

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

- ≠ 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)

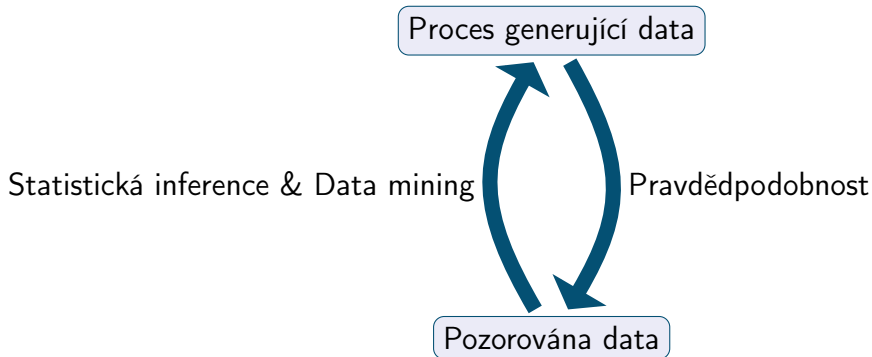
= Matematická statistika + Algoritmické výpočty

- **Matematická statistika** = odůvodněná statistika

! ne pouze “Jako?”, ale také “Proč?”

Data mining (dolování dat?)

+ Statistická inference = Stochastické modelování



Machine learning (strojové učení)

⊂ Statistical learning (statistické učení)

Matematická statistika

odhadování

klasifikace

shlukování

data

kovariáty

klasifikátor

hypotéza

interval spolehlivosti

Strojové učení

učení

učení s učitelem

uční bez učitele

trénovací soubor

znaky

hypotéza

—

—

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 Ω je množina výsledků experimentu. Prvky $\omega \in \Omega$ se nazývají **elementární jevy**, **výsledky** nebo **realizace**. Podmnožiny Ω 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^Ω představuje její potenční množinu. Podmnožina $\mathcal{A} \subseteq 2^\Omega$ se nazývá σ -*algebra* nebo σ -*pole*, pokud splňuje:

- (i) $\emptyset \in \mathcal{A}$;
- (ii) pokud $A \in \mathcal{A}$, pak $A^c \in \mathcal{A}$;
- (iii) pokud $A_1, A_2, \dots \in \mathcal{A}$, pak $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

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 necht' 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) $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ je *pravděpodobnostní míra*, pokud splňuje:

- (i) $\mathbb{P}(\Omega) = 1$;
- (ii) pokud $A_1, A_2, \dots \in \mathcal{A}$ 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: $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$,
 $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

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)

Bud' $A_n \uparrow A$ nebo $A_n \downarrow A$, pak

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

Konečné výběrové prostory

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

Příklad 1.7 (Hod dvěma kostkami)

Ω má 36 prvků: $\Omega = \{(i, j) : i, j \in \{1, \dots, 6\}\}$. 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 Ω 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 hoďu padne orel. Pak

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

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^c
$+$	.009	.099
$-$	.001	.891

Podle definice podmíněné pravděpodobnosti:

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

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 $A_1, A_2, \dots$ je disjunkt ní spočetný rozklad Ω 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 $A_1, A_2, \dots$ je disjunkt ní spočetný rozklad Ω 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 $.7 + .2 + .1 = 1$. Necht' 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 ω reálné číslo $X(\omega)$. To znamená, že

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

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

Příklad 2.2 (Hod mincí desetkrát)

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

Rozdělení pravděpodobnosti

Definice 2.3 (Rozdělení náhodné veličiny)

Rozdělením náhodné veličiny $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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 $(\mathcal{M}, \mathcal{M})$ a $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathcal{M}, \mathcal{M})$. 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,

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

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

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: μ, ν jsou σ -konečné míry na $(\mathcal{M}, \mathcal{M})$:

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

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

Definice 2.5 (Hustota náhodné veličiny vzhledem k míře)

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

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 σ -konečné míře μ , potom $\mathbb{P}[X \in B] = \int_B f_X d\mu$.

Distribuční funkce

Definice 2.7 (Distribuční funkce)

Nechť $(\Omega, \mathcal{A}, \mathbb{P})$ je pravděpodobnostní prostor. Distribuční funkce (CDF) náhodné veličiny X je funkce $F_X : \mathbb{R} \rightarrow [0, 1]$ 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 psních prostorech, nebo na stejném psním 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 μ 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

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

- μ_{aspoj} je **absolutně spojitá** míra vzhledem k Lebesgueově míře λ

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

- μ_{singn} je **singulární neatomická** míra vzhledem k Lebesgueově míře λ

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

- μ_{diska} je **diskrétní (atomická)** míra vzhledem k Lebesgueově míře λ

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

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 δ_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 ν *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 ν 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 $f_X(x) = \begin{cases} 1/4, & x = 0; \\ 1/2, & x = 1; \\ 1/4, & x = 2; \\ 0, & \text{jinak.} \end{cases}$

Spojité rozdělení

Definice 2.12 (Spojité náhodné veličiny)

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

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

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

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

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

Důsledky definice distribuční funkce

Lemma 2.13 (Důsledky definice distribuční funkce)

Necht' 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

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

- 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 $X \stackrel{d}{=} Y$ – 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 $\{1, \dots, k\}$ právě tehdy, když

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

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 $X \sim \text{Alt}(p)$ nebo $X \sim \text{Be}(p)$.

- 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 $X \sim \text{Bi}(n, p)$.

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

Příklady – Diskrétní náhodné veličiny III

Příklad 2.19 (Geometrické rozdělení)

X má geometrické rozdělení s parametrem $p \in (0, 1)$ ($X \sim \text{Geo}(p)$) 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 $p \in (0, 1)$ ($X \sim \text{NB}(n, p)$) 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.
- $\text{NB}(1, p)$ jest $\text{Geo}(p)$.

Příklady – Diskrétní náhodné veličiny IV

Příklad 2.21 (Poissonovo rozdělení)

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

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

- Pokud $X \sim \text{Po}(\lambda_X)$ a $Y \sim \text{Po}(\lambda_Y)$ jsou nezávislé ($\perp$), potom $X + Y \sim \text{Po}(\lambda_X + \lambda_Y)$ (později snadné).
- Neplatí bez $\perp$ (zkuste kontrapříklad).
- $\Leftarrow$ netriviální (Raikovova věta).
- Jestliže $n \rightarrow \infty$ a $np \rightarrow \lambda < \infty$, pak $\text{Bi}(n, p)$ konverguje k $\text{Po}(\lambda)$ (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 $X \sim U(a, b)$ nebo $X \sim R(a, b)$.

Příklad 2.23 (Normální nebo Gaussovo rozdělení)

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

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

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

Příklady – Spojité náhodné veličiny II

- Distribuční funkce $N(0, 1)$: $\Phi(x) = \int_{-\infty}^x \phi(t)dt \dots$ **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 $X \sim N(3, 5)$. 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 $\Gamma(p) = \int_0^\infty t^{p-1} \exp\{-t\} dt$ 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 $X \sim \text{Beta}(\alpha, \beta)$ nebo $X \sim B(\alpha, \beta)$.

Příklad 2.28 (χ^2 -rozdělení)

X má χ^2 -rozdělení s p stupni volnosti ($X \sim \chi_p^2$) 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 $X_1, \dots, X_p \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, pak $\sum_{i=1}^p X_i^2 \sim \chi_p^2$.

Příklady – Spojité náhodné veličiny V

Příklad 2.29 (Studentovo t -rozdělení)

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

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

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

Příklad 2.30 (Cauchyho rozdělení)

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

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

Mnohorozměrná náhodnost

Definice 2.31 (Náhodný vektor)

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

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

Definice 2.32 (Rozdělení náhodného vektoru)

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

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

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

Sdružená distribuční funkce

Definice 2.33 (Sdružená distribuční funkce)

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

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) \equiv \mathbb{P}\left(\bigcap_{\ell=1}^d \{X_\ell \leq x_\ell\}\right), \quad \forall \mathbf{x} = [x_1, \dots, x_d]^\top \in \mathbb{R}^d.$$

Věta 2.34 (Vlastnosti sdružené distribuční funkce)

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

- (i) F je po složkách neklesající a zprava spojitá;
- (ii)

$$\lim_{x_\ell \downarrow -\infty} F(\mathbf{x}) = 0 \quad \text{pro každé } \ell = 1, \dots, d$$

a

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

Marginály

Věta 2.35 (Marginální distribuční funkce)

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

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

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

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

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

Rozdělení podvektorů

Definice 2.36 (Rozdělení náhodného podvektoru – Marginální rozdělení)

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

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

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

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

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

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

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

Spojité náhodné vektory

Definice 2.38 (Spojitý náhodný vektor)

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

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

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

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

$$F_{\mathbb{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_{\mathbb{X}}(t_1, \dots, t_d) dt_d \dots dt_1, \quad \forall \mathbf{x} = [x_1, \dots, x_d]^\top \in \mathbb{R}^d$$

Hustota mnohorozměrného rozdělení

Věta 2.39 (Hustota vzhledem k součinové referenční míře)

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

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

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

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

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

Pravděpodobnostní funkce pro mnohorozměrné

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

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

Příklad 2.40

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

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

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

Hustota pro mnohorozměrné rozdělení

Příklad 2.41

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

Událost $B = \{X < 1/2, Y < 1/2\}$ odpovídá podmnožině jednotkového čtverce. Integrace $f_{[X, Y]^T}$ 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 spojitě

Důsledek 2.42 (Marginální rozdělení pro dvourozměrné náhodné vektory)

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

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

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

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

- Lze snadno zobecnit pro vícerozměrné náhodné vektory – Věta 2.39

Nezávislost II

Všimněte si, že

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

Definice 2.43 (Nezávislé náhodné veličiny)

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

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

kde $\mathbf{X} = [X_1, \dots, X_d]^\top$.

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 $\mathbb{X}_1, \dots, \mathbb{X}_d$ jsou **nezávislé**, pokud

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

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

Diskrétní a spojitá nezávislost

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

Věta 2.45 (Ekvivalentní charakterizace nezávislosti)

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

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

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

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

Kartézský nosič a nezávislost

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

Důsledek 2.46 (Kartézský součin nosičů a nezávislost)

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

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

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

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

pro nějaké měřitelné funkce g, h , a pro λ^2 -s.v. $[x, y]^{\top} \in \mathcal{S}([X, Y]^{\top})$, 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 $X_1, \dots, X_d$ jsou **nezávislé** právě tehdy, když

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

kde $\mathbf{X} = [X_1, \dots, X_d]^\top$.

Podmíněné rozdělení

Definice 2.48 (Podmíněná pravděpodobnostní funkce a hustota)

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

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

Pro spojitý náhodný vektor $[X, Y]^T$ 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 psťní 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 ... $\mathbb{P}[X \in A | Y = y] = \int_A f_{X|Y}(x|y) dx$

Mnohorozměrné normální rozdělení

Definice 2.49 (Mnohorozměrné normální rozdělení)

Náhodný vektor $\mathbf{X} = [X_1, \dots, X_d]^\top$ 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 $\mathbf{Y} = [Y_1, \dots, Y_k]^\top$ a matice $\mathbf{A} \in \mathbb{R}^{d \times k}$ takové, že

- (i) $\perp \{Y_1, \dots, Y_k\}$;
- (ii) $Y_j \sim \mathcal{N}(0, 1)$, $\forall j \in \{1, \dots, k\}$;
- (iii) $\mathbf{A}\mathbf{A}^\top = \Sigma$;
- (iv) $\mathbf{X} = \mathbf{A}\mathbf{Y} + \boldsymbol{\mu}$.

- Značení: $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$
- Speciální případ: *standardní d-rozměrné normální rozdělení* $\boldsymbol{\mu} = \mathbf{0}$ a $\Sigma = I_d$, tj.
 $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, I_d)$
- Matice $\mathbf{A}$ 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 spojitě* 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

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

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

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

kde t^{-1} určuje všechny vzory

Transformace spojitých náhodných vektorů

- **Transformace** $Z = t(X, Y)$, kde $t : \mathbb{R}^2 \rightarrow \mathbb{R}$
- Mějme *známé spojitě* rozdělení vektoru $[X, Y]^\top$, tj. $f_{[X, Y]^\top}(x, y)$
- 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

$$\begin{aligned} F_Z(z) &= \mathbb{P}[Z \leq z] = \mathbb{P}[t(X, Y) \leq z] = \mathbb{P}[\omega : t([X, Y]^\top(\omega)) \leq z] \\ &= \iint_{\mathcal{T}(z)} f_{[X, Y]^\top}(x, y) dx dy \end{aligned}$$

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

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

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 spojitě veličiny

Pomocí *Radon–Nikodýmovy věty*:

Věta 3.3 (Diskrétní a spojitá střední hodnota)

Střední hodnota veličiny X je

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

za předpokladu, že pravá strana je dobře definovaná.

Pozor na střední hodnotu

Příklad 3.4 (Cauchyho rozdělení)

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

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

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

Střední hodnota a transformace

Transformace $Y = t(X)$.

Věta 3.5 (Pravidlo líného statistika)

Bud' $t : \mathbb{R} \rightarrow \mathbb{R}$ měřitelná funkce a necht' $Y = t(X)$. Pak

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

pokud pravá strana existuje.

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

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 ℓ -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 $\mathbb{E}(\sum_{\ell=1}^d a_\ell X_\ell) = \sum_{\ell=1}^d a_\ell \mathbb{E}X_\ell$.

Věta 3.12 (Násobení při nezávislosti)

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

Rozptyl

Rozptyl měří “**variabilitu**” (disperzi) rozdělení.

Definice 3.13 (Rozptyl a směrodatná odchylka)

Rozptyl X je definován jako

$$\text{Var} X = \mathbb{E}(X - \mathbb{E}X)^2,$$

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

$$\text{sd}(X) = \sqrt{\text{Var} X}.$$

- 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:

- $$\mathrm{Var} X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \geq 0;$$
- *pokud $a, b \in \mathbb{R}$, pak*
$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var} X;$$
- *pokud $X_1, \dots, X_d \in \mathcal{L}^2$ jsou **nezávislé** a $a_1, \dots, a_d$ jsou reálné konstanty, pak*

$$\mathrm{Var}\left(\sum_{\ell=1}^d a_{\ell} X_{\ell}\right) = \sum_{\ell=1}^d a_{\ell}^2 \mathrm{Var} X_{\ell}.$$

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

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

Pokud $\text{Var}(X)\text{Var}(Y) > 0$, pak je *korelace* mezi X a Y definována jako

$$\rho_{X,Y} \equiv \mathbb{C}\text{orr}(X, Y) = \frac{\mathbb{C}\text{ov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

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

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

Vlastnosti kovariance a korelace

Věta 3.16 (Vlastnosti kovariance a korelace)

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

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

Důsledek 3.17 (Rozpyl součtu)

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

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

Varianční-kovarianční matice

Definice 3.18 (Mnohorozměrná střední hodnota)

Střední hodnota náhodného vektoru $\mathbf{X} = [X_1, \dots, X_d]^\top$ je definována jako

$$\mathbb{E}\mathbf{X} = [\mathbb{E}X_1, \dots, \mathbb{E}X_d]^\top.$$

Všimněte si, že $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ a $\text{Cov}(X, X) = \text{Var}X$.

Definice 3.19 (Varianční-kovarianční matice)

Varianční-kovarianční matice náhodného vektoru $\mathbf{X} = [X_1, \dots, X_d]^\top$ je definována jako

$$\text{Var}\mathbf{X} = \begin{bmatrix} \text{Var}X_1 & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_d) \\ \text{Cov}(X_2, X_1) & \text{Var}X_2 & \dots & \text{Cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_d, X_1) & \text{Cov}(X_d, X_2) & \dots & \text{Var}X_d \end{bmatrix}.$$

Vlastnosti varianční-kovarianční matice

Věta 3.20 (Vlastnosti varianční-kovarianční matice)

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

$$\mathbb{E}(\mathbf{a}^\top \mathbf{X} + b) = \mathbf{a}^\top \mathbb{E}\mathbf{X} + b \quad \text{a} \quad \text{Var}(\mathbf{a}^\top \mathbf{X} + b) = \mathbf{a}^\top (\text{Var}\mathbf{X}) \mathbf{a}.$$

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

$$\mathbb{E}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\mathbb{E}\mathbf{X} + \mathbf{b} \quad \text{a} \quad \text{Var}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}(\text{Var}\mathbf{X})\mathbf{A}^\top.$$

Momentová vytvořující funkce

Definice 3.21 (MGF)

Momentová vytvořující funkce (MGF) X je definována jako

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

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 $X_1, \dots, X_d$ 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 $X \stackrel{d}{=} Y$.

Charakteristická funkce

Definice 3.23 (CF)

Charakteristická funkce (CF) X je definována jako

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

- 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) φ_X existuje pro *jakékoli* rozdělení X
- (ii) $\varphi_X(0) = 1$
- (iii) $|\varphi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}$
- (iv) φ_X je *stejněměrně 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) $\varphi_X(t) \in \mathbb{R} \quad \forall t \in \mathbb{R} \iff \mathbb{P}[X > x] = \mathbb{P}[X < -x] \quad \forall x \geq 0$ (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 $a < b$,

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

CF jednoznačně určuje rozdělení.

Důsledek 3.26 (Jednoznačná charakterizace rozdělení)

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

Charakteristická funkce pro náhodný vektor

Definice 3.27 (CF pro náhodný vektor)

Charakteristická funkce náhodného vektoru $\mathbf{X} = [X_1, \dots, X_d]^\top$ je definována jako

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

- 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 $\varepsilon > 0$,

$$\mathbb{P}[|X - \mathbb{E}X| \geq \varepsilon] \leq \frac{\text{Var}[X]}{\varepsilon^2}.$$

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

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}X \text{Var}Y}.$$

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 $X_1, X_2, \dots$ je posloupnost náhodných veličin a necht' X je jiná náhodná veličina. Necht' F_n označuje distribuční funkci X_n a necht' F označuje distribuční funkci X .

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

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

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

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

Typy stochastické konvergence II

Definice 5.2 (Konvergence v Lebesgueových prostorech a skoro jistě)

Nechť $X_1, X_2, \dots$ je posloupnost náhodných veličin a necht' X je jiná náhodná veličina.

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

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

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

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

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)

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

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

$$(c) \quad p \geq q \geq 1: X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{L_q} X$$

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

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

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 $2^n + m$, $m = 0, 1, \dots, 2^n - 1$ 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 $\mathbb{P}[|X_{2^n+m}| > \varepsilon] = 2^{-n} \rightarrow 0$ když $n \rightarrow \infty$. 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. $X_n \not\xrightarrow{\mathbb{P}\text{-s.j.}}$.

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 $\mathbb{P}[|X_{2^n+m}| > \varepsilon] = 2^{-n} \rightarrow 0$ když $n \rightarrow \infty$. 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. $X_n \not\xrightarrow{\mathbb{L}_1}$. Tedy, $X_n \not\xrightarrow{\mathbb{L}_p}$, $p \geq 1$.

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

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

Obrácené implikace neplatí IV

Příklad 5.7 (Konvergence v distribuci neimplikuje konvergenci v pravděpodobnosti)

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

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. $X_n \not\xrightarrow{P}$.

Věta o spojitém zobrazení

Věta 5.8 (Věta o spojitém zobrazení (CMT))

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

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

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

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)

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

Příklad 5.12 (CF normálního rozdělení)

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

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 $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}X_1$ když $n \rightarrow \infty$.

- 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 $\text{Var}X_1 > 0$, pak

$$Z_n := \sqrt{n} \frac{\bar{X}_n - \mathbb{E}X_1}{\sqrt{\text{Var}X_1}} \xrightarrow{\mathbb{D}} Z \sim \text{N}(0, 1), \quad n \rightarrow \infty.$$

Jinými slovy,

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

Zkráceně,

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

Mnohorozměrná centrální limitní věta

Důsledek 5.17 (Cramér–Woldova věta)

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

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

Věta 5.18 (Mnohorozměrná CLT)

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

$$\sqrt{n}(\bar{\mathbf{X}}_n - \mathbb{E}\mathbf{X}_1) \xrightarrow{\mathbb{D}} N_d(\mathbf{0}, \text{Var}\mathbf{X}_1), \quad n \rightarrow \infty.$$

Delta metoda

Věta 5.19 (Delta metoda)

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

$$\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{\mathbb{D}} N\left(0, (g'(\mu))^2 \sigma^2\right), \quad n \rightarrow \infty.$$

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) $X_1, \dots, X_n \stackrel{IID}{\sim} F$, jak odvodíme F ?

Náhodný výběr

Definice 6.1 (Náhodný výběr a rozsah výběru)

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

$$X_1, \dots, X_n \stackrel{IID}{\sim} F.$$

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

Data z experimentu

- **Uvažujeme** / myslíme si / předpokládáme:
 - měřitelné zobrazení $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, \dots, n$
- **Pozorujeme** / měříme / získáváme:
 - reálná data $X_i(\omega) \in \mathbb{R}$, $i = 1, \dots, n$ 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 μ a σ^2
- Obecně,

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

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

Neparametrické modely

Příklad 6.3 (Model Sobolevova prostoru)

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

- 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 $\hat{\theta}_n$ parametru θ je měřitelná funkce t náhodných veličin $X_1, \dots, X_n$:

$$\hat{\theta}_n = t(X_1, \dots, X_n).$$

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

Konzistentní odhad

- $\hat{\theta}_n$ je **nestranný**, pokud $\mathbb{E}[\hat{\theta}_n] = \theta$ 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 $\text{bias}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n] - \theta$
- Rozumným požadavkem na odhad je, aby “konvergoval” k pravé hodnotě parametru s *rostoucím množstvím dat*

Definice 6.5 (Konzistentní odhad)

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

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

Standardní chyba

- Rozdělení $\hat{\theta}_n$ se nazývá *výběrové rozdělení*
- Standardní odchylka $\hat{\theta}_n$ se nazývá **standardní chyba**: $\text{se}(\hat{\theta}_n) = \sqrt{\text{Var}\hat{\theta}_n}$
- 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 $\hat{\text{se}}$ (nebo odhad standardní chyby)

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

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

Střední kvadratická chyba

- Kvalita bodového odhadu se někdy posuzuje pomocí **střední kvadratické chyby**

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

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

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

které generovalo data

- Kvantitativní vlastnost*

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

$$\text{MSE}(\widehat{\theta}_n) = \text{bias}^2(\widehat{\theta}_n) + \text{Var}(\widehat{\theta}_n)$$

Slabá konzistence

Věta 6.8 (Postačující podmínka pro konzistenci)

$\text{bias}(\hat{\theta}_n) \rightarrow 0$ & $\text{Var}(\hat{\theta}_n) \rightarrow 0 \Rightarrow \hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ (*konzistentní estimátor*)

Příklad 6.9 (Konzistence – Hod mincí)

Příklad 6.6: $\mathbb{E}(\hat{p}_n) = p$ & $\text{Var}(\hat{p}_n) = p(1-p)/n \rightarrow 0 \Rightarrow \hat{p}_n \xrightarrow{\mathbb{P}} p$

Definice 6.10 (Asymptoticky standardně normální odhad)

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

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

Konfidenční množiny

Definice 6.11 (Konfidenční interval – Interval spolehlivosti)

$(1 - \alpha)$ -konfidenční interval pro parametr θ je interval $C_n = (a, b)$, kde $a = a(X_1, \dots, X_n)$ a $b = b(X_1, \dots, X_n)$ 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 θ je interval C_n takový, že

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

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

CI Normální model

Příklad 6.12 (CI – Normalita)

Příklad 6.2: $X_1, \dots, X_n \stackrel{IID}{\sim} N(\mu, \sigma^2)$ 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 μ : $\hat{\mu}_n = n^{-1} \sum_{i=1}^n X_i \equiv \bar{X}_n$
- Linearita normálních rozdělení: $\sqrt{n}(\hat{\mu}_n - \mu)/\sigma \sim N(0, 1)$, $\forall n \in \mathbb{N}$
- Kvantily standardního normálního rozdělení:

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

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

Délka konfidenčního intervalu

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

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

- 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 ? ... $n \geq \left\lceil \frac{4u_{1-\alpha/2}^2 \sigma^2}{d^2} \right\rceil + 1$

Konfidenční interval založený na normalitě

Věta 6.13 (CI založený na normalitě)

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

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

Pak,

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

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

CLT konfidenční interval

Příklad 6.14 (CI – Hod mincí)

Příklad 6.9:

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

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

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

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

- $X_1, \dots, X_n \stackrel{IID}{\sim} F$ 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. $\#\{X_i \leq x\}/n$

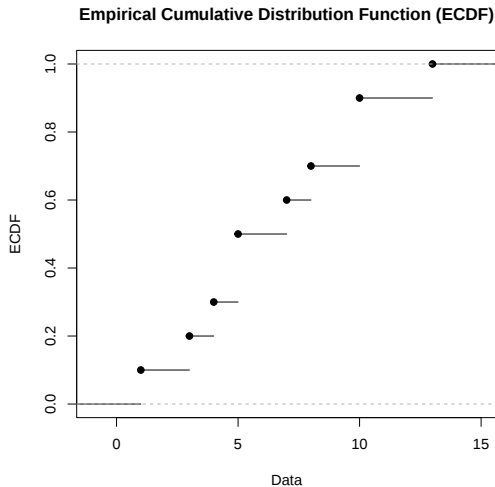
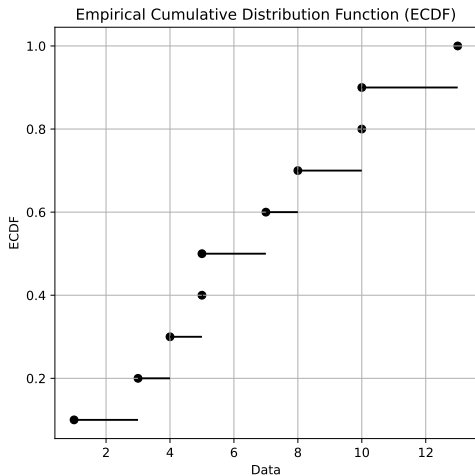
ECDF – Python

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

ECDF – R

```
1 data <- c(1, 3, 4, 5, 5, 7, 8, 10, 10, 13)
2 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 $x \in \mathbb{R}$,

- $\mathbb{E} \left[\widehat{F}_n(x) \right] = F(x);$
- $\text{Var} \left[\widehat{F}_n(x) \right] = \frac{F(x)\{1-F(x)\}}{n};$
- $\text{MSE} \left(\widehat{F}_n(x) \right) = \frac{F(x)\{1-F(x)\}}{n} \rightarrow 0 \text{ když } n \rightarrow \infty;$
- $\widehat{F}_n(x) \xrightarrow{\mathbb{P}} F(x) \text{ když } n \rightarrow \infty.$

Statistický funkcionál

Funkcionál je jen funkce funkce.

Definice 7.3 (Funkcionál)

Funkcionál je zobrazení $T : \mathcal{F} \rightarrow \mathbb{R}$, 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)

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

Lineární statistický funkcionál

Definice 7.6 (Lineární statistický funkcionál)

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

Důvod, proč se $T(\mathbb{P}_X) = \int r(x) d\mathbb{P}_X(x)$ 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 $X_1, \dots, X_n$ je *náhodný výběr* z F s rozsahem výběru n , kde $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ pro $i = 1, \dots, n$. 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 $\{X_1, \dots, X_n\}$.
- $\widehat{\mathbb{P}}_n$ je vážený součet (směs) Diracových měr v **náhodných** bodech $\{X_1, \dots, X_n\}$ 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 $\hat{\theta}_n = T(\hat{\mathbb{P}}_n)$.

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

Příklad 7.9 (ECDF)

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

Plug-in odhad pro lineární statistický funkcionál

Definice 7.10 (Empirický odhad pro lineární statistický funkcionál)

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

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

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

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

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

$$\mu = T(\mathbb{P}_X) = \int x d\mathbb{P}_X(x) \Rightarrow \hat{\mu}_n = \int x d\hat{\mathbb{P}}_n(x) = \bar{X}_n$$

Empirický rozptyl

Příklad 7.13 (Rozptyl)

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

$$\begin{aligned}\hat{\sigma}_n^2 &= \int x^2 d\hat{\mathbb{P}}_n(x) - \left(\int x d\hat{\mathbb{P}}_n(x) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(X_i - \bar{X}_n \right)^2\end{aligned}$$

Další rozumný odhad σ^2 je *výběrový rozptyl*

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

Empirická korelace

Příklad 7.14 (Korelace)

Nechť $Z = [X, Y]^T$ a necht' $\rho = T(\mathbb{P}_{[X, Y]^T}) = \mathbb{E}(X - \mu_X)(Y - \mu_Y)/(\sigma_X \sigma_Y)$ označuje korelaci mezi X a Y . Můžeme napsat

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

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

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

$$T_3(\mathbb{P}_{[X, Y]^T}) = \int xy d\mathbb{P}_{[X, Y]^T}(x, y);$$

$$T_4(\mathbb{P}_{[X, Y]^T}) = \int x^2 d\mathbb{P}_{[X, Y]^T}(x, y);$$

$$T_5(\mathbb{P}_{[X, Y]^T}) = \int y^2 d\mathbb{P}_{[X, Y]^T}(x, y);$$

$$a(t_1, t_2, t_3, t_4, t_5) = \frac{t_3 - t_1 t_2}{\sqrt{(t_4 - t_1^2)(t_5 - t_2^2)}}.$$

Empirická korelace (pokračování)

Příklad 7.15 (Korelace (pokračování))

Nahradíme $\mathbb{P}_{[X,Y]^\tau}$ pomocí $\widehat{\mathbb{P}}_n$ v

$T_1(\mathbb{P}_{[X,Y]^\tau})$, $T_2(\mathbb{P}_{[X,Y]^\tau})$, $T_3(\mathbb{P}_{[X,Y]^\tau})$, $T_4(\mathbb{P}_{[X,Y]^\tau})$, $T_5(\mathbb{P}_{[X,Y]^\tau})$ a vezměme

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

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

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

Nyní definujeme

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

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):

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

kde Θ je **parametrický prostor** a $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ 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 $T(\theta)$
- Například, pokud $X_i \stackrel{IID}{\sim} N(\mu, \sigma^2)$, pak parametr je $\theta = (\mu, \sigma^2)$
- Pokud naším cílem je odhadnout μ , pak $\mu = T(\theta)$ je nazýván **parametrem zájmu** a σ^2 je nazýván **nežádoucím / rušivým parametrem**
- Parametr zájmu může být *komplikovanou* funkcí θ

Parametr zájmu (pokračování)

Příklad 8.1 (Gamma model)

Připomeňme si, že X má rozdělení $\text{Gamma}(a, p)$, pokud

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

kde $a, p > 0$ a

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

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

Parametrické odhady

Uvažujeme náhodný výběr $X_1, \dots, X_n \stackrel{IID}{\sim} F_\theta \in \mathcal{F}$

Příklad 8.2 (Pouze konečná střední hodnota)

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

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

Příklad 8.3 (Pouze konečný rozptyl)

$$\mathcal{F} = \{F(\sigma^2) : \text{Var}_{F(\sigma^2)} = \sigma^2 \text{ \& } \sigma^2 < \infty\}$$

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

Parametrický odhad (pokračování)

Příklad 8.4 (Poissonův model)

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

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

Příklad 8.5 (Poissonův model – Zvláštní případ)

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

- Jediný *nestranný* odhad θ 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
- Uvažujeme *náhodný výběr* $X_1, \dots, X_n \stackrel{IID}{\sim} \mathbb{P}_X^\theta \sim F_\theta \in \mathcal{F}$
- Definujme **k -tý (necentrální) moment**

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

- 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 $\hat{\theta}_n$ je definován jako hodnota θ , pro kterou platí

$$\mu'_1(\hat{\theta}_n) = \hat{\mu}'_1, \quad \mu'_2(\hat{\theta}_n) = \hat{\mu}'_2, \quad \dots \quad \mu'_d(\hat{\theta}_n) = \hat{\mu}'_d.$$

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

Příklad 8.7 (Odhad metodou momentů pro alternativní distribuci)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Be}(p)$. Pak, $\mu'_1 = \mathbb{E}_p X_1 = p$ a $\hat{\mu}'_1 = \bar{X}_n$. Rovnost těchto hodnot nám dává odhad

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

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

Odhady MoM (pokračování)

Příklad 8.8 (Odhad metodou momentů pro normální rozdělení)

$X_1, \dots, X_n \stackrel{IID}{\sim} N(\mu, \sigma^2)$. Pak, $\mu'_1 = \mathbb{E}_{(\mu, \sigma^2)} X_1 = \mu$ a $\mu'_2 = \mathbb{E}_{(\mu, \sigma^2)} X_1^2 = \text{Var}_{(\mu, \sigma^2)} X_1 + (\mathbb{E}_{(\mu, \sigma^2)} X_1)^2 = \sigma^2 + \mu^2$. Musíme vyřešit rovnice

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

Toto je soustava dvou rovnic se dvěma neznámými. Řešení je

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

Odhady MoM (pokračování II)

Příklad 8.9 (MoM v Gamma modelu)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Gamma}(a, p)$, kde

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

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

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

jsou *konzistentní* a *AN*.

Odhady MoM (pokračování III)

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

$X_1, \dots, X_n \stackrel{IID}{\sim} U(\theta_1, \theta_2)$. Odhady MoM

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

jsou *konzistentní* a *AN*.

Příklad 8.11 (MoM v Beta modelu)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Beta}(\alpha, \beta)$. Odhady MoM

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

jsou *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 *sdužená hustota pozorování*, s tím že ji **považujeme za funkci parametru $\boldsymbol{\theta}$**
- Tedy, $L_n : \Theta \rightarrow [0, \infty)$
- 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 $\hat{\theta}_n$, je hodnota parametru θ , která maximalizuje $L_n(\theta)$.

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

Příklad 8.14 (Odhad metodou maximální věrohodnosti pro alternativní rozdělení)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Alt}(p)$. 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 $\partial \ell_n(p) / \partial p \stackrel{!}{=} 0$.
MLE je $\hat{p}_n = \bar{X}_n$.

Příklad 8.15 (Odhad metodou maximální věrohodnosti pro normální rozdělení)

$X_1, \dots, X_n \stackrel{IID}{\sim} N(\mu, \sigma^2)$. 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 $\partial \ell_n(\mu, \sigma^2) / \partial (\mu, \sigma^2)^\top \stackrel{!}{=} \mathbb{0}_2$. MLE je $\hat{\mu}_n = \bar{X}_n$ a $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

MLE (pokračování)

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

$X_1, \dots, X_n \stackrel{IID}{\sim} U(0, \theta)$. Potom,

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

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

MLE (pokračování II)

Příklad 8.17 (MLE v exponenciálním modelu)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Exp}(\lambda)$. MLE je $\hat{\lambda}_n = 1/\bar{X}_n$.

Příklad 8.18 (MLE v Gamma modelu)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Gamma}(a, p)$. MLE $\hat{p}_n$ pro parametr p je řešením *nelineární rovnice*

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

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

Vlastnosti maximálně věrohodních odhadů

MLE je:

- *konzistentní*
- *ekvivariantní* ... $\theta \rightsquigarrow \hat{\theta}_n \leftrightarrow g(\theta) \rightsquigarrow g(\hat{\theta}_n)$
- *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 $f(x; \theta) = f(x; \vartheta)$ pro s.v. $x \in \mathbb{R}$ implikuje $\theta = \vartheta$.

- 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 $\tau = g(\theta)$ je měřitelná funkce θ . Nechť $\hat{\theta}_n$ je MLE pro θ . Pak $\hat{\tau}_n := g(\hat{\theta}_n)$ je MLE pro τ .

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

Příklad 8.21

$X_1, \dots, X_n \stackrel{IID}{\sim} N(\theta, 1)$. MLE pro θ je $\hat{\theta}_n = \bar{X}_n$. Nechť $\tau = \exp\{\theta\}$, pak MLE pro τ je $\hat{\tau}_n = \exp\{\bar{X}_n\}$.

Skóre a informace

Definice 8.22 (Skórová funkce a Fisherova informace)

Skórová funkce je definována jako

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

Fisherova informace je definována jako

$$I_n(\theta) = \text{Var}_\theta \left[\sum_{i=1}^n S(X_i; \theta) \right] = \sum_{i=1}^n \text{Var}_\theta [S(X_i; \theta)].$$

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

Podmínky regularity

- Necht' náš parametrický model $\mathcal{F} = \{f(\cdot; \theta) : \theta \in \Theta \subseteq \mathbb{R}^d\}$ je *identifikovatelný*
- Necht' θ_0 je *skutečná neznámá hodnota* parametru θ

Definice 8.23 (Podmínky regularity)

Podmínky regularity jsou:

- (i) $\theta_0 \in \text{int } \Theta$ a $\mathcal{S}(X)$ nezávisí na θ ;
- (ii) $f(x; \theta)$ je dvakrát spojitě diferencovatelná podle θ na $\text{int } \Theta$;
- (iii) $\|\log f(x; \theta)\| \leq d(x)$ pro $\forall \theta \in \text{int } \Theta$, $\forall x \in \mathcal{S}(X)$ a $\mathbb{E}_{\theta_0}[d(X)] < \infty$;
s.v.
- (iv) $\|\partial f(x; \theta)/\partial \theta\| \leq e(x)$ pro $\forall \theta \in \text{int } \Theta$, $\forall x \in \mathcal{S}(X)$ a $\int e(x) d\lambda(x) < \infty$;
s.v.
- (v) $\|\partial^2 \log f(x; \theta)/\partial \theta^\top \partial \theta\| \leq g(x)$ pro $\forall \theta \in \text{int } \Theta$, $\forall x \in \mathcal{S}(X)$ a $\mathbb{E}_{\theta_0}[g(X)] < \infty$;
s.v.
- (vi) $\|\partial^2 f(x; \theta)/\partial \theta^\top \partial \theta\| \leq h(x)$ pro $\forall \theta \in \text{int } \Theta$, $\forall x \in \mathcal{S}(X)$ a $\int h(x) d\lambda(x) < \infty$;
s.v.
- (vii) $I(\theta_0)$ je pozitivně definitní.

Score a informace pod podmínkami regularity

- *Od této chvíle budeme předpokládat, že podmínky regularity platí*
- Označme $[\cdot]^{\otimes 2} := [\cdot][\cdot]^\top$
- Střední hodnotu matice definujeme jako matici středních hodnot jejích složek

Lemma 8.24 (Střední hodnota a rozptyl skóre)

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

Lemma 8.25 (Alternativní definice informace)

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

Důsledek 8.26 (Limitní informace)

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

Asymptotická normalita MLE

Věta 8.27 (AN MLE)

Nechť θ_0 je skutečná hodnota neznámého parametru θ . Nechť platí *podmínky regularity*. Pak je MLE $\hat{\theta}_n$ pro θ *asymptoticky normální*, tak že

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathbb{D}} N_d(\mathbb{0}_d, I^{-1}(\theta_0)).$$

Definice 8.28 (Asymptotický rozptyl)

Za podmínek regularity je *asymptotický rozptyl* MLE $\hat{\theta}_n$ rovný $\text{avar}(\hat{\theta}_n) = n^{-1}I^{-1}(\theta_0)$.

- Asymptotický rozptyl NENÍ limitní rozptyl
- Neformálně, $\hat{\theta}_n \overset{as}{\approx} N_d(\theta_0, n^{-1}I^{-1}(\theta_0))$
- Interval spolehlivosti založené na normálním rozdělení ($d = 1$) ... Theorem 6.13

Příklady MLE

Příklad 8.29 (MLE ve specifickém případě Beta modelu)

$X_1, \dots, X_n \stackrel{IID}{\sim} f(x; \theta) = \theta(1-x)^{\theta-1} \mathbb{1}\{x \in (0, 1)\}$ a necht' θ_0 je skutečná hodnota neznámého parametru $\theta > 1$. Pak,

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

Příklad 8.30 (MLE v modelu normálního rozdělení)

$X_1, \dots, X_n \stackrel{IID}{\sim} f(x; \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{(x-\theta)^2}{2\theta}\right\}$, $x \in \mathbb{R}$ a necht' θ_0 je skutečná hodnota neznámého parametru $\theta > 0$. Pak,

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

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 σ^2 je *známé*. MLE pro μ je $\hat{\mu}_n = \bar{X}_n$. Dostáváme

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

Potom

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

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

Příklady MLE (pokračování III)

Příklad 8.33 (MLE v Poissonově modelu)

$X_1, \dots, X_n \stackrel{IID}{\sim} \text{Po}(\lambda)$, kde $\lambda > 0$. MLE pro λ je $\hat{\lambda}_n = \bar{X}_n$. Dostáváme

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

Potom

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

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

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 Θ na dvě disjunktní množiny Θ_0 a Θ_1 , tj.
 $\Theta = \Theta_0 \cup \Theta_1$ a $\Theta_0 \cap \Theta_1 = \emptyset$
- Chceme testovat

$$H_0 : \theta \in \Theta_0 \quad \text{proti} \quad H_1 : \theta \in \Theta_1$$

- 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 **zamítací 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)
- Zamítací 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 (Sílofunkce a hladina významnosti)

Síla testu se zamítacím oborem R je definována jako

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

Hladina testu je definována jako

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta).$$

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

Typy hypotéz a testů

- Hypotéza ve tvaru $\theta = \theta_0$, tj. $\theta \in \Theta_0 = \{\theta_0\}$, se nazývá **jednoduchá hypotéza**
- Hypotéza ve tvaru $\theta \in \Theta_0$, kde Θ_0 obsahuje více než jeden prvek, se nazývá **složená hypotéza**
- Test ve tvaru $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ se nazývá **oboustranný test**
- Test ve tvaru $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ nebo $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$ 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)

$X_1, \dots, X_n \stackrel{IID}{\sim} N(\mu, \sigma^2)$, kde $\sigma^2 > 0$ je známé. Chceme testovat $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. Tedy $\Theta_0 = (-\infty, 0]$ a $\Theta_1 = (0, +\infty)$. Uvažujme test

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

kde $T(\mathbb{X}) = \bar{X}_n$. Zamítací obor je

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

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

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

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 μ (viz obrázek na dalším slajdu). Proto *hladina* testu je

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

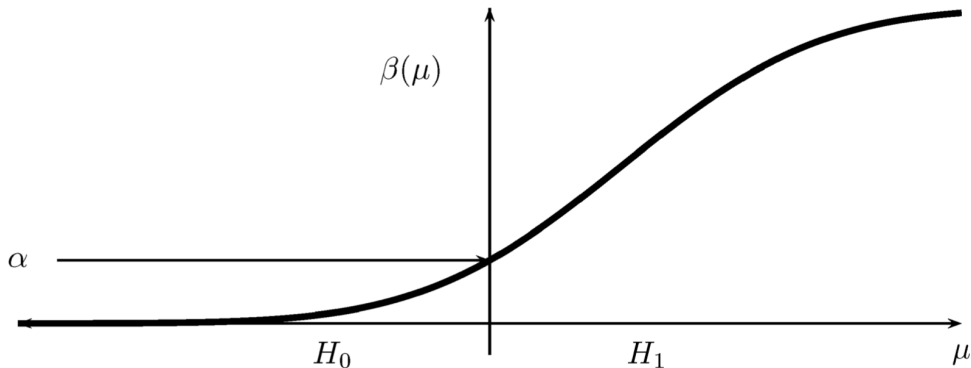
Nastavíme tuto hodnotu rovnu α a řešíme pro c , abychom získali

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

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

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

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 θ je *skalární* parametr, $\hat{\theta}$ je odhad θ a $\widehat{\text{se}}(\hat{\theta})$ je odhad směrodatné chyby $\hat{\theta}$.

Definice 9.6 (Waldův test)

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

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

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

Věta 9.7 (Asymptotická hladina Waldova testu)

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

Síla Waldova testu

- Alternativní verzí Waldovy testové statistiky je $W = (\hat{\theta} - \theta_0)/\text{se}_0(\hat{\theta})$, kde $\text{se}_0(\hat{\theta})$ 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 θ je $\theta_1 \neq \theta_0$. Síla $\beta(\theta_1)$ – pravděpodobnost správného zamítnutí nulové hypotézy – je přibližně

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

- Připomeňme, že $\widehat{\text{se}}(\hat{\theta})$ se zmenšuje s rostoucím rozsahem výběru.
- Síla je velká, pokud je θ_1 daleko od θ_0 .
- Síla je velká, pokud je rozsah výběru velký.

Testování dvou pravděpodobností – Porovnání dvou vzorků

Příklad 9.9 (Porovnání dvou predikčních algoritmů (pokračování))

Pokračujeme 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 $X \sim \text{Bi}(m, p_1)$ a $Y \sim \text{Bi}(n, p_2)$. 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 $\hat{\delta} = \hat{p}_1 - \hat{p}_2$ s odhadem směrodatné chyby

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

Waldův test na hladině α je zamítnutí H_0 , pokud $|W| > u_{1-\alpha/2}$, kde $W = (\hat{\delta} - 0)/\widehat{\text{se}}(\hat{\delta})$. 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))

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

Testování dvou očekávání – Problém dvou vzorků

Příklad 9.11 (Porovnání dvou středních hodnot)

Nechť $X_1, \dots, X_m$ a $Y_1, \dots, Y_n$ jsou dva nezávislé výběry z populací s konečnými kladnými variancemi a středními hodnotami μ_1 a μ_2 . Testujeme nulovou hypotézu, že $\mu_1 = \mu_2$. Zapišeme ji jako $H_0 : \delta = 0$ proti $H_1 : \delta \neq 0$, kde $\delta = \mu_1 - \mu_2$. Empirický odhad δ je $\hat{\delta} = \bar{X}_m - \bar{Y}_n$ s odhadem směrodatné chyby

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

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

Dualita v testování hypotéz a intervalech spolehlivosti

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

Věta 9.12 (Ekvivalence testování hypotéz a intervalu spolehlivosti)

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

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

p -hodnoty

- Rozhodnout “zamítnutí H_0 ” nebo “nezamítnutí H_0 ” není příliš informativní
- Místo toho bychom se mohli ptát, zda pro každé α test zamítá na této úrovni
- Obecně platí, že pokud test zamítá na úrovni α , zamítá i na úrovni $\alpha' > \alpha$
- Proto existuje **nejmenší α , 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 α se zamítacím oborem R_α . 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 α má tvar

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

Potom

$$p\text{-hodnota} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta [T(\mathbb{X}) > T(\mathfrak{x})],$$

kde $\mathfrak{x}$ je pozorovaná hodnota $\mathbb{X}$. Pokud $\Theta_0 = \{\theta_0\}$, potom

$$p\text{-hodnota} = \mathbb{P}_{\theta_0} [T(\mathbb{X}) > T(\mathfrak{x})].$$

- 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)

Necht' $w = \frac{\hat{\theta}(\mathbf{x}) - \theta_0}{\widehat{\text{se}}(\hat{\theta})}$ je pozorovaná hodnota Waldovy statistiky W . p -hodnota je dána vztahem

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

Permutační test

- Neparametrická metoda pro testování, zda jsou dvě rozdělení stejná
- **Exaktní test**, což znamená, že není založen na aproximacích teorie velkých vzorků
- $X_1, \dots, X_m \stackrel{IID}{\sim} F_X \perp\!\!\!\perp Y_1, \dots, Y_n \stackrel{IID}{\sim} F_Y$
- $H_0 : F_X = F_Y$ versus $H_1 : F_X \neq F_Y$
- Nechť T je nějaká testová statistika, například
$$T(X_1, \dots, X_m, Y_1, \dots, Y_n) = |\bar{X}_m - \bar{Y}_n|$$
- Nechť $N = m + n$ a uvažujme všechny $N!$ permutace dat $X_1, \dots, X_m, Y_1, \dots, Y_n$
- Pro každou permutaci vypočítejme testovou statistiku T
- Tyto hodnoty označíme jako $T_1, \dots, T_{N!}$
- Pod nulovou hypotézou jsou všechny tyto hodnoty stejně pravděpodobné
- Přesněji řečeno, pod H_0 , vzhledem k uspořádaným datům jsou $X_1, \dots, X_m, Y_1, \dots, Y_n$ 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 výběráním* z množiny permutací
- Podíl případů, kdy $T_j > t_{obs}$, mezi těmito vzorky aproximuje p -hodnotu

1. Vypočítejme pozorovanou hodnotu testové statistiky

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

2. Náhodně promíchejme data a znovu vypočítejme statistiku pomocí permutovaných dat

3. Opakujme předchozí krok B krát a označme výsledné hodnoty jako $T_1, \dots, T_B$

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 0.262 0.217 0.240 0.230 0.229 0.235 0.217

Z deseti Snodgrassových esejí máme:

0.209 0.205 0.196 0.210 0.202 0.207 0.224 0.223 0.220 0.201

Připomeňme si příklad 9.11.

Mark Twain dle Abrahama Walda v Pythonu I

```
1 X = [0.225,0.262,0.217,0.240,0.230,0.229,0.235,0.217]
2 Y = [0.209,0.205,0.196,0.210,0.202,0.207,0.224,0.223,0.220,0.201]
3
4 import numpy as np
5 from scipy.stats import norm
6
7 np.random.seed(2024)
8
9 X = np.array(X)
10 Y = np.array(Y)
11
12 x_hat = X.mean()
13 y_hat = Y.mean()
14
15 diff_hat = x_hat - y_hat
16 se_hat = np.sqrt(X.var(ddof=1)/len(X) + Y.var(ddof=1)/len(Y))
```

Mark Twain dle Abrahama Walda v Pythonu II

```
1 u = norm.ppf(0.975)
2 confidence_interval = (diff_hat - u * se_hat, diff_hat + u * se_hat)
3
4 w = diff_hat / se_hat
5 p_value = 2 * (1 - norm.cdf(abs(w)))
6
7 print('Estimated difference of means:\t %.3f' % diff_hat)
8 print('Estimated SE: \t\t\t %.3f' % se_hat)
9 print('95%% confidence interval:\t (%.3f, %.3f)' % confidence_interval)
10 print('Wald statistic: \t\t %.3f' % w)
11 print('Wald test p-value: \t\t %.4f' % p_value)
```

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

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

Mark Twain skrz permutační test v Pythonu

```
1 # Permutation test using random shuffling
2 B = 1000000
3 full_series = np.concatenate([X, Y])
4 nx = len(X)
5 diff_boot_count = 0
6 for i in range(B):
7     np.random.shuffle(full_series)
8     xx, yy = full_series[:nx], full_series[nx:]
9     diff_boot = xx.mean() - yy.mean()
10    if diff_boot > diff_hat:
11        diff_boot_count += 1
12
13    p_value_perm = diff_boot_count / B
14    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 $H_0 : \boldsymbol{\theta} \in \Theta_0$ proti $H_1 : \boldsymbol{\theta} \notin \Theta_0$. Statistika poměru věrohodností je

$$\lambda = 2 \log \frac{\sup_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta_0} L_n(\boldsymbol{\theta})} = 2 \log \frac{L_n(\hat{\boldsymbol{\theta}})}{L_n(\hat{\boldsymbol{\theta}}_0)},$$

kde $\hat{\boldsymbol{\theta}}$ je MLE (maximálně věrohodný odhad) a $\hat{\boldsymbol{\theta}}_0$ je MLE, když je $\boldsymbol{\theta}$ omezeno na Θ_0 .

Věta 9.18 (Test poměrem věrohodností)

Za předpokladu platnosti nulové hypotézy

$H_0 : \boldsymbol{\theta} \in \Theta_0 = \{\boldsymbol{\theta} \in \mathbb{R}^d : [\theta_{q+1}, \dots, \theta_d]^\top = [\theta_{q+1,0}, \dots, \theta_{d,0}]^\top\}$ platí, že

$$\lambda \xrightarrow[n \rightarrow \infty]{\mathbb{D}} \chi_{d-q}^2,$$

kde $d - q$ je rozdíl dimenze Θ a dimenze Θ_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`