 2.42 (Marginální rozdělení pro dvourozměrné náhodné vektory)**

*Pokud diskrétní náhodný vektor* $[X, Y]^\intercal$ *má sdruženou pravděpodobnostní funkce* $f_{[X,Y]^\intercal}$, *pak marginální pravděpodobnostní funkce pro* $X$ *je*

$$f_X(x) = \mathbb{P}[X = x] = \sum_y \mathbb{P}[X = x, Y = y] = \sum_y f_{[X,Y]^\intercal}(x, y).$$

*Pokud spojitý náhodný vektor* $[X, Y]^\intercal$ *má sdruženou hustotu* $f_{[X,Y]^\intercal}$, *pak marginální hustota pro* $X$ *je*

$$f_X(x) = \int f_{[X,Y]^\intercal}(x, y)\,dy.$$

- Lze snadno zobecnit pro vícerozměrné náhodné vektory – Věta 2.39

# Nezávislost II

Všimněte si, že

$$\lim_{x_j \uparrow +\infty \forall j \in \{1,\dots,k\} \setminus \{\ell\}} F_{\mathbb{X}}(\mathbb{x}) = \lim_{x_j \uparrow +\infty \forall j \in \{1,\dots,k\} \setminus \{\ell\}} \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d]$$
$$= \mathbb{P}[X_\ell \leq x_\ell] = F_{X_\ell}(x_\ell).$$

**Definice 2.43 (Nezávislé náhodné veličiny)**

Náhodné veličiny $X_1, \dots, X_d$ jsou nezávislé, pokud

$$F_{\mathbb{X}}(\mathbb{x}) = \prod_{\ell=1}^{d} F_{X_\ell}(x_\ell) \quad \text{pro každé } \mathbb{x} = [x_1, \dots, x_d]^\intercal \in \mathbb{R}^d,$$

kde $\mathbb{X} = [X_1, \dots, X_d]^\intercal$.

Pokud jsou $X$ a $Y$ nezávislé náhodné veličiny, píšeme $X \perp\!\!\!\perp Y$. Pokud nejsou nezávislé (jsou závislé), píšeme $X \not\perp\!\!\!\perp Y$.

# Nezávislé náhodné vektory

Analogie s předchozí Definicí 2.43

**Definice 2.44 (Nezávislé náhodné vektory)**

Náhodné vektory $\mathbb{X}_1, \ldots, \mathbb{X}_d$ jsou nezávislé, pokud

$$F_{\mathbb{X}}(\mathbb{x}) = \prod_{\ell=1}^{d} F_{\mathbb{X}_\ell}(\mathbb{x}_\ell) \quad \text{pro každé } \mathbb{x} = [\mathbb{x}_1^\intercal, \ldots, \mathbb{x}_d^\intercal]^\intercal \in \mathbb{R}^{\sum_{\ell=1}^{d} d_\ell},$$

kde $\mathbb{X} = [\mathbb{X}_1^\intercal, \ldots, \mathbb{X}_d^\intercal]^\intercal$ a $\mathbb{X}_\ell$ je $d_\ell$-rozměrný náhodný vektor pro všechna $\ell \in \{1, \ldots, d\}$.

# Diskrétní a spojitá nezávislost

- Nosič diskrétní náhodné veličiny $X$ ... $\mathcal{S}(X) = \{x \in \mathbb{R} : \mathbb{P}[X = x] > 0\}$
- Nosič spojité náhodné veličiny $X$ ... $\mathcal{S}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$

**Věta 2.45 (Ekvivalentní charakterizace nezávislosti)**

*Nechť sdružená pravděpodobnostní funkce diskrétního náhodného vektoru* $\mathbb{X} = [X_1, \dots, X_d]^\top$ *je* $f_{\mathbb{X}}(\mathbb{x}) = \mathbb{P}[\mathbb{X} = \mathbb{x}]$. *Pak platí* $\perp\!\!\!\perp \{X_1, \dots, X_d\}$ *právě tehdy, když*

$$\mathbb{P}[\mathbb{X} = \mathbb{x}] = \prod_{\ell=1}^{d} \mathbb{P}[X_\ell = x_\ell] \quad \textit{pro všechna } \mathbb{x} = [x_1, \dots, x_d]^\top \in \times_{\ell=1}^{d} \mathcal{S}(X_\ell).$$

*Nechť sdružená hustota spojitého náhodného vektoru* $\mathbb{X} = [X_1, \dots, X_d]^\top$ *je* $f_{\mathbb{X}}(\mathbb{x})$. *Pak platí* $\perp\!\!\!\perp \{X_1, \dots, X_d\}$ *právě tehdy, když*

$$f_{\mathbb{X}}(\mathbb{x}) = \prod_{\ell=1}^{d} f_{X_\ell}(x_\ell) \quad \lambda^d\textit{-s.v. } \mathbb{x} = [x_1, \dots, x_d]^\top \in \times_{\ell=1}^{d} \mathcal{S}(X_\ell).$$

# Kartézský nosič a nezávislost

- Nosič diskrétního náhodného vektoru $\mathbb{X}$ ... $\mathcal{S}(\mathbb{X}) = \{\mathbb{x} : \mathbb{P}[\mathbb{X} = \mathbb{x}] > 0\}$
- Nosič spojitého náhodného vektoru $\mathbb{X}$ ... $\mathcal{S}(\mathbb{X}) = \{\mathbb{x} : f_{\mathbb{X}}(\mathbb{x}) > 0\}$

**Důsledek 2.46 (Kartézský součin nosičů a nezávislost)**

*Předpokládejme, že $\mathcal{S}([X, Y]^\intercal) = \mathcal{S}(X) \times \mathcal{S}(Y)$. Pokud pro sdruženou pravděpodobnostní funkci diskrétního náhodného vektoru $[X, Y]^\intercal$ platí*

$$\mathbb{P}[X = x, Y = y] = g(x)h(y)$$

*pro nějaké měřitelné funkce $g$, $h$, a pro všechna $[x, y]^\intercal \in \mathcal{S}([X, Y]^\intercal)$, pak $X \perp\!\!\!\perp Y$. Pokud pro sdruženou hustotu spojitého náhodného vektoru $[X, Y]^\intercal$ platí*

$$f_{[X,Y]^\intercal}(x, y) = g(x)h(y)$$

*pro nějaké měřitelné funkce $g$, $h$, a pro $\lambda^2$-s.v. $[x, y]^\intercal \in \mathcal{S}([X, Y]^\intercal)$, pak $X \perp\!\!\!\perp Y$.*

- $g$ a $h$ *nemusí* být nutně hustoty nebo pravděpodobnostní funkce, ale $\bar{g} := \frac{g}{\int g}$ již ano

# Alternativní definice nezávislosti

**Věta 2.47 (Alternativní charakterizace nezávislosti)**

*Náhodné veličiny $X_1, \ldots, X_d$ jsou nezávislé právě tehdy, když*

$$\mathbb{P}_{\mathbb{X}} = \bigotimes_{\ell=1}^{d} \mathbb{P}_{X_\ell},$$

*kde $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$.*

# Podmíněné rozdělení

**Definice 2.48 (Podmíněná pravděpodobnostní funkce a hustota)**

Pro diskrétní náhodný vektor $[X, Y]^\intercal$ je *podmíněná pravděpodobnostní funkce* $X$ za podmínky $Y = y$:

$$f_{X|Y}(x|y) \equiv \mathbb{P}[X = x|Y = y] := \frac{\mathbb{P}[X = x, Y = y]}{\mathbb{P}[Y = y]} \equiv \frac{f_{[X,Y]^\intercal}(x, y)}{f_Y(y)} \text{ pokud } \mathbb{P}[Y = y] > 0.$$

Pro spojitý náhodný vektor $[X, Y]^\intercal$ je *podmíněná hustota* $X$ za podmínky $Y = y$:

$$f_{X|Y}(x|y) := \frac{f_{(X,Y)}(x, y)}{f_Y(y)} \text{ pokud } f_Y(y) > 0.$$

- Podmíněné pstní fce a hustota jsou *funkce* argumentu $x$ s parametrem $y$
- Jinak definovány *libovolně*
- Diskrétní ... $\mathbb{P}[X \in A|Y = y] = \sum_{x \in A} \mathbb{P}[X = x|Y = y]$
- Spojité ... $\mathbb{P}[X \in A|Y = y] = \int_A f_{X|Y}(x|y)\mathrm{d}x$

# Mnohorozměrné normální rozdělení

**Definice 2.49 (Mnohorozměrné normální rozdělení)**

Náhodný vektor $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ má *$d$-rozměrné normální rozdělení* s parametry $\boldsymbol{\mu} \in \mathbb{R}^d$ a $\Sigma \in \mathbb{R}^{d \times d}$, pokud existuje $k$-rozměrný náhodný vektor $\mathbb{Y} = [Y_1, \ldots, Y_k]^\intercal$ a matice $\mathbb{A} \in \mathbb{R}^{d \times k}$ takové, že

(i) $\perp\!\!\!\perp \{Y_1, \ldots, Y_k\}$;

(ii) $Y_j \sim \mathsf{N}(0, 1)$, $\forall j \in \{1, \ldots, k\}$;

(iii) $\mathbb{A}\mathbb{A}^\intercal = \Sigma$;

(iv) $\mathbb{X} = \mathbb{A}\mathbb{Y} + \boldsymbol{\mu}$.

- Značení: $\mathbb{X} \sim \mathsf{N}_d(\boldsymbol{\mu}, \Sigma)$
- Speciální případ: *standardní* $d$-rozměrné normální rozdělení $\boldsymbol{\mu} = \mathbb{0}$ a $\Sigma = I_d$, tj. $\mathbb{X} \sim \mathsf{N}_d(\mathbb{0}, I_d)$
- Matice $\mathbb{A}$ je deterministická (nenáhodná); vektor $\boldsymbol{\mu}$ je též deterministický (nenáhodný)

# Transformace diskrétních náhodných veličin

- Transformace $Y = t(X)$ (ne nutně monotónní)
- Mějme *známé diskrétní* rozdělení veličiny $X$, tj. $\mathbb{P}[X = x]$
- Cíl je $\mathbb{P}[Y = y]$:

$$\mathbb{P}[Y = y] = \mathbb{P}[t(X) = y] = \mathbb{P}[\omega : t(X(\omega)) = y] = \mathbb{P}[X \in t^{-1}(y)] = \sum_{x}^{t(x)=y} \mathbb{P}[X = x],$$

kde $t^{-1}$ určuje všechny *vzory*

# Transformace spojitých náhodných veličin

- Transformace $Y = t(X)$ (ne nutně monotónní)
- Mějme *známé spojité* rozdělení veličiny $X$, tj. $f_X(x)$
- Cíl je $f_Y(y)$:

1. Pro každé $y$ najdeme množinu $\mathcal{T}(y) = \{x : t(x) \leq y\}$
2. Distribuční funkce je

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[t(X) \leq y] = \mathbb{P}[\omega : t(X(\omega)) \leq y] = \int_{\mathcal{T}(y)} f_X(x)\mathrm{d}x$$

3. Pokud je $F_Y$ absolutně spojitá, hustota je $f_Y(y) = F'_Y(y)$

# Transformace diskrétních náhodných vektorů

- Transformace $Z = t(X, Y)$, kde $t : \mathbb{R}^2 \to \mathbb{R}$
- Mějme *známé diskrétní* rozdělení vektoru $[X, Y]^\intercal$, tj. $\mathbb{P}[X = x, Y = y]$
- Cíl je $\mathbb{P}[Z = z]$:

$$\mathbb{P}[Z = z] = \mathbb{P}[t(X, Y) = z] = \mathbb{P}[\omega : t([X, Y]^\intercal(\omega)) = z] = \mathbb{P}[[X, Y]^\intercal \in t^{-1}(z)]$$
$$= \sum_{[x,y]}^{t(x,y)=z} \mathbb{P}[X = x, Y = y],$$

kde $t^{-1}$ určuje všechny *vzory*

# Transformace spojitých náhodných vektorů

- Transformace $Z = t(X, Y)$, kde $t : \mathbb{R}^2 \to \mathbb{R}$
- Mějme *známé spojité* rozdělení vektoru $[X, Y]^\intercal$, tj. $f_{[X,Y]^\intercal}(x, y)$
- Cíl je $f_Z(z)$:

1. Pro každé $z$ najděte množinu $\mathcal{T}(z) = \{[x, y] : t(x, y) \leq z\}$
2. Distribuční funkce je

$$F_Z(z) = \mathbb{P}[Z \leq z] = \mathbb{P}[t(X, Y) \leq z] = \mathbb{P}[\omega : t([X, Y]^\intercal(\omega)) \leq z]$$
$$= \iint_{\mathcal{T}(z)} f_{[X,Y]^\intercal}(x, y) \mathrm{d}x \mathrm{d}y$$

3. Pokud je $F_Z$ absolutně spojitá, hustota je $f_Z(z) = F'_Z(z)$

# Agenda


## 3. Střední hodnota

3.1 Definice
3.2 Diskrétní a spojitý případ
3.3 Střední hodnota transformace
3.4 Moment
3.5 Vlastnosti
3.6 Rozptyl
3.7 Kovariance a korelace
3.8 Varianční-kovarianční matice
3.9 Momentová vytvořující funkce
3.10 Charakteristická funkce

# Střední hodnota nebo očekávaná hodnota

**Definice 3.1 (Střední hodnota)**

*Střední hodnota* náhodné veličiny $X$ je

$$\mathbb{E}[X] \equiv \mathbb{E}X = \int X \mathrm{d}\mathbb{P} \equiv \int X(\omega)\mathrm{d}\mathbb{P}(\omega),$$

pokud pravá strana existuje.

- Nepraktické pro výpočet, ale máme větu o přenosu integrace
- Pomocí *obrazu pravděpodobnostní míry* $\mathbb{P}_X(\cdot) := \mathbb{P}[X \in \cdot] \equiv \mathbb{P}[X^{-1}(\cdot)]$, kde $\mathbb{P}_X$ je rozdělení náhodné veličiny $X$:

**Věta 3.2 (Výpočet střední hodnoty)**

*Střední hodnota náhodné veličiny $X$ je $\mathbb{E}X = \int x d\mathbb{P}_X(x)$, pokud pravá strana existuje.*

# Střední hodnota pro diskrétní a spojité veličiny

Pomocí *Radon–Nikodýmovy věty*:

**Věta 3.3 (Diskrétní a spojitá střední hodnota)**

*Střední hodnota veličiny $X$ je*

$$\mathbb{E}X = \begin{cases} \int_{-\infty}^{+\infty} x f_X(x)\,dx & \text{pokud je } X \text{ spojitá;} \\ \sum_{x \in \mathcal{S}(X)} x \mathbb{P}[X = x] & \text{pokud je } X \text{ diskrétní,} \end{cases}$$

*za předpokladu, že pravá strana je dobře definovaná.*

# Pozor na střední hodnotu

**Příklad 3.4 (Cauchyho rozdělení)**

Pokud $X \sim$ Cauchy, pak $\mathbb{E}X$ neexistuje. Pomocí per partes,

$$\int_0^\infty \frac{x}{\pi(1+x^2)}\mathrm{d}x = [x\arctan(x)]_0^\infty - \int_0^\infty \arctan(x)\mathrm{d}x = \infty.$$

Tedy pro $\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\mathrm{d}x$ není $\infty - \infty$ definováno.

# Střední hodnota a transformace

Transformace $Y = t(X)$.

**Věta 3.5 (Pravidlo líného statistika)**

*Buď $t : \mathbb{R} \to \mathbb{R}$ měřitelná funkce a nechť $Y = t(X)$. Pak*

$$\mathbb{E}\, Y = \int t(x) d\mathbb{P}_X(x),$$

*pokud pravá strana existuje.*

- Také známé jako *Zákon nevědomého statistika*
- Diskrétní ... $\mathbb{E}\, Y = \sum_{x \in \mathcal{S}(X)} t(x) \mathbb{P}[X = x]$
- Spojitá ... $\mathbb{E}\, Y = \int_{-\infty}^{+\infty} t(x) f_X(x) \mathrm{d}x$

# Moment

**Definice 3.6 (Moment)**

$k$-tý *moment* $X$ je definován jako $\mathbb{E}[X^k]$ za předpokladu, že $\mathbb{E}[|X|^k] < \infty$.

**Věta 3.7 (Vyšší momenty)**

*Pokud existuje $k$-tý moment, pak existuje $\ell$-tý moment pro jakékoli $0 < \ell \leq k$.*

**Příklad 3.8 (Studentovo $t$-rozdělení jako "kontra" příklad)**

$t$-rozdělení s $\nu = 3$ stupni volnosti: $\mathbb{E}X = 0$, $\mathbb{E}X^2 = \infty$, $\mathbb{E}X^3$ neexistuje.

**Definice 3.9 (Absolutní moment)**

$k$-tý *absolutní moment* $X$ je definován jako $\mathbb{E}[|X|^k]$ za předpokladu, že existuje.

# Vlastnosti střední hodnoty

**Definice 3.10 ($\mathcal{L}^p$-prostory náhodných veličin)**

$X \in \mathcal{L}^p \equiv \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, pokud $\mathbb{E}[|X|^p] < \infty$.

**Věta 3.11 (Linearita)**

*Pokud* $X_1, \dots, X_d \in \mathcal{L}^1$ *a* $a_1, \dots, a_d$ *jsou konstanty, pak* $\mathbb{E}\left(\sum_{\ell=1}^d a_\ell X_\ell\right) = \sum_{\ell=1}^d a_\ell \mathbb{E} X_\ell$.

**Věta 3.12 (Násobení při nezávislosti)**

*Pokud* $X_1, \dots, X_d \in \mathcal{L}^1$ *jsou nezávislé náhodné veličiny, pak* $\mathbb{E}\left(\prod_{\ell=1}^d X_\ell\right) = \prod_{\ell=1}^d \mathbb{E} X_\ell$.

# Rozptyl

Rozptyl měří "variabilitu" (disperzi) rozdělení.

**Definice 3.13 (Rozptyl a směrodatná odchylka)**

*Rozptyl* $X$ je definován jako

$$\mathbb{V}\mathrm{ar}\, X = \mathbb{E}(X - \mathbb{E}X)^2,$$

za předpokladu, že pravá strana je dobře definovaná. Pak *směrodatná odchylka* $X$ je

$$\mathrm{sd}(X) = \sqrt{\mathbb{V}\mathrm{ar}\, X}.$$

- Nemůžeme použít $\mathbb{E}(X - \mathbb{E}X)$ jako míru variability, protože $\mathbb{E}(X - \mathbb{E}X) = 0$.
- Můžeme a někdy používáme $\mathbb{E}|X - \mathbb{E}X|$ jako *míru variability*, ale častěji používáme rozptyl.

# Vlastnosti rozptylu

**Věta 3.14 (Vlastnosti rozptylu)**

*Za předpokladu, že uvažované druhé momenty jsou konečné, potom:*

- $$\mathbb{V}\mathrm{ar}\, X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \geq 0;$$
- *pokud* $a, b \in \mathbb{R}$, *pak*
$$\mathbb{V}\mathrm{ar}(aX + b) = a^2 \mathbb{V}\mathrm{ar}\, X;$$
- *pokud* $X_1, \ldots, X_d \in \mathcal{L}^2$ *jsou nezávislé a* $a_1, \ldots, a_d$ *jsou reálné konstanty, pak*
$$\mathbb{V}\mathrm{ar}\left(\sum_{\ell=1}^{d} a_\ell X_\ell\right) = \sum_{\ell=1}^{d} a_\ell^2 \mathbb{V}\mathrm{ar}\, X_\ell.$$

# Kovariance

Kovariance a korelace mezi $X$ a $Y$ měří, jak silný je lineární vztah mezi $X$ a $Y$.

**Definice 3.15 (Kovariance a korelace)**

*Kovariance* mezi $X$ a $Y$ je definována jako

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

Pokud $\mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y) > 0$, pak je *korelace* mezi $X$ a $Y$ definována jako

$$\rho_{X,Y} \equiv \mathbb{C}\mathrm{orr}(X, Y) = \frac{\mathbb{C}\mathrm{ov}(X, Y)}{\sqrt{\mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y)}}.$$

Všimněte si, že $\mathbb{E}[t(X, Y)] = \int_{\mathbb{R}^2} t(x, y)\mathrm{d}\mathbb{P}_{[X,Y]^\intercal}(x, y)$ a tedy,

$$\mathbb{E}XY = \begin{cases} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y & \text{pokud } X \text{ a } Y \text{ jsou spojité,} \\ \sum_{x\in\mathcal{S}(X), y\in\mathcal{S}(Y)} xy\mathbb{P}[X = x, Y = y] & \text{pokud } X \text{ a } Y \text{ jsou diskrétní} \end{cases}$$

# Vlastnosti kovariance a korelace

## Věta 3.16 (Vlastnosti kovariance a korelace)

- $\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$.
- $-1 \leq \mathbb{C}\mathrm{orr}(X, Y) \leq 1$.
- $|\mathbb{C}\mathrm{orr}(X, Y)| = 1 \iff \exists\ (a \neq 0\ \&\ b \in \mathbb{R}) :\ Y = aX + b$ *s pravděpodobností 1.*
- *Pokud jsou $X$ a $Y$ nezávislé, pak* $\mathbb{C}\mathrm{ov}(X, Y) = 0$.

Obecně platí, že $\mathbb{C}\mathrm{ov}(X, Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$.

## Důsledek 3.17 (Rozpyl součtu)

*Pokud $X_1, \ldots, X_d \in \mathcal{L}^2$ a $a_1, \ldots, a_d$ jsou reálné konstanty, pak*

$$\mathbb{V}\mathrm{ar}\left(\sum_{\ell=1}^{d} a_\ell X_\ell\right) = \sum_{\ell=1}^{d} a_\ell^2 \mathbb{V}\mathrm{ar} X_\ell + 2 \sum_{1 \leq j < \ell \leq d}\sum a_j a_\ell \mathbb{C}\mathrm{ov}(X_j, X_\ell).$$

# Varianční-kovarianční matice

**Definice 3.18 (Mnohorozměrná střední hodnota)**

*Střední hodnota* náhodného vektoru $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ je definována jako

$$\mathbb{E}\mathbb{X} = [\mathbb{E}X_1, \ldots, \mathbb{E}X_d]^\intercal.$$

Všimněte si, že $\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{C}\mathrm{ov}(Y, X)$ a $\mathbb{C}\mathrm{ov}(X, X) = \mathbb{V}\mathrm{ar}X$.

**Definice 3.19 (Varianční-kovarianční matice)**

*Varianční-kovarianční matice* náhodného vektoru $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ je definována jako

$$\mathbb{V}\mathrm{ar}\mathbb{X} = \begin{bmatrix} \mathbb{V}\mathrm{ar}X_1 & \mathbb{C}\mathrm{ov}(X_1, X_2) & \ldots & \mathbb{C}\mathrm{ov}(X_1, X_d) \\ \mathbb{C}\mathrm{ov}(X_2, X_1) & \mathbb{V}\mathrm{ar}X_2 & \ldots & \mathbb{C}\mathrm{ov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}\mathrm{ov}(X_d, X_1) & \mathbb{C}\mathrm{ov}(X_d, X_2) & \ldots & \mathbb{V}\mathrm{ar}X_d \end{bmatrix}.$$

# Vlastnosti varianční-kovarianční matice

**Věta 3.20 (Vlastnosti varianční-kovarianční matice)**

- *Pokud je* $\mathbb{X}$ *náhodný vektor,* $\mathbb{a}$ *je reálný vektor a* $b$ *je reálné číslo, pak*

$$\mathbb{E}(\mathbb{a}^\intercal \mathbb{X} + b) = \mathbb{a}^\intercal \mathbb{E}\mathbb{X} + b \quad a \quad \mathbb{V}\mathrm{ar}(\mathbb{a}^\intercal \mathbb{X} + b) = \mathbb{a}^\intercal (\mathbb{V}\mathrm{ar}\mathbb{X})\mathbb{a}.$$

- *Pokud je* $\mathbb{X}$ *náhodný vektor,* $\mathbb{A}$ *je reálná matice a* $\mathbb{b}$ *je reálný vektor, pak*

$$\mathbb{E}(\mathbb{A}\mathbb{X} + \mathbb{b}) = \mathbb{A}\mathbb{E}\mathbb{X} + \mathbb{b} \quad a \quad \mathbb{V}\mathrm{ar}(\mathbb{A}\mathbb{X} + \mathbb{b}) = \mathbb{A}(\mathbb{V}\mathrm{ar}\mathbb{X})\mathbb{A}^\intercal.$$

# Momentová vytvořující funkce

## Definice 3.21 (MGF)

*Momentová vytvořující funkce* (MGF) $X$ je definována jako

$$\psi_X(t) = \mathbb{E}[\exp\{tX\}] = \int_{\mathbb{R}} \exp\{tx\} \mathrm{d}\mathbb{P}_X(x), \quad t \in \mathbb{R},$$

pokud pravá strana existuje.

$\psi_X(-t)$ se nazývá *Laplaceova transformace* $X$

## Věta 3.22 (Vlastnosti MGF)

- $\exists \epsilon > 0\ \forall t \in (-\epsilon, \epsilon)\ \exists \psi_X(t) \implies \psi_X^{(m)}(0) = \mathbb{E}X^m$, $m \in \mathbb{N}_0$.
- *Pokud* $Y = aX + b$, *pak* $\psi_Y(t) = \exp\{bt\}\psi_X(at)$.
- *Pokud* $X_1, \ldots, X_d$ *jsou nezávislé a* $Y = \sum_{\ell=1}^d X_\ell$, *pak* $\psi_Y(t) = \prod_{\ell=1}^d \psi_{X_\ell}(t)$.

Pokud $\psi_X(t) = \psi_Y(t)$ pro všechna $t$ v otevřeném intervalu kolem 0, pak $X \overset{d}{=} Y$.

# Charakteristická funkce

**Definice 3.23 (CF)**

*Charakteristická funkce* (CF) $X$ je definována jako

$$\varphi_X(t) = \mathbb{E}[\exp\{itX\}] = \int_{\mathbb{R}} \exp\{itx\} \mathrm{d}\mathbb{P}_X(x), \quad t \in \mathbb{R}.$$

- *Charakteristická funkce* je dobře definována pro všechna $t \in \mathbb{R}$ (na rozdíl od MGF)
- *Fourierova transformace*: $\mathbb{E}[\exp\{-itX\}] = \varphi_X(-t)$
- $\varphi_X(t) = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]$

# Vlastnosti charakteristické funkce

**Věta 3.24 (Vlastnosti CF)**

(i) $\varphi_X$ *existuje pro jakékoli rozdělení* $X$

(ii) $\varphi_X(0) = 1$

(iii) $|\varphi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}$

(iv) $\varphi_X$ *je stejnoměrně spojitá* $- \forall \varepsilon > 0\ \exists \delta > 0$: $|\varphi_X(t) - \varphi_X(s)| \leq \varepsilon$ *kdykoli* $|t - s| \leq \delta$

(v) $\varphi_{a+bX}(t) = e^{iat}\varphi_X(bt) \quad \forall t \in \mathbb{R} \quad \forall a, b \in \mathbb{R}$

(vi) $\varphi_{-X}(t) = \bar{\varphi}_X(t) \quad \forall t \in \mathbb{R}$ *(komplexní sdružení)*

(vii) $\varphi_X(t) \in \mathbb{R} \quad \forall t \in \mathbb{R} \iff \mathbb{P}[X > x] = \mathbb{P}[X < -x]\ \forall x \geq 0$ *(rozdělení symetrické kolem nuly)*

(viii) $X \perp\!\!\!\perp Y \Rightarrow \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) \quad \forall t \in \mathbb{R}$

# Jednoznačná charakterizace rozdělení

**Věta 3.25 (Inverzní formule)**

*Pro jakékoli* $a < b$,

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\,dt = \mathbb{P}[a < X < b] + \frac{\mathbb{P}[X = a] + \mathbb{P}[X = b]}{2}.$$

*CF* jednoznačně určuje rozdělení.

**Důsledek 3.26 (Jednoznačná charakterizace rozdělení)**

$$\varphi_X = \varphi_Y \iff X \overset{d}{=} Y$$

# Charakteristická funkce pro náhodný vektor

**Definice 3.27 (CF pro náhodný vektor)**

*Charakteristická funkce* náhodného vektoru $\mathbb{X} = [X_1, \ldots, X_d]^\intercal$ je definována jako

$$\varphi_{\mathbb{X}}(\mathbb{t}) = \mathbb{E}[\exp\{i\mathbb{t}^\intercal \mathbb{X}\}] = \int_{\mathbb{R}^d} \exp\{i\mathbb{t}^\intercal \mathbb{x}\} \mathrm{d}\mathbb{P}_{\mathbb{X}}(\mathbb{x}), \quad \mathbb{t} \in \mathbb{R}^d.$$

- Poznamenejte, že vlastnosti jednorozměrné CF lze rozšířit na vícerozměrný případ.

# Agenda

**4. Stochastické nerovnosti**

4.1 Markovovy nerovnosti

4.2 Nerovnosti pro střední hodnotu

# Markovovy nerovnosti I

**Věta 4.1 (Markovova nerovnost)**

*Nechť $X$ je nezáporná náhodná veličina a předpokládejme, že $\mathbb{E}[X]$ existuje. Pro každé $\varepsilon > 0$,*

$$\mathbb{P}[X \geq \varepsilon] \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

**Důsledek 4.2 (Zobecněná Markovova nerovnost)**

*Nechť $X$ je nezáporná náhodná veličina a předpokládejme, že $\mathbb{E}[X^r]$ existuje pro nějaké $r > 0$. Pro každé $\varepsilon > 0$,*

$$\mathbb{P}[X \geq \varepsilon] \leq \frac{\mathbb{E}[X^r]}{\varepsilon^r}.$$

# Markovovy nerovnosti II

**Věta 4.3 (Čebyševova nerovnost)**

*Nechť $X$ je náhodná veličina a předpokládejme, že $\mathbb{E}[X]$ je existuje. Pro každé $\varepsilon > 0$,*

$$\mathbb{P}[|X - \mathbb{E}X| \geq \varepsilon] \leq \frac{\mathbb{V}\mathrm{ar}[X]}{\varepsilon^2}.$$

# Nerovnosti pro střední hodnotu

**Věta 4.4 (Cauchy–Schwarzova nerovnost)**

*Pokud mají $X$ a $Y$ konečné rozptyly, pak*

$$|\mathbb{E}XY| \leq \sqrt{\mathbb{E}X^2\mathbb{E}Y^2}$$

*a*

$$|\mathbb{C}\mathrm{ov}(X, Y)| \leq \sqrt{\mathbb{V}\mathrm{ar}\,X\,\mathbb{V}\mathrm{ar}\,Y}.$$

# Nerovnosti pro střední hodnotu (pokračování)

**Věta 4.5 (Jensenova nerovnost)**

*Pokud je $g$ konvexní, pak*

$$\mathbb{E}g(X) \geq g(\mathbb{E}X).$$

*Pokud je $g$ konkávní, pak*

$$\mathbb{E}g(X) \leq g(\mathbb{E}X).$$

# Agenda


## 5. Stochastické konvergence

5.1 Typy stochastické konvergence
5.2 Vztahy mezi typy konvergence
5.3 Věta o spojitém zobrazení
5.4 Slutského věta
5.5 Lévyho věta o spojitosti
5.6 Slabý zákon velkých čísel
5.7 Centrální limitní věta
5.8 Delta metoda

# Typy stochastické konvergence I

**Definice 5.1 (Konvergence v pravděpodobnosti a v distribuci)**

Nechť $X_1, X_2, \ldots$ je posloupnost náhodných veličin a nechť $X$ je jiná náhodná veličina. Nechť $F_n$ označuje distribuční funkci $X_n$ a nechť $F$ označuje distribuční funkci $X$.

(i) $X_n$ *konverguje k* $X$ *v pravděpodobnosti*, značíme $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$, pokud pro každé $\varepsilon > 0$,

$$\mathbb{P}[|X_n - X| > \varepsilon] \to 0 \text{ když } n \to \infty.$$

(ii) $X_n$ *konverguje k* $X$ *v distribuci*, značíme $X_n \xrightarrow[n\to\infty]{\mathbb{D}} X$, pokud

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ pro všechna } x \text{ kde je } F \text{ spojitá.}$$

# Typy stochastické konvergence II

**Definice 5.2 (Konvergence v Lebesgueových prostorech a skoro jistě)**

Nechť $X_1, X_2, \ldots$ je posloupnost náhodných veličin a nechť $X$ je jiná náhodná veličina.

(i) $X_n$ *konverguje k* $X$ *v* $\mathbb{L}_p$ pro $p \geq 1$, značíme $X_n \xrightarrow[n\to\infty]{\mathbb{L}_p} X$, pokud

$$\mathbb{E}|X_n - X|^p \to 0 \text{ když } n \to \infty.$$

(ii) $X_n$ *konverguje k* $X$ $\mathbb{P}$*-skoro jistě*, značíme $X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}s.j.} X$, pokud

$$\mathbb{P}[\lim_{n\to\infty} X_n = X] \equiv \mathbb{P}[\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)] = 1.$$

Konvergence v $L_1$ $\equiv$ konvergence v *střední hodnotě*
Konvergence v $L_2$ $\equiv$ konvergence podle *kvadratického středu*

# Vztahy mezi konvergencemi

**Věta 5.3 (Implikace mezi typy konvergence)**

(a) $X_n \xrightarrow{\mathbb{P}\text{-s.j.}} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$

(b) $p \geq 1$: $X_n \xrightarrow{\mathbb{L}_p} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$

(c) $p \geq q \geq 1$: $X_n \xrightarrow{\mathbb{L}_p} X \Rightarrow X_n \xrightarrow{\mathbb{L}_q} X$

(d) $X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathbb{D}} X$

(e) *Pokud* $X_n \xrightarrow{\mathbb{D}} X$ *a* $\mathbb{P}[X = c] = 1$ *pro nějaké* $c \in \mathbb{R}$, *pak* $X_n \xrightarrow{\mathbb{P}} X$.

# Obrácené implikace neplatí I

**Příklad 5.4 (Konvergence v pravděpodobnosti neimplikuje konvergenci skoro jistě)**

$\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbb{P} = \lambda$. Každé kladné celé číslo můžeme jednoznačně zapsat jako $2^n + m$, $m = 0, 1, \ldots, 2^n - 1$ a definovat

$$X_{2^n+m}(\omega) = \mathbb{1}\{\omega \in (m2^{-n}, (m+1)2^{-n}]\}, \quad \omega \in [0,1].$$

Například, protože $33 = 2^5 + 1$, dostaneme $X_{33}(\cdot) = \mathbb{1}\{\cdot \in (2^{-5}, 2^{-4}]\}$. Pak pro každé $\varepsilon \in (0,1)$ dostaneme $\mathbb{P}[|X_{2^n+m}| > \varepsilon] = 2^{-n} \to 0$ když $n \to \infty$. Tedy, $X_n \xrightarrow{\mathbb{P}} 0$.

Nicméně, pro každé $\omega \in (0,1]$, $X_j(\omega) = 1$ a $X_j(\omega) = 0$ pro *nekonečně mnoho* $j$ a tedy posloupnost *nekonverguje skoro jistě*, tj. $X_n \overset{\mathbb{P}\text{-}s.j.}{\not\longrightarrow}$.

# Obrácené implikace neplatí II

**Příklad 5.5 (Konvergence v pravděpodobnosti neimplikuje $\mathbb{L}_p$ konvergenci)**

$\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbb{P} = \lambda$. Můžeme definovat

$$X_{2^n+m}(\omega) = 2^n \mathbb{1}\{\omega \in ((m-1)2^{-n}, m2^{-n}]\}, \quad \omega \in [0,1].$$

Pak opět, pro každé $\varepsilon \in (0,1)$ dostaneme $\mathbb{P}[|X_{2^n+m}| > \varepsilon] = 2^{-n} \to 0$ když $n \to \infty$. Tedy, $X_n \xrightarrow{\mathbb{P}} 0$.

Nicméně, $\mathbb{E}|X_{2^n+m} - 0| = 2^n \mathbb{P}[X_{2^n+m} = 2^n] = 2^n 2^{-n} = 1$ a tedy posloupnost *nekonverguje v* $\mathbb{L}_1$, tj. $X_n \overset{\mathbb{L}_1}{\not\to}$. Tedy, $X_n \overset{\mathbb{L}_p}{\not\to}$, $p \geq 1$.

# Obrácené implikace neplatí III

**Příklad 5.6 ($\mathbb{L}_q$ konvergence neimplikuje $\mathbb{L}_p$ konvergenci, když $p > q \geq 1$)**

$\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbb{P} = \lambda$. Můžeme definovat

$$X_{2^n+m}(\omega) = 2^{\lfloor n/2 \rfloor} \mathbb{1}\{\omega \in ((m-1)2^{-n}, m2^{-n}]\}, \quad \omega \in [0,1].$$

Pak, $\mathbb{E}|X_{2^n+m} - 0| = 2^{\lfloor n/2 \rfloor}\mathbb{P}[X_{2^n+m} = 2^n] = 2^{\lfloor n/2 \rfloor}2^{-n} \to 0$ když $n \to \infty$. Tedy, $X_n \xrightarrow{\mathbb{L}_1} 0$.

Nicméně, $\mathbb{E}|X_{2^n+m} - 0|^2 = 2^n\mathbb{P}[X_{2^n+m} = 2^n] = 2^{2\lfloor n/2 \rfloor}2^{-n} \to 1$ když $n \to \infty$ a tedy posloupnost *nekonverguje v* $\mathbb{L}_2$, tj. $X_n \overset{\mathbb{L}_2}{\not\longrightarrow}$.

# Obrácené implikace neplatí IV

**Příklad 5.7 (Konvergence v distribuci neimplikuje konvergenci v pravděpodobnosti)**

$X \sim \mathrm{N}(0,1)$ a $X_n := -X$, $n \in \mathbb{N}$. Tedy, $X_n \sim \mathrm{N}(0,1)$ pro každé $n \in \mathbb{N}$. Takže, triviálně, $\lim_{n\to\infty} F_n(x) = F(x)$ pro všechna $x \in \mathbb{R}$. Tedy, $X_n \xrightarrow{\mathbb{D}} X$.

Nicméně, $\mathbb{P}[|X_n - X| > \varepsilon] = \mathbb{P}[|2X| > \varepsilon] = \mathbb{P}[|X| > \varepsilon/2] \neq 0$ (což nezávisí na $n$) a tedy posloupnost *nekonverguje v pravděpodobnosti*, tj. $X_n \overset{\mathbb{P}}{\nrightarrow}$.

# Věta o spojitém zobrazení

**Věta 5.8 (Věta o spojitém zobrazení (CMT))**

*Nechť $\mathbb{X}, \mathbb{X}_1, \mathbb{X}_2, \ldots$ jsou $d$-rozměrné náhodné vektory a $g : \mathbb{R}^d \to \mathbb{R}^m$ je spojitá v každém bodě množiny $C$ takové, že $P[\mathbb{X} \in C] = 1$.*

- $\mathbb{X}_n \xrightarrow{\mathbb{P}\text{-s.j.}} \mathbb{X} \Rightarrow g(\mathbb{X}_n) \xrightarrow{\mathbb{P}\text{-s.j.}} g(\mathbb{X})$
- $\mathbb{X}_n \xrightarrow{\mathbb{P}} \mathbb{X} \Rightarrow g(\mathbb{X}_n) \xrightarrow{\mathbb{P}} g(\mathbb{X})$
- $\mathbb{X}_n \xrightarrow{\mathbb{D}} \mathbb{X} \Rightarrow g(\mathbb{X}_n) \xrightarrow{\mathbb{D}} g(\mathbb{X})$

! Obecně: $\mathbb{X}_n \xrightarrow{\mathbb{L}_p} \mathbb{X} \not\Rightarrow g(\mathbb{X}_n) \xrightarrow{\mathbb{L}_p} g(\mathbb{X})$

# Cramér–Slutského věta

## Věta 5.9 (Slutského věta)

*Pokud* $X_n \xrightarrow{\mathbb{D}} X$ *a* $Y_n \xrightarrow{\mathbb{P}} c \in \mathbb{R}$, *pak*

- $X_n + Y_n \xrightarrow{\mathbb{D}} X + c$;
- $X_n Y_n \xrightarrow{\mathbb{D}} cX$.

## Důsledek 5.10

*Pokud* $X_n \xrightarrow{\mathbb{P}} a \in \mathbb{R}$ *a* $Y_n \xrightarrow{\mathbb{P}} b \in \mathbb{R}$, *pak*

- $X_n + Y_n \xrightarrow{\mathbb{P}} a + b$;
- $X_n Y_n \xrightarrow{\mathbb{P}} ab$.

# Konvergence v distribuci a bodová konvergence charakteristických funkcí

**Věta 5.11 (Lévyho věta o spojitosti)**

$\mathbb{X}_n \xrightarrow{\mathbb{D}} \mathbb{X} \Leftrightarrow \varphi_{\mathbb{X}_n}(\mathbb{t}) \to \varphi_{\mathbb{X}}(\mathbb{t}),\ \forall \mathbb{t} \in \mathbb{R}^d$

**Příklad 5.12 (CF normálního rozdělení)**

$X \sim \mathrm{N}(\mu, \sigma^2) \Rightarrow \varphi_X(t) = \exp\{i\mu t - \sigma^2 t^2/2\},\ t \in \mathbb{R}$

**Definice 5.13 (Posloupnost nezávislých náhodných veličin)**

$\{X_n\}_{n\in\mathbb{N}}$ je *posloupnost nezávislých* náhodných veličin, pokud

$$F_{\{X_j\}_{j\in J}}(\{x_j\}_{j\in J}) = \prod_{j\in J} F_{X_j}(x_j) \quad \forall \{x_j\}_{j\in J} \in \mathbb{R}^{|J|},\ \forall J \subseteq \mathbb{N},\ |J| < \infty.$$

# Slabý zákon velkých čísel

**Definice 5.14 (IID posloupnost; IID = Nezávislé a identicky rozdělené)**

$\{X_n\}_{n\in\mathbb{N}}$ je *posloupnost IID* náhodných veličin, pokud je to posloupnost nezávislých náhodných veličin majících stejnou distribuční funkci.

Výše uvedené lze definovat i pro *náhodné vektory*.

**Věta 5.15 (Slabý zákon velkých čísel (WLLN))**

*Pokud* $\{X_n\}_{n\in\mathbb{N}}$ *je IID posloupnost náhodných veličin s* $\mathbb{E}|X_1| < \infty$, *pak* $\overline{X}_n := n^{-1}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}X_1$ *když* $n \to \infty$.

- Silný zákon velkých čísel (SLLN): nahrazení $\xrightarrow{\mathbb{P}}$ za $\xrightarrow{\mathbb{P}\text{-}s.j.}$
- Konečný rozptyl *není* vyžadován. Ačkoli, základní důkaz by byl jednodušší, viz. Čebyševova nerovnost.
- Nezávislost lze uvolnit.
- Identické rozdělení lze uvolnit.

# Centrální limitní věta pro IID

**Věta 5.16 (Centrální limitní věta (CLT))**

*Pokud* $\{X_n\}_{n\in\mathbb{N}}$ *je IID posloupnost náhodných veličin s* $\mathbb{E}X_1^2 < \infty$ *a* $\mathbb{V}\mathrm{ar}X_1 > 0$, *pak*

$$Z_n := \sqrt{n}\frac{\overline{X}_n - \mathbb{E}X_1}{\sqrt{\mathbb{V}\mathrm{ar}X_1}} \xrightarrow{\mathbb{D}} Z \sim \mathrm{N}(0,1), \quad n \to \infty.$$

*Jinými slovy,*

$$\lim_{n\to\infty}\mathbb{P}[Z_n \le x] = \Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\exp\{-t^2/2\}dt, \quad \forall x \in \mathbb{R}.$$

*Zkráceně,*

$$Z_n \xrightarrow[n\to\infty]{\mathbb{D}} \mathrm{N}(0,1).$$

# Mnohorozměrná centrální limitní věta

**Důsledek 5.17 (Cramér–Woldova věta)**

$\mathbb{X}_n \xrightarrow{\mathbb{D}} \mathbb{X} \Leftrightarrow \mathbb{t}^\intercal \mathbb{X}_n \xrightarrow{\mathbb{D}} \mathbb{t}^\intercal \mathbb{X}$, $\forall \mathbb{t} \in \mathbb{R}^d$

Cramér–Woldova věta je triviálním důsledkem Lévyho věty o spojitosti.

**Věta 5.18 (Mnohorozměrná CLT)**

*Pokud $\{\mathbb{X}_n\}_{n\in\mathbb{N}}$ je IID posloupnost $d$-rozměrných náhodných vektorů s pozitivně definitní varianční-kovarianční maticí $\mathbb{V}\mathrm{ar}\mathbb{X}_1$, pak*

$$\sqrt{n}\left(\overline{\mathbb{X}}_n - \mathbb{E}\mathbb{X}_1\right) \xrightarrow{\mathbb{D}} \mathrm{N}_d(\mathbb{0}, \mathbb{V}\mathrm{ar}\mathbb{X}_1), \quad n \to \infty.$$

# Delta metoda

**Věta 5.19 (Delta metoda)**

*Pokud* $\sqrt{n}\big(Y_n - \mu\big) \xrightarrow{\mathbb{D}} \mathrm{N}(0, \sigma^2)$ *a* $g$ *je spojitě diferencovatelná v okolí* $\mu$ *tak, že* $g'(\mu) \neq 0$, *pak*

$$\sqrt{n}\big(g(Y_n) - g(\mu)\big) \xrightarrow{\mathbb{D}} \mathrm{N}\Big(0, \big(g'(\mu)\big)^2\sigma^2\Big), \quad n \to \infty.$$

# Agenda


## 6. Statistické učení

6.1 Náhodný výběr
6.2 Statistický experiment
6.3 Stochastické modely
6.4 Parametrické modely
6.5 Neparametrické modely
6.6 Základní pojmy v inferenci
6.7 Odhad
6.8 Standardní chyba
6.9 Střední kvadratická chyba
6.10 Konfidenční množiny

# Statistické učení

- *Statistické učení* alias Statistická inference

= Proces používání *dat k odvození rozdělení, které data generovala*

? Máme-li náhodný vzorek (výběr) $X_1, \ldots, X_n \overset{IID}{\sim} F$, jak odvodíme $F$?

# Náhodný výběr

**Definice 6.1 (Náhodný výběr a rozsah výběru)**

Pokud jsou $X_1, \ldots, X_n$ *nezávislé* a každá má *stejné marginální rozdělení* s distribuční funkcí $F$, říkáme, že $X_1, \ldots, X_n$ jsou IID (nezávislé a stejně rozdělené) a píšeme

$$X_1, \ldots, X_n \overset{IID}{\sim} F.$$

Také $X_1, \ldots, X_n$ nazýváme *náhodný výběr* rozsahu $n$ z $F$.

# Data z experimentu

- Uvažujeme / myslíme si / předpokládáme:
  - *měřitelné zobrazení* $X_i : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, \ldots, n$

- Pozorujeme / měříme / získáváme:
  - *reálná data* $X_i(\omega) \in \mathbb{R}$, $i = 1, \ldots, n$ pro jedno konkrétní $\omega \in \Omega$

# Stochastický model

- *Stochastický model* $\approx$ pravděpodobnostní model, statistický model, ...

- Parametrický model
  - Množina $\mathcal{F}$, kterou lze parametrizovat konečným počtem parametrů

- Neparametrický model
  - Množina $\mathcal{F}$, kterou nelze parametrizovat *konečným* počtem parametrů

# Parametrické modely

**Příklad 6.2 (Normální model)**

$$\mathcal{F} = \left\{ f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \ \mu \in \mathbb{R}, \sigma^2 > 0 \right\}.$$

- Data pocházejí z normálního rozdělení se *dvěma* parametry $\mu$ a $\sigma^2$
- Obecně,

$$\mathcal{F} = \left\{ f(\cdot; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d \right\}$$

- Některé notace: $\mathbb{P}_\theta[X \in A] = \int_A f(x; \theta) \mathrm{d}x$, $E_\theta[g(X)] = \int_{\mathbb{R}} g(x) f(x; \theta) \mathrm{d}x$

# Neparametrické modely

**Příklad 6.3 (Model Sobolevova prostoru)**

$$\mathcal{F} = \left\{ f : \int_{\mathbb{R}} \{f''(x)\}^2 \mathrm{d}x < \infty \right\}.$$

- Data pocházejí z rozdělení s *hustotou*, která není příliš "vlnitá"
- Rozdíl mezi parametrickými a neparametrickými modely je subtilnější, ale pro naše účely nepotřebujeme rigorózní definici
- Například, celá funkce hustoty může být považována za *nekonečně dimenzionální* parametr
- Semi-parametrické modely

# Bodový odhad

**Definice 6.4 (Odhad)**

Bodový odhad $\widehat{\theta}_n$ parametru $\theta$ je měřitelná funkce $t$ náhodných veličin $X_1, \ldots, X_n$:

$$\widehat{\theta}_n = t(X_1, \ldots, X_n).$$

- Zde obecně není požadováno, aby $X_1, \ldots, X_n$ byly nezávislé a stejně rozdělené
- Důležité:
  - *Parametr* $\theta$ je pevné reálné číslo (vektor), ale neznámé
  - *Odhad* $\widehat{\theta}_n$ je náhodná veličina (měřitelná funkce náhodných veličin), ale známá (tj., získatelná z dat)

# Konzistentní odhad

- $\widehat{\theta}_n$ je nestranný, pokud $\mathbb{E}\left[\widehat{\theta}_n\right] = \theta$ pro všechna $n \in \mathbb{N}$
- Nestrannost byla dříve velmi důležitá, ale dnes je méně relevantní
- Mnoho odhadů, které budeme používat, není nestranných
- Vychýlení odhadu je definována jako $\text{bias}(\widehat{\theta}_n) = \mathbb{E}\left[\widehat{\theta}_n\right] - \theta$
- Rozumným požadavkem na odhad je, aby "konvergoval" k pravé hodnotě parametru *s rostoucím množstvím dat*

**Definice 6.5 (Konzistentní odhad)**

Bodový odhad $\widehat{\theta}_n$ parametru $\theta$ je *konzistentní*, pokud $\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ pro $n \to \infty$.

- *(slabá) konzistence* $\Leftrightarrow$ v pravděpodobnosti $\leftrightarrow$ *silná konzistence* $\Leftrightarrow$ skoro jistě
- *Kvalitativní* vlastnost

# Standardní chyba

- Rozdělení $\widehat{\theta}_n$ se nazývá *výběrové rozdělení*
- *Standardní odchylka* $\widehat{\theta}_n$ se nazývá standardní chyba: $\mathrm{se}(\widehat{\theta}_n) = \sqrt{\mathbb{V}\mathrm{ar}\widehat{\theta}_n}$
- Často standardní chyba závisí na *neznámém* $F$
- V těchto případech je se neznámá veličina (*parametr*), ale obvykle ji můžeme odhadnout
- *Odhadnutá* standardní chyba je označena $\widehat{\mathrm{se}}$ (nebo odhad standardní chyby)

**Příklad 6.6 (Standardní chyba v alternativním modelu – Hod mincí)**

Bernoulliho náhodný výběr $X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{Be}(p)$ a parametr $p \in (0, 1)$
... $\widehat{p}_n = n^{-1} \sum_{i=1}^n X_i$ relativní četnost jako odhad ... $\widehat{\mathrm{se}}(\widehat{p}_n) = \sqrt{\widehat{p}_n(1 - \widehat{p}_n)/n}$.

# Střední kvadratická chyba

- Kvalita bodového odhadu se někdy posuzuje pomocí střední kvadratické chyby

$$\mathrm{MSE}(\widehat{\theta}_n) = \mathbb{E}_\theta\left[\widehat{\theta}_n - \theta\right]^2$$

- Mějte na paměti, že $\mathbb{E}_\theta$ se v případě nezávislých a stejně rozdělených $X$'ů vztahuje k očekávané hodnotě vzhledem k rozdělení

$$f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

  které generovalo data
- *Kvantitativní* vlastnost

**Věta 6.7 (Rozklad střední kvadratické chyby na rozptyl a vychýlení)**

$$\mathrm{MSE}(\widehat{\theta}_n) = \mathrm{bias}^2(\widehat{\theta}_n) + \mathbb{V}\mathrm{ar}(\widehat{\theta}_n)$$

# Slabá konzistence

**Věta 6.8 (Postačující podmínka pro konzistenci)**

$\text{bias}(\widehat{\theta}_n) \to 0$ & $\mathbb{V}\text{ar}(\widehat{\theta}_n) \to 0 \Rightarrow \widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ *(konzistentní estimátor)*

**Příklad 6.9 (Konzistence – Hod mincí)**

Příklad 6.6: $\mathbb{E}(\widehat{p}_n) = p$ & $\mathbb{V}\text{ar}(\widehat{p}_n) = p(1-p)/n \to 0 \Rightarrow \widehat{p}_n \xrightarrow{\mathbb{P}} p$

**Definice 6.10 (Asymptoticky standardně normální odhad)**

Odhad $\widehat{\theta}_n$ parametru $\theta$ je *asymptoticky standardně normální*, pokud

$$\frac{\widehat{\theta}_n - \theta}{\text{se}(\widehat{\theta}_n)} \xrightarrow{\mathbb{D}} \text{N}(0,1), \quad n \to \infty.$$

# Konfidenční množiny

**Definice 6.11 (Konfidenční interval – Interval spolehlivosti)**

$(1-\alpha)$-konfidenční interval pro parametr $\theta$ je interval $C_n = (a, b)$, kde $a = a(X_1, \ldots, X_n)$ a $b = b(X_1, \ldots, X_n)$ jsou měřitelné funkce dat takové, že

$$\mathbb{P}_\theta[\theta \in C_n] = 1 - \alpha \quad \text{pro všechna } \theta \in \Theta.$$

Asymptotický (přibližný) $(1-\alpha)$-konfidenční interval pro parametr $\theta$ je interval $C_n$ takový, že

$$\lim_{n\to\infty} \mathbb{P}_\theta[\theta \in C_n] = 1 - \alpha \quad \text{pro všechna } \theta \in \Theta.$$

- Slovně, $(a, b)$ zachytí $\theta$ s pravděpodobností (přibližně) $1-\alpha$
- $1-\alpha$ nazýváme *pokrytí* konfidenčního intervalu (CI)
- Pozor: $C_n$ je *náhodný* a $\theta$ je *pevné*
- $\theta \in \mathbb{R}^d$, $d > 1$: *konfidenční množina* (jako koule/elipsoid) místo intervalu
- CI není pravděpodobnostní výrok o $\theta$, protože $\theta$ je pevné, nikoliv náhodná veličina

# CI Normální model

## Příklad 6.12 (CI – Normalita)

Příklad 6.2: $X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{N}(\mu, \sigma^2)$ a parametr $\mu \in \mathbb{R}$ je *neznámý* a má být odhadnut (bodový odhad a konfidenční interval), $\sigma^2 > 0$ je předpokládáno *známé*

- Bodový odhad $\mu$: $\widehat{\mu}_n = n^{-1} \sum_{i=1}^n X_i \equiv \overline{X}_n$
- Linearita normálních rozdělení: $\sqrt{n}(\widehat{\mu}_n - \mu)/\sigma \sim \mathrm{N}(0,1)$, $\forall n \in \mathbb{N}$
- Kvantily standardního normálního rozdělení:

$$\mathbb{P}\left[-u_{1-\alpha/2} \leq \sqrt{n}\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2}} \leq +u_{1-\alpha/2}\right] = 1 - \alpha, \quad \forall n \in \mathbb{N} \text{ a } \forall \mu \in \mathbb{R}$$

- Konfidenční interval pro $\mu$: $\left(\overline{X}_n - u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \overline{X}_n + u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$

# Délka konfidenčního intervalu

*Délka* CI v normálním modelu pro $\mu$ se známým (daným) $\sigma^2 > 0$:

$$\text{CI: } \overline{X}_n \pm u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}} \quad \Rightarrow \quad \text{Délka: } 2u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}$$

- Vyšší pokrytí $1-\alpha \Rightarrow$ širší CI
- Větší rozptyl $\sigma^2 \Rightarrow$ širší CI
- Menší rozsah výběru $n \Rightarrow$ širší CI

*Kolik* pozorování potřebuji pro CI užší než $d$? ... $n \geq \left\lfloor \frac{4u_{1-\alpha/2}^2\sigma^2}{d^2} \right\rfloor + 1$

# Konfidenční interval založený na normalitě

**Věta 6.13 (CI založený na normalitě)**

*Předpokládejme, že $\widehat{\theta}_n$ je asymptoticky standardně normální odhad parametru $\theta$ a $\widehat{\mathrm{se}}(\widehat{\theta}_n)$ je konzistentní odhad $\mathrm{se}(\widehat{\theta}_n)$, tj. $\widehat{\mathrm{se}}(\widehat{\theta}_n) - \mathrm{se}(\widehat{\theta}_n) \xrightarrow{\mathbb{P}} 0$. Nechť $u_{1-\alpha/2}$ je $(1-\alpha/2)$-kvantil standardního normálního rozdělení a*

$$C_n = \left(\widehat{\theta}_n - u_{1-\alpha/2}\widehat{\mathrm{se}}(\widehat{\theta}_n), \widehat{\theta}_n + u_{1-\alpha/2}\widehat{\mathrm{se}}(\widehat{\theta}_n)\right).$$

*Pak,*

$$\mathbb{P}_\theta\left[\theta \in C_n\right] \to 1-\alpha, \quad n \to \infty.$$

- Neformálně: $\widehat{\theta}_n \approx \mathrm{N}\big(\theta, \widehat{\mathrm{se}}(\widehat{\theta}_n)\big)$
- Přibližně: Pro 95%-konfidenční intervaly je $\alpha = 0.05$ a $u_{.975} \doteq 1.96 \approx 2$ vedoucí k přibližnému 95%-konfidenčnímu intervalu $\widehat{\theta}_n \pm 2\widehat{\mathrm{se}}(\widehat{\theta}_n)$

# CLT konfidenční interval

## Příklad 6.14 (CI – Hod mincí)

Příklad 6.9:

- $\widehat{p}_n \xrightarrow{\mathbb{P}} p$ & $\mathrm{se}(\widehat{p}_n) = \sqrt{p(1-p)/n}$ & $\widehat{\mathrm{se}}(\widehat{p}_n) := \sqrt{\widehat{p}_n(1-\widehat{p}_n)/n}$ & Slutského věta $\Rightarrow$ $\widehat{\mathrm{se}}(\widehat{p}_n) - \mathrm{se}(\widehat{p}_n) \xrightarrow{\mathbb{P}} 0$
- CLT $\Rightarrow \frac{\widehat{p}_n - p}{\mathrm{se}(\widehat{p}_n)} \xrightarrow{\mathbb{D}} \mathrm{N}(0,1)$

Pak podle Slutského věty ještě jednou $\Rightarrow \frac{\widehat{p}_n - p}{\widehat{\mathrm{se}}(\widehat{p}_n)} \xrightarrow{\mathbb{D}} \mathrm{N}(0,1)$. Tedy,

$$\widehat{p}_n \pm u_{1-\alpha/2}\sqrt{\widehat{p}_n(1-\widehat{p}_n)/n}$$

je asymptotický (přibližný) $(1-\alpha)$-konfidenční interval pro $p$.

# Agenda


## 7. Statistické funkcionály

7.1 Empirická distribuční funkce
7.2 Vlastnosti ECDF
7.3 Statistický funkcionál
7.4 Lineární statistický funkcionál
7.5 Empirická míra
7.6 Plug-in odhad
7.7 Plug-in odhad pro lineární statistický funkcionál

# Empirická distribuční funkce

- $X_1, \ldots, X_n \overset{IID}{\sim} F$ je *náhodný výběr* z $F$ s *rozsahem výběru* $n$
- Odhadnout $F$ její empirickým *protějškem*

**Definice 7.1 (ECDF)**

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}, \quad x \in \mathbb{R}.$$

- ECDF přiřazuje váhu $1/n$ každému pozorování $X_i$
- *Relativní četnost* $X$ menších nebo rovných pevnému $x$, tj. $\#\{X_i \leq x\}/n$

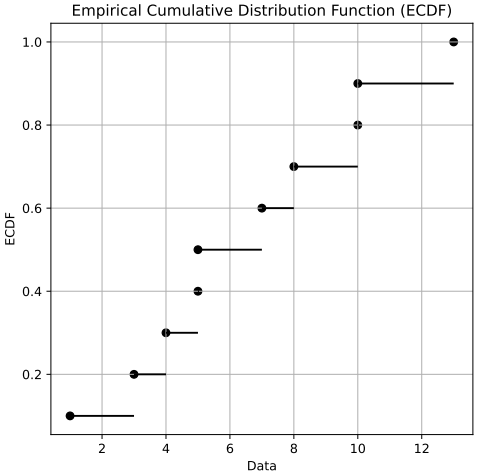
# ECDF – Python

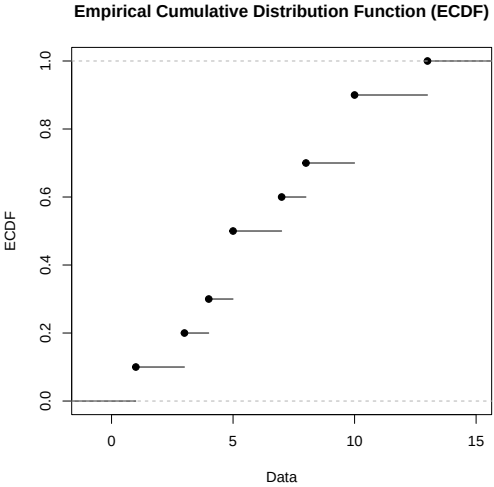
```python
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
# Example data
data = [1, 3, 4, 5, 5, 7, 8, 10, 10, 13]
# Compute ECDF
ecdf = sm.distributions.empirical_distribution.ECDF(data)
# Plot ECDF
plt.figure(figsize=(6, 6))
plt.hlines(ecdf.y[:-1],ecdf.x[:-1],ecdf.x[1:], color='black')
plt.scatter(ecdf.x, ecdf.y, marker='o', color='black')
plt.xlabel('Data')
plt.ylabel('ECDF')
plt.title('Empirical Cumulative Distribution Function (ECDF)')
plt.grid(True)
plt.show()
```

# ECDF – R

```r
data <- c(1, 3, 4, 5, 5, 7, 8, 10, 10, 13)
plot(ecdf(data), xlab='Data', ylab='ECDF', main='Empirical Cumulative
    Distribution Function (ECDF)')
```

# ECDF – Python & R



# Vlastnosti ECDF

**Věta 7.2 (Bodové vlastnosti ECDF)**

*Pro libovolné pevné* $x \in \mathbb{R}$,

- $\mathbb{E}\left[\widehat{F}_n(x)\right] = F(x)$;
- $\mathbb{V}\mathrm{ar}\left[\widehat{F}_n(x)\right] = \frac{F(x)\{1-F(x)\}}{n}$;
- $\mathrm{MSE}\left(\widehat{F}_n(x)\right) = \frac{F(x)\{1-F(x)\}}{n} \to 0$ *když* $n \to \infty$;
- $\widehat{F}_n(x) \xrightarrow{\mathbb{P}} F(x)$ *když* $n \to \infty$.

# Statistický funkcionál

Funkcionál je jen funkce funkce.

**Definice 7.3 (Funkcionál)**

Funkcionál je zobrazení $T : \mathcal{F} \to \mathbb{R}$, kde $\mathcal{F}$ je nějaká množina funkcí.

**Definice 7.4 (Statistický funkcionál)**

Statistický funkcionál je zobrazení $T$, které přiřadí rozdělení $\mathbb{P}_X$ reálné číslo.

Vektorový funkcionál lze také definovat (stačí nahradit obor hodnot $\mathbb{R}^d$).

**Příklad 7.5 (Střední hodnota, Rozptyl, Medián)**

- $\mu \equiv \mathbb{E}X = \int x \mathrm{d}\mathbb{P}_X(x)$ ... střední hodnota
- $\sigma^2 \equiv \mathbb{V}\mathrm{ar}X = \int (x-\mu)^2 \mathrm{d}\mathbb{P}_X(x)$ ... rozptyl
- $\mathrm{median}(X) = F^{-1}(1/2) \equiv \inf\{x : \mathbb{P}_X((-\infty, x]) > 1/2\}$ ... medián

# Lineární statistický funkcionál

**Definice 7.6 (Lineární statistický funkcionál)**

Pokud $T(\mathbb{P}_X) = \int r(x)\mathrm{d}\mathbb{P}_X(x)$ pro nějakou měřitelnou funkci $r$, pak $T$ se nazývá lineární statistický funkcionál.

Důvod, proč se $T(\mathbb{P}_X) = \int r(x)\mathrm{d}\mathbb{P}_X(x)$ nazývá lineární funkcionál, je ten, že $T$ splňuje

$$T(a\mathbb{P}_X + b\mathbb{P}_Y + c) = aT(\mathbb{P}_X) + bT(\mathbb{P}_Y) + c, \quad a, b, c \in \mathbb{R};$$

pokud obě strany existují. Tedy $T$ je *lineární* ve svých argumentech.

# Empirická pravděpodobnostní míra

**Definice 7.7 (Empirická míra)**

Nechť $X_1, \ldots, X_n$ je *náhodný výběr* z $F$ s *rozsahem výběru* $n$, kde $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ pro $i = 1, \ldots, n$. Pak

$$\widehat{\mathbb{P}}_n(B) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \in B\} \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(B), \quad \forall B \in \mathcal{B}(\mathbb{R})$$

se nazývá empirická (pravděpodobnostní) míra.

- Pozor! Jedná se o náhodnou pravděpodobnostní míru.
- $\widehat{\mathbb{P}}_n$ je *diskrétní rovnoměrné pravděpodobnostní rozdělení* na náhodných bodech $\{X_1, \ldots, X_n\}$.
- $\widehat{\mathbb{P}}_n$ je vážený součet (směs) Diracových měr v náhodných bodech $\{X_1, \ldots, X_n\}$ s váhou $1/n$.

# Plug-in (empirický) odhad

**Definice 7.8 (Plug-in odhad)**

Plug-in odhad neznámého parametru $\theta = T(\mathbb{P}_X)$ je $\widehat{\theta}_n = T(\widehat{\mathbb{P}}_n)$.

Stačí dosadit *empirickou míru* $\widehat{\mathbb{P}}_n$ za *neznámou míru* $\mathbb{P}_X$ ... alias empirický odhad

**Příklad 7.9 (ECDF)**

$\widehat{F}_n(x) = \widehat{\mathbb{P}}_n((-\infty, x]), \forall x \in \mathbb{R}$

# Plug-in odhad pro lineární statistický funkcionál

**Definice 7.10 (Empirický odhad pro lineární statistický funkcionál)**

*Empirický (plug-in) odhad* pro lineární statistický funkcionál $T(\mathbb{P}_X) = \int r(x)\mathrm{d}\mathbb{P}_X(x)$ je

$$T(\widehat{\mathbb{P}}_n) = \int r(x)\mathrm{d}\widehat{\mathbb{P}}_n(x).$$

**Věta 7.11 (Plug-in odhad pro lineární statistický funkcionál – výpočet)**

*Pro empirický odhad lineárního statistického funkcionálu $T(\mathbb{P}_X) = \int r(x)d\mathbb{P}_X(x)$ platí*

$$T(\widehat{\mathbb{P}}_n) = \frac{1}{n}\sum_{i=1}^{n} r(X_i).$$

**Příklad 7.12 (Střední hodnota – empirická střední hodnota – výběrový průměr)**

$\mu = T(\mathbb{P}_X) = \int x\mathrm{d}\mathbb{P}_X(x) \Rightarrow \widehat{\mu}_n = \int x\mathrm{d}\widehat{\mathbb{P}}_n(x) = \overline{X}_n$

# Empirický rozptyl

## Příklad 7.13 (Rozptyl)

$\mathbb{V}\mathrm{ar}X = \sigma^2 = T(\mathbb{P}_X) = \int (x-\mu)^2 \mathrm{d}\mathbb{P}_X(x) = \int x^2 \mathrm{d}\mathbb{P}_X(x) - \left(\int x \mathrm{d}\mathbb{P}_X(x)\right)^2 \Rightarrow$

$$\widehat{\sigma}_n^2 = \int x^2 \mathrm{d}\widehat{\mathbb{P}}_n(x) - \left(\int x \mathrm{d}\widehat{\mathbb{P}}_n(x)\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$$

Další rozumný odhad $\sigma^2$ je *výběrový rozptyl*

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$$

# Empirická korelace

## Příklad 7.14 (Korelace)

Nechť $Z = [X, Y]^\intercal$ a nechť $\rho = T(\mathbb{P}_{[X,Y]^\intercal}) = \mathbb{E}(X - \mu_X)(Y - \mu_Y)/(\sigma_X \sigma_Y)$ označuje korelaci mezi $X$ a $Y$. Můžeme napsat

$$T(\mathbb{P}_{[X,Y]^\intercal}) = a\big(T_1(\mathbb{P}_{[X,Y]^\intercal}), T_2(\mathbb{P}_{[X,Y]^\intercal}), T_3(\mathbb{P}_{[X,Y]^\intercal}), T_4(\mathbb{P}_{[X,Y]^\intercal}), T_5(\mathbb{P}_{[X,Y]^\intercal})\big),$$

$$T_1(\mathbb{P}_{[X,Y]^\intercal}) = \int x \mathrm{d}\mathbb{P}_{[X,Y]^\intercal}(x, y); \qquad T_2(\mathbb{P}_{[X,Y]^\intercal}) = \int y \mathrm{d}\mathbb{P}_{[X,Y]^\intercal}(x, y);$$

$$T_3(\mathbb{P}_{[X,Y]^\intercal}) = \int xy \mathrm{d}\mathbb{P}_{[X,Y]^\intercal}(x, y); \qquad T_4(\mathbb{P}_{[X,Y]^\intercal}) = \int x^2 \mathrm{d}\mathbb{P}_{[X,Y]^\intercal}(x, y);$$

$$T_5(\mathbb{P}_{[X,Y]^\intercal}) = \int y^2 \mathrm{d}\mathbb{P}_{[X,Y]^\intercal}(x, y); \qquad a(t_1, t_2, t_3, t_4, t_5) = \frac{t_3 - t_1 t_2}{\sqrt{(t_4 - t_1^2)(t_5 - t_2^2)}}.$$

# Empirická korelace (pokračování)

**Příklad 7.15 (Korelace (pokračování))**

Nahradíme $\mathbb{P}_{[X,Y]^\intercal}$ pomocí $\widehat{\mathbb{P}}_n$ v $T_1(\mathbb{P}_{[X,Y]^\intercal})$, $T_2(\mathbb{P}_{[X,Y]^\intercal})$, $T_3(\mathbb{P}_{[X,Y]^\intercal})$, $T_4(\mathbb{P}_{[X,Y]^\intercal})$, $T_5(\mathbb{P}_{[X,Y]^\intercal})$ a vezměme

$$\widehat{\rho} = T(\widehat{\mathbb{P}}_n) = a\big(T_1(\widehat{\mathbb{P}}_n), T_2(\widehat{\mathbb{P}}_n), T_3(\widehat{\mathbb{P}}_n), T_4(\widehat{\mathbb{P}}_n), T_5(\widehat{\mathbb{P}}_n)\big),$$

Dostaneme

$$\widehat{\rho} = \frac{\sum_i (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_i (X_i - \overline{X}_n)^2 \sum_i (Y_i - \overline{Y}_n)^2}},$$

což se nazývá výběrová korelace.

# Empirické kvantily

**Příklad 7.16 (Kvantily)**

Pro $p \in (0,1)$ je *$p$-tý kvantil* definován jako

$$T(F) = F^{-1}(p) = \inf\Big\{x : \ F(x) > p\Big\}.$$

Nyní definujeme

$$T(\widehat{F}_n) = \widehat{F}_n^{-1}(p) = \inf\Big\{x : \ \widehat{F}_n(x) > p\Big\}$$

a nazýváme to $p$-tý výběrový kvantil. Tedy, výběrový medián je $\widehat{F}_n^{-1}(1/2)$. Navíc, *mezikvartilové rozpětí (IQR)* $\widetilde{T}(F) = F^{-1}(3/4) - F^{-1}(1/4)$ lze odhadnout pomocí výběrového mezikvartilového rozpětí $\widetilde{T}(\widehat{F}_n) = \widehat{F}_n^{-1}(3/4) - \widehat{F}_n^{-1}(1/4)$.

# Agenda

## 8. Parametrická inference


8.1 Parametrická rodina
8.2 Metoda momentů
8.3 Metoda maximální věrohodnosti
8.4 Asymptotické vlastnosti maximálně věrohodního odhadu
8.5 Invariance
8.6 Skórová funkce a Fisherova informace
8.7 Asymptotická normalita MLE

# Parametrická rodina

- Parametrická rodina modelů (nebo rodina parametrických modelů):

$$\mathcal{F} = \left\{ f(\cdot; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d \right\},$$

  kde $\boldsymbol{\Theta}$ je parametrický prostor a $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ je parametr

? Jak bychom vůbec věděli, že distribuce, která generovala data, patří do nějakého parametrického modelu?

! Rozumná *aproximace* & *diagnostika* (testy dobré shody, ...)

# Parametr zájmu

- Zajímáme se pouze o nějakou funkci $T(\boldsymbol{\theta})$
- Například, pokud $X_i \overset{IID}{\sim} \mathrm{N}(\mu, \sigma^2)$, pak parametr je $\boldsymbol{\theta} = (\mu, \sigma^2)$
- Pokud naším cílem je odhadnout $\mu$, pak $\mu = T(\boldsymbol{\theta})$ je nazýván parametrem zájmu a $\sigma^2$ je nazýván nežádoucím / rušivým parametrem
- Parametr zájmu může být *komplikovanou* funkcí $\boldsymbol{\theta}$

# Parametr zájmu (pokračování)

**Příklad 8.1 (Gamma model)**

Připomeňme si, že $X$ má rozdělení Gamma$(a, p)$, pokud

$$f_X(x; a, p) = \frac{a^p}{\Gamma(p)} x^{p-1} \exp\{-ax\}, \quad x > 0;$$

kde $a, p > 0$ a

$$\Gamma(p) = \int_0^\infty y^{p-1} \exp\{-y\} \mathrm{d}y$$

je Gamma funkce. Parametr je $\boldsymbol{\theta} = (a, p)$. Gamma rozdělení se někdy používá pro modelování délek života lidí, zvířat a elektronických zařízení. Představme si, že chceme odhadnout průměrnou délku života. Pak, $T(a, p) = \mathbb{E}_{\boldsymbol{\theta}} X = p/a$.

# Parametrické odhady

Uvažume *náhodný výběr* $X_1, \ldots, X_n \overset{IID}{\sim} F \in \mathcal{F}$

**Příklad 8.2 (Pouze konečná střední hodnota)**

$\mathcal{F} = \left\{F(\mu) : \mathbb{E}_{F(\mu)} = \mu \ \& \ |\mu| < \infty\right\}$

- $\overline{X}_n$ je *konzistentní* a *nestranný* odhad $\mu$
- $X_1$ je *nestranný*, ale *ne konzistentní* odhad $\mu$

**Příklad 8.3 (Pouze konečný rozptyl)**

$\mathcal{F} = \left\{F(\sigma^2) : \mathbb{V}\mathrm{ar}_{F(\sigma^2)} = \sigma^2 \ \& \ \sigma^2 < \infty\right\}$

- $\widehat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ je *konzistentní*, ale ne *nestranný* odhad $\sigma^2$
- $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ je *konzistentní* a *nestranný* odhad $\sigma^2$

# Parametrický odhad (pokračování)

## Příklad 8.4 (Poissonův model)

$\mathcal{F} = \{\mathrm{Po}(\lambda),\ \lambda > 0\}$ a $\theta = \mathbb{P}[X_i = 0]$

- $\widehat{\theta}_n = n^{-1} \sum_{i=1}^{n} \mathbb{1}\{X_i = 0\}$ je *konzistentní* a *nestranný* odhad $\lambda$
- $\widetilde{\theta}_n = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$ je také *konzistentní* a *nestranný* odhad $\lambda$

## Příklad 8.5 (Poissonův model – Zvláštní případ)

$\mathcal{F} = \{\mathrm{Po}(\lambda),\ \lambda > 0\}$ a $\theta = \exp\{-2\lambda\}$

- Jediný *nestranný* odhad $\theta$ je $(-1)^{X_1}$, nicméně nikdy nedosáhne přípustné hodnoty $\exp\{-2\lambda\}$

# Metoda momentů

- Metoda pro získání parametrických odhadů
- Odhady MoM nejsou optimální, ale často jsou snadno vypočítatelné
- Jsou také užitečné jako *počáteční hodnoty pro jiné metody*, které vyžadují iterativní numerické algoritmy
- Definujme $k$-tý (necentrální) moment

$$\mu'_k \equiv \mu'_k(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} X^k = \int x^k \mathrm{d}F_{\boldsymbol{\theta}}(x)$$

- Definujme $k$-tý výběrový moment

$$\widehat{\mu}'_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

# Odhady MoM

## Definice 8.6

Odhad metodou momentů $\widehat{\boldsymbol{\theta}}_n$ je definován jako hodnota $\boldsymbol{\theta}$, pro kterou platí

$$\mu_1'(\widehat{\boldsymbol{\theta}}_n) = \widehat{\mu}_1', \quad \mu_2'(\widehat{\boldsymbol{\theta}}_n) = \widehat{\mu}_2', \quad \dots \quad \mu_d'(\widehat{\boldsymbol{\theta}}_n) = \widehat{\mu}_d'.$$

Alternativně mohou být použity *centrované $k$-té momenty* spolu s jejich empirickými protějšky.

## Příklad 8.7 (Odhad metodou momentů pro alternativní distribuci)

$X_1, \dots, X_n \overset{IID}{\sim} \mathrm{Be}(p)$. Pak, $\mu_1' = \mathbb{E}_p X_1 = p$ a $\widehat{\mu}_1' = \overline{X}_n$. Rovnost těchto hodnot nám dává odhad

$$\widehat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

a je to opět stejný *plug-in (empirický) odhad*.

# Odhady MoM (pokračování)

**Příklad 8.8 (Odhad metodou momentů pro normální rozdělení)**

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{N}(\mu, \sigma^2)$. Pak, $\mu_1' = \mathbb{E}_{(\mu,\sigma^2)} X_1 = \mu$ a
$\mu_2' = \mathbb{E}_{(\mu,\sigma^2)} X_1^2 = \mathbb{V}\mathrm{ar}_{(\mu,\sigma^2)} X_1 + (\mathbb{E}_{(\mu,\sigma^2)} X_1)^2 = \sigma^2 + \mu^2$. Musíme vyřešit rovnice

$$\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{a} \quad \widehat{\mu}_n^2 + \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Toto je soustava dvou rovnic se dvěma neznámými. Řešení je

$$\widehat{\mu}_n = \overline{X}_n \quad \text{a} \quad \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

# Odhady MoM (pokračování II)

## Příklad 8.9 (MoM v Gamma modelu)

$X_1, \ldots, X_n \overset{IID}{\sim}$ Gamma$(a, p)$, kde

$$f_X(x; a, p) = \frac{a^p}{\Gamma(p)} x^{p-1} \exp\{-ax\}, \quad x > 0;$$

kde $a, p > 0$ a $\Gamma(p) = \int_0^\infty y^{p-1} \exp\{-y\} \mathrm{d}y$ je Gamma funkce. Pak, odhady MoM

$$\widehat{a}_n = \frac{\overline{X}_n}{\widehat{\sigma}_n^2} \quad \text{a} \quad \widehat{p}_n = \frac{\overline{X}_n^2}{\widehat{\sigma}_n^2}$$

jsou *konzistentní* a *AN*.

# Odhady MoM (pokračování III)

## Příklad 8.10 (MoM v modelu rovnoměrného rozdělení – oba koncové body neznámé)

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{U}(\theta_1, \theta_2)$. Odhady MoM

$$\widehat{\theta}_{1,n} = \overline{X}_n - \sqrt{3\widehat{\sigma}_n^2} \quad \text{a} \quad \widehat{\theta}_{2,n} = \overline{X}_n + \sqrt{3\widehat{\sigma}_n^2}$$

jsou *konzistentní* a *AN*.

## Příklad 8.11 (MoM v Beta modelu)

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{Beta}(\alpha, \beta)$. Odhady MoM

$$\widehat{\alpha}_n = \overline{X}_n \left\{ \frac{\overline{X}_n(1 - \overline{X}_n)}{\widehat{\sigma}_n^2} - 1 \right\} \quad \text{a} \quad \widehat{\beta}_n = (1 - \overline{X}_n) \left\{ \frac{\overline{X}_n(1 - \overline{X}_n)}{\widehat{\sigma}_n^2} - 1 \right\}$$

sou *konzistentní* a *AN*.

# Metoda maximální věrohodnosti

- Nejběžnější metoda pro odhad parametrů v parametrickém modelu je metoda maximální věrohodnosti

**Definice 8.12**

Funkce věrohodnosti je definována jako

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i; \boldsymbol{\theta}).$$

Funkce logaritmické věrohodnosti je definována jako $\ell_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta})$.

- Funkce věrohodnosti je vlastně *sdružená hustota pozorování*, s tím že ji považujeme za funkci parametru $\boldsymbol{\theta}$
- Tedy, $L_n : \boldsymbol{\Theta} \to [0, \infty)$
- Funkce věrohodnosti není hustota: obecně není pravda, že $L_n(\boldsymbol{\theta})$ se integruje na 1 (vzhledem k $\boldsymbol{\theta}$)

# Odhad metodou maximální věrohodnosti

**Definice 8.13 (MLE)**

Odhad metodou maximální věrohodnosti MLE, označovaný jako $\widehat{\boldsymbol{\theta}}_n$, je hodnota parametru $\boldsymbol{\theta}$, která maximalizuje $L_n(\boldsymbol{\theta})$.

- Maximum funkce $L_n(\boldsymbol{\theta})$ nastává *na stejném místě* jako maximum funkce $\ell_n(\boldsymbol{\theta})$

**Příklad 8.14 (Odhad metodou maximální věrohodnosti pro alternativní rozdělení)**

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{Alt}(p)$. Hustoty je $f(x; p) = p^x(1-p)^{1-x}$, $x \in \{0, 1\}$. Potom,

$$L_n(p) = \prod_{i=1}^{n} f(X_i; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i}.$$

Tedy, $\ell_n(p) = (\log p)\sum_{i=1}^n X_i + \{\log(1-p)\}(n - \sum_{i=1}^n X_i)$. Položíme $\partial \ell_n(p)/\partial p \overset{!}{=} 0$.
MLE je $\widehat{p}_n = \overline{X}_n$.

# MLE

**Příklad 8.15 (Odhad metodou maximální věrohodnosti pro normální rozdělení)**

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{N}(\mu, \sigma^2)$. Potom,

$$L_n(\mu, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\right\}.$$

Položíme $\partial \ell_n(\mu, \sigma^2)/\partial(\mu, \sigma^2)^\intercal \overset{!}{=} \mathbb{0}_2$. MLE je $\widehat{\mu}_n = \overline{X}_n$ a $\widehat{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

# MLE (pokračování)

**Příklad 8.16 (Odhad metodou maximální věrohodnosti pro rovnoměrné rozdělení – Komplikovaný případ)**

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{U}(0, \theta)$. Potom,

$$L_n(\theta) = \theta^{-n} \mathbb{1}\{\theta \geq X_{(n)}\},$$

což není diferencovatelná funkce podle $\theta$. Nicméně, může být maximalizována, a tedy MLE je $\widehat{\theta}_n = X_{(n)} \equiv \max_{1 \leq i \leq n}\{X_i\}$.

# MLE (pokračování II)

## Příklad 8.17 (MLE v exponenciálním modelu)

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{Exp}(\lambda)$. MLE je $\widehat{\lambda}_n = 1/\overline{X}_n$.

## Příklad 8.18 (MLE v Gamma modelu)

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{Gamma}(a, p)$. MLE $\widehat{p}_n$ pro parametr $p$ je řešením *nelineární rovnice*

$$\log \widehat{p}_n - \frac{\Gamma'(\widehat{p}_n)}{\Gamma(\widehat{p}_n)} = \frac{\overline{X}_n}{\sqrt[n]{\prod_{i=1}^n X_i}}.$$

MLE $\widehat{a}_n$ pro parametr $a$ je pak $\widehat{a}_n = \widehat{p}_n / \overline{X}_n$.

# Vlastnosti maximálně věrohodních odhadů

MLE je:

- *konzistentní*
- *ekvivariantní* ... $\boldsymbol{\theta} \rightsquigarrow \widehat{\boldsymbol{\theta}}_n \leftrightarrow g(\boldsymbol{\theta}) \rightsquigarrow g(\widehat{\boldsymbol{\theta}}_n)$
- *asymptoticky normální*
- *asymptoticky optimální* nebo *efektivní* ... to znamená, že mezi všemi dobře chovajícími se odhady má MLE nejmenší rozptyl, alespoň pro velké rozsahy výběrů
- přibližně *Bayesovský* odhad

# Identifikovatelnost

**Definice 8.19 (Identifikovatelný model)**

Řekneme, že model $\mathcal{F}$ je identifikovatelný, pokud $f(x; \boldsymbol{\theta}) = f(x; \boldsymbol{\vartheta})$ pro s.v. $x \in \mathbb{R}$ implikuje $\boldsymbol{\theta} = \boldsymbol{\vartheta}$.

- To znamená, že *různé hodnoty parametru* odpovídají *různým rozdělením*
- *Od této chvíle* budeme předpokládat pouze identifikovatelné modely

# Invariance MLE odhadů

**Věta 8.20 (MLE invariance)**

*Nechť $\boldsymbol{\tau} = g(\boldsymbol{\theta})$ je měřitelná funkce $\boldsymbol{\theta}$. Nechť $\widehat{\boldsymbol{\theta}}_n$ je MLE pro $\boldsymbol{\theta}$. Pak $\widehat{\boldsymbol{\tau}}_n := g(\widehat{\boldsymbol{\theta}}_n)$ je MLE pro $\boldsymbol{\tau}$.*

Princip invariance pro maximální věrohodnost tedy říká: $\widehat{g(\boldsymbol{\theta})}_n = g(\widehat{\boldsymbol{\theta}}_n)$

**Příklad 8.21**

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{N}(\theta, 1)$. MLE pro $\theta$ je $\widehat{\theta}_n = \overline{X}_n$. Nechť $\tau = \exp\{\theta\}$, pak MLE pro $\tau$ je $\widehat{\tau}_n = \exp\{\overline{X}_n\}$.

# Skóre a informace

**Definice 8.22 (Skórová funkce a Fisherova informace)**

Skórová funkce je definována jako

$$S(X;\boldsymbol{\theta}) = \frac{\partial \log f(X;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Fisherova informace je definována jako

$$I_n(\boldsymbol{\theta}) = \mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}}\left[\sum_{i=1}^{n} S(X_i;\boldsymbol{\theta})\right] = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}}[S(X_i;\boldsymbol{\theta})].$$

- Pro $n = 1$ někdy píšeme $I(\boldsymbol{\theta})$ místo $I_1(\boldsymbol{\theta})$
- $S(X;\boldsymbol{\theta})$ je $d$-rozměrný náhodný vektor
- $I_n(\boldsymbol{\theta})$ je $(d \times d)$-ropzměrná deterministická matice

# Podmínky regularity

- Nechť náš parametrický model $\mathcal{F} = \left\{f(\cdot;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d\right\}$ je *identifikovatelný*
- Nechť $\boldsymbol{\theta}_0$ je *skutečná neznámá hodnota* parametru $\boldsymbol{\theta}$

**Definice 8.23 (Podmínky regularity)**

Podmínky regularity jsou:

(i) $\boldsymbol{\theta}_0 \in \operatorname{int}\boldsymbol{\Theta}$ a $\mathcal{S}(X)$ nezávisí na $\boldsymbol{\theta}$;

(ii) $f(x;\boldsymbol{\theta})$ je dvakrát spojitě diferencovatelná podle $\boldsymbol{\theta}$ na $\operatorname{int}\boldsymbol{\Theta}$;

(iii) $\|\log f(x;\boldsymbol{\theta})\| \leq d(x)$ pro $\forall\boldsymbol{\theta} \in \operatorname{int}\boldsymbol{\Theta}$, $\underset{a.e.}{\forall}\, x \in \mathcal{S}(X)$ a $\mathbb{E}_{\boldsymbol{\theta}_0}[d(X)] < \infty$;

(iv) $\|\partial f(x;\boldsymbol{\theta})/\partial\boldsymbol{\theta}\| \leq e(x)$ pro $\forall\boldsymbol{\theta} \in \operatorname{int}\boldsymbol{\Theta}$, $\underset{a.e.}{\forall}\, x \in \mathcal{S}(X)$ a $\int e(x)\mathrm{d}\nu(x) < \infty$;

(v) $\|\partial^2\log f(x;\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}\| \leq g(x)$ pro $\forall\boldsymbol{\theta} \in \operatorname{int}\boldsymbol{\Theta}$, $\underset{a.e.}{\forall}\, x \in \mathcal{S}(X)$ a $\mathbb{E}_{\boldsymbol{\theta}_0}[g(X)] < \infty$;

(vi) $\|\partial^2 f(x;\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\theta}\| \leq h(x)$ pro $\forall\boldsymbol{\theta} \in \operatorname{int}\boldsymbol{\Theta}$, $\underset{a.e.}{\forall}\, x \in \mathcal{S}(X)$ a $\int h(x)\mathrm{d}\nu(x) < \infty$;

(vii) $I(\boldsymbol{\theta}_0)$ je pozitivně definitní.

# Score a informace pod podmínkami regularity

- *Od této chvíle* budeme předpokládat, že podmínky regularity platí
- Označme $[\cdot]^{\otimes 2} := [\cdot][\cdot]^\top$

**Lemma 8.24 (Střední hodnota a rozptyl skóre)**

$$\mathbb{E}_{\boldsymbol{\theta}}[S(X;\boldsymbol{\theta})] = \mathbb{0}_d \quad a \quad \mathbb{V}\mathrm{ar}_{\boldsymbol{\theta}}[S(X;\boldsymbol{\theta})] = \mathbb{E}_{\boldsymbol{\theta}}[S^{\otimes 2}(X;\boldsymbol{\theta})].$$

**Lemma 8.25 (Alternativní definice informace)**

$$I_n(\boldsymbol{\theta}) = nI(\boldsymbol{\theta}) \quad a \quad I(\boldsymbol{\theta}) = -\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log f(X;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}}\right].$$

**Důsledek 8.26 (Limitní informace)**

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(X_i;\boldsymbol{\theta}) \xrightarrow[n\to\infty]{\mathbb{D}} \mathrm{N}_d\big(\mathbb{0}_d, I(\boldsymbol{\theta})\big).$$

# Asymptotická normalita MLE

## Věta 8.27 (AN MLE)

*Nechť $\boldsymbol{\theta}_0$ je skutečná hodnota neznámého parametru $\boldsymbol{\theta}$. Nechť platí podmínky regularity. Pak je MLE $\widehat{\boldsymbol{\theta}}_n$ pro $\boldsymbol{\theta}$ asymptoticky normální, tak že*

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\big) \xrightarrow[n\to\infty]{\mathbb{D}} \mathrm{N}_d\big(\mathbb{0}_d, I^{-1}(\boldsymbol{\theta}_0)\big).$$

## Definice 8.28 (Asymptotický rozptyl)

Za podmínek regularity je *asymptotický rozptyl* MLE $\widehat{\boldsymbol{\theta}}_n$ rovný $\operatorname{avar}(\widehat{\boldsymbol{\theta}}_n) = n^{-1} I^{-1}(\boldsymbol{\theta}_0)$.

- Asymptotický rozptyl NENÍ limitní rozptyl
- Neformálně, $\widehat{\boldsymbol{\theta}}_n \overset{as}{\approx} \mathrm{N}_d\big(\boldsymbol{\theta}_0, n^{-1} I^{-1}(\boldsymbol{\theta}_0)\big)$
- Intervaly spolehlivosti založené na normálním rozdělení $(d = 1)$ ... Theorem 6.13

# Příklady MLE

## Příklad 8.29 (MLE ve specifickém případě Beta modelu)

$X_1, \ldots, X_n \overset{IID}{\sim} f(x;\theta) = \theta(1-x)^{\theta-1}\mathbb{1}\{x \in (0,1)\}$ a nechť $\theta_0$ je skutečná hodnota neznámého parametru $\theta > 1$. Pak,

$$\sqrt{n}\big(\widehat{\theta}_n - \theta_0\big) \xrightarrow[n\to\infty]{\mathbb{D}} \mathrm{N}(0, \theta_0^2).$$

## Příklad 8.30 (MLE v modelu normálního rozdělení)

$X_1, \ldots, X_n \overset{IID}{\sim} f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}}\exp\big\{-\frac{(x-\theta)^2}{2\theta}\big\}$, $x \in \mathbb{R}$ a nechť $\theta_0$ je skutečná hodnota neznámého parametru $\theta > 0$. Pak,

$$\sqrt{n}\big(\widehat{\theta}_n - \theta_0\big) \xrightarrow[n\to\infty]{\mathbb{D}} \mathrm{N}\big(0, 2\theta_0^2/(2\theta_0+1)\big).$$

# Příklady MLE (pokračování)

**Příklad 8.31 (MLE v Bernoulliho modelu – Házení mincí)**

Uvažujme příklad 8.14. Dostáváme

$$S(X;p) = \frac{X}{p} - \frac{1-X}{1-p} \quad \text{a} \quad S'(X;p) = -\frac{X}{p^2} - \frac{1-X}{(1-p)^2}.$$

Takže

$$I(p) = \frac{1}{p(1-p)} \quad \text{a} \quad \widehat{\text{se}}(\widehat{p}_n) = \frac{1}{\sqrt{I_n(\widehat{p}_n)}} = \sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}.$$

Přibližný 95% interval spolehlivosti je $\overline{X}_n \pm u_{.975}\widehat{\text{se}}(\widehat{p}_n)$.

# Příklady MLE (pokračování II)

## Příklad 8.32 (MLE v normálním modelu)

Uvažujme příklad 8.15, kde $\sigma^2$ je *známé*. MLE pro $\mu$ je $\widehat{\mu}_n = \overline{X}_n$. Dostáváme

$$S(X;\mu) = \frac{X-\mu}{\sigma^2} \quad \text{a} \quad S'(X;\mu) = -\frac{1}{\sigma^2}.$$

Potom

$$I(\mu) = \frac{1}{\sigma^2} \quad \text{a} \quad \widehat{\text{se}}(\widehat{\mu}_n) = \frac{1}{\sqrt{I_n(\widehat{\mu}_n)}} = \sqrt{\frac{\sigma^2}{n}}.$$

Takže $\widehat{\mu}_n \approx \mathrm{N}(\mu, \sigma^2/n)$ (neformálně; pro velké $n$). Normální aproximace je ve skutečnosti *přesná*, tj. $\widehat{\mu}_n \sim \mathrm{N}(\mu, \sigma^2/n)$.

# Příklady MLE (pokračování III)

## Příklad 8.33 (MLE v Poissonově modelu)

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{Po}(\lambda)$, kde $\lambda > 0$. MLE pro $\lambda$ je $\widehat{\lambda}_n = \overline{X}_n$. Dostáváme

$$S(X;\lambda) = \frac{X}{\lambda} - 1 \quad \text{a} \quad S'(X;\lambda) = -\frac{X}{\lambda^2}.$$

Potom

$$I(\lambda) = \frac{1}{\lambda} \quad \text{a} \quad \widehat{\mathrm{se}}(\widehat{\lambda}_n) = \frac{1}{\sqrt{I_n(\widehat{\lambda}_n)}} = \sqrt{\frac{\overline{X}_n}{n}}.$$

Přibližný 95% interval spolehlivosti je $\overline{X}_n \pm u_{.975}\sqrt{\overline{X}_n/n}$.

# Agenda


## 9. Testování hypotéz

9.1 Nulová hypotéza a alternativní hypotéza
9.2 Statistický test
9.3 Zamítnout nebo ne?
9.4 Hladina a síla testu
9.5 Typy hypotéz a testů
9.6 Waldův test
9.7 Dualita v testování hypotéz a intervalech spolehlivosti
9.8 $p$-hodnota
9.9 Permutační testy
9.10 Test podílem věrohodností

# Testování – Motivační příklad

**Příklad 9.1 (Porovnání dvou predikčních algoritmů)**

Testuje se výkon dvou predikčních algoritmů. Algoritmus 1 testujeme na prvním testovacím souboru a algoritmus 2 na druhém testovacím souboru.

Nulová hypotéza: Pravděpodobnost správné predikce je pro oba algoritmy stejná.

Alternativní hypotéza: Pravděpodobnost správné predikce není stejná.

$$H_0 : \ p_1 = p_2 \quad \text{versus} \quad H_1 : \ p_1 \neq p_2.$$

# Formální rámec testování

- Zůstáváme u *parametrických modelů*
- Rozdělujeme prostor parametrů $\boldsymbol{\Theta}$ na dvě disjunktní množiny $\boldsymbol{\Theta}_0$ a $\boldsymbol{\Theta}_1$, tj. $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$ a $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \emptyset$
- Chceme testovat

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \quad \text{proti} \quad H_1 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_1$$

- $H_0$ nazýváme *nulovou hypotézou* a $H_1$ *alternativní hypotézou*

# Rozhodnutí pomocí testování hypotéz

- Nechť $\mathbb{X}$ je náhodný vektor (např. náhodný výběr) a nechť $\mathcal{X}$ je obor hodnot $\mathbb{X}$
- Hypotézu testujeme tím, že najdeme vhodnou podmnožinu výsledků $R \subset \mathcal{X}$, nazývanou kritický obor
- Rozhodnutí je následující:
  - $\mathbb{X} \in R \Rightarrow$ zamítnout $H_0$
  - $\mathbb{X} \notin R \Rightarrow$ nezamítnout $H_0$ proti $H_1$ (zachovat $H_0$)
- Odmítací oblast $R$ lze obvykle přepsat ve formě

$$R = \{\mathbb{x} : \ T(\mathbb{x}) \in C\},$$

kde $T$ je testová statistika a $C$ je kritický obor

**Definice 9.2 (Statistický test)**

Statistický test $(T, C)$ je dvojice sestávající z testové statistiky $T$ a kritického oboru $C$.

# Výsledky testování hypotéz

- Existují dvě možné chyby, které můžeme udělat:
  - Zamítnutí $H_0$, když je $H_0$ pravdivá, se nazývá chyba typu I
  - Zachování $H_0$, když je $H_1$ pravdivá, se nazývá chyba typu II
  - Zamítnout nebo ne?

| | Nezamítnout nulovou $H_0$ | Zamítnout nulovou $H_0$ proti alternativě $H_1$ |
|---|---|---|
| $H_0$ pravdivá | ✓ | chyba typu I |
| $H_1$ pravdivá | chyba typu II | ✓ |

- Nezamítáme $H_0$, pokud není silný důkaz pro zamítnutí $H_0$

# Hladina – Síla

**Definice 9.3 (Silofunkce a hladina významnosti)**

*Síla testu* s kritickým oborem $R$ je definována jako

$$\beta(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}[\mathbb{X} \in R].$$

*Hladina testu* je definována jako

$$\alpha = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta(\boldsymbol{\theta}).$$

- Alternativně platí $\beta(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}[T(\mathbb{X}) \in C]$, kde $(T, C)$ je statistický test

# Typy hypotéz a testů

- Hypotéza ve tvaru $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, tj. $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 = \{\boldsymbol{\theta}_0\}$, se nazývá jednoduchá hypotéza
- Hypotéza ve tvaru $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, kde $\boldsymbol{\Theta}_0$ obsahuje více než jeden prvek, se nazývá složená hypotéza
- Test ve tvaru $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ se nazývá oboustranný test
- Test ve tvaru $H_0 : \boldsymbol{\theta} \leq \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} > \boldsymbol{\theta}_0$ nebo $H_0 : \boldsymbol{\theta} \geq \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} < \boldsymbol{\theta}_0$ se nazývá jednostranný test
- Nejčastější testy jsou *oboustranné*

# Funkce síly pro parametr střední hodnoty I

## Příklad 9.4 (Síla testu pro střední hodnotu v normálním modelu)

$X_1, \ldots, X_n \overset{IID}{\sim} \mathrm{N}(\mu, \sigma^2)$, kde $\sigma^2 > 0$ je známé. Chceme testovat $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. Tedy $\boldsymbol{\Theta}_0 = (-\infty, 0]$ a $\boldsymbol{\Theta}_1 = (0, +\infty)$. Uvažujme test

$$\text{zamítneme } H_0 \text{ pokud } T(\mathbb{X}) > c,$$

kde $T(\mathbb{X}) = \overline{X}_n$. Kritický obor je

$$R = \{(x_1, \ldots, x_n) : \; T(x_1, \ldots, x_n) > c\}.$$

Protože $\sqrt{n}(\overline{X}_n - \mu)/\sigma \sim \mathrm{N}(0, 1)$, *síla testu* je

$$\beta(\mu) = \mathbb{P}_\mu \left[\overline{X}_n > c\right] = \mathbb{P}_\mu \left[\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} > \sqrt{n}\frac{c - \mu}{\sigma}\right] = 1 - \Phi\left(\sqrt{n}\frac{c - \mu}{\sigma}\right).$$

# Funkce síly pro parametr střední hodnoty II

**Příklad 9.5 (Síla testu pro střední hodnotu v normálním modelu (pokračování))**

Funkce síly $\beta(\mu)$ je rostoucí v $\mu$ (viz obrázek na dalším slajdu). Proto *hladina* testu je

$$\sup_{\mu \leq 0} \beta(\mu) = \beta(0) = 1 - \Phi\left(\sqrt{n}\frac{c}{\sigma}\right).$$

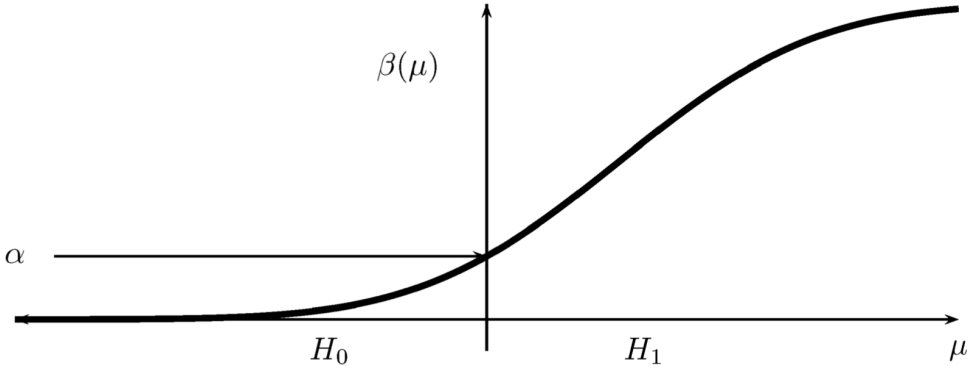
Nastavíme tuto hodnotu rovnu $\alpha$ a řešíme pro $c$, abychom získali

$$c = \frac{\sigma \Phi^{-1}(1-\alpha)}{\sqrt{n}}.$$

Zamítáme, pokud $\overline{X}_n > \sigma\Phi^{-1}(1-\alpha)/\sqrt{n}$. Všimněme si, že $\mu_0 = 0$. Ekvivalentně zamítáme, pokud

$$\sqrt{n}\frac{\overline{X}_n - 0}{\sigma} > u_{1-\alpha} \equiv \Phi^{-1}(1-\alpha).$$

# Graf síly testu pro parametr střední hodnoty




- Hladina testu je největší pravděpodobnost zamítnutí $H_0$, když $H_0$ je pravdivá
- K tomu dochází při $\mu = 0$ a proto hladina je $\beta(0)$
- Kritickou hodnotu $c$ volíme tak, aby $\beta(0) = \alpha$

# Waldův test

- Nechť $\theta$ je *skalární* parametr, $\widehat{\theta}$ je odhad $\theta$ a $\widehat{\mathrm{se}}(\widehat{\theta})$ je odhad směrodatné chyby $\widehat{\theta}$.

**Definice 9.6 (Waldův test)**

Uvažujme *testování* $H_0 : \theta = \theta_0$ proti $H_1 : \theta \neq \theta_0$. Předpokládejme, že $\widehat{\theta}$ je *asymptoticky normální*:

$$W := \frac{\widehat{\theta} - \theta_0}{\widehat{\mathrm{se}}(\widehat{\theta})} \xrightarrow[n\to\infty]{\mathbb{D}} \mathrm{N}(0,1).$$

Waldův test na hladině $\alpha$: *zamítni* $H_0$, pokud $|W| > u_{1-\alpha/2}$.

**Věta 9.7 (Asymptotická hladina Waldova testu)**

$$\mathbb{P}_{\theta_0}[|W| > u_{1-\alpha/2}] \to \alpha, \quad n \to \infty.$$

# Síla Waldova testu

- Alternativní verzí Waldovy testové statistiky je $W = (\widehat{\theta} - \theta_0)/\mathrm{se}_0(\widehat{\theta})$, kde $\mathrm{se}_0(\widehat{\theta})$ je směrodatná chyba počítaná při $\theta = \theta_0$.
- Obě verze testu jsou platné.

**Věta 9.8 (Síla Waldova testu při alternativě)**

*Předpokládejme, že skutečná hodnota $\theta$ je $\theta_1 \neq \theta_0$. Síla $\beta(\theta_1)$ – pravděpodobnost správného zamítnutí nulové hypotézy – je přibližně*

$$1 - \Phi\left(\frac{\theta_0 - \theta_1}{\widehat{\mathrm{se}}(\widehat{\theta})} + u_{1-\alpha/2}\right) + \Phi\left(\frac{\theta_0 - \theta_1}{\widehat{\mathrm{se}}(\widehat{\theta})} - u_{1-\alpha/2}\right).$$

- Připomeňme, že $\widehat{\mathrm{se}}(\widehat{\theta})$ se zmenšuje s rostoucím rozsahem výběru.
- Síla je velká, pokud je $\theta_1$ daleko od $\theta_0$.
- Síla je velká, pokud je rozsah výběru velký.

# Testování dvou pravděpodobností – Porovnání dvou vzorků

## Příklad 9.9 (Porovnání dvou predikčních algoritmů (pokračování))

Pokračujme s motivačním příkladem 9.1. Testujeme predikční Algoritmus 1 na testovací množině velikosti $m$ a Algoritmus 2 na druhé testovací množině velikosti $n$. Nechť $X$ je počet správných predikcí pro Algoritmus 1 a $Y$ počet správných predikcí pro Algoritmus 2. Potom $X \sim \mathrm{Bi}(m, p_1)$ a $Y \sim \mathrm{Bi}(n, p_2)$. Pro testování nulové hypotézy $p_1 = p_2$ píšeme $H_0 : \delta = 0$ proti $H_1 : \delta \neq 0$, kde $\delta = p_1 - p_2$. MLE je $\widehat{\delta} = \widehat{p}_1 - \widehat{p}_2$ s odhadem směrodatné chyby

$$\widehat{\mathrm{se}}(\widehat{\delta}) = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{m} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n}}.$$

Waldův test na hladině $\alpha$ je zamítnutí $H_0$, pokud $|W| > u_{1-\alpha/2}$, kde $W = (\widehat{\delta} - 0)/\widehat{\mathrm{se}}(\widehat{\delta})$. Síla testu je největší, pokud je $p_1$ daleko od $p_2$ a rozsahy výběrů jsou velké.

# Testování dvou pravděpodobností – Párové porovnání

## Příklad 9.10 (Porovnání dvou predikčních algoritmů (upravené))

Pokračujme s příkladem 9.9. Co kdybychom použili *stejnou testovací* množinu pro oba algoritmy? Vzorky již nejsou *nezávislé*. Použijeme následující strategii. Nechť $X_i = 1$, pokud Algoritmus 1 předpoví správně na testovací množině $i$, a $X_i = 0$ jinak. Nechť $Y_i = 1$, pokud Algoritmus 2 předpoví správně na testovací množině $i$, a $Y_i = 0$ jinak. Definujme $D_i = X_i - Y_i$. Nechť

$$\delta = \mathbb{E}D_i = \mathbb{E}X_i - \mathbb{E}Y_i = \mathbb{P}[X_i = 1] - \mathbb{P}[Y_i = 1] = p_1 - p_2.$$

Empirický odhad $\delta$ je $\widehat{\delta} = \overline{D}_n = n^{-1}\sum_{i=1}^n D_i$ a $\widehat{\mathrm{se}}(\widehat{\delta}) = S/\sqrt{n}$, kde $S = n^{-1}\sum_{i=1}^n (D_i - \overline{D}_n)^2$. Pro testování $H_0 : \delta = 0$ proti $H_1 : \delta \neq 0$ použijeme $W = \widehat{\delta}/\widehat{\mathrm{se}}(\widehat{\delta})$ a zamítneme $H_0$, pokud $|W| > u_{1-\alpha/2}$.

# Testování dvou očekávání – Problém dvou vzorků

## Příklad 9.11 (Porovnání dvou středních hodnot)

Nechť $X_1, \ldots, X_m$ a $Y_1, \ldots, Y_n$ jsou dva nezávislé výběry z populací s konečnými kladnými variancemi a středními hodnotami $\mu_1$ a $\mu_2$. Testujeme nulovou hypotézu, že $\mu_1 = \mu_2$. Zapíšeme ji jako $H_0 : \delta = 0$ proti $H_1 : \delta \neq 0$, kde $\delta = \mu_1 - \mu_2$. Empirický odhad $\delta$ je $\widehat{\delta} = \overline{X}_m - \overline{X}_n$ s odhadem směrodatné chyby

$$\widehat{\mathrm{se}}(\widehat{\delta}) = \sqrt{\frac{\widehat{\sigma}_1^2}{m} + \frac{\widehat{\sigma}_2^2}{n}},$$

kde $\widehat{\sigma}_1^2 = m^{-1} \sum_{i=1}^m (X_i - \overline{X}_m)^2$ a $\widehat{\sigma}_2^2 = n^{-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2$ jsou výběrové rozptyly. Waldův test velikosti $\alpha$ je zamítnutí $H_0$, pokud $|W| > u_{1-\alpha/2}$, kde $W = (\widehat{\delta} - 0)/\widehat{\mathrm{se}}(\widehat{\delta})$.

# Dualita v testování hypotéz a intervalech spolehlivosti

- Existuje vztah mezi Waldovým testem na hladině $\alpha$ a $(1-\alpha)$ asymptotickým intervalem spolehlivosti $\widehat{\theta} \pm u_{1-\alpha/2}\widehat{\mathrm{se}}(\widehat{\theta})$.
- Testování hypotézy je ekvivalentní rozhodnutí, zda nulová hodnota spadá do intervalu spolehlivosti.

**Věta 9.12 (Ekvivalence testování hypotéz a intervalu spolehlivosti)**

*Waldův test hladině $\alpha$ zamítá $H_0 : \theta = \theta_0$ proti $H_1 : \theta \neq \theta_0$ právě tehdy, když*

$$\theta_0 \notin \left(\widehat{\theta} - u_{1-\alpha/2}\widehat{\mathrm{se}}(\widehat{\theta}), \widehat{\theta} + u_{1-\alpha/2}\widehat{\mathrm{se}}(\widehat{\theta})\right).$$

# $p$-hodnoty

- Rozhodnout "zamítnutí $H_0$" nebo "nezamítnutí $H_0$" není příliš informativní
- Místo toho bychom se mohli ptát, zda pro každé $\alpha$ test zamítá na této úrovni
- Obecně platí, že pokud test zamítá na úrovni $\alpha$, zamítá i na úrovni $\alpha' > \alpha$
- Proto existuje nejmenší $\alpha$, při které test zamítá, a toto číslo nazýváme $p$-hodnota

**Definice 9.13 ($p$-hodnota)**

Pro konkrétní $\omega \in \Omega$ nechť pozorované hodnoty jsou $\mathbb{X}(\omega) = \mathbb{x}$. Předpokládejme, že pro každé $\alpha \in (0, 1)$ máme test na hladině $\alpha$ s kritickým oborem $R_\alpha$. Potom

$$p\text{-hodnota} = \inf \{\alpha : \mathbb{x} \in R_\alpha\}.$$

To znamená, že $p$-hodnota je *nejmenší hladina, při které můžeme zamítnout* $H_0$.

# $p$-hodnoty (pokračování)

- Neformálně je $p$-hodnota mírou evidence proti $H_0$: čím menší $p$-hodnota, tím silnější důkaz proti $H_0$
- Typicky výzkumníci používají následující škálu evidence:
  - $< .01$ ... velmi silný důkaz proti $H_0$
  - $.01 - .05$ ... silný důkaz proti $H_0$
  - $.05 - .10$ ... slabý důkaz proti $H_0$
  - $> .1$ ... malý nebo žádný důkaz proti $H_0$

! Velká $p$-hodnota NENÍ silný důkaz ve prospěch $H_0$

- Velká $p$-hodnota může nastat ze dvou důvodů:
  (i) $H_0$ je pravdivá nebo
  (ii) $H_0$ je nepravdivá, ale test má malou sílu
- Nezaměňujeme $p$-hodnotu s $\mathbb{P}[H_0|Data]$

! $p$-hodnota NENÍ pravděpodobnost, že nulová hypotéza je pravdivá

# Jak vypočítat $p$-hodnotu

**Věta 9.14 (Výpočet $p$-hodnoty)**

*Předpokládejme, že test na hladině $\alpha$ má tvar*

$$\textit{zamítnutí } H_0 \textit{ pokud } T(\mathbb{X}) > c_\alpha.$$

*Potom*

$$p\text{-}\textit{hodnota} = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \mathbb{P}_{\boldsymbol{\theta}} \left[ T(\mathbb{X}) > T(\mathbb{x}) \right],$$

*kde $\mathbb{x}$ je pozorovaná hodnota $\mathbb{X}$. Pokud $\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta}_0\}$, potom*

$$p\text{-}\textit{hodnota} = \mathbb{P}_{\boldsymbol{\theta}_0} \left[ T(\mathbb{X}) > T(\mathbb{x}) \right].$$

- $p$-hodnota je pravděpodobnost (za $H_0$) pozorování hodnoty testové statistiky stejné nebo extrémnější, než byla skutečně pozorována

# $p$-hodnota pro Waldův test

**Věta 9.15 ($p$-hodnota pro Waldův test)**

*Nechť* $w = \frac{\widehat{\theta}(\mathbb{x}) - \theta_0}{\widehat{\mathsf{se}}(\widehat{\theta})}$ *je pozorovaná hodnota Waldovy statistiky* $W$. $p$*-hodnota je dána vztahem*

$$p\text{-hodnota} = \mathbb{P}_{\theta_0}\left[|W| > |w|\right] \to 2\Phi(-|w|), \quad n \to \infty.$$

# Permutační test

- Neparametrická metoda pro testování, zda jsou dvě rozdělení stejná
- Exaktní test, což znamená, že není založen na aproximacích teorie velkých vzorků
- $X_1, \ldots, X_m \overset{IID}{\sim} F_X \perp\!\!\!\perp Y_1, \ldots, Y_n \overset{IID}{\sim} F_Y$
- $H_0 : F_X = F_Y$ versus $H_1 : F_X \neq F_Y$
- Nechť $T$ je nějaká testová statistika, například $T(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = |\overline{X}_m - \overline{Y}_n|$
- Nechť $N = m + n$ a uvažujme všechny $N!$ permutace dat $X_1, \ldots, X_m, Y_1, \ldots, Y_n$
- Pro každou permutaci vypočítejme testovou statistiku $T$
- Tyto hodnoty označíme jako $T_1, \ldots, T_{N!}$
- Pod nulovou hypotézou jsou všechny tyto hodnoty stejně pravděpodobné
- Přesněji řečeno, pod $H_0$, vzhledem k uspořádaným datům jsou $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ rovnoměrně rozděleny přes $N!$ permutací dat

# Permutační test (pokračování)

- Rozdělení $\mathbb{P}_0$, které dává váhu $1/N!$ každému $T_j$, se nazývá permutace rozdělení $T$
- Nechť $t_{obs}$ je pozorovaná hodnota testové statistiky
- Předpokládáme-li, že zamítáme, když je $T$ velké,

$$p\text{-hodnota} = \mathbb{P}_0[\,T > t_{obs}] = \frac{1}{N!}\sum_{j=1}^{N!} \mathbb{1}\{\,T_j > t_{obs}\}$$

# Algoritmus pro permutační test

- Obvykle není praktické počítat všech $N!$ permutací
- $p$-hodnotu můžeme *aproximovat náhodným výbíraním* z množiny permutací
- Podíl případů, kdy $T_j > t_{obs}$, mezi těmito vzorky aproximuje $p$-hodnotu

1. Vypočítejme pozorovanou hodnotu testové statistiky

$$t_{obs} = T(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$$

2. Náhodně promíchejme data a znovu vypočítejme statistiku pomocí permutovaných dat
3. Opakujme předchozí krok $B$ krát a označme výsledné hodnoty jako $T_1, \ldots, T_B$
4. Přibližná $p$-hodnota je

$$p\text{-hodnota} = \frac{1}{B} \sum_{j=1}^{B} \mathbb{1}\{T_j > t_{obs}\}$$

# Waldovy vs permutační testy – dle Marka Twaina

## Příklad 9.16 (Mark Twain)

V roce 1861 se v deníku New Orleans Daily Crescent objevilo 10 esejí. Byly podepsány ‘Quintus Curtuis Snodgrass‘ a někteří lidé měli podezření, že je ve skutečnosti napsal Mark Twain. Abychom to prozkoumali, budeme se zabývat *podílem třípísmenných slov* v autorových textech.

Z osmi Twainových esejí máme:

$$0.225 \quad 0.262 \quad 0.217 \quad 0.240 \quad 0.230 \quad 0.229 \quad 0.235 \quad 0.217$$

Z deseti Snodgrassových esejí máme:

$$0.209 \quad 0.205 \quad 0.196 \quad 0.210 \quad 0.202 \quad 0.207 \quad 0.224 \quad 0.223 \quad 0.220 \quad 0.201$$

Připomeňme si příklad 9.11.

# Mark Twain dle Abrahama Walda v Pythonu I

```
X = [0.225,0.262,0.217,0.240,0.230,0.229,0.235,0.217]
Y = [0.209,0.205,0.196,0.210,0.202,0.207,0.224,0.223,0.220,0.201]

import numpy as np
from scipy.stats import norm

np.random.seed(2024)

X = np.array(X)
Y = np.array(Y)

x_hat = X.mean()
y_hat = Y.mean()

diff_hat = x_hat - y_hat
se_hat = np.sqrt(X.var(ddof=1)/len(X) + Y.var(ddof=1)/len(Y))
```

# Mark Twain dle Abrahama Walda v Pythonu II

```python
u = norm.ppf(0.975)
confidence_interval = (diff_hat - u * se_hat, diff_hat + u * se_hat)

w = diff_hat / se_hat
p_value = 2 * (1 - norm.cdf(abs(w)))

print('Estimated difference of means:\t %.3f' % diff_hat)
print('Estimated SE: \t\t\t %.3f' % se_hat)
print('95%% confidence interval:\t (%.3f, %.3f)' % confidence_interval)
print('Wald statistic: \t\t %.3f' % w)
print('Wald test p-value: \t\t %.4f' % p_value)
```

$$\widehat{\mu}_T - \widehat{\mu}_S = 0.022, \quad \widehat{\mathrm{se}}(\widehat{\mu}_T - \widehat{\mu}_S) = 0.006, \quad C_n^{AN}(95\%) = (0.010, 0.034)$$

$$W = 3.704, \quad p\text{-value} = 0.0002$$

# Mark Twain skrz permutatační test v Pythonu

```
# Permutation test using random shuffling
B = 1000000
full_series = np.concatenate([X, Y])
nx = len(X)
diff_boot_count = 0
for i in range(B):
np.random.shuffle(full_series)
xx, yy = full_series[:nx], full_series[nx:]
diff_boot = xx.mean() - yy.mean()
if diff_boot > diff_hat:
diff_boot_count += 1

p_value_perm = diff_boot_count / B
print('Permutation test p-value: \t\t %.4f' % p_value_perm)
```

$p$-value = 0.0005

# Test podílem věrohodností

**Definice 9.17 (Statistika testu podílu věrohodností)**

Uvažujme *testování* $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ proti $H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$. *Statistika poměru věrohodností* je

$$\lambda = 2\log\frac{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} L_n(\boldsymbol{\theta})} = 2\log\frac{L_n(\widehat{\boldsymbol{\theta}})}{L_n(\widehat{\boldsymbol{\theta}}_0)},$$

kde $\widehat{\boldsymbol{\theta}}$ je MLE (maximálně věrohodný odhad) a $\widehat{\boldsymbol{\theta}}_0$ je MLE, když je $\boldsymbol{\theta}$ omezeno na $\boldsymbol{\Theta}_0$.

**Věta 9.18 (Test poměrem věrohodností)**

*Za předpokladu platnosti nulové hypotézy*
$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 = \{\boldsymbol{\theta} \in \mathbb{R}^d : [\theta_{q+1}, \ldots, \theta_d]^\top = [\theta_{q+1,0}, \ldots, \theta_{d,0}]^\top\}$ *platí, že*

$$\lambda \xrightarrow[n\to\infty]{\mathbb{D}} \chi^2_{d-q},$$

*kde $d - q$ je rozdíl dimenze $\boldsymbol{\Theta}$ a dimenze $\boldsymbol{\Theta}_0$.*

# Agenda

10. Literatura

# Seznam literatury


Casella, G. and Berger, R.L. (2001)
*Statistical Inference*, 2nd Edition
Pacific Grove, CA: Duxbury

Chung, K.L. (2001)
*A Course in Probability Theory*, 3rd Edition
San Diego, CA: Academic Press

Dupač, V. and Hušková, M. (2013)
*Pravděpodobnost a matematická statistika*
Praha, CZ: Karolinum

Resnick, S.I. (2013)
*A Probability Path*, 2014th Edition
Basel, CH: Birkhäuser

Rosenthal, J.S. (2006)
*A First Look at Rigorous Probability Theory*, 2nd Edition
Singapore, SG: World Scientific

Ross, S.M. (2020)
*A First Course in Probability*, 10th Edition
London, UK: Pearson

Wasserman, L. (2013)
*All of Statistics: A Concise Course in Statistical Inference*
New York, NY: Springer

# Konec

michal.pesta@mff.cuni.cz